IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received November 22, 2020, accepted November 30, 2020, date of publication December 7, 2020,
date of current version December 28, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.3043182

Signal Denoising Method Combined With Variational Mode Decomposition, Machine Learning Online Optimization and the Interval Thresholding Technique

ZHENXING LIU[1,3], JIANHUA CHANG[1,2], HONGXU LI[1], LUYAO ZHANG[1], AND SICHENG CHEN[1]

[1]Jiangsu Key Laboratory of Meteorological Observation and Information Processing, Nanjing University of Information Science and Technology, Nanjing 210044, China
[2]Collaborative Innovation Center of Atmospheric Environment and Equipment Technology, Nanjing University of Information Science and Technology, Nanjing 210044, China
[3]Department of Information Technology, Taizhou Polytechnic College, Taizhou 225300, China

Corresponding author: Jianhua Chang (jianhuachang@nuist.edu.cn)

This work was supported in part by the National Natural Science Foundation of China under Grant 61875089, in part by the Project Funded by the Priority Academic Program Development of Jiangsu Higher Education Institutions, China, under Grant 1081080015001, in part by the Postgraduate Research and Practice Innovation Program of Jiangsu Province under Grant SJCX19_0308, and in part by Jiangsu Universities "Qing Lan" project.

ABSTRACT The signal-to-noise ratio of lidar signals decreases rapidly with an increase in distance, which seriously affects the application of lidar detection technology. Variational mode decomposition (VMD) has performed optimality in dealing with noise, but the number of modes, K, and the penalty parameter, α, must be preset. Therefore, a novel lidar signal denoising method that combines VMD with machine learning online optimization (MLOO) and the interval thresholding (IT) technique, named VMD-MLOO-IT, is proposed in this article. The proposed method defines new fitness functions to evaluate the result of VMD-based denoising, and selects the optimal parameters by the model which development by MLOO. In addition, IT is used to deal with the recovered signal. The experimental results demonstrate the superiority of the presented method over the other empirical mode decomposition-based and VMD-based denoising methods.

INDEX TERMS LiDAR signal denoising, machine learning online optimization, Gaussian process, variational mode decomposition, fitness function.

I. INTRODUCTION

Lidar is an active remote sensing instrument that has been widely used to study the optical and physical characteristics of aerosols in the atmosphere [1]–[5]. The parameter knowledge of aerosols can be retrieved by using specialized algorithms and lidar data [6]. However, lidar signals are strongly affected by background noise, atmospheric turbulence and detector noise, which severely limit the detection distance and accuracy of lidar as well as subsequent related applications [7]–[9]. Removing the noise from lidar signals is very important for the development of lidar remote sensing detection technology.

The associate editor coordinating the review of this manuscript and approving it for publication was Sotirios Goudos.

Various methods have been proposed for signal denoising, the most recent of which are based on empirical mode decomposition (EMD) and variational mode decomposition (VMD). EMD is a recursive decomposition model that can decompose a signal into a finite number of intrinsic mode functions (IMFs) with an iterative process called sifting [10], [11]. The decomposition is related to the characteristics of the signal itself without the requirement of any prior basis functions [12]. With these advantages, many EMD-based denoising methods have been proposed, such as EMD-HT, EMD-ST [13], and EMD-STRP [14]. In [13], the position of the noisy parts of the corrupted IMFs is altered in a random way and EMD multiple iterative is applied to selected interval thresholding. In [14], the relevant and irrelevant intrinsic mode functions are distinguished by correlation coefficient,

the soft thresholding technique and the roughness penalty technique are applied to the irrelevant and relevant modes respectively to extract useful information effectively. Despite the success of these works, EMD-based methods still have some inherent limitations; for example: (a) there is a lack of mathematical theory; (b) the methods may produce mode mixing; and (c) the methods are sensitive to noise [15], [16].

In contrast to EMD, VMD, which was proposed by Dragomiretskiy and Zosso in 2014 [17], is a nonrecursive decomposition model in which the modes are extracted concurrently. VMD is a very simple and fast method for signal adaptive decomposition that can decompose a real value signal into an ensemble of modes with a central frequency, and it has optimality in dealing with noise because of its close relations to the Wiener filter [17]. However, penalty parameter α and mode number K in VMD need to be predefined, and these two values have an important impact on VMD [18]. Although the parameters can be determined by the trial-and-error method or through experience, these options are time-consuming and inefficient because of the wide range of choices. Thus, the most challenging task is to select the optimal values of α and K. In [19], detrended fluctuation analysis (DFA) is used to select K by estimating the scaling exponent of each mode and comparing it with the threshold. However, this method will affected by the relational model between K and the scaling exponent of the input signal. In [20], parameter K is selected with EMD. There could be serious distortion in the modes by not considering the different decomposition principles of EMD and VMD. In [16], the optimized two parameters are obtained by the particle swarm optimization (PSO). This approach raises the question of how to find a proper balance between exploration and exploitation due to the stochastic nature of the optimization process, and the maximum value of the fitness function does not indicate the best result of mode decomposition. In [18], whale optimization algorithm (WOA) is utilized to search for the optimal parameter combination of K and α. This method can obtain the appropriate parameter value, but not always work because of the lowest energy entropy is used as the fitness function. Besides, we cannot know how the parameters affect the VMD results.

Machine learning, which can be used to build a model to learn the accurate correlation between the input parameters and output target, is a powerful tool for parameter optimization [21]–[23]. Furthermore, Gaussian processes (GP) based on Bayesian statistics have characteristics relevant to machine learning, and their flexible nonparametric nature and computational simplicity have attracted researchers from many fields [24]–[27]. Therefore, in this article, a new VMD-based lidar signal denoising method is proposed. The proposed method designs the fitness functions to evaluate the result of VMD-based denoising. In addition, GP is used to construct the model of the relationship between parameters and the fitness functions, and the model is refined by machine learning online optimization (MLOO). Accordingly, the optimal parameters are identified. Finally, the reconstructed signal is denoised again using an interval thresholding (IT) technique. To verify the effectiveness of the proposed VMD-MLOO-IT, numerical simulation and real data tests are carried out, and comparisons with EMD-HT, EMD-STRP, VMD-EMD, VMD-DFA and VMD-WOA are made. The main contributions of this article are:

- For selecting the parameters, optimization algorithms has the best performance of global search. However, the key part of optimization algorithms based on the variable mode decomposition is the selection of the fitness function. In this article, we propose new fitness functions to evaluate the results of VMD-based denoising by analyzing the essence of VMD.
- The common thresholding techniques not take into account the distribution characteristics of a signal, which limits the denoising effect. In this article, by employing the interval thresholding (IT) technique to the reconstructed signal, the performance improves.
- A novel lidar signal denoising method named VMD-MLOO-IT is proposed. In contrast to existing VMD parameter optimization methods, VMD-MLOO-IT built a statistical model that relates the result of VMD-based denoising with the parameters. It can be seen from the model how the parameters affect the result of VMD-based denoising, which can provide guidance for our next experiment. The method is also different from common machine learning whose goal is to construct an accurate model. VMD-MLOO-IT implemented optimization, and the learning was performed in real time. Thus, a small sample was sufficient.

The rest of this article is organized as follows. In Section II, the basic ideas of VMD and the GP are introduced. The proposed VMD denoising method combined with MLOO and IT is described in Section III. The results and discussion of the proposed method applied to the simulation and real signal are presented in Section IV. The last section is the conclusion.

II. RELATED WORK

A. VARIATIONAL MODE DECOMPOSITION

VMD can decompose a real valued input signal y into an ensemble of band-limited intrinsic mode functions (BLIMFS), which are amplitude-modulated-frequency-modulated (AM-FM) signals, as follows:

$$u_k(t) = A_k(t)\cos(\Phi_k(t)) \tag{1}$$

where $u_k(t)$ is the kth mode and a pure harmonic signal, $A_k(t)$ is the nonnegative amplitude, $A_k(t) \geq 0$, $\Phi_k(t)$ is the phase, and the instantaneous frequency is $\omega_k(t) = \Phi_k'(t)$, where $\omega_k(t) \geq 0$.

For each mode $u_k(t)$, VMD computes the associated analytic signal by means of the Hilbert transform and obtains the unilateral frequency spectrum. Then, the mode's spectrum is shifted to the baseband by mixing with an exponential of the respective estimated center frequency. After that, the bandwidth of each mode is assessed through the Gaussian smooth

demodulation signal. Thus, the constrained variational problem minimizes the sum of the estimated bandwidths for each mode and is written as

$$\min_{\{u_k\},\{\omega_k\}} \left\{ \sum_{k=1}^{K} \left\| \partial_t \left[\left(\delta(t) + \frac{j}{\pi t} \right) * u_k(t) \right] e^{-j\omega_k t} \right\|_2^2 \right\}$$
$$s.t. \quad \sum_{k=1}^{K} u_k = y \tag{2}$$

where $\{\omega_k\}$ is the center frequency ensemble corresponding to mode ensemble $\{u_k\}$, K is the number of modes, ∂_t ∂_t is the gradient with respect to t, δ is the Dirac distribution, and y is the input signal that is decomposed by VMD.

