SCIENCE CHINA
Technological Sciences

•Article•

https://doi.org/10.1007/s11431-024-2764-5

A systematic framework of constructing surrogate model for slider track peeling strength prediction

DONG XingJian[1*], CHEN Qian[1*], LIU WenBo[2], WANG Dong[1,3], PENG ZhiKe[1] & MENG Guang[1]

[1] *State Key Laboratory of Mechanical System and Vibration, Shanghai Jiao Tong University, Shanghai 200240, China;*
[2] *KEIPER Seating Mechanisms Co., Ltd., Soochow 215500, China;*
[3] *Department of Industrial Engineering and Management, Shanghai Jiao Tong University, Shanghai 200240, China*

Received April 7, 2024; accepted July 26, 2024; published online September 14, 2024

Peeling strength can comprehensively reflect slider track safety and is crucial in car seat safety assessments. Current methods for determining slider peeling strength are primarily physical testing and numerical simulation. However, these methods encounter the potential challenges of high costs and overlong time consumption which have not been adequately addressed. Therefore, the efficient and low-cost surrogate model emerges as a promising solution. Nevertheless, currently used surrogate models suffer from inefficiencies and complexity in data sampling, lack of robustness in local model predictions, and isolation between data sampling and model prediction. To overcome these challenges, this paper aims to set up a systematic framework for slider track peeling strength prediction, including sensitivity analysis, dataset sampling, and model prediction. Specifically, the interpretable linear regression is performed to identify the sensitivity of various geometric variables to peeling strength. Based on the variable sensitivity, a distance metric is constructed to measure the disparity of different variable groups. Then, the sparsity-targeted sampling (STS) is proposed to formulate a representative dataset. Finally, the sequentially selected local weighted linear regression (SLWLR) is designed to achieve accurate track peeling strength prediction. Additionally, a quantitative cost assessment of the supplementary dataset is proposed by utilizing the minimum adjacent sample distance as a mediator. Experimental results validate the efficacy of sequential selection and the weighting mechanism in enhancing localization robustness. Furthermore, the proposed SLWLR method surpasses similar approaches and other common surrogate methods in terms of prediction performance and data quantity requirements, achieving an average absolute error of 3.3 kN in the simulated test dataset.

slider track, peeling strength, surrogate model, sensitivity analysis, data sampling, local weighted linear regression

Citation: Dong X J, Chen Q, Liu W B, et al. A systematic framework of constructing surrogate model for slider track peeling strength prediction. Sci China Tech Sci, 2024, 67, https://doi.org/10.1007/s11431-024-2764-5

1 Introduction

The slider track serves as a crucial connecting component between the car seat and the chassis, fulfilling essential functions such as fixation, adjustment, and locking for the seat. As depicted in Figure 1(a), it typically comprises upper rail, lower rail, locking module, bearing, and carriages. Slider track plays a crucial role in automotive safety assessments [1], adhering to specific national standards [2]. Common metrics for evaluating the slider track strength include peeling strength, unlocking strength, free clearance, etc. Notably, as shown in Figure 1(b), peeling strength, representing the maximum force before deformation and failure, could comprehensively reflect the slider track's safety performance. Therefore, obtaining peeling strength is vital for track profile design, seat performance evaluation, and

*Corresponding authors (email: donxj@sjtu.edu.cn; chenqian2020@sjtu.edu.cn)

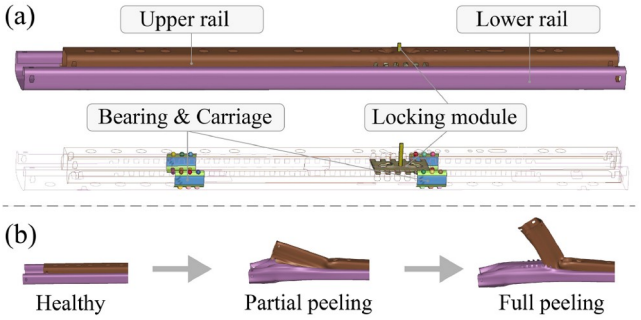

Figure 1 (Color online) Illustration of the structure of the slider track and the peeling process. (a) The slider track structure; (b) the peeling process.

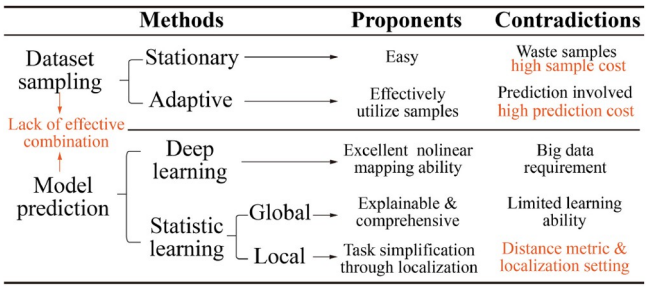

Figure 2 (Color online) A simple introduction of dataset sampling and model prediction.

passenger safety assurance [3,4].

The peeling strength of the slider track exhibits nonlinear dependence on multiple coupled geometric variables. Currently, various methods are employed to evaluate the peeling strength. However, the majority, if not all, rely on physical testing and finite element analysis (FEA). Physical testing conducts multiple standard physical peeling experiments on slider tracks to acquire precise peeling strength data. Conversely, FEA [5–7] constructs numerical models of the slider track under corresponding constraints and loads to derive peeling strength. For example, Liu et al. [8] established a slider track model to optimize the geometric parameters for specific application scenarios. While FEA reduces resource and time costs compared to physical testing, it still encounters challenges in offering a comprehensive assessment of slider track strength under diverse conditions, which may not align with the iterative calculation requirements inherent in the slider track design process. Therefore, the surrogate model emerges as a viable approach [9–11], which could offer a swift, cost-effective, and precise solution for predicting slider track peeling strength.

The surrogate model construction [9,12,13] consists of two main parts: dataset sampling and model prediction. Specifically, a training dataset is obtained through numerical simulation, such as finite element analysis (FEA), and utilized to train the surrogate model for predicting the target output. Unlike direct numerical simulation, the surrogate model eliminates the need for real-time calculations, relying solely on numerical simulation during the dataset sampling phase. This streamlined approach significantly reduces time costs and computational resources, making it well-suited for the time-consuming assessment of slider track peeling strength.

An appropriate dataset sampling strategy can maintain the quality of the surrogate model while mitigating prohibitive sampling costs. As illustrated in Figure 2, dataset sampling is classified into stationary and adaptive methods [9]. Stationary methods, based upon patterns [14] like Latin hypercube sampling [15,16] or design of experiments (DOE) literature [17] like orthogonal sampling, are straightforward but inefficient due to the lack of feedback from prediction task, resulting in high sample costs. In contrast, adaptive sampling [18–22] selects samples sequentially to enhance the surrogate model's precision with the consideration of the current dataset and feedback from the prediction task. However, these methods entail a high prediction cost and require careful tuning due to multiple evaluations of prediction performance required in each iteration [23–25]. Therefore, it is necessary to develop a dataset sampling method that balances the effectiveness (i.e., considering prediction task feedback) and the ease of implementation (i.e., reducing reliance on prediction model).

Regarding model prediction, current methods can be broadly categorized into two types: deep learning [26] and statistical learning. Deep learning methods [27–29] exhibit potent non-linear mapping capabilities but demand substantial training data, rendering them less suitable for scenarios with high data costs. Conversely, prevalent statistical learning methods, such as response surface [30,31] and support vector regression [32], typically use the entire training dataset to construct the surrogate model, known as global methods. However, predicting slider peeling strength is a complicated task with multiple factors, high coupling, and intricate non-linearity. It proves very challenging to establish an effective general global model, particularly when data is limited. To address this challenge, local methods, adeptly leveraging adjacent samples of the test data to craft more focused local models, are well-suited for intricate prediction tasks, and local weighted linear regression (LWLR) [33] is a typical one. For instance, Zhang et al. [34] utilized LWLR to predict the damping ratio of a dominant mode in real power systems, while Xu et al. [33] applied LWLR to an ensemble of surrogate models, achieving dynamic response prediction for submerged floating tunnels. Despite the demonstrated effectiveness of local methods across various applications [33–35], different localization settings and distance metrics can result in different selection of adjacent samples, and eventually affect the stability of prediction results [36,37]. Consequently, mitigating sensitivity to the localization settings (i.e., the number of adjacent samples k) [38,39] and choosing appropriate distance metrics [40,41] emerge as crucial challenges in refining local pre-

diction methods.

Furthermore, the current surrogate model construction also faces the challenge of the isolation between data sampling and model prediction. Specifically, although scholars have developed detailed research on data sampling and model prediction, these investigations often focus on a single aspect, lacking an effective combination of both [9,12]. This overlooks that model prediction might lay the groundwork for improved data sampling, or alternatively, data sampling could assess the supplementary cost for model performance. Therefore, proposing a systematic framework for constructing surrogate models that effectively combinates data sampling and model prediction is also a pivotal concern in current research on surrogate models.

To address the aforementioned challenges related to dataset sampling and model prediction, we present a systematic and engineering-oriented framework comprising sensitivity analysis, sparsity-targeted sampling (STS), sequentially-selected local weighted linear regression (SLWLR), and the associated quantitative cost assessment of supplementary dataset. This framework ultimately establishes a cost-effective, efficient, accurate, and incrementally learning solution for slider track peeling strength prediction. The innovations of this work can be summarized as follows.

(1) Regarding dataset sampling, a sparsity-targeted sampling (STS) method is established based on the sensitivity analysis of the slider track. STS not only effectively reduces the sample cost to construct a representative dataset, but also avoids the involvement of prediction models, ensuring ease of implementation.

