

Research Article

Step-by-Step Penalized Blind Kriging Methods for Surrogate Modeling

Ziheng Feng ¹, Min Li ¹, Xiao Wang ¹, Jianxin Zhao,² and Xinmin Li ¹

¹School of Mathematics and Statistics, Qingdao University, Qingdao 266 071, China

²Basic Courses Department, Navy Submarine Academy, Qingdao, China

Correspondence should be addressed to Xinmin Li; xmli@qdu.edu.cn

Received 17 May 2021; Revised 23 July 2021; Accepted 24 December 2021; Published 21 January 2022

Academic Editor: Suzanne M. Shontz

Copyright © 2022 Ziheng Feng et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

When building a surrogate model, it is important to identify a proper mean function for the kriging model. The commonly used variable selection method is the penalized blind kriging (PBK) method. But this method could lead to a low time efficiency, which is not suitable for experiments with time-sensitive data. In this paper, we propose three step-by-step approaches for constructing an appropriate mean function to improve the prediction accuracy and time efficiency of the PBK method. Several functions and two engineering examples are used to prove the effectiveness of the proposed methods. From simulation results, we can see that Method 1 (M1) and Method 2 (M2) have been significantly improved in both the prediction accuracy and the time efficiency compared with PBK. Especially, in the Test function, compared with the traditional PBK method, the prediction accuracy of M2 is improved by 69.08% and 26.13%, respectively, under the penalty of Lasso and Elastic Net, and the time efficiency of M1 is improved by 85.15% and 90.33%, respectively, under the penalty of Lasso and Elastic Net. In addition, Method 3 (M3) has been significantly improved in prediction accuracy compared with PBK.

1. Introduction

The surrogate model has been widely used to replace computation-intensive engineering simulation models or black-box systems [1–3]. There are several commonly used surrogate models, such as kriging, support vector regression, and radial basis functions [4]. Their application fields involve the structural dynamics of aeroengine casings, dynamic reliability estimation of turbine blisk, and computationally expensive constrained optimization problems.

The kriging model can be used to fit computer experiments because of its interpolate character. Its error estimation can be extended to many aspects such as failure probability estimation [5], robust design optimization [6], uncertainty analysis [7], and global sequential sampling [8]. As the combination of mean function and stochastic Gaussian process (GP), the kriging model is called ordinary kriging (OK) if using a constant mean to fit the overall trend or universal kriging (UK) if supposing some specified variables in the mean function. However, their prediction accuracy will be reduced if the mean function is specified

inappropriately [9]. Computer experiments usually include a large number of input variables, inactive variables which hardly have an impact on the response tending to reduce prediction accuracy; thus, it is a matter to identify active variables into a mean function [10].

In this paper, we consider the problem of variable selection for the kriging model. We want to use the method of variable selection to construct an interpretable kriging model which can select the active variables and remove inactive variables with high prediction accuracy and high time efficiency. The commonly used variable selection method in the kriging model is the penalty likelihood method. For the penalty likelihood method, Hung [11] proposed the penalized blind kriging (PBK) approach to select variables. This method, however, suffers from the limitation of low time efficiency and prediction accuracy. In the field of engineering, because of the time sensitive of data, researchers always want to improve time efficiency greatly while maintaining preferable prediction accuracy. Therefore, efficient variable selection methods are called for.

The idea here is to improve the time efficiency and prediction accuracy of PBK using step-by-step approaches. Three new approaches are proposed for variable selection. The first method is to select the variables using a linear model at the first and then estimate the parameters in the kriging model. The second method is first to estimate the parameters with the OK model and then select the variables via the penalized kriging model and finally refit the kriging model. The third method is an improvement of PBK, which is a refitting based on the variable selection of PBK. These methods are proved to improve the prediction accuracy and time efficiency by several analytic functions and two examples.

The remainder of the paper is organized as follows: in Section 2, three new methods are proposed for variable selection. Section 3 illustrates the effectiveness of proposed methods with four simulated examples and compares the prediction accuracy, identification, and time efficiency. Section 4 implements the proposed methods for two engineering examples. Section 5 provides some discussion and concluding remarks.

2. Step-by-Step Variable Selection Methods

In this section, we first review some of the details; then, three methods are proposed for variable selection based on PBK methods.