To solve the above constrained variational problem, a quadratic penalty term and Lagrangian multiplier are introduced, which transform the constrained problem with an unconstrained problem. Then, Equation (2) can be expressed as

$$\begin{aligned} &L(\{u_k\}, \{\omega_k\}, \lambda) \\ &:= \alpha \sum_k \left\| \partial_t \left[\left(\delta(t) + \frac{j}{\pi t} \right) * u_k(t) \right] e^{-j\omega_k t} \right\|_2^2 \\ &\quad + \left\| y(t) - \sum_k u_k(t) \right\|_2^2 + \left\langle \lambda(t), y(t) - \sum_k u_k(t) \right\rangle \end{aligned} \tag{3}$$

where α is the quadratic penalty parameter and λ is the Lagrangian multiplier.

Now, Equation (2) is solved by finding the saddle point of Equation (3) by iterative suboptimizations called the alternate direction method of multipliers (ADMM).

u_k is updated as

$$\hat{u}_k^{n+1}(\omega) = \frac{\hat{y}(\omega) - \sum_{i \neq k} \hat{u}_i(\omega) + \hat{\lambda}(\omega)/2}{1 + 2\alpha(\omega - \omega_k)^2} \tag{4}$$

ω_k is updated as

$$\omega_k^{n+1} = \frac{\int_0^\infty \omega \left| \hat{u}_k(\omega) \right|^2 d\omega}{\int_0^\infty \left| \hat{u}_k(\omega) \right|^2 d\omega} \tag{5}$$

For denoising, only the quadratic penalty term is retained and the Lagrangian multiplier is dropped. The details of VMD can be found in [17].

B. GAUSSIAN PROCESS FOR MACHINE LEARNING

The Gaussian process is a special example of a stochastic process in which random variables obey a Gaussian distribution, and their joint distributions are Gaussian [21], [28]. The properties of the Gaussian process are completely determined by mean value function $m(x)$ and covariance function $k(x, x')$ and can be written as

$$f(x) \sim \mathrm{GP}(m(x), k(x, x')) \tag{6}$$

where $m(x) = \mathrm{E}[f(x)]$ and $k(x, x') = \mathrm{E}[(f(x) - m(x))(f(x') - m(x'))]$.

Gaussian process machine learning is a method for solving machine learning tasks by using Gaussian process models based on Bayesian statistics. A Gaussian process model used for regression estimation is

$$\boldsymbol{y} = f(\boldsymbol{x}) + \varepsilon \tag{7}$$

where $\boldsymbol{x}$ is the input vector, $f(\boldsymbol{x})$ is the function value, $\boldsymbol{y}$ is the observed target value and ε is the additive Gaussian noise that follows an independent, identically distribution with zero mean and variance σ_n^2, $\varepsilon \sim N(0, \sigma_n^2)$.

Assume there is a training dataset D of n samples, $D = \{(\boldsymbol{x}_i; y_i)_{i=1}^{\mathrm{n}}\} = \{\boldsymbol{X}, \boldsymbol{y}\}$, $\boldsymbol{X}$ is an $n \times d$ matrix, and $\boldsymbol{y}$ is a column vector with n rows. d is the dimension of $\boldsymbol{x}$. Our goal is to predict a latent function f^* on test set $\boldsymbol{X}^*$.

Starting from a zero mean Gaussian process f, we can obtain a prior on the noisy observations with

$$\boldsymbol{y} \left| \boldsymbol{X} \sim N(0, K(\boldsymbol{X}, \boldsymbol{X}) + \sigma_n^2 I) \right. \tag{8}$$

The covariance functions that we use have some free parameters called hyperparameters θ, which have an important impact on the GP prediction [29]. In general, the hyperparameters are unknown a priori. The hyperparameters are usually set by optimizing the (log) marginal likelihood. First, we can obtain the log marginal likelihood by training the dataset as follows:

$$\begin{aligned} \log p(\boldsymbol{y} | \boldsymbol{X}, \boldsymbol{\theta}) = &-\frac{1}{2} \boldsymbol{y}^T (K_\theta + \sigma_n^2 I)^{-1} \boldsymbol{y} \\ &- \frac{1}{2} \log \left| K_\theta + \sigma_n^2 I \right| - \frac{n}{2} \log 2\pi \end{aligned} \tag{9}$$

where covariance matrix $K_{\boldsymbol{\theta}} = K(\boldsymbol{X}, \boldsymbol{X}; \theta)$.

Then, the negative log marginal likelihood is minimized as follows:

$$\boldsymbol{\theta}^* = \arg\min_{\boldsymbol{\theta}} - \log p(\boldsymbol{y} | \boldsymbol{X}, \boldsymbol{\theta}) \tag{10}$$

The GP model is specified by $\boldsymbol{\theta}^*$. After they have been determined, we can use the GP models to perform prediction distributions at $\boldsymbol{X}^*$.

$$f^* | \boldsymbol{X}^*, D \sim N(\mu(\boldsymbol{X}^*), \mathrm{cov}(\boldsymbol{X}^*)) \tag{11}$$
$$\mu(\boldsymbol{X}^*) = K(\boldsymbol{X}^*, \boldsymbol{X})^T [K(\boldsymbol{X}, \boldsymbol{X}) + \sigma_n^2 I]^{-1} \boldsymbol{y} \tag{12}$$
$$\begin{aligned} \mathrm{cov}(\boldsymbol{X}^*) = &K(\boldsymbol{X}^*, \boldsymbol{X}^*) - K(\boldsymbol{X}^*, \boldsymbol{X})[K(\boldsymbol{X}, \boldsymbol{X}) \\ &+ \sigma_n^2 I]^{-1} K(\boldsymbol{X}, \boldsymbol{X}^*) \end{aligned} \tag{13}$$

where $\mu(\boldsymbol{X}^*)$ and $\mathrm{cov}(\boldsymbol{X}^*)$ are the mean function and covariance function, respectively, of $f(\boldsymbol{X}^*)$.

III. PROPOSED VMD DENOISING METHOD COMBINED WITH MLOO AND IT

A. PRINCIPLE OF VMD-BASED DENOISING

Without loss of generality, a signal can be expressed as

$$\boldsymbol{y}(n) = \boldsymbol{f}(n) + \boldsymbol{W}(n), \quad n = 1, 2, \cdots, N \tag{14}$$

where $\boldsymbol{f}(n)$ is the ideal noise-free signal corresponding to $\boldsymbol{y}(n)$. $\boldsymbol{W}(n)$ is additive white Gaussian noise (AWGN) with zero mean and variance σ_N^2, N is the length of $\boldsymbol{f}(n)$.

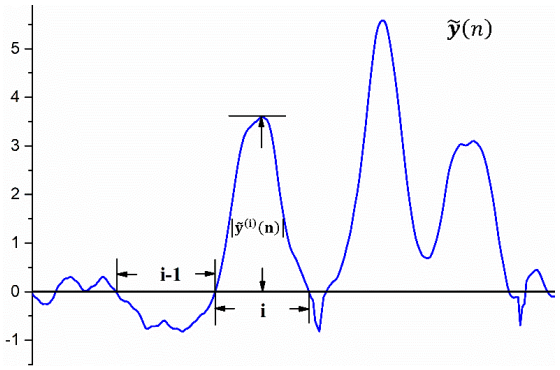

FIGURE 1. Illustration of interval thresholding.

Denoising is performed to find an estimate $\tilde{\boldsymbol{y}}(n)$ that is close to $\boldsymbol{f}(n)$. From previous research [16], [18], [19], VMD-based denoising is a recovered signal with partial BLIMFs, as follows:

$$\tilde{\boldsymbol{y}}(n) = \sum_{k=1}^{J} u_k(n), \quad 1 \le J \le K \tag{15}$$

where K and J denote the total number of BLIMFs and the number of BLIMFs that are relevant to $\mathbf{y}(n)$, respectively.

The method that we use to distinguish the relevant and irrelevant BLIMFs is to measure the Bhattacharyya distance (BD) of the probability density functions (PDFs) of each mode and input signal as follows [18]:

$$\begin{aligned} C(k) &= BD\left[PDF(u_k(n)), PDF(\mathbf{y}(n))\right] \\ s.t.\ \ \mathbf{y}(n) &= \sum_{k=1}^{K} u_k(n) \end{aligned} \tag{16}$$

The relevant BLIMFs can be determined by evaluating the slope of two adjacent BDs. The maximum slope can be written as

$$J = \arg\max_{k} \left| C(k+1) - C(k) \right|, \quad 1 \le k \le K \tag{17}$$

Therefore, u_k $(k = 1,\ldots,J)$ is the relevant BLIMF, and the others are irrelevant modes.

B. PARAMETER OPTIMIZATION BASED ON MLOO

Research shows that MLOO can use the GP to develop a relationship model between parameters and the experiment target value and implement optimization quickly in real time [30], [31]. Inspired by this, we try to use MLOO to build the relationship model between parameters α and K and the fitness function. Thus, the key part of VMD-MLOO is the selection of the fitness function. However, it is difficult to find such a fitness function that simultaneously selects the optimal parameters α and K and finds the best result of VMD without prior information about the signal, which is the case in the real world.