(2) Concerning model prediction, the sensitivity analysis is leveraged to formulate the distance metric and the sequentially-selected local weighted linear regression (SLWLR) method is proposed to improve the robustness of localization.

(3) Besides, a systematic framework is established for slider track peeling strength prediction, incorporating a quantitative supplementary cost assessment derived from sampling estimation and prediction experience.

2 Experimental

2.1 Problem defination

Based on prior knowledge, specific geometric variables of the slider track, recognized as influential on peeling strength, are selected to form the design variable group $\mathbf{x} = [1, x_1, x_2, \cdots, x_p] \in \mathbb{R}^{p+1}$. Here, the first element 1 is employed as a bias term in linear regression, p represents the number of chosen geometric variables, and the corresponding peeling strength is denoted as the label y to be predicted. Subsequently, the peeling strength prediction task is delineated into two primary phases: dataset sampling and model prediction.

During dataset sampling, the design variable groups $\{\mathbf{x}\}_{i=1}^{n}$ are determined through a sampling algorithm, and the corresponding peeling strengths $\{y\}_{i=1}^{n}$ are acquired via FEA. Consequently, the training dataset $D^{\text{train}} = \{(\mathbf{x}_i, y_i)\}_{i=1}^{n}$ is assembled by combining these pairs of variables and labels. For the test dataset $D^{\text{test}} = \{(\mathbf{x}_i, y_i)\}_{i=1}^{m}$, a similar procedure is followed, except that the design variable groups are randomly sampled for model evaluation. Here, n and m represent the number of samples in the training and test dataset, respectively.

As for model prediction, a predictive model $f: \mathbb{R}^{p+1} \rightarrow \mathbb{R}$ is developed using the training dataset D^{train}, and subsequently employed to estimate the peeling strength for the samples within D^{test}.

2.2 Local weighted linear regression

Local weighted linear regression (LWLR), an extension of traditional local linear regression (LLR) incorporating a weighted mechanism, is a widely utilized regression method in diverse fields, including hydrology [33,42–44], computer science [45,46], medical imaging [47], and others. As depicted in Figure 3(a), traditional LLR treats all adjacent samples equally, assigning them equivalent weights. However, this approach overlooks the fact that excessively distant samples may not contribute sufficient information and can even impair the predictive accuracy of the model. Hence, LLR is extremely sensitive to the number of adjacent sample k, as distant samples do not provide adequate support for accurate predictions.

Addressing this limitation, LWLR intuitively assigns varying weights to each adjacent sample based on its proximity to the input sample. This approach optimally utilizes information from nearby neighbors, mitigating potential harm to model prediction accuracy caused by distant neighbors. As illustrated in Figure 3(b), LWLR assigns higher weights to nearby adjacent samples and lower weights

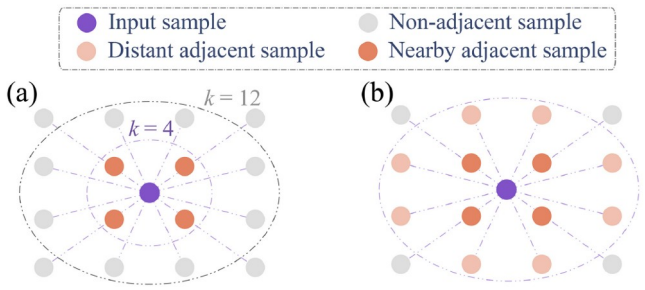

Figure 3 (Color online) Illustration of the comparison between LLR and LWLR. (a) LLR treats all adjacent samples equally by assigning them equivalent weights; (b) LWLR assigns different weights to each adjacent sample based on their proximity to the input sample.

to distant adjacent samples, enhancing the model's adaptability and robustness.

Mathematically, the standardized weight $\widetilde{w}_i$ assigned to each neighbor $\mathbf{x}_i$ of the input sample $\bar{\mathbf{x}}$ can be denoted as

$$\widetilde{w}_i = \frac{w_i}{\sum_{j\in D_k} w_j},\ w_j = \frac{1}{d(\mathbf{x}_j,\bar{\mathbf{x}})+\epsilon}, \tag{1}$$

where w_j represents the non-standardized weight, D_k represents the adjacent sample set containing k samples, $d(\cdot,\cdot)$ is the distance metric to measure the disparity between two samples, and ϵ is a small positive constant introduced to prevent singularity in the division operation.

Denoting the sample matrix $\mathbf{X}$ and the label vector $\mathbf{y}$ as

$$\begin{aligned}&\mathbf{X} = [\mathbf{x}_1^T, \mathbf{x}_2^T, \cdots, \mathbf{x}_n^T] \in \mathbb{R}^{n\times(p+1)},\\ &\mathbf{y} = [y_1, y_2, \cdots, y_n]^T \in \mathbb{R}^{n\times 1}.\end{aligned} \tag{2}$$

Then, the predicted values $\hat{\mathbf{y}}$ of a linear regression model (including LLR and LWLR) could be expressed as

$$\hat{\mathbf{y}} = \mathbf{X}\boldsymbol{\theta}, \tag{3}$$

where $\boldsymbol{\theta} = [b, \theta_1, \theta_2, \cdots, \theta_p] \in \mathbb{R}^{p+1}$ is the regression coefficient vector to be optimized, and b is the bias of linear regression.

Incorporating both sample weighting $\widetilde{\mathbf{w}} = [\widetilde{w}_1, \widetilde{w}_2, \cdots, \widetilde{w}_k] \in \mathbb{R}^k$ and variable regularization $\mathbf{r} = [0, r_1, r_2, \cdots, r_p] \in \mathbb{R}^{p+1}$, the optimization objective of LWLR can be formulated as

$$\begin{aligned}&\min_{\theta} L(\boldsymbol{\theta}, D_k, \widetilde{\mathbf{w}}, \mathbf{s})\\ &= \min_{\theta} \sum_{i\in D_k}^{k} \widetilde{w}_i(\hat{y}_i - y_i)^2 + \alpha\cdot\sum_{j=1}^{p} r_j\theta_j^2\\ &= \min_{\theta} (\mathbf{X}\boldsymbol{\theta}-\mathbf{y})^T\widetilde{\mathbf{W}}(\mathbf{X}\boldsymbol{\theta}-\mathbf{y}) + \alpha\cdot\boldsymbol{\theta}^T\mathbf{R}\boldsymbol{\theta},\end{aligned} \tag{4}$$

where α is the regularization coefficient, and $\widetilde{\mathbf{W}} = \mathrm{diag}(\widetilde{\mathbf{w}})$ and $\mathbf{R} = \mathrm{diag}(\mathbf{r})$ denote the diagonal matrices of $\widetilde{\mathbf{w}}$ and $\mathbf{r}$, respectively. The first term of L is weighted square errors to ensure prediction accuracy, while the second term is the variable regularization to prevent overfitting.

The derivative of the above optimization objective L can be obtained as

$$\begin{aligned}\frac{\partial L}{\partial \boldsymbol{\theta}} &= \frac{\partial}{\partial \boldsymbol{\theta}}[\boldsymbol{\theta}^T(\mathbf{X}^T\widetilde{\mathbf{W}}\mathbf{X}+\alpha\mathbf{R})\boldsymbol{\theta} - 2\mathbf{y}^T\widetilde{\mathbf{W}}\mathbf{X}\boldsymbol{\theta} + \mathbf{y}^T\mathbf{y}]\\ &= 2(\mathbf{X}^T\widetilde{\mathbf{W}}\mathbf{X}+\alpha\mathbf{R})\boldsymbol{\theta} - 2\mathbf{X}^T\widetilde{\mathbf{W}}\mathbf{y}.\end{aligned} \tag{5}$$

Since the optimization objective L is convex, setting eq. (5) to zero leads to the optimal regression coefficient $\hat{\boldsymbol{\theta}}$ as

$$\frac{\partial L}{\partial \boldsymbol{\theta}} = 0 \rightarrow \hat{\boldsymbol{\theta}} = (\mathbf{X}^T\widetilde{\mathbf{W}}\mathbf{X}+\alpha\mathbf{R})^{-1}\mathbf{X}^T\widetilde{\mathbf{W}}\mathbf{y}. \tag{6}$$

In model prediction, bringing $\hat{\boldsymbol{\theta}}$ back into eq. (3), the predicted value $\hat{y}^{\mathrm{pre}}$ for a new sample $\mathbf{x}^{\mathrm{pre}}$ can be expressed as

$$\hat{y}^{\mathrm{pre}} = (\mathbf{x}^{\mathrm{pre}})^T\hat{\boldsymbol{\theta}}. \tag{7}$$

3 Methodology

3.1 Framework

The holistic systematic framework for slider peeling strength prediction is illustrated in Figure 4. This framework comprises four main components: sensitivity analysis (SA), sparsity-targeted sampling (STS), sequentially-selected locally weighted linear regression (SLWLR), and supplementary cost assessment. The workflow of this framework is as follows.