2.1. Kriging Preliminaries. The kriging model [12] assumes that the response $y(x)$ is a realization from the stochastic process:

$$y(x) = \sum_{i=1}^p \beta_i f_i(x) + z(x) = f(x)^T \beta + z(x), \quad (1)$$

where $f(x) = (f_1(x), \dots, f_p(x))^T$ is the basis functions, and $\beta = (\beta_1, \dots, \beta_p)^T$ is a vector of unknown coefficients, respectively. The stochastic process $z(x)$ has mean 0 and covariance $\sigma^2 R$, where $R(\theta, x_i, x_j)$ is a correlation function between x_i and x_j depending on parameter θ . In most engineering applications, the most commonly used correlation function is the stationary Gaussian correlation function $R_{ij}(\theta) = R(\theta, x_i, x_j) = \prod_{k=1}^d \exp(-\theta_k (x_{ik} - x_{jk})^2)$, which is chosen in our research.

Given some sampled points $x_i \in R^d$ and the responses $y_i \in R, i = 1, \dots, n$, the best linear unbiased predictor (BLUP) can be obtained as follows [12]:

$$\hat{y}(x) = f(x)^T \hat{\beta} + r(x)^T R^{-1} (y - F \hat{\beta}), \quad (2)$$

where $r = [R(\theta, x_1, x), \dots, R(\theta, x_n, x)]^T$ is a correlation between the unsampled point x and the sampled points $x_i, i = 1, \dots, n$; the matrix $F = [f(x_1), \dots, f(x_n)]^T$ is constructed by $f(x)$ at the sampled points.

The log-likelihood function can be written as follows:

$$L(\beta, \sigma^2, \theta | y) = -\frac{n}{2} \log(2\pi) - \frac{1}{2} \log|\sigma^2 R| - \frac{1}{2\sigma^2} (y - F\beta)^T R^{-1} (y - F\beta). \quad (3)$$

Then, the parameters β , σ^2 , and θ could be estimated via maximum likelihood estimation (MLE). For given correlation parameters θ , the MLE of β is as follows:

$$\hat{\beta} = (F^T R^{-1} F)^{-1} F^T R^{-1} y, \quad (4)$$

and the MLE of σ^2 is as follows:

$$\hat{\sigma}^2 = \frac{1}{n} (y - F \hat{\beta})^T R^{-1} (y - F \hat{\beta}). \quad (5)$$

Since there is no analytical solution for the parameters θ , optimization methods are needed to obtain as follows:

$$\hat{\theta} = \operatorname{argmin}_{\theta} (n \log \hat{\sigma}^2 + \log|R|). \quad (6)$$

Finally, the estimated parameters $\hat{\beta}, \hat{\sigma}^2, \hat{\theta}$ are substituted into equation (2) to obtain the predictor of responses for the unsampled points.

2.2. Step-by-Step Variable Selection Methods. Although the kriging model has a character of interpolator due to its stochastic process, penalty methods still could be used to select the important variables from the candidates to improve prediction accuracy [13].

Two penalty functions are used for variable selection in the PBK model by Hung [11], i.e., Lasso [14] and adaptive Lasso [15]. It estimates the regression coefficients β by minimizing the negative penalized log-likelihood function.

$$Q(\beta, \theta, \sigma^2) = \frac{(y - F\beta)^T R^{-1} (y - F\beta)}{2\sigma^2} + \sum_{j=1}^p P_{\lambda}(|\beta_j|), \quad (7)$$

where $P_{\lambda}(|\beta_j|)$ is the regularization term called ‘‘penalty function,’’ and λ is the regularization parameter. In order to efficiently obtain the estimation of parameters in the penalty likelihood function, Hung [11] proposed the iteratively reweighted least angle regression (IRLARS) algorithm for Lasso penalty and adaptive Lasso penalty.

Besides above penalty methods, Elastic Net penalty proposed by Zou and Hastie [16] also could be used for variable selection. It has regularization parameters λ and α . The Elastic Net estimation of the parameters β can be solved by the following formulation:

$$\hat{\beta}_{EN} = \operatorname{argmin}_{\beta} \left(\frac{(y - F\beta)^T R^{-1} (y - F\beta)}{2\sigma^2} + \lambda \alpha \|\beta\|_1 + \lambda (1 - \alpha) \|\beta\|_2^2 \right), \quad (8)$$

where $\|\beta\|_1 = \sum_{j=1}^p |\beta_j|$ and $\|\beta\|_2^2 = \sum_{j=1}^p \beta_j^2$. Thus, it can be seen that the penalty function of Elastic Net is a convex combination of ridge regression and Lasso. Particularly, the Elastic Net penalty becomes ridge regression when $\alpha = 0$,

and the Elastic Net penalty becomes Lasso penalty when $\alpha = 1$. According to the IRLARS algorithm for Lasso penalty, we describe the IRLARS algorithm for Elastic Net penalty in Algorithm 1.

Regularization parameters are obtained based on 10-fold cross validation, and the estimation of parameters β, θ, σ^2 are obtained based on the IRLARS algorithm in this research.