Here, we perform the analysis for the selection of parameters. According to [17], the essence of VMD is to decompose a signal in the frequency domain into band-limited signals with K center frequencies, which relate to the signal. Based on this feature, for a certain signal, K is also certain. The function of α improves convergence and makes the reconstructed signal have high fidelity simultaneously. As mentioned in [17], when α is small, too few modes mean that the lost modes will be shared by their neighboring modes. In addition, the decomposed mode will contain a large amount of noise in too many modes. Therefore, parameter selection can be divided into two stages. First, to determining K when α is small, set $\alpha = 100$ as in [17]. Then, α is

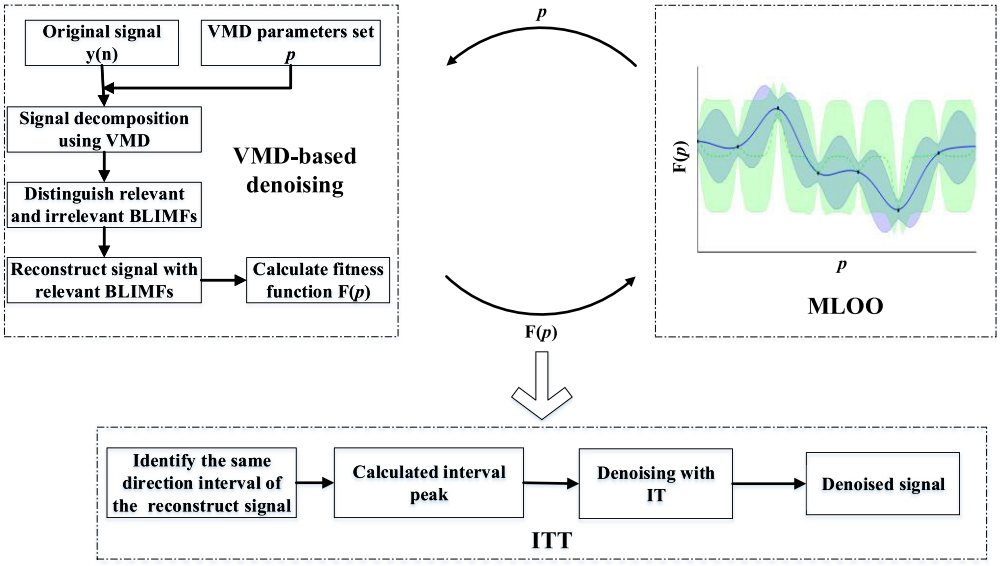

FIGURE 2. Flowchart of VMD-MLOO-IT.

TABLE 1. The execution time of different denoising methods (s).

N	2^8	2^9	2^{10}	2^{11}	2^{12}
EMD-HT	0.3267	0.3502	0.3768	0.4006	1.4864
EMD-STRP	0.3325	0.2612	0.3830	0.4031	1.6235
VMD-DFA	2.1389	2.1465	10.2833	24.0509	99.5202
VMD-EMD	0.7359	1.0159	6.5298	8.2686	32.4043
VMD-WOA	5.3261	11.6895	47.9435	199.2972	492.5864
VMD-MLOO	4.7699	12.0147	38.1417	141.4904	570.2450
VMD-MLOO-IT	4.7877	12.0339	38.1657	141.5098	570.2669

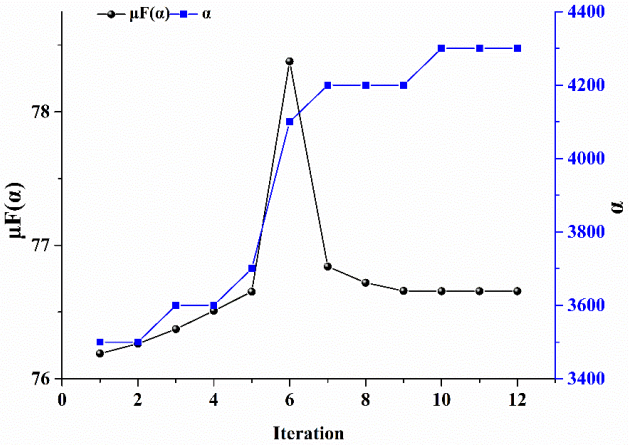

FIGURE 3. The MLOO convergence curve for VMD parameter optimization.

determined by K. Equation (18) and Equation (19) are used as fitness functions in the first and second stages, respectively, as follows:

$$F(K) = \arg\min_{K} \left\{ \frac{mean(C)}{var(C)} \right\} \tag{18}$$

where C is the cross-correlation coefficient between the decomposed mode component and the input signal as in [16]. In the first stages, the best K is obtained with the smallest fitness functions, which is the opposite of the process described in [16]. As mentioned above, the smallest fitness function indicates the best result of VMD, since α is small.

$$F(\alpha) = \arg\max_{\alpha} \sum_{n=1}^{N} \left\{ \frac{y(n) - \tilde{y}(n)}{K - J} \right\} \tag{19}$$

where $\tilde{\boldsymbol{y}}(n)$ is the recovered signal in (15), y(n) is the original signal, and J is the number of relevant BLIMFs. It can be seen in (19) that the fitness function is the average energy of the irrelevant modes. Clearly, the useful signal and noise content of the original signal is certain, which is represented by the relevant modes and irrelevant modes. In addition, the best K is selected, which means that it is currently in a good decomposition state. In this case, the larger the energy of the irrelevant BLIMFs is, the more noise removed, implying a better result of VMD-based denoising. However, the value range of α is very wide, so it is difficult to select the optimal parameter. Consequently, MLOO is used to build the relationship model between α and the fitness functions. The detailed steps of parameter optimization based on MLOO are as follows:

Step 1. Building a relationship model. Assume we have a dataset of input/output pairs $D = \{\boldsymbol{p}, \mathrm{F}(\boldsymbol{p})\}$, $\boldsymbol{p} = \{\alpha_i\}$ $i = 1, \ldots, N$. GP can be used to develop a statistical model for the relationship between $\boldsymbol{p}$ and $\mathrm{F}(\boldsymbol{p})$, as in (10).

Step 2. Predict the distribution of $\mathrm{F}(p^*)$, $p^* = \alpha^*$. For any p^* in Ω, which is the space of all possible parameter values, we can predict the distribution of $\mathrm{F}(p^*)$ with (12) and (13) based on the model built in step 1.

Step 3. Select the optimal parameter. We can select parameter $p^*_{\mathrm{opt}}(\alpha^*)$ corresponding to the largest $\mu_{\mathrm{F}}(p^*)$, using an 'optimizer' as follows:

$$p^*_{opt} = \arg\max_{\boldsymbol{p}^* \in \Omega} \mu_F(p^*) \tag{20}$$

Step 4. Refine the model. The model based on current dataset D may not be sufficient to accurately describe the relationship between parameter $\boldsymbol{p}$ and fitness function $\mathrm{F}(\boldsymbol{p})$ because the dataset is too small, and this 'optimizer' could become trapped in local minima. However, we can select parameter $p^*_{\mathrm{L}}(\alpha^*)$ corresponding to the largest $\mathrm{cov}_{\mathrm{F}}(p^*)$, using a 'learner' as follows:

$$p^*_L = \arg\max_{\boldsymbol{p}^* \in \Omega} \mathrm{cov}_F(p^*) \tag{21}$$

A larger $\mathrm{cov}_{\mathrm{F}}(p^*)$ indicates a greater difference between the true value and the predicted value of $\mathrm{F}(p^*)$. To improve the performance of the model, we can calculate the real value of $\mathrm{F}(p^*_{\mathrm{L}})$ in the experiment and merge $(p^*_{\mathrm{L}}; \mathrm{F}(p^*_{\mathrm{L}}))$ into dataset D. Next, return to step 1.

In the above steps, the 'learner' best refines our model, but it learns in the whole parameter space and has a large cost, which is unnecessary. Therefore, the 'optimizer' works at the same time and is used to monitor the learning process and obtain a global minimum. The optimal parameter $p^*_{\mathrm{opt}}(\alpha^*)$ is selected when the optimization process converges.

C. INTERVAL THRESHOLDING TECHNIQUE

Conventional thresholding techniques, such as hard thresholding and soft thresholding, do not take into account the distribution characteristics of a signal, which limits the denoising effect. Interval thresholding, which considers the same direction interval as the whole to implement threshold operations, has better performance [32]. In contrast to the conventional methods, we perform interval thresholding on the reconstructed signal after VMD denoising as follows:

$$\tilde{y}^{(i)}(n) = \begin{cases} \tilde{y}^{(i)}(n), & \left|\tilde{y}^{(i)}(n)\right| \geq \mathrm{T} \\ 0, & \left|\tilde{y}^{(i)}(n)\right| \leq \mathrm{T} \end{cases}$$
$$\mathrm{T} = \sigma\sqrt{2\ln(\mathrm{N})}, \sigma = median(|\tilde{y}(n)|)/0.6745 \tag{22}$$

TABLE 2. The denoising results compared with the different fitness functions for $\alpha = 50$, $\alpha = 75$, $\alpha = 100$ and $\alpha = 200$.