Sensitivity analysis: Acquire a mini batch dataset D^{SA} and calculate the sensitivity $\mathbf{s} \in \mathbb{R}^p$ for each variable in the slider track, which forms the basis for the weighted distance metric

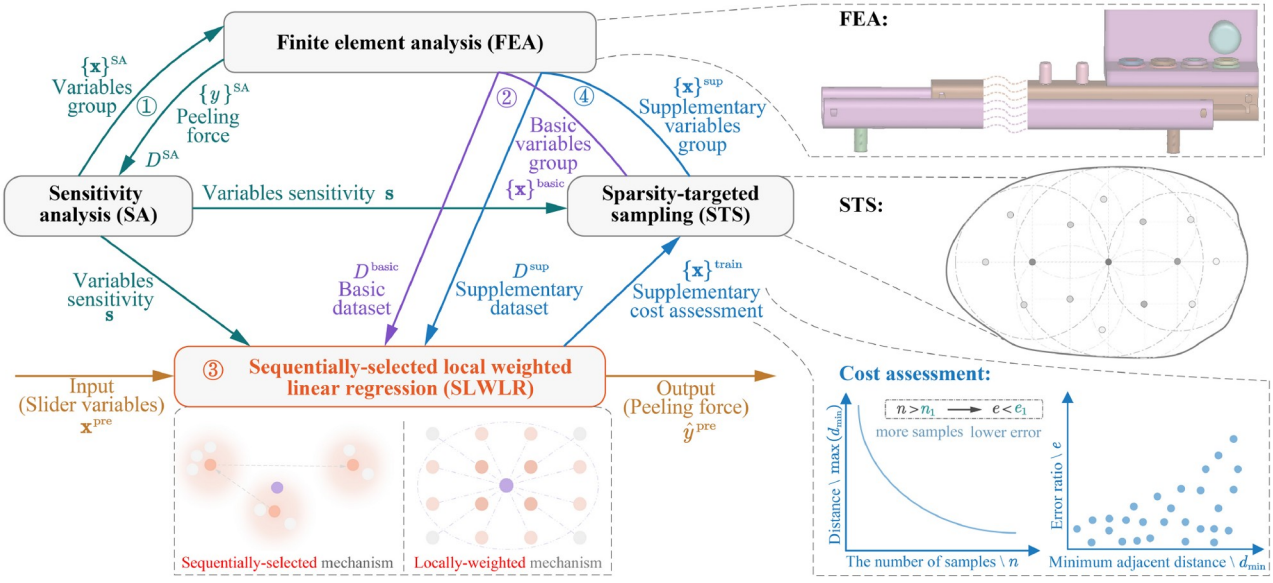

Figure 4 (Color online) The holistic systematic framework of constructing surrogate model for slider track peeling strength prediction.

$d_{\mathbf{s}}(\cdot,\cdot)$ utilized in subsequent STS and SLWLR algorithms.

Sparsity-targeted sampling: With the weighted distance metric $d_{\mathbf{s}}(\cdot,\cdot)$ and current training samples $\{\mathbf{x}\}^{\text{train}}$, STS iteratively explores sparse regions in the variable space and conducts sampling within these regions, progressively reducing the potential sparsity and aiming to construct a representative dataset.

Sequentially-selected local weighted linear regression: With the weighted distance metric $d_{\mathbf{s}}(\cdot,\cdot)$, current training dataset D^{train}, and test dataset D^{test}, SLWLR considers the repulsive effects from the already selected samples and sequentially selects k neighbor samples $\widetilde{D}^{k}(\mathbf{x}^{\text{pre}})$ for each test sample $\mathbf{x}^{\text{pre}} \in D^{\text{test}}$. Subsequently, SLWLR assigns different weights and predicts the peeling strength $\tilde{y}^{\text{pre}}$ of $\mathbf{x}^{\text{pre}}$ based on $\widetilde{D}^{k}(\mathbf{x}^{\text{pre}})$, similar to the approach utilized by LWLR.

Supplementary cost assessment: Leveraging the minimum adjacent sample distance as a mediator, this assessment establishes the relationship between the performance improvement of predicting peeling strengths and the number of required supplementary samples. This constitutes a quantitative assessment between the supplementary cost and the targeted accuracy.

3.2 Sensitivity analysis

Sensitivity analysis aims to obtain the sensitivity of each variable in the slider track, which is a prerequisite for subsequent STS and SLWLR algorithms. The sensitivity s_i quantifies the impact of each variable on the output and is typically defined as the partial derivative, formulated as

$$s_i = \left|\frac{\partial y}{\partial x_i}\right|. \tag{8}$$

Linear regression, as a typical interpretable statistical learning model, not only achieves effective predictions but also elucidates the sensitivity of corresponding variables via the optimized regression coefficient $\hat{\boldsymbol{\theta}}$. The relationship between linear regression and variable sensitivity is demonstrated through a case study with the following linear regression model:

$$y \approx \hat{y} = \mathbf{x}^{T}\hat{\boldsymbol{\theta}} = 3x_1 - 8x_2 - 1, \tag{9}$$

where $\hat{\boldsymbol{\theta}} = [b, \hat{\theta}_1, \hat{\theta}_2] = [-1, 3, -8]$. Then, we could get the sensitivity s_i from the optimized regression coefficient $\hat{\boldsymbol{\theta}}$ as

$$\begin{aligned} s_1 &= \left|\frac{\partial y}{\partial x_1}\right| \approx \left|\frac{\partial \hat{y}}{\partial x_1}\right| = \left|\hat{\theta}_1\right| = 3, \\ s_2 &= \left|\frac{\partial y}{\partial x_2}\right| \approx \left|\frac{\partial \hat{y}}{\partial x_2}\right| = \left|\hat{\theta}_2\right| = 8. \end{aligned} \tag{10}$$

Based on the above explanation, the sensitivity analysis could be conducted as follows.

Initially, acquire a small batch of variable groups $\{\mathbf{x}\}^{\text{SA}}$ using specific basic sampling methods such as Latin hypercube sampling (LHS). Then, calculate corresponding peeling strengths $\{y\}^{\text{SA}}$ via FEA and formulate a mini dataset $D^{\text{SA}} = \{(\mathbf{x}, y), \cdots\}^{\text{SA}}$ for sensitivity analysis.

Secondly, train a linear regression model using D^{SA} and obtain the optimized regression coefficient $\hat{\boldsymbol{\theta}}^{\text{SA}}$ as discussed in Section 2.2 with $\widetilde{\mathbf{W}} = \mathbb{I}$ and $\mathbf{R} = \mathbb{I}$.

Finally, the sensitivity of each variable $\mathbf{s}$ can be obtained from $\hat{\boldsymbol{\theta}}^{\text{SA}}$ as

$$\mathbf{s} = [s_1, s_2, \cdots, s_p] \in \mathbb{R}^{p}, \text{ where } s_i = \left|\hat{\theta}_i^{\text{SA}}\right|. \tag{11}$$

3.3 Sparsity-targeted sampling

Aiming to construct a representative dataset, sparsity-targeted sampling (STS) sequentially samples the most sparse potential variable group based on the current training samples $\{\mathbf{x}\}^{\text{train}}$ and distance metric $d(\cdot,\cdot) : \mathbb{R}^{p+1} \times \mathbb{R}^{p+1} \to \mathbb{R}$. This guarantees that the resultant dataset offers adequately adjacent samples for the input samples in prediction tasks.

Here, the weighted Euclidean distance $d_s(\mathbf{x}_i, \mathbf{x}_j)$ based on the variable sensitivity $\mathbf{s}$ serves as the distance metric, which can be formulated as

$$\begin{aligned} d_s(\mathbf{x}_i, \mathbf{x}_j) &= \sqrt{\sum_{k=1}^{p} s_k \cdot (\mathbf{x}_{ik} - \mathbf{x}_{jk})^2} \\ &= \sqrt{(\mathbf{x}_i - \mathbf{x}_j)^{T} \operatorname{diag}(\mathbf{s})(\mathbf{x}_i - \mathbf{x}_j)}. \end{aligned} \tag{12}$$

Denoting current training sample set as $P = \{\mathbf{x}\}^{\text{train}}$ ($P = \varnothing$ in initialization), the workflow of STS to construct or supplement the training dataset can be represented as Algorithm 1.

The core idea of STS can be divided into two steps. Firstly,

Algorithm 1 Sparsity-targeted sampling (STS)

Intput: Variable constraint space S, current sample set P, sensitivity $\mathbf{s}$, candidate sample number n_1, and supplementary sample number n_2

Output: Supplemented sample set $\widehat{P}$

Initializtion: $\widehat{P} \leftarrow P$, $count \leftarrow 0$

1: **while** $count < n_2$ **do**

2: $C \leftarrow \text{RandomSample}(S, n_1)$ ▷ Randomly get n_1 samples from S as candidate sample set C

3: $\mathbf{x}^{\text{new}} \leftarrow \underset{\mathbf{x}_i \in C}{\operatorname{argmax}}\left(\min_{\mathbf{x}_j \in \widehat{P}} d_{\mathbf{s}}(\mathbf{x}_i, \mathbf{x}_j)\right)$ ▷ Select the sparsest sample from C compared to $\widehat{P}$

4: $\widehat{P} \leftarrow \widehat{P} \cup \{\mathbf{x}^{\text{new}}\}$ ▷ Add $\mathbf{x}^{\text{new}}$ to $\widehat{P}$

5: $count \leftarrow count + 1$

6: **end while**

a candidate set C of variable groups is obtained from the variable constraint space S through Monte Carlo uniform sampling. Secondly, STS selects the variable group from the candidate set C that has the most sparsity relative to the current sample set $\hat{P}$. This process is repeated until the specified number of supplementary samples is obtained. The process of STS selecting the next variable group $\mathbf{x}^{\text{new}}$ can be expressed as

$$\begin{aligned} d_{\min}(\mathbf{x}_i, \hat{P}, \mathbf{s}) &= \min_{\mathbf{x}_j \in \hat{P}} d_{\mathbf{s}}(\mathbf{x}_i, \mathbf{x}_j), \\ \mathbf{x}^{\text{new}} &= \operatorname*{argmax}_{\mathbf{x}_i \in C} \left(d_{\min}(\mathbf{x}_i, \hat{P}, \mathbf{s}) \right), \end{aligned} \tag{13}$$

where $d_{\min}$ is the minimum adjacent distance of the input sample compared to the training dataset. Eq. (13) first calculates the minimum adjacent distance $d_{\min}$ for each candidate sample. Then, the candidate sample with the maximum distance $d_{\min}$ is selected, meaning that this sample is the most sparse within the candidate sample set C relative to the current sample set $\hat{P}$.