Although PBK effectively reduces the prediction error, further improvements are needed for low time efficiency. Thus, we propose three step-by-step variable selection methods to improve the prediction accuracy and time efficiency. We describe them in detail as follows:

2.2.1. Method One (M1). We integrate penalty into the linear model $y(x) = f(x)^T \beta + \varepsilon$ firstly, where $\varepsilon \sim N(0, \sigma^2)$; then, we select active variables by solving the penalized likelihood function for β :

$$\hat{\beta} = \operatorname{argmin}_{\beta} \left(\|y - F\beta\|^2 + \sum_{j=1}^p P_{\lambda}(|\beta_j|) \right), \quad (9)$$

where λ is obtained based on cross validation. Then, selected basis functions $f'_i(x), i = 1, \dots, m$ corresponding nonzero regression coefficients β are substituted into the Kriging model (1) as new variables, and correlation parameters θ are updated as θ' . $\beta', \sigma'^2, \theta'$ are estimated by formulas (4)–(6), respectively. Finally, the estimated parameters $\hat{\beta}', \hat{\sigma}'^2, \hat{\theta}'$ are used to predict the responses of unsampled points by substituting into formula (2), and the predicted model can be expressed as follows:

$$\hat{y}'(x) = f'(x)^T \hat{\beta}' + \hat{r}'(x)^T \hat{R}'^{-1} (y - F' \hat{\beta}'). \quad (10)$$

The proposed method M1 is given in Algorithm 2.

2.2.2. Method Two (M2). Firstly, we estimate the parameters $\hat{\theta}_{OK}$ and $\hat{\sigma}_{OK}^2$ with the ordinary kriging model $y(x) = \mu + z(x)$ by solving the following formulas:

$$\hat{\mu}_{OK} = (\mathbf{1}_n^T R^{-1} \mathbf{1}_n)^{-1} \mathbf{1}_n^T R^{-1} y, \quad (11)$$

$$\hat{\sigma}_{OK}^2 = \frac{1}{n} (y - \mathbf{1}_n \hat{\mu}_{OK})^T R^{-1} (y - \mathbf{1}_n \hat{\mu}_{OK}), \quad (12)$$

$$\hat{\theta}_{OK} = \operatorname{argmin}_{\theta} (n \log \hat{\sigma}_{OK}^2 + \log |R|), \quad (13)$$

where $\mathbf{1}_n$ is an $n \times 1$ vector of ones. Then, we can select active variables by minimizing the negative penalized log-likelihood function $Q(\beta, \hat{\theta}_{OK}, \hat{\sigma}_{OK}^2)$ and update correlation parameters. Finally, the new kriging model is fitted similarly to that in M1. The proposed method M2 is given in Algorithm 3.

2.2.3. Method Three (M3). Firstly, the PBK method is used to select active variables β based on the kriging model via

minimizing the negative penalized log-likelihood function (7). It considers correlation between sample points, which is different from M1. Then, the selected variables based on PBK were substituted into the model (1) as new variables, and correlation parameters are updated. Finally, the estimated parameters $\hat{\beta}', \hat{\sigma}'^2, \hat{\theta}'$ estimated by formulas (4)–(6) are used to predict the responses of unsampled points by substituting into (2).

The variable selection part of M3 is using IRLARS of PBK in Algorithm 1. The fitting steps are the same as Steps 2–4 in Algorithm 2 after selecting variables. We will not go into details here.

Theoretically, M1 and M2 do not need the iterative step of parameter estimation of β such as PBK in variable selection; thus, they can improve time efficiency greatly. In addition, the last step of the proposed three methods is to update θ and refit the kriging model, so the prediction accuracy could be improved.

3. Simulation

The effectiveness of the proposed methods is proved in terms of simulation performance. We validate the simulation performance of the methods compared with back propagation (BP) neural network, generalized regression (GR) neural network, and PBK method according to time efficiency and prediction accuracy. We use the following indicators to measure time and prediction accuracy.

- (a) The root mean square prediction error (RMSPE) is defined as

$$RMSPE = \sqrt{\frac{1}{N} \sum_{i=1}^N (y(x_i^{\text{test}}) - \hat{y}(x_i^{\text{test}}))^2}, \quad (18)$$

where N is the number of testing samples. The mean of RMSPE (MRMSPE) of the simulations is calculated to evaluate the prediction performance.

- (b) The time of consumption (CPU) shows the uptime of the system. Especially, we need identification to measure their accuracy of variable selection in linear functions using the following indicators:
- (c) The average number of inactive effect identified rate (IEIR).
- (d) The average number of active effect identified rate (AEIR).
- (e) The average size of identified mean function (MEAN).