α =50	SNRout/dB	RMSE	F_ cross	F_ entropy
K=2	12.0334	0.4504	**13.8731**	0.0894
K=3	13.4399	0.3831	10.0249	0.1283
K=4	14.4549	0.3408	8.7970	0.1417
K=5	15.1651	0.3141	8.1648	0.1215
K=6	16.2303	0.2778	8.2785	0.1146
***K*=7**	**16.3681**	**0.2735**	**8.0125**	0.0841
K=8	16.0825	0.2826	8.1742	0.0809
K=9	15.5816	0.2994	8.0835	0.0700
K=10	13.9395	0.3617	8.2617	0.0584
K=11	14.1296	0.3538	8.3560	0.0521
K=12	15.5472	0.3006	9.0160	0.0516
K=13	15.7488	0.2937	9.1248	0.0515
K=14	15.7472	0.2937	8.9532	0.0435
K=15	16.2790	0.2763	9.1575	**0.0428**
α =75	SNRout/dB	RMSE	F_ cross	F_ entropy
K=2	16.6020	0.4219	**13.5576**	0.0804
K=3	13.7324	0.3704	9.9408	0.1217
K=4	14.6165	0.3345	8.7501	0.1378
K=5	15.2431	0.3113	8.1269	0.1197
K=6	16.2534	0.2771	8.2674	0.1133
***K*=7**	**16.3682**	**0.2734**	**8.0033**	0.0836
K=8	15.3387	0.3079	8.4023	0.0813
K=9	14.4751	0.3400	8.3492	0.0712
K=10	8.8188	0.6521	8.3237	0.0583
K=11	14.1008	0.3550	8.2208	0.0505
K=12	14.8629	0.3252	8.4418	0.0438
K=13	15.4834	0.3028	8.6541	0.0410
K=14	15.6005	0.2987	8.5952	0.0364
K=15	15.6942	0.2955	8.5355	**0.0357**
α =100	SNRout/dB	RMSE	F_ cross	F_ entropy
K=2	13.0601	0.4002	**13.2545**	0.0738
K=3	13.9835	0.3598	9.8445	0.1160
K=4	14.7629	0.3290	8.7010	0.1342
K=5	15.3139	0.3087	8.0875	0.1179
K=6	16.2688	0.2766	8.2530	0.1120
***K*=7**	**16.3647**	**0.2736**	**7.9893**	0.0831
K=8	14.238	0.3494	8.6244	0.0815
K=9	8.7457	0.6577	8.6975	0.0692
K=10	14.1965	0.3511	8.6631	0.0622
K=11	14.7508	0.3294	8.6180	0.0512
K=12	15.2604	0.3106	8.6412	0.0397
K=13	15.4975	0.3023	8.5025	0.0364
K=14	15.5781	0.2995	8.3947	0.0357
K=15	15.637	0.2975	8.3390	**0.0325**
α =200	SNRout/dB	RMSE	F_ cross	F_ entropy
K=2	14.2566	0.3487	**11.9960**	0.0599

TABLE 2. The denoising results compared with the different fitness functions for $\alpha = 50$, $\alpha = 75$, $\alpha = 100$ and $\alpha = 200$.

K=3	14.7329	0.3301	9.4397	0.0998
K=4	15.2265	0.3119	8.4940	0.1224
K=5	15.5543	0.3003	7.9388	0.1118
K=6	16.2921	0.2759	8.1874	0.1072
***K*=7**	**16.3311**	**0.2746**	**7.9205**	0.0811
K=8	14.3418	0.3453	9.5504	0.0809
K=9	15.5641	0.3000	9.8872	0.0799
K=10	15.6237	0.2979	9.6302	0.0669
K=11	15.6937	0.2955	9.3871	0.0630
K=12	15.7217	0.2946	9.1063	0.0520
K=13	15.7358	0.2941	8.8902	0.0444
K=14	15.7556	0.2934	8.6695	0.0365
K=15	15.7701	0.2929	8.5280	**0.0271**

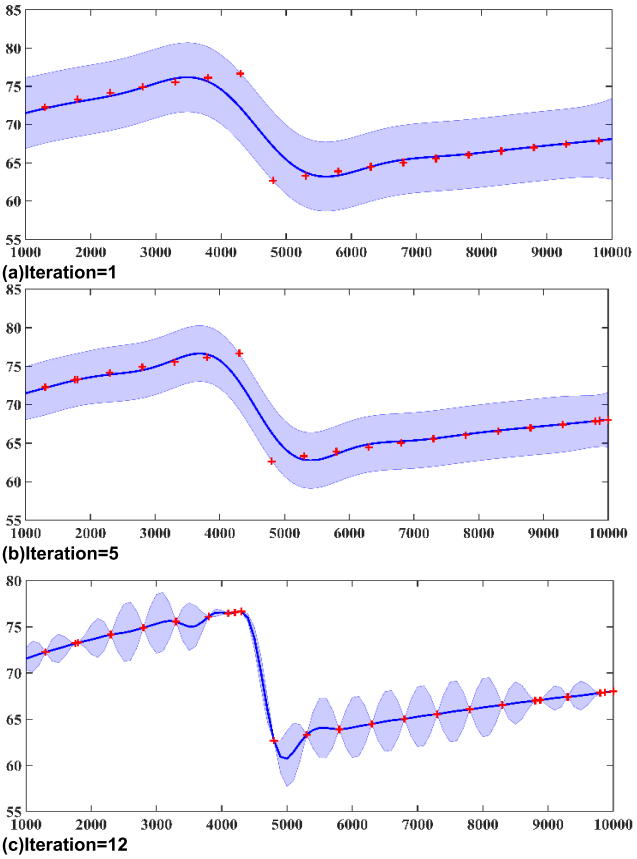

FIGURE 4. The predicted distribution diagram where the optimal parameters occur.

where $\tilde{\boldsymbol{y}}(n)$ is the reconstructed signal, $\tilde{\boldsymbol{y}}^{(i)}(n)$ indicates the ith interval that has the same positive or negative value, $\left|\tilde{\boldsymbol{y}}^{(i)}(n)\right|$ is the peak of the ith interval, as shown in Fig. 1, and T is the threshold of $\tilde{\boldsymbol{y}}(n)$. Equation (22) denotes the component of the ith interval set to zero when the peak $\left|\tilde{\boldsymbol{y}}^{(i)}(n)\right|$ is less than T. Otherwise, it remains constant.

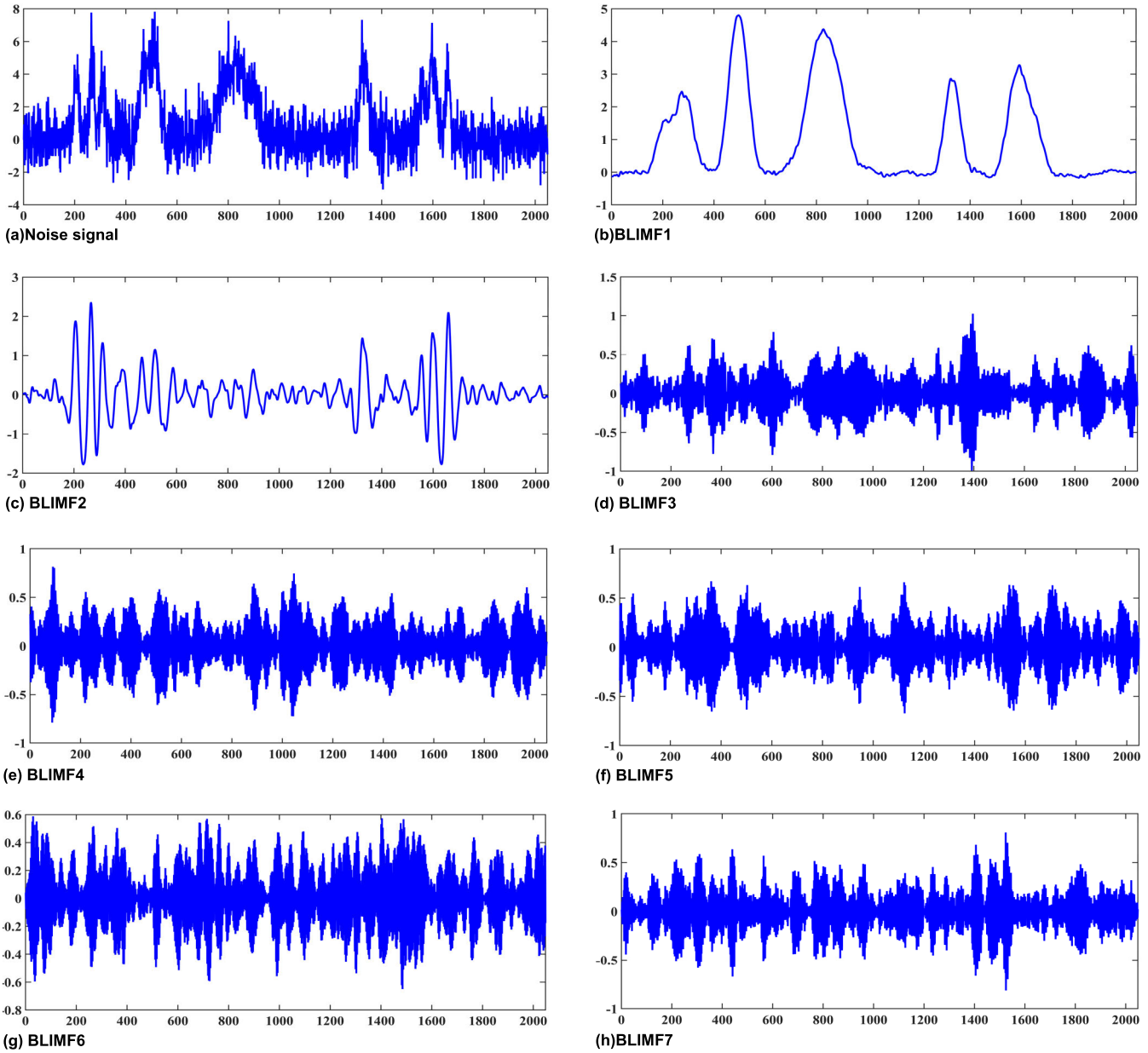

FIGURE 5. Decomposition result for the noisy Bumps signal.