The sampling process of STS is further elucidated in Figure 5. In the illustration, the solid circular line represents the variable constraint space S. Assuming the number of candidate samples n_1 is sufficiently large, the Monte Carlo sampling approach nearly covers all points. STS will first identify the sparsest sample #1, then consider the influence of the selected sample #1 to determine the next most sparse sample #2. This process continues to determine sample #3, #4, #5, and so on, until the specified number of samples n_2 is reached.

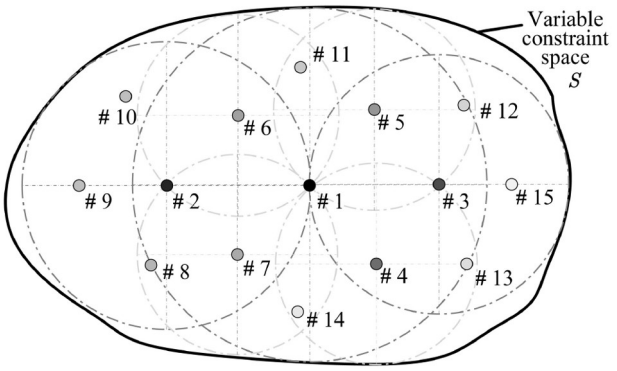

Figure 5 The process of sparsity-targeted sampling (STS).

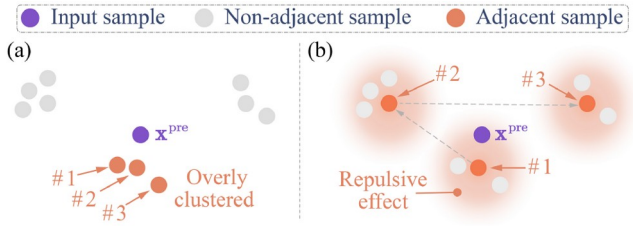

Figure 6 (Color online) Illustration of the comparison between traditional selection and sequential selection. (a) Traditional selection directly chooses k nearest samples as the adjacent samples; (b) sequential selection considers the repulsive effect of the already selected samples on subsequent sample selection.

3.4 Sequentially-selected local weighted linear regression

Traditional local methods directly select k nearest samples as the adjacent samples for model training. However, as illustrated in Figure 6(a), such methods tend to choose overly clustered adjacent samples, leading to singularity issues during subsequent local regression optimization, significantly impacting prediction accuracy.

To avoid the aforementioned issues, we consider the repulsive effect of the already selected samples on subsequent sample selection and propose a sequential selection method to determine adjacent samples. In this approach, denoting the set of already selected q adjacent samples as $\tilde{D}^q$, the selection criterion for the next adjacent sample $\tilde{\mathbf{x}}_{q+1}$ can be expressed as

$$\begin{aligned} \tilde{\mathbf{x}}_{q+1} &= \text{SeqSelect}(D^{\text{train}}, \tilde{D}^q, \mathbf{x}^{\text{pre}}, \mathbf{s}, \sigma, \beta) \\ &= \operatorname*{argmin}_{\mathbf{x}_i \in (D^{\text{train}} - \tilde{D}^q)} \left(d_{\mathbf{s}}(\mathbf{x}_i, \mathbf{x}^{\text{pre}}) + \beta \sum_{\tilde{\mathbf{x}}_j \in \tilde{D}^q} \frac{d_{\mathbf{s}}(\mathbf{x}_i, \tilde{\mathbf{x}}_j)}{\sigma^2} \right), \end{aligned} \tag{14}$$

where σ is the scale factor of the Gaussian kernel function controlling the range of the repulsive effect, and β is the repulsive factor controlling the strength of the repulsive effect. This selection criterion consists of two parts: the first part aims to make the candidate sample $\mathbf{x}_i$ as close as possible to the predicted sample $\mathbf{x}^{\text{pre}}$, while the second part considers the repulsive effect of the already selected samples $\tilde{\mathbf{x}}_j$ on $\mathbf{x}_i$.

The process of sequential selection method is illustrated in Figure 6(b). The sequential selection method firstly selects sample #1 that closest to the input predicted sample $\mathbf{x}^{\text{pre}}$. It then considers the repulsive effect of the selected sample #1 and chooses samples #2 and #3 that are relatively farther away. This leads to a broader span of adjacent samples, aiming to mitigate the singularity issues encountered in traditional local methods as elaborated in Figure 6(a).

After determining the adjacent samples $\tilde{D}^k$ through the sequential selection method, the complete SLWLR method can be summarized as shown in Algorithm 2. Its core idea can be divided into three steps: the first step is to sequentially determine adjacent samples $\tilde{D}^k$ by eq. (14). The second step is to assign different weights to $\tilde{D}^k$ base on their distances to $\mathbf{x}^{\text{pre}}$ by eq. (1). Finally, a weighted linear regression model is trained based on $\tilde{D}^k$ to estimate the peeling strength $\hat{y}^{\text{pre}}$ of $\mathbf{x}^{\text{pre}}$ as discussed in Section 2.2.

Algorithm 2 Sequentially-selected local weighted linear regression (SLWLR)

Intput: Training dataset D^{train}, input sample $\mathbf{x}^{\text{pre}}$, sensitivity $\mathbf{s}$, number of adjacent samples k, Gaussian scale factor σ, and repulsive factor β

Output: Predicted peeling strength $\hat{y}^{\text{pre}}$

Initializtion: $\widetilde{D}^k \leftarrow \varnothing$

1: **while** $|\widetilde{D}^k| < k$ **do** ▷ eq. (14)

2: $\tilde{\mathbf{x}} = \text{SeqSelect}(D^{\text{train}}, \widetilde{D}^k, \mathbf{x}^{\text{pre}}, \mathbf{s}, \sigma, \beta)$

3: $\widetilde{D}^k \leftarrow \widetilde{D}^k \cup \{\tilde{\mathbf{x}}\}$

4: **end while**

5: $\mathbf{X}^k, \mathbf{y}^k \leftarrow \widetilde{D}^k$

6: $\tilde{w}_i = \dfrac{w_i}{\sum_{\mathbf{x}_j \in \widetilde{D}^k} w_j}$, $w_i = \dfrac{1}{d_{\mathbf{s}}(\mathbf{x}_i, \mathbf{x}^{\text{pre}}) + \epsilon}$ ▷ eq. (1)

7: $\mathbf{r} = [0, r_1, r_2, \cdots, r_p]$, $r_i = \dfrac{1}{s_i + \epsilon}$ ▷ Assign regularization

8: $\hat{\boldsymbol{\theta}} = (\mathbf{X}^T \tilde{\mathbf{W}} \mathbf{X} + \alpha \mathbf{R})^{-1} \mathbf{X}^T \tilde{\mathbf{W}} \mathbf{y}$ ▷ eq. (6)

9: $\hat{y}^{\text{pre}} = (\mathbf{x}^{\text{pre}})^T \hat{\boldsymbol{\theta}}$ ▷ eq. (7)

3.5 Supplementary cost assessment

SLWLR searches the training dataset D^{train} for adjacent samples $\widetilde{D}^k$ at each prediction, making it a natural incremental learning model. In incremental learning, it only requires supplementing the current training dataset D^{train} with supplementary dataset D^{sup}.

Additionally, as a local method, the prediction of SLWLR is only influenced by adjacent samples. Therefore, increasing the sample numbers allows the input sample to find closer adjacent samples, and the prediction accuracy of SLWLR should certainly be improved.

Despite the predictive improvement brought by supplementation, acquiring data (i.e., time-consuming calculation of FEA) is quite expensive. One common issue is the assessment of supplementary cost, which quantitatively measures the relationship between model performance improvement and the number of required data samples.

To assess the supplementary cost, we use the minimum adjacent distance $d_{\min}$ defined in eq. (13) as a mediator. Specifically, as the sampling estimation depicted in Figure 7(a), the relationship between the potential minimum adjacent distance $\max(d_{\min})$ and the number of required samples n could be estimated during the sampling process. Then, the relationship between the error ratio e and the minimum adjacent distance $d_{\min}$ could be established during prediction process as the prediction experience depicted in Figure 7(b). Finally, by merging the above sampling estimation and prediction experience together, a quantitative assessment between the supplementary cost and the targeted accuracy could be formulated.

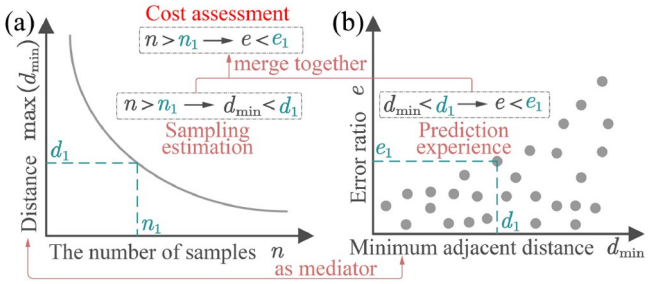

Figure 7 (Color online) Illustration of the cost assessment of supplementary dataset by using the minimum adjacent distance $d_{\min}$ as a mediator. (a) Sampling estimation depicts the relationship between the potential minimum adjacent distance $\max(d_{\min})$ and the number of required samples n; (b) prediction experience depicts the relationship between the error ratio e and the minimum adjacent distance $d_{\min}$.