Obviously, when AEIR is larger, MRMSPE, IEIR, and CPU are smaller, and if MEAN is closer to the target, the model is better.

Four analytical functions are displayed in the following equations:

Step 1: set the initial values for $\beta^{(0)}, \theta^{(0)}, \sigma^{(0)}$
 Step 2: with $R(\hat{\theta}^{(k)})^{-1}/\hat{\sigma}^{2(k)} = C^T C$, $y' = Cy$, and $F' = CF$, solve penalized likelihood function for λ and α obtained by cross validation, $\hat{\beta}^{(k+1)} = \operatorname{argmin}_{\beta} (\|y' - F'\beta\|^2 + \lambda\alpha \sum_{j=1}^p |\beta_j| + \lambda(1-\alpha) \sum_{j=1}^p |\beta_j|^2)$
 Step 3: estimate σ^2, θ via maximizing the log-likelihood function (3), $\hat{\sigma}^{2(k+1)} = 1/n(y - F\hat{\beta}^{(k+1)})^T R^{-1}(y - F\hat{\beta}^{(k+1)})$, $\hat{\theta}^{(k+1)} = \operatorname{argmin}_{\theta} (n \log \hat{\sigma}^{2(k+1)} + \log|R|)$
 Step 4: if the convergence is attained, get the estimation $\hat{\beta}, \hat{\theta}, \hat{\sigma}^2$; otherwise, return to step 2

ALGORITHM 1: (IRLARS algorithm for Elastic Net penalty).

(1) A known function [11]:

$$y(x) = 0.4x_1 + 0.3x_2 + 0.2x_3 + 0.1x_4 + 0.05x_5 + 0.01x_6 + z(x). \quad (14)$$

(2) Test function [17]:

$$y(x) = 9 + 15 \sum_{i=1}^6 x_i + \frac{25}{2} \sum_{i=1}^5 x_i x_{i+1} + 10 \sum_{i=1}^6 x_i^2. \quad (15)$$

(3) Gramacy and Lee function [13]:

$$y(x) = \exp[\sin(0.9(x_1 + 0.48)^{10})] + x_2 x_3 + x_4. \quad (16)$$

(4) Borehole function [18]:

$$y(x) = 2\pi x_3 (x_4 - x_6) \times \log\left(\frac{x_2}{x_1}\right) \left(1 + 2 \frac{x_3 x_4}{\log(x_2/x_1) x_1^2 x_8}\right)^{-1}. \quad (17)$$

A known function (KF) includes twelve-dimensional inputs, where the first six variables (x_1, x_2, \dots, x_6) decrease the effects on the responses, the remaining variables (x_7, x_8, \dots, x_{12}), are unrelated (i.e., zero coefficients) to the outputs, and $z(x)$ has mean 0 and covariance $\sigma^2 R(\theta)$, where $\sigma^2 = 0.05$ and $\theta = 1_6$. The Test function (TF) includes twelve-dimensional inputs, where the last six variables are irrelevant to the outputs. The Gramacy and Lee function (GL) includes six-dimensional inputs, where the last two variables are irrelevant to the outputs. The Borehole function (BF) includes eight-dimensional inputs. And the nonlinearity gradually gets more obvious from the function KF to function BF. In function KF, we consider 60, 80, and 100 sample points for training in order to compare the effect of different sample sizes. In function TF, GL, and BF, we consider 100 sample points for training in order to compare the effect of nonlinearity. 1,000 sample points are randomly selected as testing data. We evaluate the accuracy of prediction, identification, and time efficiency based on 500 simulations.

In order to compare the effectiveness of the proposed methods, simulation studies are conducted to evaluate the performance of the BP neural network and GR neural network methods as other surrogate models firstly. The number of neurons in the hidden layer of the BP neural network is the optimal number of neurons among the 15 candidate neurons. The transfer functions from the input layer to the hidden layer and the hidden layer to the output layer choose “logsig” and “purelin,” respectively. The simulation results for four functions are given in Table 1.

Due to the lack of interpretability of a neural network, neural network methods cannot select variables. In this paper, we will focus on variable selection for the kriging model. We carry out simulation studies using PBK, three methods in Section 2 with Lasso penalty and Elastic Net penalty. The first-order polynomial basis function and the stationary Gaussian correlation function are used. And Latin Hypercube designs [19] are used to select sample points for training data in the design domain because they can be produced with minimal computational cost, and the design space is fulfilled well. The simulation results for four functions are given in Tables 2–6, respectively.