D. PRINCIPLE OF VMD-MLOO-IT

As shown in Fig.2, the 'learner' produces a parameter set $\boldsymbol{p}$, for the VMD-based denoising to test, the denoised signal is used to calculate the fitness function based on its quality as a resource F($\boldsymbol{p}$). F($\boldsymbol{p}$) is back to the 'learner'. Then, the model of the relationship between parameters $\boldsymbol{p}$ and F($\boldsymbol{p}$) is developed, the 'optimizer' select the optimal parameter p^* base on this model, and the next parameters p is picked for the VMD-based denoising. The loop is iterated until the predicted value $\mu_{\mathrm{F}}(p^*)$ tends to converge, and the signal is reconstructed with relevant modes which is obtained by VMD with optimal parameters. Finally, the same direction interval of the reconstruct signal is identified, and the denoised signal is picked up by removing noise using the interval thresholding technique (ITT).

E. COMPUTATIONAL COMPLEXITY OF VMD-MLOO-IT

The computational complexity of an algorithm itself should be concerned except for its performance. Yang *et al*. [33] analyzed the computational complexity of EMD-based denoising. Liu *et al*. [19] analyzed the computational complexity of VMD-based denoising. On the basis of these works, we analyze the computational complexity of VMD-MLOO-IT.

Here, an input with a length of N through VMD is decomposed into a given number K mode, α is assigned in the interval (800–10,000), the number of search agents in VMD-WOA

TABLE 3. Denoising performance with different SNRin for the Bumps signal.

	SNRin/dB	10	5	3	1	-1
EMD-HT	SRNout/dB	17.6620	14.7848	13.7122	11.7561	10.1685
	RMSE	0.2356	0.3281	0.3713	0.4650	0.5583
	MAE	0.1848	0.2552	0.2701	0.3608	0.4300
EMD-STRP	SRNout/dB	18.8389	15.5093	13.1225	11.8644	10.1768
	RMSE	0.2507	0.3019	0.3973	0.4593	0.5578
	MAE	0.1613	0.2433	0.2843	0.3555	0.4297
VMD-DFA	SRNout/dB	19.1816	15.6988	15.5479	13.0235	11.6937
	RMSE	0.1949	0.2901	0.3004	0.4044	0.4610
	MAE	0.1571	0.2179	0.2331	0.3152	0.3362
VMD-EMD	SRNout/dB	19.6218	16.0759	14.5100	12.1852	12.0282
	RMSE	0.1880	0.2828	0.3387	0.4426	0.4507
	MAE	0.1502	0.2225	0.2656	0.3509	0.3390
VMD-WOA	SRNout/dB	19.1696	16.0825	15.4593	12.9487	12.1306
	RMSE	0.1981	0.2826	0.3036	0.4054	0.4538
	MAE	0.1573	0.2175	0.2367	0.3167	0.3562
VMD-MLOO	SRNout/dB	19.7301	16.2568	15.7533	13.4532	12.2096
	RMSE	0.1852	0.2770	0.2935	0.3825	0.4414
	MAE	0.1498	0.2172	0.2287	0.2921	0.3539
VMD-MLOO-IT	SRNout/dB	**21.1550**	**18.6349**	**17.8283**	**16.5891**	**15.2093**
	RMSE	**0.1781**	**0.2106**	**0.2311**	**0.2666**	**0.3125**
	MAE	**0.1143**	**0.1214**	**0.1347**	**0.1412**	**0.1920**

is 30, the number of input/output pairs used in MLOO is m, m = 19. The time complexity of VMD is $O(2N\log_2(2N))$, and the time complexity of MLOO is $O(m^2)$ [34]. The arithmetic operations involved include addition (ADD), subtraction (SUB), multiplication (MUL), division (DIV) and comparison (COMP), assuming they take the same amount of time. Then, the time complexity of IT is $N\cdot$(1COMP). From Fig.2, it can be found that the time complexity of VMD-MLOO-IT is $m\cdot O(2N\log_2(2N))+ O(m^2)$, which is bigger than the VMD and EMD.

In order to verify the correctness of proposed algorithm complexity, a series of experiments are carried out. The test signal is Bumps with 5 dB of Gaussian white noise, where the length N ranges from 2^8 to 2^{12}. The configuration of the computer is as follows: Intel (R) Pentium (R) G3260 @3.30GHz CPU and 4.00 GB RAM memory running windows 7. The execution time is shown in Table 1. As analyzed the time complexity of VMD-based denoising methods are bigger than the EMD-based denoising methods. The smallest time is the EMD-HT. The bigger time is the VMD-WOA, VMD-MLOO and VMD-MLOO-IT because of these methods search the wide range of parameter α. The others select parameters α by empirical, which is sometimes invalid, and the cost time will be even bigger if they use trial and error. The future work will be reducing the complexity of the proposed algorithms, especially for MLOO.

IV. RESULTS AND DISCUSSION

To assess the validity of the proposed VMD-MLOO-IT method, this section includes experiments with simulated and real noisy signals. The results are compared with different denoising methods, such as EMD-HT, EMD-STRP, VMD-DFA, VMD-EMD, and VMD-WOA.

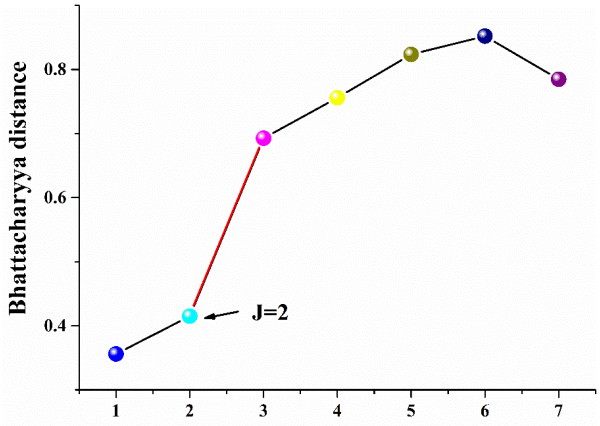

FIGURE 6. Bhattacharyya distance between the modes and the Bumps signal.

A. SIMULATION EXPERIMENT ON TYPICAL SIGNAL

In this experiment, we generate a typical signal, "bumps", using the "wnoise" function in MATLAB and a lidar simulation signal generated by the lidar equation [35]. These signals are corrupted by additive Gaussian noise. The output signal-to-noise ratio (SNRout), root mean square error (RMSE), and mean absolute error (MAE) are used as performance indices.

$$SNRout = 10\log\frac{\sum_{n=1}^{N} f^2(n)}{\sum_{n=1}^{N}\left[\tilde{y}(n)-f(n)\right]^2} \tag{23}$$

$$RMSE = \sqrt{\frac{1}{N}\sum_{n=1}^{N}\left[\tilde{y}(n)-f(n)\right]^2} \tag{24}$$

$$MAE = \frac{1}{N}\sum_{n=1}^{N}\left|\tilde{y}(n)-f(n)\right| \tag{25}$$

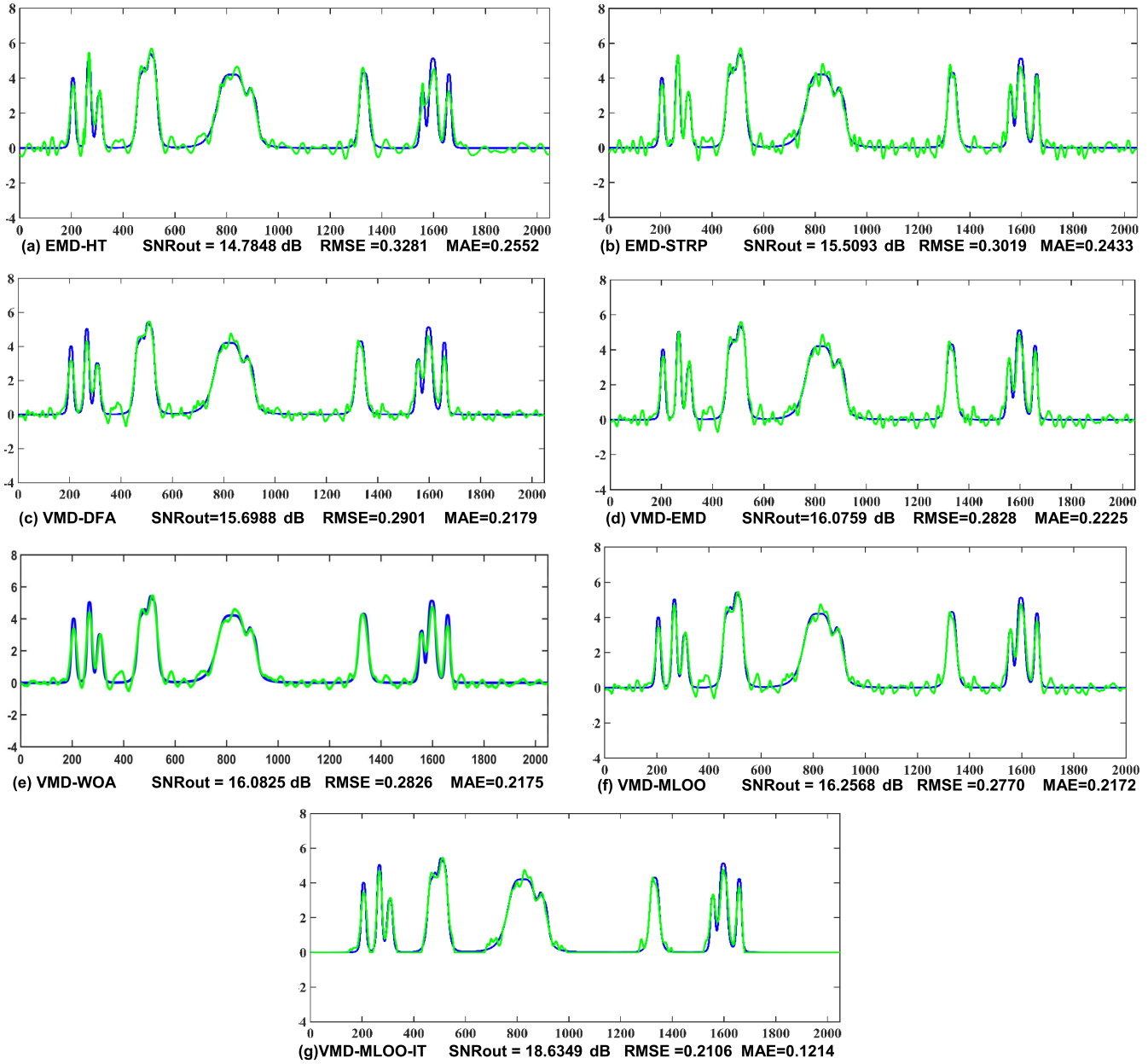

FIGURE 7. Comparisons of different denoising techniques for the noisy Bumps signal.