4 Experiment

4.1 Finite element analysis and sensitivity analysis of slider track

To simulate the peeling process of the slider track, we constructed a model as depicted in Figure 8, using LS-DYNA, a commercial FEA software. This model comprises a 550 mm upper rail, a 600 mm lower rail, a locking module, two bearings with carriages, along with a loading beam and other auxiliary components.

The upper and lower rails of the slider track are discretized into tetrahedral meshes with dimensions of 3.5 mm × 3.5 mm. These meshes are configured as 5-point integration, plane-symmetric shell elements, facilitating a balance between simulation precision and computational efficiency. The material of the upper and lower rails is alloy steel, characterized by a density of 7.85 g/cm^3, Young's modulus of 210 GPa, and a Poisson's ratio of 0.3. Concerning contact modeling, a global contact friction coefficient of 0.1 and a global viscosity damping coefficient of 20 are specified. Fixed constraints are applied at the front and rear mounting points of the lower slider track. Displacement excitation is imposed along the direction of the loading beam, and LS-DYNA solver is invoked for explicit calculation.

The simulation is executed on a server platform equipped with two Intel Xeon E5-2680 v3 processors (48 cores in total), 64 GB of memory. The computation is time-consuming, requiring approximately 16 h with parallel computation utilizing 8 cores.

Numerous geometric variables have different influences on the peeling strength of the slider track. Key geometric variables, such as the relative position of the upper and lower rails and the loading direction, significantly affect track

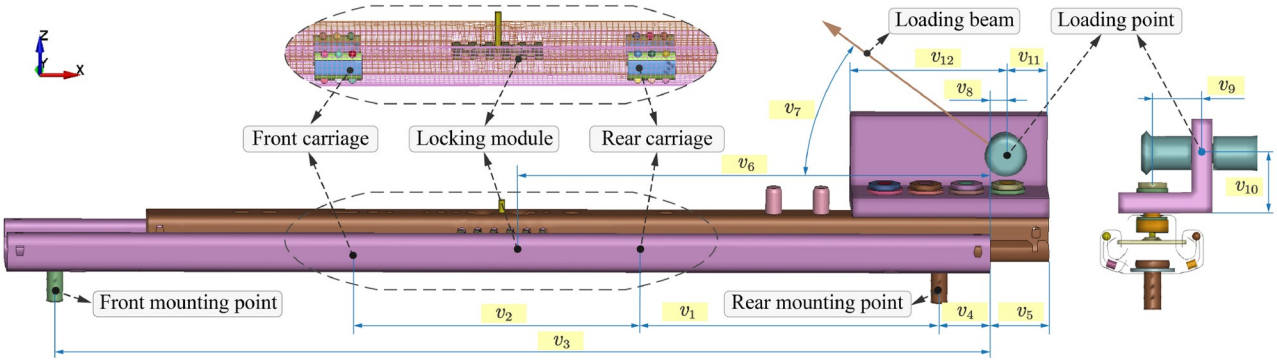

Figure 8 (Color online) The finite element model of the slider track and the geometric variables selected as the design variable group **x**.

peeling strength. Conversely, variables like the cage and mounting positions have less pronounced effects. Guided by the prior knowledge, 12 geometric variables of the slider track are identified as design variables as illustrated in Figure 8.

To validate the effectiveness of the slider track FEA model, a benchmark experiment is conducted, and the experimental result is compared with the FEA result. The geometric variables for both are consistent, and the experimental settings, along with the peeling strength curve, are illustrated in Figure 9. The peeling strength of the physical slider track is measured at 48.5 kN, while the corresponding FEA simulation yields a value of 50.6 kN. The resulting error ratio under these conditions is 4.3%. Furthermore, the peeling strength curves of them are closely aligned, affirming the reliability of the FEA simulation.

In the sensitivity analysis, 12 design variables of the slider track are considered as depicted in Figure 8. An orthogonal experimental design [17] is implemented to create the sensitivity analysis dataset D^{SA}. Specifically, we utilize an orthogonal experimental table $L_{27}(12^3)$ with 12 factors and 3 levels, yielding a total of 27 variable groups. The specific values are outlined in Table 1.

Taking the above 27 variable groups $L_{27}(12^3)$ as the input matrix $\mathbf{X}^{SA}$ and the corresponding peeling strength calculated by FEA as the output vector $\mathbf{y}^{SA}$, the sensitivity of each variable **s** can be obtained as Table 2 according to the sensitivity analysis proposed in Section 3.2 with specific regularization. The result highlights that variables v_5 (relative distance between upper and lower tracks), v_7 (loading direction), v_8 (loading position), and v_9 (eccentricity) significantly influence peeling strength. This observation aligns with our prior knowledge of the slider track.

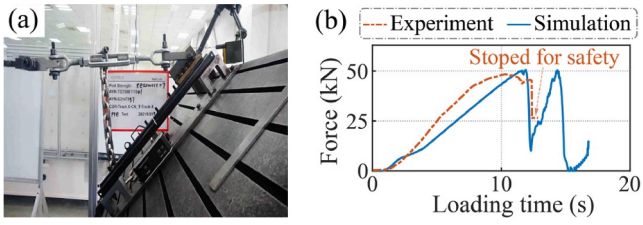

Figure 9 (Color online) Experimental rig and peeling strength curve. (a) Experimental rig of physical slider track; (b) peeling strength curve of the physical slider peeling experiment and its corresponding FEA simulation.

4.2 SLWLR result with basic dataset

After completing the sensitivity analysis, the STS method described in Section 3.3 is employed to sample 120 variable groups as the basic dataset D^{basic} for model training. Additionally, another 150 variable groups are randomly sampled as the test dataset D^{test}. Certainly, the construction of both datasets requires the calculation of the peeling strengths **y** corresponding to these variable groups **X** through the FEA of slider track.

As for the prediction evaluation, three common metrics

Table 1 The specific values of the orthogonal experimental table $L_{27}(12^3)$ used for sensitivity analysis

Variable	Value	Variable	Value
v_1 (mm)	150|170|190	v_7 (°)	15|45|75
v_2 (mm)	180|210|240	v_8 (mm)	80|–20|–120
v_3 (mm)	400|450|500	v_9 (mm)	10|15|20
v_4 (mm)	30|50|70	v_{10} (mm)	25|30|35
v_5 (mm)	v_8+40|140|210	v_{11} (mm)	20|25|30
v_6 (mm)	200|220|240	v_{12} (mm)	50|70|90

Table 2 The sensitivity of each variable calculated by the sensitivity analysis

Variable sensitivity	Value	Variable sensitivity	Value
v_1 (mm)	0	v_7 (°)	51.1
v_2 (mm)	40.1	v_8 (mm)	297.8
v_3 (mm)	24.4	v_9 (mm)	177.3
v_4 (mm)	17.2	v_{10} (mm)	4.6
v_5 (mm)	102.1	v_{11} (mm)	0
v_6 (mm)	5.8	v_{12} (mm)	21.4

[48] are used to measure the accuracy between the predicted peeling strength and FEA result. They are the coefficient of determination R^2, mean relative error (MRE), and mean absolute error (MAE), which are defined as

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2},$$
$$\mathrm{MRE} = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{y_i - \hat{y}_i}{y_i}\right|, \quad (15)$$
$$\mathrm{MAE} = \frac{1}{n}\sum_{i=1}^{n}\left|y_i - \hat{y}_i\right|,$$

where n is the number of test samples, y_i is the FEA result of the i-th test sample, $\hat{y}_i$ is the predicted peeling strength of the i-th test sample, and $\bar{y}$ is the mean of all FEA results.

Based on the basic dataset D^{basic} of 120 samples, the SLWLR model proposed in Section 3.4 is established with a parameter setting of $k = 15$. The predictive performance on the testing dataset comprising 150 samples is illustrated in Figure 10, which depicts the relationship between the prediction error ratio e and the minimum distance of its adjacent samples $d_{\min}$. Moreover, the corresponding values of R^2, MRE, and MAE are annotated above the graph.

From Figure 10, the maximum relative error is approximately 40%, and there is a clear positive correlation between the prediction error ratio e and the minimum adjacent distance $d_{\min}$. In other words, as the minimum adjacent distance $d_{\min}$ decreases, the maximum prediction error ratio e would also decrease. This phenomenon indicates the feasibility of improving model prediction accuracy by supplementing more dataset to reduce the minimum adjacent distance $d_{\min}$. It aligns with the inherent logic of data-driven approaches, suggesting that more data leads to better model performance. Overall, the predictive results based on the basic dataset D^{basic} are not ideal. Nevertheless, considering the positive correlation between e and $d_{\min}$, it is possible to enhance model prediction performance by supplementing additional dataset.

4.3 Cost assessment and SLWLR result with supplemented dataset

To conduct supplementary cost assessment and simultaneously validate the effectiveness of the STS method described in Section 3.3, random sampling (RS), Latin hypercube sampling (LHS), and STS are employed to estimate the relationship between the potential minimum adjacent distance $\max(d_{\min})$ and the number of samples n. It is crucial to highlight that, at this stage, only geometric parameter groups $\mathbf{X}$ are determined to assess the supplementary cost, and time-consuming FEA calculations for the corresponding peeling strengths $\mathbf{y}$ are not needed or performed.

As shown in Figure 11, the potential minimum adjacent distance $\max(d_{\min})$ for all three sampling methods exhibits an exponential decrease with the increase of the number of samples n. However, the descent rate of STS is significantly higher than that of RS and LHS. This is attributed to the capability of STS to actively identify and fill sparse regions, thereby obtaining smaller $\max(d_{\min})$ with fewer samples and constructing a more representative dataset. The above result thoroughly confirms the effectiveness of the proposed STS method in dataset formulation.