The principal results of these simulations are as follows:

- (1) According to Tables 2–6, the prediction accuracy of the proposed three methods is better than that of the PBK, BP, and GR neural networks, where M3 is the better one. In addition, their high prediction accuracy is obvious with escalating nonlinearity. Particularly, we take Table 4 as an example; the RMSPE for the predictors with Lasso penalty and Elastic Net in M3 gets low to (69.30% = (0.3528 – 0.1083) / 0.3528) and 58.53% smaller than the PBK predictors, respectively.
- (2) According to simulation results of the function KF (i.e., Tables 2 and 3), the method M1 could identify more variables, which lead to its highest AEIR and IEIR. Although AEIR of M2 is not high, its IEIR is the lowest and MEAN is close to the target. As to M3, its identification is basically the same as PBK. It indicates that M2 and M3 have better identification in variable selection.
- (3) In the aspect of time, the efficiency of M1 and M2 is obviously better than that of PBK, BP, and GR neural networks, where M1 is the most efficient. But the efficiency of M3 is little worse than that of PBK. Particularly, we take $n = 100$ in Tables 2 and 3 as an example the time of M1 with Lasso penalty and Elastic Net penalty gets low to 79.16% and 86.50% smaller than PBK, respectively. For other cases, the optimization of time efficiency of M1 is almost higher than 80%.

In summary, the BP neural network is mostly worse than the traditional PBK method in the accuracy of prediction, and it is not faster than the PBK method with Lasso penalty in the time efficiency. Although the GR neural network is faster than the PBK method in the time efficiency, it is mostly worse than the traditional PBK method in the accuracy of prediction. We

Step 1: Screen variables using formula (9).
 Step 2: Select the new basis function $f'_i(x)$ as new variables. Update the correlation parameter to θ' .
 Step 3: Estimate the parameters $\beta', \sigma'^2, \theta'$ via formulas (4)–(6).
 Step 4: The estimated parameters $\hat{\beta}', \hat{\sigma}'^2, \hat{\theta}'$ are used to predict the responses by substituting into formula (10).

ALGORITHM 2: (variable selection algorithm for M1).

Step 1: give $\mu^{(0)}, \theta^{(0)}$, and $\sigma^{2(0)}$ as initial variables
 Step 2: use formulas (12) and (13) for optimal model parameters $\hat{\theta}_{OK}, \hat{\sigma}_{OK}^2$
 Step 3: with $R(\hat{\theta}_{OK})^{-1}/\hat{\sigma}_{OK}^2 = C^T C, y^* = Cy$, and $F^* = CF$ solve the penalized likelihood function for $\lambda, \hat{\beta} = \operatorname{argmin}_{\beta} (\|y^* - F^* \beta\|^2 + \sum_{j=1}^p P_{\lambda}(|\beta_j|))$, where λ is obtained based on cross validation
 Step 4: the remaining steps of fitting are the same as Steps 2–4 in Algorithm 2 and will not be repeated here

ALGORITHM 3: (variable selection algorithm for M2).

TABLE 1: Data simulation results of neural network.

Function	Sample Size	BP		GR	
		MRMSPE	CPU(s)	MRMSPE	CPU(s)
KF	$n = 60$	0.110 6	3008.31	0.114 5	2142.63
	$n = 80$	0.108 5	3509.58	0.110 5	2236.98
	$n = 100$	0.107 7	3794.20	0.108 9	2246.46
TF	$n = 100$	3.972 7	5112.79	14.469 4	2429.14
GL	$n = 100$	0.668 5	4354.56	0.672 7	2306.72
BF	$n = 100$	4.341 1	3597.09	22.362 2	2405.54

TABLE 2: Data simulation results of the function KF with Lasso.

Sample Size	Method	MRMSPE	IEIR(%)	AEIR(%)	MEAN	CPU(s)
$n = 60$	PBK	0.067 8	37.60	92.47	7.80	2815.65
	M1	0.067 5	55.93	96.53	9.15	589.02
	M2	0.067 4	4.37	79.20	5.01	687.74
	M3	0.066 9	37.67	92.43	7.81	2879.55
$n = 80$	PBK	0.066 7	20.53	91.70	6.73	3101.15
	M1	0.066 5	58.17	97.00	9.31	618.83
	M2	0.066 3	5.67	81.20	5.21	953.33
	M3	0.065 7	20.77	91.60	6.74	3384.24
$n = 100$	PBK	0.065 8	12.53	93.17	6.34	3525.52
	M1	0.065 7	58.90	97.50	9.38	734.71
	M2	0.065 0	7.77	82.40	5.41	1128.58
	M3	0.064 9	14.20	93.17	6.44	3881.17

TABLE 3: Data simulation results of the function KF with Elastic Net.