where $f(n)$ is the ideal noise-free signal, $\tilde{\boldsymbol{y}}(n)$ is the denoised signal, and N is the length of the signal. The comparison experiment of the denoising method in this article has gone through 30 experiments and the experimental data results are averaged.

Here, we illustrate the proposed algorithm with an example of the Bumps signal with 5 dB of Gaussian white noise and signal length $N = 2048$. Table 2 shows the denoising results when α is set to a small value ($\alpha = 50$, $\alpha = 75$, $\alpha = 100$, and $\alpha = 200$) and is compared with the different fitness functions mentioned in [16], [18], which are expressed as F_ cross and F_ entropy. F_ cross is the ratio of the mean value of the cross-correlation coefficient and the variance, and F_ entropy is the lowest energy entropy. However, it can be seen that the SNRout and RMSE change with K, and the largest F_ cross (red bold) and lowest F_ entropy (red bold) never obtain the largest SNRout and smallest RMSE. In contrast, when the F_ entropy is minimized (black bold), the largest SNRout and smallest RMSE are found, which is in accordance with the analysis in Section III.B. Consequently, it is appropriate to use the minimum F_ entropy as the fitness function for (18) to obtain the optimal parameter $K = 7$.

Fig. 3 shows the search process for $(\alpha^*, \mu_F(\alpha^*))$. In the initial stage, point (3500, 76.1887) is selected. However, due to the small number of samples, the model cannot accurately describe the relationship between parameters α and

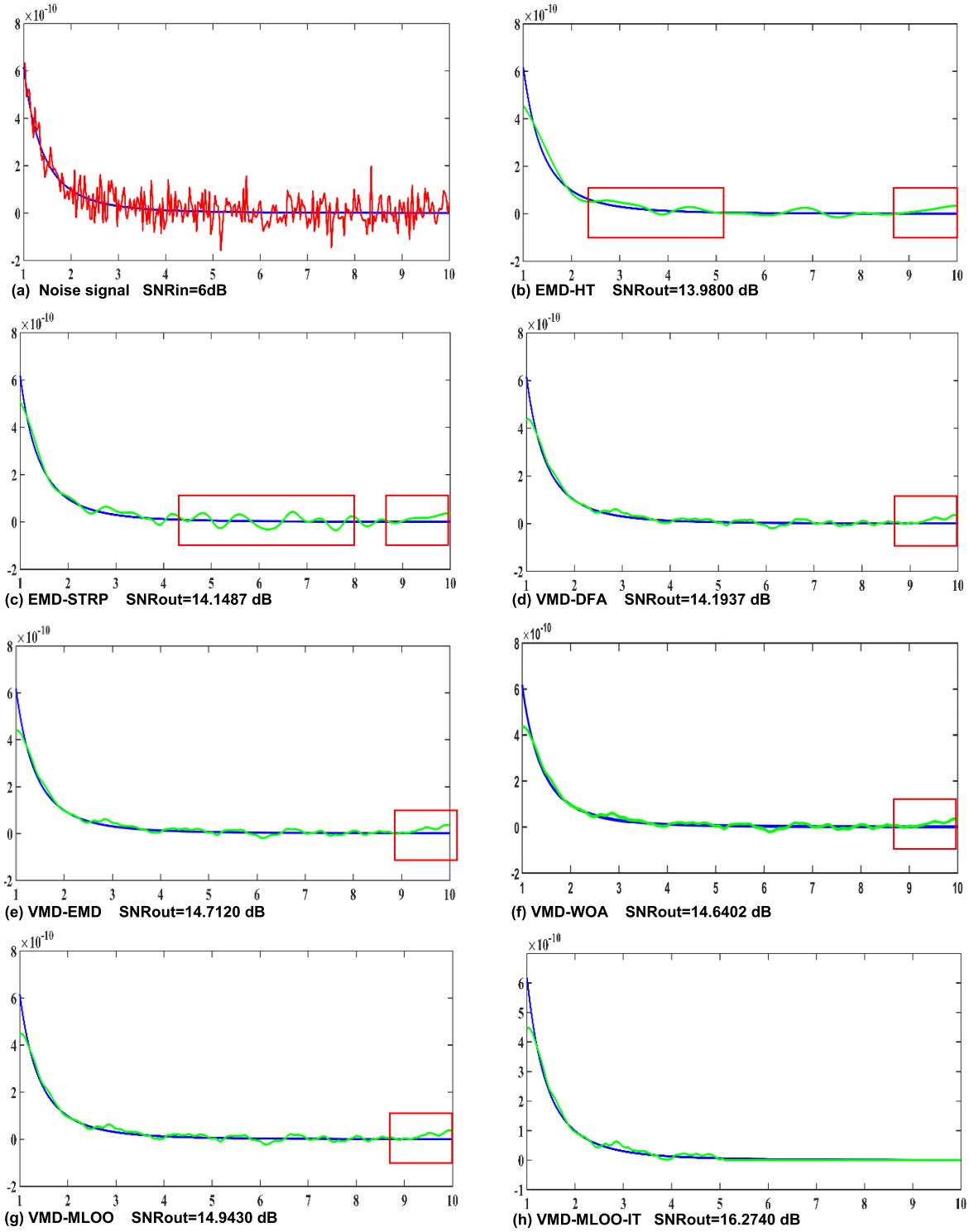

FIGURE 8. Results of denoising for lidar simulation signals.

fitness function F(α), and the prediction may be inaccurate. Therefore, the model continues to learn and is modified. The learning continues until $\mu_{\mathrm{F}}(\alpha^*)$ is found and converges, at which point the largest $\mu_{\mathrm{F}}(\alpha^*)$ is 76.6556, and optimal parameter α^* is 4300. In particular, it should be noted that parameter α is the same from iterations 7 to 9, but the predicted values are different, which indicates the adjustment ability of MLOO, and this is also the difference from the general optimization algorithm. Fig. 4 is the predicted distribution diagram where the largest $\mu_{\mathrm{F}}(\alpha^*)$ is obtained. It can be

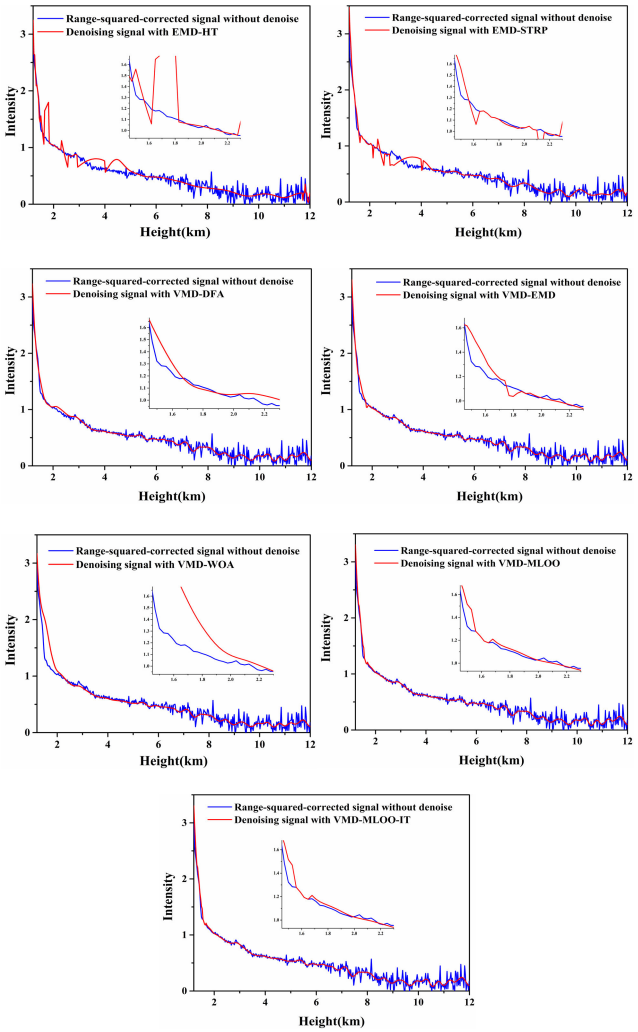

FIGURE 9. Comparison of the denoising results from the proposed VMD-MLOO method, VMD-MLOO -IT method and other methods for the range squared corrected signal.

seen from the figure how parameter α affects fitness function F(α), which can provide guidance for our next experiment, and the accuracy of the prediction increased with the number of training points (red cross).