Using the minimum adjacent distance $d_{\min}$ as a mediator, we can establish an indirect relationship between the prediction error ratio e and the number of training samples n, achieving a quantitative assessment of the supplementary cost. The sampling estimation depicted in Figure 11 reveals that increasing n to 240 results in $\max(d_{\min})$ being approximately 182. The prediction experience in Figure 10 further indicates that when $d_{\min}$ is below 180, e can be controlled within 20%. Therefore, combining the sampling estimation and the prediction experience together, we decide to supplement additional 120 samples D^{sup} through STS into the basic dataset D^{basic} and perform FEA calculations to determine their corresponding peeling strengths. The supple-

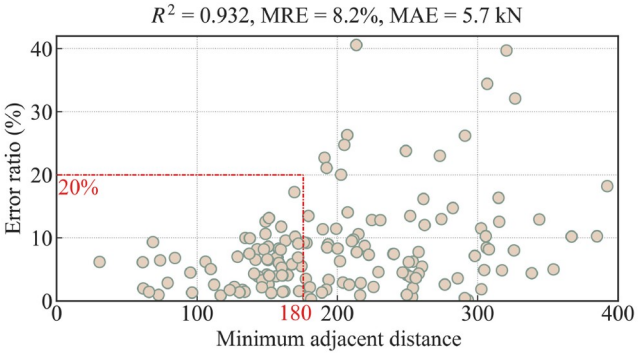

Figure 10 (Color online) The predictive performance of SLWLR trained on the basic dataset comprising 120 samples.

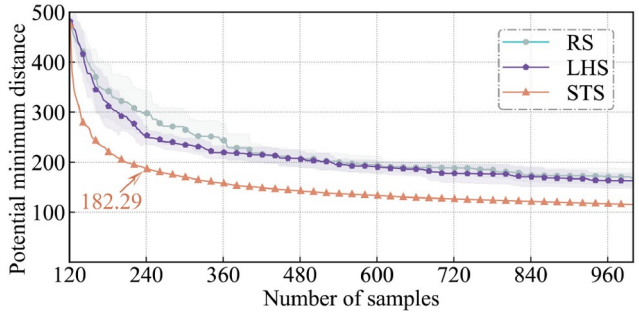

Figure 11 (Color online) The relationship between the potential minimum adjacent distance $\max(d_{\min})$ and the number of samples n under three different sampling strategies: random sampling (RS), Latin hypercube sampling (LHS), and sparsity-targeted sampling (STS).

mented dataset D^{train} comprises a total of 240 operating condition samples.

With the supplemented dataset D^{train} of 240 training samples and the experimental settings consistent with Section 4.2, the predictive performance of SLWLR on the testing dataset comprising 150 samples is illustrated in Figure 12. The minimum adjacent distances $d_{\min}$ of all test data are within 180 and the maximum relative error is below 20%. The coefficient of determination R^2, mean relative error (MRE), and mean absolute error (MAE) all surpass the corresponding metrics in Figure 10. This improvement underscores that supplementing additional dataset significantly enhances the prediction accuracy of SLWLR. In conclusion, after supplementing 120 additional samples D^{sup}, the minimum adjacent distance $d_{\min}$ aligns with the previous supplementary cost assessment, and the predictive performance of SLWLR has been effectively elevated.

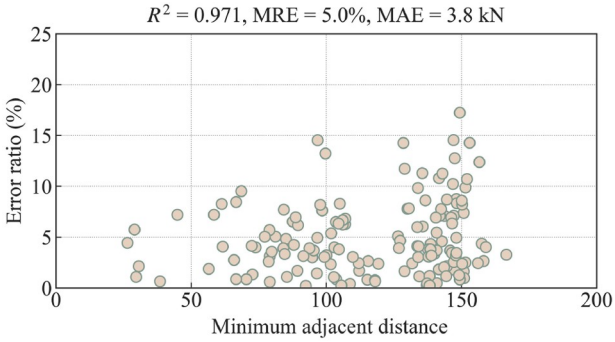

Figure 12 (Color online) The predictive performance of SLWLR trained on the full dataset comprising 240 samples.

5 Discussion

5.1 Ablation study

SLWLR incorporates a sequential selection of adjacent samples in addition to the distance-based weighting mechanism of LWLR, aiming to address the robustness issue associated with the number of adjacent samples k. Here, ablation study is conducted to validate the effectiveness of the distance-based weighting mechanism and the proposed sequential selection.

The experimental setup remains consistent with the previous section, where the full training dataset D^{train} comprises 240 samples, and the test dataset D^{test} consists of 150 samples. This ablation study compares the predictive performance of four local methods under different adjacent sample numbers k: local linear regression (LLR) [33], sequentially-selected local linear regression (SLLR), LWLR, and the proposed SLWLR method. The results are presented in Figure 13, and the conclusions drawn are as follows.

(1) **The effect of distance-based weighting mechanism**: LWLR, incorporating the distance-based weighting mechanism, outperforms LLR in terms of prediction results and is less affected by the increase of adjacent sample number k. It indicates that the distance-based weighting mechanism effectively utilizes information from nearby samples, avoiding interference from distant samples in predicting results. This conclusion similarly applies to the comparison between SLWLR and SLLR. Experimental results demonstrate that the distance-based weighting mechanism significantly enhances model prediction performance and mitigates the impact of the adjacent sample number k especially when k is large.

(2) **The effect of sequential selection of adjacent samples:** With sequential selection, SLLR proves effective in selecting appropriate samples, resulting in better predictive results than LLR when the number of k is small. However, as k increases, the sequential selection method and traditional method gradually converge in the selection of adjacent samples, causing the predictive performance of SLLR and LLR to approach each other. This phenomenon also holds true for LWLR and SLWLR. In conclusion, experimental results indicate that the selection of adjacent samples is particularly effective in improving the predictive performance when the adjacent sample number k is small.

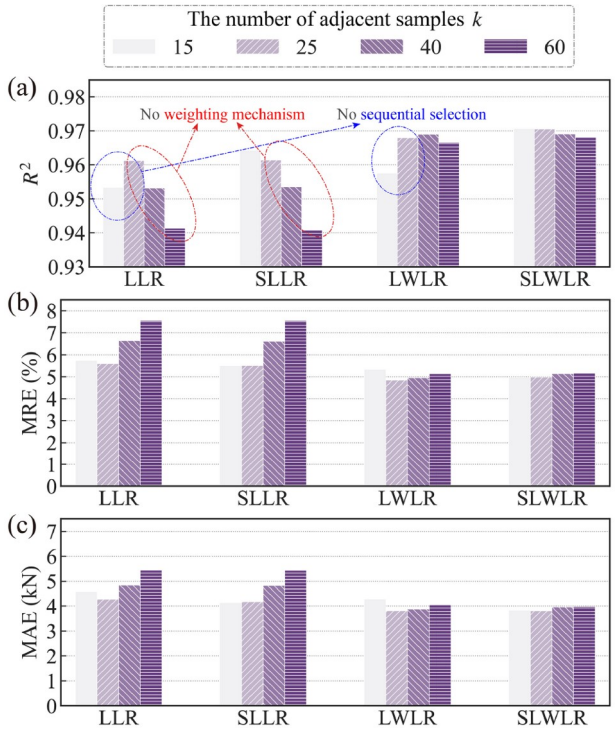

Figure 13 (Color online) The performance of different local methods trained on the full dataset comprising 240 samples. (a) R^2; (b) MRE; (c) MAE.

(3) **Superiority of SLWLR:** SLWLR, incorporating both distance-based weighting and sequential selection together, adeptly avoids the issue of overly clustered adjacent samples

when k is small through the sequential selection approach, and mitigates the negative impact of distant samples simultaneously when k is large through distance-based weighting. SLWLR consistently achieves the best predictive performance among the four methods, demonstrating robustness in the selection of the adjacent sample number k, which emphasizes the superiority of the proposed SLWLR method.

5.2 Comparison with other methods

To validate the superiority of the proposed SLWLR method, it is compared with current popular surrogate model construction methods in track peeling strength prediction task. The methods for comparison include as follows.

(1) **K-nearest neighbor (kNN)** [36]: kNN is a non-parametric method that directly uses the nearest k samples to predict the peeling strength of the input sample. The number of adjacent samples k is set to 5 and the distance metric is the same as that used in SLWLR.

(2) **Support vector regression (SVR)** [32,33]: SVR is a kernel-based regression method that uses the kernel function to map the input sample to a high-dimensional feature space and then performs linear regression in the feature space. We choose Gaussian kernel with the same distance metric used in SLWLR.

(3) **Radical basis function (RBF)** [33,49,50]: RBF is a typical interpolation method but can also be applied to prediction tasks. After parameter optimization, RBF employs a Gaussian kernel with a scaling factor σ set to 800, and the distance metric is the same as that used in SLWLR.

(4) **Multi-layer perceptron (MLP)** [49]: MLP is a basic neural network that usually performs well in simple scenarios with a small number of samples. After parameter optimization, a 2-layer MLP is selected with the number of hidden neurons set to 8, the activation function set to *Sigmoid*, and Bayesian regularization employed for model optimization.

(5) **Kriging** [33,50]: Kriging is an interpolating model that is a linear combination of a known function which is added to a realization of a stochastic process. The kernel function chosen is the squared exponential kernel, and the optimizer selected is LBFGS.