Sample Size	Method	MRMSPE	IEIR(%)	AEIR(%)	MEAN	CPU(s)
$n = 60$	PBK	0.067 8	44.20	93.40	8.26	10 315.95
	M1	0.067 5	63.00	97.00	9.60	1126.76
	M2	0.067 0	6.33	81.57	5.27	1315.17
	M3	0.067 2	43.40	93.63	8.22	10 980.71
$n = 80$	PBK	0.066 7	25.17	92.27	7.05	10 787.02
	M1	0.066 5	63.10	97.27	9.62	1277.98
	M2	0.066 1	8.33	83.27	5.50	1685.08
	M3	0.065 8	25.37	92.53	7.07	11 161.00
$n = 100$	PBK	0.065 7	17.37	93.63	6.66	11 135.97
	M1	0.065 6	63.57	97.53	9.67	1503.19
	M2	0.065 1	10.73	83.40	5.65	1914.47
	M3	0.065 0	17.90	93.90	6.71	11 761.19

TABLE 4: Data simulation results of the function TF.

Method	Lasso		Elastic Net	
	MRMSPE	CPU(s)	MRMSPE	CPU(s)
PBK	0.352 8	3896.31	0.355 9	13 416.90
M1	0.205 1	578.45	0.268 5	1297.63
M2	0.109 1	888.38	0.262 9	2286.61
M3	0.108 3	3924.61	0.147 6	14 977.26

TABLE 5: Data simulation results of the function GL.

Method	Lasso		Elastic Net	
	MRMSPE	CPU(s)	MRMSPE	CPU(s)
PBK	0.755 5	2015.58	0.755 8	7034.73
M1	0.641 6	242.03	0.640 2	766.81
M2	0.633 8	600.29	0.632 6	1237.27
M3	0.634 2	2181.55	0.634 6	8150.91

TABLE 6: Data simulation results of the function BF.

Method	Lasso		Elastic Net	
	MRMSPE	CPU(s)	MRMSPE	CPU(s)
PBK	0.877 9	2198.33	0.878 7	10 522.03
M1	0.710 4	367.71	0.871 3	1234.14
M2	0.687 0	696.51	0.827 8	1459.26
M3	0.616 4	2775.03	0.796 4	12 315.72

take $n = 100$ in function KF as an example; the RMSPE for BP neural network, the GR neural network and PBK is 0.107 7, 0.108 9, and 0.065 8, respectively, and their time is 3794.20 s, 2246.46 s, and 3525.52 s. In addition, due to the lack of interpretability of BP and GR neural networks, these methods do not have identification. Thus, the neural network is not the preferred candidate method in variable selection methods compared with PBK. Although M1 greatly improves time efficiency by reducing the iterative steps of parameter estimation of β , it identified more inactive variables than PBK, which lead to its high AEIR and IEIR, as the correlation between sample points is not considered in the variable selection. As to M2, it not only maintains the advantage of M1 in time efficiency but also makes up for the shortcomings of M1 by considering the correlation between sample points. Therefore, M2 is superior to M1 in both identification and prediction accuracy and can be used to replace PBK. Although M3 is refitted on the basis of the variable selection of PBK, which leads to its low time efficiency, it has a significant improvement in the prediction accuracy. Therefore, we can use M3 as an improved method of PBK.

4. Empirical Application

4.1. Circuit Simulation. The proposed methods are applied to the circuit-simulation code [12]. Six input variables x_1, \dots, x_6 and one response y are in this experiment. The simulation experiments were conducted 32 times. This data are also used to analyze the relationship between output and input [20]. It can be useful for illustrating the prediction accuracy and time efficiency. The candidate variable set includes all linear effects. To compare the prediction

accuracy and time efficiency of different methods, the dataset was separated into two sections: half of the data are used for training and the other half for testing. The experimental results with Lasso and Elastic Net penalty are listed in Tables 7 and 8, respectively.

From the second and third columns of Tables 7 and 8, we could find the prediction accuracy of M1 (0.2207 and 0.2207), M2 (0.2207 and 0.2207), and M3 (0.1520 and 0.1591) which can achieve better prediction than PBK (0.2869 and 0.2740), and the prediction of M1 and M2 are all improved by 23.07% and 19.45%, those of M3 are improved by 47.02% and 41.93% relative to PBK.

As displayed in the last two columns of Tables 7 and 8, we could find the operation time of M1 (0.37 s and 0.81 s) and M2 (0.43 s and 1.08 s) which are less than those of PBK (2.26 s and 8.65 s) and M3 (2.29 s and 9.01 s), and the approximation efficiency of M1 is improved by 83.63% and 90.64%, those of M2 are improved by 80.97% and 87.51% relative to PBK.

Therefore, M1 and M2 can be used as variable selection methods to replacing the traditional PBK method according to prediction accuracy and time efficiency for fitting small samples. And M3 can be used as an improved method of PBK due to better prediction accuracy.