Fig. 5 is the decomposition result for the noisy Bumps signal by VMD with the optimal parameter (7, 4300). The seven BLIMFs are distributed from low frequency to high frequency. Fig. 6 shows the Bhattacharyya distance (BD) between the BLIMFs and the Bumps signal calculated by (16). We can see that the slope between BD2 and BD3 is the greatest, as shown by the red line. In other words, BLIMF1 and BLIMF2 are chosen as the relevant BLIMFs.

Fig. 7 shows the denoised results for the seven denoising methods on a noisy Bump signal with an input signal-to-noise ratio SNRin = 5 dB. The blue lines are the true signals, and the green lines represent the denoised signals. Notably, the best results are provided by the VMD-MLOO-IT, which has the highest SNRout = 18.6349 dB, the lowest RMSE = 0.2106, and the lowest MAE = 0.1214, and VMD-MLOO is the second best. As seen from the noisy Bumps signal in Fig. 5(a), both EMD-based denoising methods and VMD-based denoising methods can remove noise effectively, but the latter is superior to the former because of its close relations to the Wiener filter. However, the denoised signal obtained by VMD-DFA, VMD-EMD, VMD-WOA, and VMD-MLOO still has some amplitude distortion at both ends and in the middle, whereas the proposed VMD-MLOO-IT overcomes these amplitude distortions. The main reason is that the IT technique eliminates the faint jitter in the signal and improves performance.

Table 3 lists the denoising results of EMD-HT, EMD-STRP, VMD-DFA, VMD-EMD, VMD-WOA, VMD-MLOO, and VMD-MLOO-IT with input signal-to-noise ratio (SNRin) varying from −1 to 10dB. The highest SNRout, lowest RMSE, and lowest MAE are shown in bold. The denoising results of VMD-based methods are better than those of EMD-based methods, and VMD-MLOO-IT obtains the best result.

Fig. 8 shows the denoising results of these seven methods for lidar simulation signals. The blue line and red line represent a noise-free signal and a noise signal, respectively, with SNRin = 6dB. The green line denotes the denoised signal. In addition, the ordinate represents the intensity of the signal, and the abscissa represents the distance. It can be seen from Fig. 8 that the denoised signal obtained by EMD-HT and EMD-STRP has obvious amplitude distortion, which is much better in other VMD-based methods. However, there is a noticeable jitter in the x-coordinate range of 9 to 10 for all methods except VMD-MLOO-IT. Table 4 compares the denoising performance of these seven methods, among which the best performance is shown in bold. The denoised result in Fig. 8 and Table 4 indicates that the VMD-MLOO-IT obtains the best results and is much more robust to noise.

TABLE 4. Performance of the seven denoising methods.

	SNRout/dB	RMSE($\times 10^{-11}$)	MAE($\times 10^{-11}$)
EMD-HT	13.9800	0.7928	1.3462
EMD-STRP	14.1487	0.7776	1.6069
VMD-DFA	14.1937	0.7288	1.1012
VMD-EMD	14.7120	0.7288	1.1012
VMD-WOA	14.6402	0.7348	1.1141
VMD-MLOO	14.9430	0.7096	1.1350
VMD-MLOO-IT	**16.2740**	**0.6088**	**0.6819**

B. REAL EXPERIMENT ON A LIDAR ECHO SIGNAL

To further show the capacity of the proposed VMD-MLOO-IT method, we test it on a real lidar signal. The lidar station is at Nanjing University of Information Science & Technology (118.7° N, 32.2° E). Data were collected using a Rayleigh-Raman-Mie lidar based on a diode-pumped Nd: YAG laser emitting at 532 nm with a total pulse energy of 200mJ. The lidar echo signal is received by the photomultiplier tube and sampled at a range resolution of 30 m. More details related to the instrument are given in [35].

To eliminate the influence of distance, the measured lidar backscattering signal will be corrected with range squared by multiplying the square of the distance. It can be seen from Fig. 9 that the intensity of the range squared corrected signal decreases with the height, and at heights above 8 km, the signal without denoising is almost masked by noise. Consequently, the detection range of the lidar instrument is severely limited, which affects its application in related fields. However, the noise can be removed by EMD-based and VMD-based denoising, as shown in Fig. 9. We propose that VMD-MLOO and VMD-MLOO-IT achieve the best result. The signal is smoother and the detection range is increased to nearly 12 km after denoising, while the range squared corrected signal denoised by EMD-HT and EMD-STRP has serious distortion at heights of 2-4 km and the signal denoised by VMD-DFA and VMD-EMD has a large jitter at heights of 2 km.

V. CONCLUSION

In this article, we proposed a novel method for lidar signal denoising, which we called VMD-MLOO-IT. The proposed method designed a new fitness function to evaluate the results of VMD-based denoising and utilized the GP to develop a relational model between the parameters and the fitness function. Next, the model was optimized online with machine learning so that the optimal parameters could be predicted accurately. Finally, the signal was reconstructed with the relevant BLIMFs and denoised by IT. In contrast to existing VMD parameter optimization methods, VMD-MLOO-IT built a statistical model that relates the result of VMD with the parameters and predicted the outcome given any parameters. The method is also different from common machine learning, whose goal is to construct an accurate model. VMD-MLOO-IT implemented optimization, and the learning was performed in real time. Thus, a small sample was sufficient.

We tested VMD-MLOO-IT and VMD-MLOO on a typical signal, called ''bumps'', with different input signal-to-noise ratios and found that they were superior to other denoising methods (EMD-HT, EMD-STRP, VMD-DFA, VMD-EMD, and VMD-WOA). We also experimented on a lidar simulation signal, which was corrupted with 6 dB of additive Gaussian noise. The results show that VMD-MLOO-IT achieves the best performance. Finally, we tested VMD-MLOO-IT and VMD-MLOO on a real lidar backscattering signal corrected with range squared. The experimental results demonstrated that the proposed method not only produced favorable signal denoising performance but also extended the detection range from 8 km to 12 km.

REFERENCES

[1] S. Garbarino, A. Sorrentino, A. M. Massone, A. Sannino, A. Boselli, X. Wang, N. Spinelli, and M. Piana, ''Expectation maximization and the retrieval of the atmospheric extinction coefficients by inversion of Raman lidar data,'' *Opt. Exp.*, vol. 24, no. 19, p. 21497, Sep. 2016, doi: 10.1364/oe.24.021497.

[2] M. de Graaf, A. Apituley, and D. P. Donovan, ''Feasibility study of integral property retrieval for tropospheric aerosol from Raman lidar data using principal component analysis,'' *Appl. Opt.*, vol. 52, no. 10, p. 2173, Apr. 2013, doi: 10.1364/ao.52.002173.

[3] A. Comerón, C. Muñoz-Porcar, F. Rocadenbosch, A. Rodríguez-Gómez, and M. Sicard, ''Current research in lidar technology used for the remote sensing of atmospheric aerosols,'' *Sensors*, vol. 17, no. 6, p. 1450, Jun. 2017, doi: 10.3390/s17061450.

[4] C. Böckmann and A. Kirsche, ''Iterative regularization method for lidar remote sensing,'' *Comput. Phys. Commun.*, vol. 174, no. 8, pp. 607–615, Apr. 2006, doi: 10.1016/j.cpc.2005.12.019.

[5] Y. Song, L. Yue, Y. Wang, H. Di, F. Gao, S. Li, Y. Zhou, and D. Hua, ''Research on BP network for retrieving extinction coefficient from mie scattering signal of lidar,'' *Measurement*, vol. 164, Nov. 2020, Art. no. 108028, doi: 10.1016/j.measurement.2020.108028.

[6] E. Chemyakin, S. Burton, A. Kolgotin, D. Müller, C. Hostetler, and R. Ferrare, ''Retrieval of aerosol parameters from multiwavelength lidar: Investigation of the underlying inverse mathematical problem,'' *Appl. Opt.*, vol. 55, no. 9, p. 2188, Mar. 2016, doi: 10.1364/ao.55.002188.

[7] S. Wu, Z. Liu, and B. Liu, ''Enhancement of lidar backscatters signal-to-noise ratio using empirical mode decomposition method,'' *Opt. Commun.*, vol. 267, no. 1, pp. 137–144, Nov. 2006, doi: 10.1016/j.optcom.2006.05.069.

[8] I.-Y. Choi, S.-H. Baik, N.-G. Park, H.-Y. Kang, and J.-H. Kim, ''Improvement of the measuring accuracy of the Raman lidar for remote detection of the hydrogen gas,'' *Int. J. Precis. Eng. Manuf.*, vol. 19, no. 7, pp. 967–973, Jul. 2018, doi: 10.1007/s12541-018-0114-z.

[9] C. Li, Z. Pan, F. Mao, W. Gong, S. Chen, and Q. Min, ''De-noising and retrieving algorithm of mie lidar data based on the particle filter and the fernald method,'' *Opt. Exp.*, vol. 23, no. 20, p. 26509, Oct. 2015, doi: 10.1364/oe.23.026509.

[10] R. Sharma, L. Vignolo, G. Schlotthauer, M. A. Colominas, H. L. Rufiner, and S. R. M. Prasanna, ''Empirical mode decomposition for adaptive AM-FM analysis of speech: A review,'' *Speech Commun.*, vol. 88, pp. 39–64, Apr. 2017, doi: 10.1016/j.specom.2016.12.004.