The above five methods and the proposed SLWLR method are trained on the full training dataset D^{train} comprises 240 operating conditions, and their perdictive performance on the test dataset D^{test} is shown in Figure 14. From the results, it can be observed that kNN performs the worst, while SVR and RBF show significant improvement. MLP and Kriging perform better than the first three methods, and the proposed SLWLR method surpasses all other methods in terms of accuracy metrics (i.e., R^2, MRE, and MAE) and aligns more closely with the ground truth (i.e., the peeling strength of FEA).

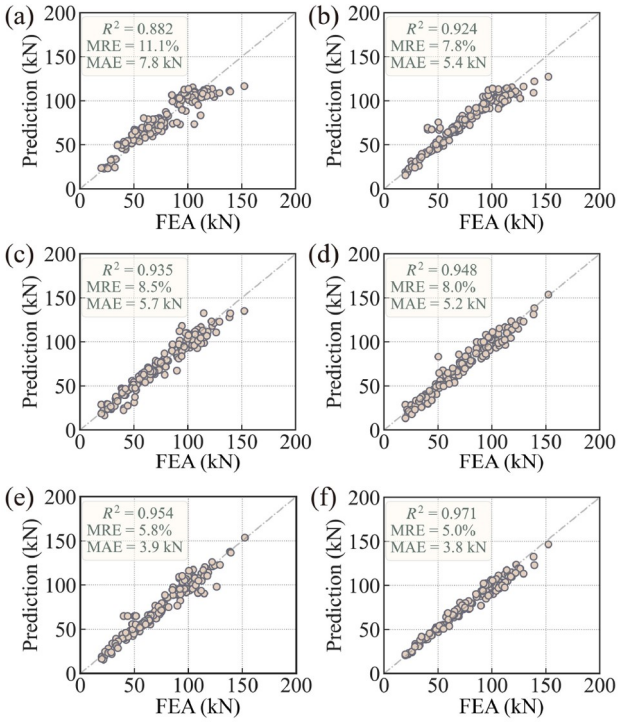

Figure 14 (Color online) The performance of different methods trained on the full dataset comprising 240 samples. (a) kNN; (b) SVR; (c) RBF; (d) MLP; (e) Kriging; (f) SLWLR.

Additionally, we explore the predictive performance of these methods under different numbers of training samples n as shown in Figure 15. The results indicate that the predictive performance of all methods gradually converges as the number of training samples increases (i.e., converges when $k > 150$), while the performance of SLWLR consistently stands out as superior and more robust than other methods. Notably, the performance of SVR, Kriging, and MLP fluctuates when n is small, and even degradates when n = 60, 95, 90, respectively. This underscores that the proposed SLWLR method exhibits greater robustness to the number of training samples n than other five methods, and have superior performance especially when the training samples are limited.

In conclusion, slider peeling strength prediction is a problem with multiple factors, high coupling, and complex nonlinearity, and the cost of data acquisition limits the richness of the dataset, making it challenging for other regression methods to achieve the outstanding predictive performance as the proposed SLWLR method does. Therefore, SLWLR outperforms other regression methods in terms of predictive performance metrics and data quantity requirements.

5.3 Effectiveness of STS

In this subsection, an experiment is conducted to validate the

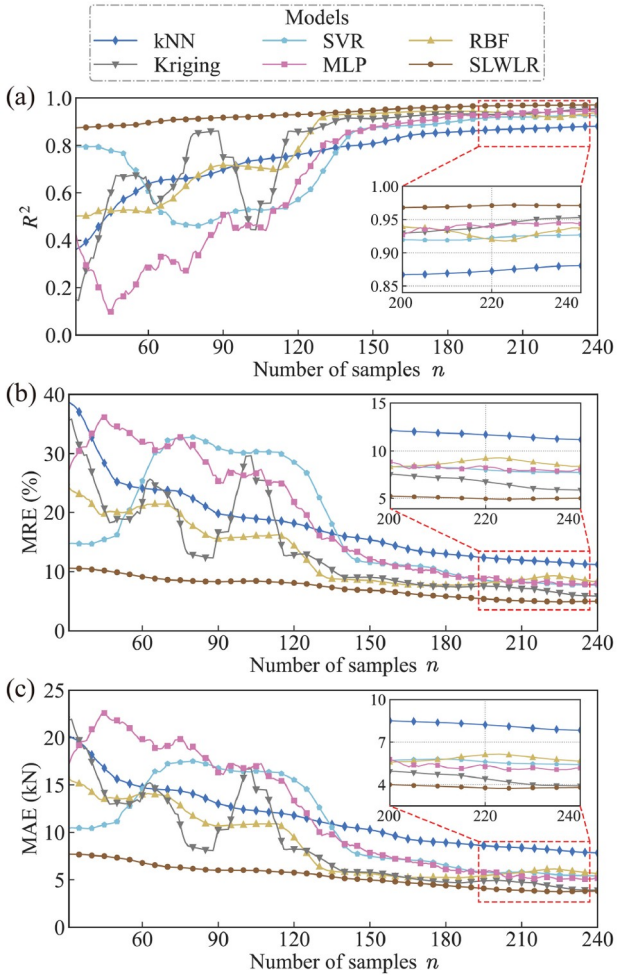

Figure 15 (Color online) The performance of different local methods under different numbers of training samples n. (a) R^2; (b) MRE; (c) MAE.

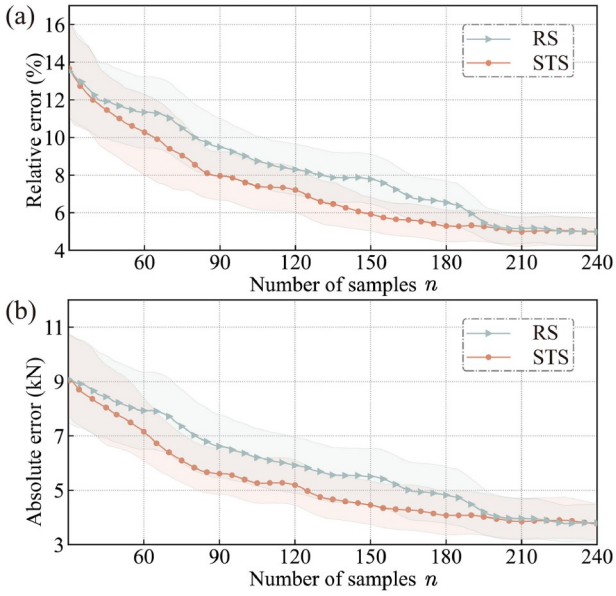

Figure 16 (Color online) The performance of SLWLR with sparsity-targeted sampling (STS) and random sampling (RS) in different number of training samples, where the data points and fillings represent the mean and standard deviation of the predicted results, respectively. Smaller error and error ratio indicate better predictive performance of the model. (a) Relative error; (b) absolute error.

effectiveness of the proposed sparse-targeted sampling (STS) method in dataset formulation. Considering the prohibitive cost of data acquisition (i.e., 16 h for a single simulation), an indirect comparison is chosen. Specifically, this experiment initially selects 30 samples from the complete dataset of 240 samples. Then, STS and random sampling (RS) are independently employed to sequentially select the remaining 210 samples. The performance of the proposed SLWLR model with the training dataset obtained from these two sampling methods is shown in Figure 16. If the prediction performance of STS surpasses that of RS, it can be concluded that the proposed STS method is effective in dataset formulation.

The relationship between the prediction performance of SLWLR and the number of training samples n under RS and STS is illustrated in Figure 16. At the beginning (n = 30) or end (n = 240), there is no significant difference between these two sampling methods in dataset formulation because this is an indirect comparison, and their corresponding prediction performances are consistent. However, notably, when n is from 60 to 180, STS outperforms RS in terms of relative error and absolute error obviously, i.e., STS exhibits lower error metrics than RS. In summary, although this direct comparison was not feasible due to the cost of data acquisition, the indirect comparative experiment provides ample evidence of the effectiveness of the proposed STS method in dataset formulation.

5.4 Physical experiment verification

To validate the predictive performance of the proposed framework in the real-world, we compare the results of our SLWLR model (trained on 240 FEA samples) with FEA simulation and the physical slider peeling experiment. The peeling strengths obtained from these three methods under three typical working conditions–retraction (Rear), centering (Middle), and overhanging (Overhang)–are presented in Table 3. The results indicate that the peeling strength error obtained by SLWLR is within 2 kN, with a relative error of within 4%. This effectively confirms the feasibility and effectiveness of the proposed framework in predicting the peeling strength of slider tracks.

6 Conclusions

To overcome the high cost, time-consuming nature, and insufficiently comprehensive assessment involved in obtaining slider track peeling strength, we propose a systematic sur-

Table 3 The peeling strengths of physical experiment, FEA simulation model, and the proposed SLWLR method under three typical working conditions: retraction (Rear), centering (Middle), and overhanging (Overhang)

Method	Rear (kN)	Middle (kN)	Overhang (kN)
Physic	49.3	48.5	40.2
FEA	50.23	50.57	40.94
SLWLR	51.40	51.14	39.53

rogate model construction framework encompassing sensitivity analysis, dataset sampling, and model prediction. Through a series of experiments, we validate the effectiveness of the proposed method and draw the following conclusions.

(1) SLWLR, through sequential selection of adjacent samples and distance-based weighting mechanism, effectively overcomes the sensitivity issue to the localization (i.e., k), and its prediction performance is significantly better than similar local methods.

(2) Slider track peeling strength prediction is a complex, multi-factor, highly coupled, and nonlinear problem. Considering the data limitation due to exorbitant computational cost, the SLWLR method outperforms other surrogate model methods in terms of prediction accuracy and data quantity requirements.