4.2. Piston Clap Noise. Piston secondary motion would cause an engine noise which is unwanted. It is the departure of a piston from the nominal motion prescribed by the mechanism. Six variables were used to minimize the piston clap noise. The variables are the cylinder liner x_1 , location of peak pressure x_2 , skirt length x_3 , skirt profile x_4 , skirt ovality x_5 , and pin offset x_6 , respectively.

TABLE 7: Experimental results for circuit simulation with Lasso.

Method	RMSPE	Improved (%)	CPU(s)	Improved (%)
PBK	0.2869	—	2.26	—
M1	0.2207	23.07	0.37	83.63
M2	0.2207	23.07	0.43	80.97
M3	0.1520	47.02	2.29	-1.33

TABLE 8: Experimental results for circuit simulation with Elastic Net.

Method	RMSPE	Improved (%)	CPU(s)	Improved (%)
PBK	0.2740	—	8.65	—
M1	0.2207	19.45	0.81	90.64
M2	0.2207	19.45	1.08	87.51
M3	0.1591	41.93	9.01	-4.16

TABLE 9: Experimental results for piston clap noise with Lasso.

Method	RMSPE	Improved (%)	CPU(s)	Improved (%)
PBK	1.0069	—	18.50	—
M1	0.8316	17.41	2.11	88.59
M2	0.5466	45.71	5.20	71.89
M3	0.6357	36.87	21.66	-17.08

TABLE 10: Experimental results for piston clap noise with Elastic Net.

Method	RMSPE	Improved (%)	CPU(s)	Improved (%)
PBK	0.9942	—	189.45	—
M1	0.7108	28.51	5.75	96.96
M2	0.5466	45.02	16.89	91.08
M3	0.5466	45.02	207.76	-9.66

The relevant dataset of the example is derived from Fang et al. [2]. The data include 100 observations, 6 input variables, and the candidate variable set includes all linear effects. In this paper, we divide the 100 data into 5 sets, select one of the sets as the test set for calculating RMSPE, and others as the train sets; this process is repeated 5 times and train sets are different every time. PBK, M1, M2, and M3 are compared, and the running time is recorded. The simulation results with Lasso and Elastic Net penalty are shown in Tables 9 and 10, respectively.

From the second and third columns of Tables 9 and 10, we could find the prediction accuracy of M2 (0.5466 and 0.5466) which can achieve better prediction than PBK (1.0069 and 0.9942), M1 (0.8316 and 0.7108), and M3 (0.6357 and 0.5466), and the prediction of M1 is improved by 17.41% and 28.51%, those of M2 are improved by 45.71% and 45.02%, and those of M3 are improved by 36.87% and 45.02% relative to PBK.

As displayed in the last two columns of Tables 9 and 10, we can find that the operation time of the M1 (2.11 s and 5.75 s) and M2 (5.20 s and 16.89 s) which are less than those of PBK (18.50 s and 189.45 s) and M3 (21.66 s and 207.76 s), and the approximation efficiency of M1 is improved by 88.59% and 96.96%, and those of M2 are improved by 71.89% and 91.08% relative to PBK.

Therefore, M1 and M2 can achieve the goal which replace the PBK method both in terms of prediction accuracy and time efficiency. And M3 can be viewed as an improvement of PBK.

5. Conclusion and Discussion

In the field of engineering, we always use a simple surrogate model to fit the complex true model for time saving and interpretability [21]. In order to improve the accuracy of prediction and save the cost of time for the PBK method, three new methods are proposed to obtain the trend function to improve the prediction accuracy and time efficiency. M1 is to select the variables using the linear model at the first and then to estimate the parameters in the kriging model. M2 is first to estimate the parameters with the OK model and then to select the variables via the penalized kriging model and finally refit kriging model. M3 is an improvement of PBK, which is a refitting based on the variable selection of PBK.

Several analysis functions with escalating nonlinearity and two engineering examples with different sample sizes are used to test our proposed step-by-step variable selection methods. The experimental results show that M2 which is superior to PBK, BP, and GR neural networks and other

methods can be used as a favorable method of variable selection in both prediction accuracy and time efficiency. If we want to pursue higher time efficiency while maintaining certain prediction accuracy, M1 might be a good choice. As well, M3 could be used as an improved method of PBK due to good prediction accuracy.

Data Availability

The data used to support the findings of this study are included in this paper.

Disclosure

Ziheng Feng and Min Li are co-first authors.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

This work was supported by the National Natural Science Foundation of China (Nos. 11871294 and 12101346) and Shandong Provincial Natural Foundation, China (Nos. ZR2021QA053 and ZR2021QA044).