[11] Z. Wu and N. E. Huang, ''A study of the characteristics of white noise using the empirical mode decomposition method,'' *Proc. Roy. Soc. London. Ser. A, Math., Phys. Eng. Sci.*, vol. 460, no. 2046, pp. 1597–1611, Jun. 2004, doi: 10.1098/rspa.2003.1221.

[12] J. L. Gómez and D. R. Velis, ''A simple method inspired by empirical mode decomposition for denoising seismic data,'' *Geophysics*, vol. 81, no. 6, pp. 403–413, Nov. 2016, doi: 10.1190/GEO2015-0566.1.

[13] M. Li, L.-H. Jiang, and X.-L. Xiong, ''A novel EMD selecting thresholding method based on multiple iteration for denoising LIDAR signal,'' *Opt. Rev.*, vol. 22, no. 3, pp. 477–482, Jun. 2015, doi: 10.1007/s10043-015-0086-5.

[14] J. Chang, L. Zhu, H. Li, F. Xu, B. Liu, and Z. Yang, ''Noise reduction in lidar signal using correlation-based EMD combined with soft thresholding and roughness penalty,'' *Opt. Commun.*, vol. 407, pp. 290–295, Jan. 2018, doi: 10.1016/j.optcom.2017.09.063.

[15] M. A. Colominas, G. Schlotthauer, and M. E. Torres, ''Improved complete ensemble EMD: A suitable tool for biomedical signal processing,'' *Biomed. Signal Process. Control*, vol. 14, pp. 19–29, Nov. 2014, doi: 10.1016/j.bspc.2014.06.009.

[16] C. Yi, Y. Lv, and Z. Dang, ''A fault diagnosis scheme for rolling bearing based on particle swarm optimization in variational mode decomposition,'' *Shock Vib.*, vol. 2016, Jun. 2016, Art. no. 9372691, doi: 10.1155/2016/9372691.

[17] K. Dragomiretskiy and D. Zosso, ''Variational mode decomposition,'' *IEEE Trans. Signal Process.*, vol. 62, no. 3, pp. 531–544, Feb. 2014, doi: 10.1109/TSP.2013.2288675.

[18] H. Li, J. Chang, F. Xu, Z. Liu, Z. Yang, L. Zhang, S. Zhang, R. Mao, X. Dou, and B. Liu, ''Efficient lidar signal denoising algorithm using variational mode decomposition combined with a whale optimization algorithm,'' *Remote Sens.*, vol. 11, no. 2, p. 126, Jan. 2019, doi: 10.3390/rs11020126.

[19] Y. Liu, G. Yang, M. Li, and H. Yin, ''Variational mode decomposition denoising combined the detrended fluctuation analysis,'' *Signal Process.*, vol. 125, pp. 349–364, Aug. 2016, doi: 10.1016/j.sigpro.2016.02.011.

[20] S. Lahmiri and M. Boukadoum, ''Biomedical image denoising using variational mode decomposition,'' in *Proc. IEEE Biomed. Circuits Syst. Conf. (BioCAS)*, Oct. 2014, pp. 340–343, doi: 10.1109/BioCAS.2014.6981732.

[21] C. E. Rasmussen and C. K. I. Williams, *Gaussian Processes for Machine Learning*. Cambridge, MA, USA: MIT Press, 2006, pp. 1–30.

[22] R. Yan, T. Wang, X. Jiang, Q. Zhong, X. Huang, L. Wang, and X. Yue, "Design of high-performance plasmonic nanosensors by particle swarm optimization algorithm combined with machine learning," *Nanotechnology*, vol. 31, no. 37, Sep. 2020, Art. no. 375202, doi: 10.1088/1361-6528/ab95b8.
[23] L. Dai and H. Zhang, "Propagation-Model-Free base station deployment for mobile networks: Integrating machine learning and heuristic methods," *IEEE Access*, vol. 8, pp. 83375–83386, 2020, doi: 10.1109/ACCESS.2020.2990631.
[24] X. Zeng, W. Guo, K. Yang, and M. Xia, "Noise reduction and retrieval by modified lidar inversion method combines joint retrieval method and machine learning," *Appl. Phys. B, Lasers Opt.*, vol. 124, no. 12, p. 238, Dec. 2018, doi: 10.1007/s00340-018-7095-6.
[25] R. B. Gramacy and N. G. Polson, "Particle learning of Gaussian process models for sequential design and optimization," *J. Comput. Graph. Statist.*, vol. 20, no. 1, pp. 102–118, Jan. 2011, doi: 10.1198/jcgs.2010.09171.
[26] X. Chen, Y. Tian, T. Zhang, and J. Gao, "Differential evolution based manifold Gaussian process machine learning for microwave Filter's parameter extraction," *IEEE Access*, vol. 8, pp. 146450–146462, 2020, doi: 10.1109/ACCESS.2020.3015043.
[27] W. Yu, Y. Liu, Z. Ma, and J. Bi, "Improving satellite-based PM2.5 estimates in China using Gaussian processes modeling in a Bayesian hierarchical setting," *Sci. Rep.*, vol. 7, no. 1, Dec. 2017, Art. no. 7048, doi: 10.1038/s41598-017-07478-0.
[28] A. G. Wilson, "Covariance kernels for fast automatic pattern discovery and extrapolation with Gaussian processes," Ph.D.dissertation, Dept. Trinity., Cambridge Univ., Cambridge, U.K.:, 2014.
[29] Z. Chen and B. Wang, "How priors of initial hyperparameters affect Gaussian process regression models," *Neurocomputing*, vol. 275, pp. 1702–1710, Jan. 2018, doi: 10.1016/j.neucom.2017.10.028.
[30] F. Sahli Costabal, K. Matsuno, J. Yao, P. Perdikaris, and E. Kuhl, "Machine learning in drug development: Characterizing the effect of 30 drugs on the QT interval using Gaussian process regression, sensitivity analysis, and uncertainty quantification," *Comput. Methods Appl. Mech. Eng.*, vol. 348, pp. 313–333, May 2019, doi: 10.1016/j.cma.2019.01.033.
[31] P. B. Wigley, P. J. Everitt, A. van den Hengel, J. W. Bastian, M. A. Sooriyabandara, G. D. McDonald, K. S. Hardman, C. D. Quinlivan, P. Manju, C. C. N. Kuhn, I. R. Petersen, A. N. Luiten, J. J. Hope, N. P. Robins, and M. R. Hush, "Fast machine-learning online optimization of ultra-cold-atom experiments," *Sci. Rep.*, vol. 6, no. 1, May 2016, Art. no. 25890, doi: 10.1038/srep25890.
[32] Y. Duan and C. Song, "Relevant modes selection method based on spearman correlation coefficient for laser signal denoising using empirical mode decomposition," *Opt. Rev.*, vol. 23, no. 6, pp. 936–949, Dec. 2016, doi: 10.1007/s10043-016-0275-x.
[33] G. Yang, Y. Liu, Y. Wang, and Z. Zhu, "EMD interval thresholding denoising based on similarity measure to select relevant modes," *Signal Process.*, vol. 109, pp. 95–109, Apr. 2015, doi: 10.1016/j.sigpro.2014.10.038.
[34] J. Ko and D. Fox, "GP-BayesFilters: Bayesian filtering using Gaussian process prediction and observation models," *Auto. Robots*, vol. 27, no. 1, pp. 75–90, Jul. 2009, doi: 10.1007/s10514-009-9119-x.
[35] H. Li, J. Chang, F. Xu, B. Liu, Z. Liu, L. Zhu, and Z. Yang, "An RBF neural network approach for retrieving atmospheric extinction coefficients based on lidar measurements," *Appl. Phys. B, Lasers Opt.*, vol. 124, no. 9, p. 184, Sep. 2018, doi: 10.1007/s00340-018-7055-1.

ZHENXING LIU was born in Hunan, China. He is currently pursuing the Ph.D. degree with the Jiangsu Key Laboratory of Meteorological Observation and Information Processing, Nanjing University of Information Science and Technology, Nanjing. He is currently with the Department of Information Technology, Taizhou Polytechnic College. His current research interests include LIDAR measurements and signal processing.

JIANHUA CHANG was born in Jiangsu, China. He received the Ph.D. degree from the Department of Electronic Engineering, Southeast University. He was a Postdoctoral Fellow with the Anhui Institute of Optical Mechanics, Chinese Academy of Sciences, and McMaster University, Canada. He is the Deputy Dean with the School of Electronics and Information Engineering, Nanjing University of Information Technology. His current research interests include photonics and optical devices, all-solid-state lasers, lidar, and photoelectric sensing. He is a member of the Optoelectronic Technology Committee of the China Optical Society and a Senior Member of the China Electronics Society.

HONGXU LI was born in Henan, China. He is currently pursuing the Ph.D. degree with the Jiangsu Key Laboratory of Meteorological observation and Information Processing, Nanjing University of Information Science and Technology, Nanjing. His current research interests include LIDAR measurements and aerosol detection.

LUYAO ZHANG was born in Jiangsu, China. She is currently pursuing the master's degree with the Jiangsu Key Laboratory of Meteorological observation and Information Processing, Nanjing University of Information Science and Technology, Nanjing. Her current research interests include LIDAR measurements and signal processing.

SICHENG CHEN was born in Jiangsu, China. He is currently pursuing the master's degree with the Jiangsu Key Laboratory of Meteorological observation and Information Processing, Nanjing University of Information Science and Technology, Nanjing. His current research interests include LIDAR measurements and signal processing.