(3) Considering data acquisition costs, the effectiveness of the proposed STS method in dataset construction is validated through sampling estimation and indirect comparison.

(4) Finally, the established surrogate model demonstrates feasibility and effectiveness in the slider track peeling strength prediction task. In the simulated test dataset of 150 randomly samples, the average absolute error is 3.3 kN, the average ralative error is 4.3%, and in three physical test scenarios, the absolute error is within 2 kN, and the ralative error is within 4%.

This work was supported by the National Natural Science Foundation of China (Grant Nos. 12272219 and 12121002).

1 Zhang Z, Jin K, Li F, et al. Effects of adjustment devices on the fore-and-aft mode of an automobile seat system: Headrest, height adjuster, recliner and track slide. Proc Institution Mech Engineers Part D-J Automobile Eng, 2016, 230: 1140–1152
2 Ministry of Industry and Information Technology. GB 15083-2019: Strength requirement and test method of automobile seats, their anchorages and any head restraints. China: https://www.chinesestandard.net, 2019
3 Kongwat S, Homsnit T, Padungtree C, et al. Safety assessment and crash compatibility of heavy quadricycle under frontal impact collisions. Sustainability, 2022, 14: 13458
4 Lopes R, Ramos N V, Cunha R, et al. Coach crashworthiness and failure analysis during a frontal impact. Eng Fail Anal, 2023, 151: 107369
5 Zhang J, Wu Q, Wang G Y, et al. Numerical analysis on propulsive efficiency and predeformated optimization of a composite marine propeller. Sci China Tech Sci, 2020, 63: 2562–2574
6 Zhang X M, Yu X D, Chen J X, et al. Vibration properties and transverse shear characteristics of multibody molded beetle elytron plates. Sci China Tech Sci, 2022, 63: 2584–2592
7 Kong C, Xiao M, Yuan Q T. Implicit nonlinear FEM for steel sets in tunnels. Sci China Tech Sci, 2023, 66: 771–783
8 Liu F, Shang J, Luo Z, et al. Design of space sliding rails and mechanical analysis. In: Proceedings of International Conference on Mechatronics, Materials and Manufacturing (ICMMM). Chengdu, 2014. 207–212
9 Kudela J, Matousek R. Recent advances and applications of surrogate models for finite element method computations: A review. Soft Comput, 2022, 26: 13709–13733
10 Li K, Fu T, Zhang T, et al. CMS: A novel surrogate model with hierarchical structure based on correlation mapping. Eng Comput, 2022, 38: 4589–4604
11 Wang Y, Li K, Li Q, et al. Multi-fidelity information fusion with hierarchical surrogate guided by feature mapping. Knowledge-Based Syst, 2023, 275: 110693
12 Alizadeh R, Allen J K, Mistree F. Managing computational complexity using surrogate models: A critical review. Res Eng Des, 2020, 31: 275–298
13 Xu K K, Meng H. Analyses of surrogate models for calculating thermophysical properties of aviation kerosene RP-3 at supercritical pressures. Sci China Tech Sci, 2015, 58: 510–518
14 Novák L, Vořechovský M, Sadílek V, et al. Variance-based adaptive sequential sampling for polynomial chaos expansion. Comput Meth Appl Mech Eng, 2021, 386: 114105
15 McKay M D, Beckman R J, Conover W J. A comparison of three methods for selecting values of input variables in the analysis of output from a computer code. Technometrics, 1979, 21: 239
16 Bogoclu C, Roos D, Nestorović T. Local Latin hypercube refinement for multi-objective design uncertainty optimization. Appl Soft Computing, 2021, 112: 107807
17 Deming S N, Morgan S L. Experimental Design: A Chemometric Approach. 2nd ed. Elsevier, 1993
18 Lämmle S, Bogoclu C, Cremanns K, et al. Gradient and uncertainty enhanced sequential sampling for global fit. Comput Methods Appl Mech Eng, 2023, 415: 116226
19 Cai X, Ruan G, Yuan B, et al. Complementary surrogate-assisted differential evolution algorithm for expensive multi-objective problems under a limited computational budget. Inf Sci, 2023, 632: 791–814
20 Viana F A C, Gogu C, Goel T. Surrogate modeling: Tricks that endured the test of time and some recent developments. Struct Multidisc Optim, 2021, 64: 2881–2908
21 Zhai Z, Li H, Wang X. An adaptive sampling method for Kriging surrogate model with multiple outputs. Eng Comput, 2020, 38: 277–295
22 Hu C, Zeng S, Li C. A framework of global exploration and local exploitation using surrogates for expensive optimization. Knowledge-Based Syst, 2023, 280: 111018
23 Xu Y, Renteria A, Wang P. Adaptive surrogate models with partially observed information. Reliability Eng Syst Saf, 2022, 225: 108566
24 Garud S S, Karimi I A, Kraft M. Smart sampling algorithm for surrogate model development. Comput Chem Eng, 2017, 96: 103–114
25 Eason J, Cremaschi S. Adaptive sequential sampling for surrogate model generation with artificial neural networks. Comput Chem Eng, 2014, 68: 220–232
26 Chen Q, Dong X, Tu G, et al. TFN: An interpretable neural network with time-frequency transform embedded for intelligent fault diagnosis. Mech Syst Signal Processing, 2024, 207: 110952
27 Zhang S, Li K, Wang S, et al. Recursive surrogate model based on generalized regression neural network. Appl Soft Computing, 2023, 145: 110576
28 Sun Y, Sengupta U, Juniper M. Physics-informed deep learning for

simultaneous surrogate modeling and PDE-constrained optimization of an airfoil geometry. Comput Methods Appl Mech Eng, 2023, 411: 116042
29 Conti P, Guo M, Manzoni A, et al. Multi-fidelity surrogate modeling using long short-term memory networks. Comput Methods Appl Mech Eng, 2023, 404: 115811
30 Box G E, Draper N R. Empirical Model-Building and Response Surfaces. John Wiley & Sons, 1987
31 Deaton J D, Grandhi R V. A survey of structural and multidisciplinary continuum topology optimization: Post 2000. Struct Multidiscip Optim, 2014, 49: 1–38
32 Shi M, Lv L, Sun W, et al. A multi-fidelity surrogate model based on support vector regression. Struct Multidisc Optim, 2020, 61: 2363–2375
33 Xu G, Wei H, Wang J, et al. A local weighted linear regression (LWLR) ensemble of surrogate models based on stacking strategy: Application to hydrodynamic response prediction for submerged floating tunnel (SFT). Appl Ocean Res, 2022, 125: 103228
34 Zhang J, Chung C Y, Han Y. Online damping ratio prediction using locally weighted linear regression. IEEE Trans Power Syst, 2016, 31: 1954–1962
35 Wu J, Fang L C, Dong G Z, et al. State of health estimation for lithium-ion batteries in real-world electric vehicles. Sci China Tech Sci, 2023, 66: 47–56
36 Huang M, Lin R, Huang S, et al. A novel approach for precipitation forecast via improved K-nearest neighbor algorithm. Adv Eng Inf, 2017, 33: 89–95
37 Geler Z, Kurbalija V, Radovanović M, et al. Comparison of different weighting schemes for the kNN classifier on time-series data. Knowl Inf Syst, 2016, 48: 331–378
38 Zhang S, Li X, Zong M, et al. Efficient kNN classification with different numbers of nearest neighbors. IEEE Trans Neural Netw Learn Syst, 2018, 29: 1774–1785
39 Zhang S, Li X, Zong M, et al. Learning *k* for kNN classification. ACM Trans Intell Syst Technol, 2017, 8: 43
40 Ye H J, Zhan D C, Jiang Y. Fast generalization rates for distance metric learning. Mach Learn, 2019, 108: 267–295
41 Rodrigues É O. Combining Minkowski and Chebyshev: New distance proposal and survey of distance metrics using k-nearest neighbours classifier. Pattern Recognition Lett, 2018, 110: 66–71
42 Ahmadianfar I, Jamei M, Chu X. A novel hybrid wavelet-locally weighted linear regression (W-LWLR) model for electrical conductivity (EC) prediction in surface water. J Contam Hydrol, 2020, 232: 103641
43 Elbeltagi A, Salam R, Pal S C, et al. Groundwater level estimation in northern region of Bangladesh using hybrid locally weighted linear regression and Gaussian process regression modeling. Theor Appl Climatol, 2022, 149: 131–151
44 Kisi O, Ozkan C. A new approach for modeling sediment-discharge relationship: Local weighted linear regression. Water Resour Manage, 2016, 31: 1–23
45 Dong X, Chen J, Zhang K, et al. Privacy-preserving locally weighted linear regression over encrypted millions of data. IEEE Access, 2020, 8: 2247–2257
46 Yu X, Lin J, Jiang F, et al. A cross-domain collaborative filtering algorithm based on feature construction and locally weighted linear regression. Comput Intell Neurosci, 2018, 2018: 1–12
47 Zhang W, Zhao S, Pan H, et al. A locally weighted linear regression look-up table-based iterative reconstruction method for dual spectral CT. IEEE Trans Biomed Eng, 2022, 70: 3028–3039
48 Hastie T, Tibshirani R, Friedman J H, et al. The Elements of Statistical Learning: Data Mining, Inference, and Prediction. Springer, 2009
49 Yilmaz I ı, Kaynar O. Multiple regression, ANN (RBF, MLP) and ANFIS models for prediction of swell potential of clayey soils. Expert Syst Appl, 2011, 38: 5958–5966
50 Bouhlel M A, Hwang J T, Bartoli N, et al. A Python surrogate modeling framework with derivatives. Adv Eng Software, 2019, 135: 102662