References

- [1] T. J. Santner, B. J. Williams, and W. I. Notz, *The Design and Analysis of Computer Experiments*, Springer, New York, 2003.
- [2] K. T. Fang, R. Li, and A. Sudjianto, *Design and Modeling for Computer Experiments*, CRC Press, Florida, 2005.
- [3] H. Dong, P. Wang, C. Fu, and B. Song, "Kriging-assisted teaching-learning-based optimization (KTLBO) to solve computationally expensive constrained problems," *Information Sciences*, vol. 556, pp. 404–435, 2021.
- [4] H. Dong, P. Wang, B. Song, Y. Zhang, and X. An, "Kriging-assisted Discrete Global Optimization (KDGO) for black-box problems with costly objective and constraints," *Applied Soft Computing*, vol. 94, p. 106429, 2020.
- [5] B. Echard, N. Gayton, and M. Lemaire, "Ak-mcs: an active learning reliability method combining Kriging and Monte Carlo simulation," *Structural Safety*, vol. 33, no. 2, pp. 145–154, 2011.
- [6] S. Zhang, P. Zhu, W. Chen, and P. Arendt, "Concurrent treatment of parametric uncertainty and metamodeling uncertainty in robust design," *Structural and Multidisciplinary Optimization*, vol. 47, no. 1, pp. 63–76, 2013.
- [7] P. Kersaudy, B. Sudret, N. Varsier, O. Picon, and J. Wiart, "A new surrogate modeling technique combining Kriging and polynomial chaos expansions - application to uncertainty analysis in computational dosimetry," *Journal of Computational Physics*, vol. 286, pp. 103–117, 2015.
- [8] H. Liu, Y.-S. Ong, and J. Cai, "A survey of adaptive sampling for global metamodeling in support of simulation-based complex engineering design," *Structural and Multidisciplinary Optimization*, vol. 57, no. 1, pp. 393–416, 2018.
- [9] J. D. Martin and T. W. Simpson, "Use of kriging models to approximate deterministic computer models," *AIAA Journal*, vol. 43, no. 4, pp. 853–863, 2005.
- [10] V. R. Joseph, Y. Hung, and A. Sudjianto, "Blind kriging: a new method for developing metamodels," *Journal of Mechanical Design*, vol. 130, no. 3, pp. 350–353, 2008.
- [11] Y. Hung, "Penalized blind kriging in computer experiments," *Statistica Sinica*, vol. 21, no. 3, pp. 1171–1190, 2011.
- [12] J. Sacks, W. J. Welch, T. J. Mitchell, and H. P. Wynn, "Design and analysis of computer experiments," *Statistical Science*, vol. 4, no. 4, pp. 409–423, 1989.
- [13] Y. Zhang, W. Yao, S. Ye, and X. Chen, "A regularization method for constructing trend function in Kriging model," *Structural and Multidisciplinary Optimization*, vol. 59, no. 4, pp. 1221–1239, 2019.
- [14] R. Tibshirani, "Regression shrinkage and selection via the lasso," *Journal of the Royal Statistical Society: Series B*, vol. 58, no. 1, pp. 267–288, 1996.
- [15] H. Zou, "The adaptive lasso and its oracle properties," *Journal of the American Statistical Association*, vol. 101, no. 476, pp. 1418–1429, 2006.
- [16] H. Zou and T. Hastie, "Regularization and variable selection via the elastic net," *Journal of the Royal Statistical Society: Series B*, vol. 67, no. 2, pp. 301–320, 2005.
- [17] Y. Zhang, W. Yao, X. Chen, and S. Ye, "A penalized blind likelihood Kriging method for surrogate modeling," *Structural and Multidisciplinary Optimization*, vol. 61, no. 2, pp. 457–474, 2020.
- [18] M. D. Morris, T. J. Mitchell, and D. Ylvisaker, "Bayesian design and analysis of computer experiments: use of derivatives in surface prediction," *Technometrics*, vol. 35, no. 3, pp. 243–255, 1993.
- [19] M. D. McKay, R. J. Beckman, and W. J. Conover, "Comparison of three methods for selecting values of input variables in the analysis of output from a computer code," *Technometrics*, vol. 21, no. 2, pp. 239–245, 1979.
- [20] W. J. Welch, R. J. Buck, J. Sacks, H. P. Wynn, T. J. Mitchell, and M. D. Morris, "Screening, predicting, and computer experiments," *Technometrics*, vol. 34, no. 1, pp. 15–25, 1992.
- [21] M. Plumlee and V. R. Joseph, "Orthogonal Gaussian process models," *Statistica Sinica*, vol. 28, no. 2, pp. 601–619, 2018.