Ensemble Empirical Mode Decomposition Based on Sparse Bayesian Learning with Mixed Kernel for Landslide Displacement Prediction

Ping Jiang[1]*, Jiejie Chen[2]
School of Computer, Hubei PolyTechnic University, Hangshi, China[1]
College of Computer Science and Technology, Hubei Normal University, Huangshi, China[2]

***Abstract*—Inspired by the principles of decomposition and ensemble, we introduce an Ensemble Empirical Mode Decomposition (EEMD) method that incorporates Sparse Bayesian Learning (SBL) with Mixed Kernel, referred to as EEMD-SBLMK, specifically tailored for landslide displacement prediction. EEMD and Mutual Information (MI) techniques were jointly employed to identify potential input variables for our forecast model. Additionally, each selected component was trained using distinct kernel functions. By minimizing the number of Relevance Vector Machine (RVM) rules computed, we achieved an optimal balance between kernel functions and selected parameters. The EEMD-SBLMK approach generated final results by summing the prediction values of each subsequence along with the residual function associated with the corresponding kernel function. To validate the performance of our EEMD-SBLMK model, we conducted a real-world case study on the Liangshuijing (LSJ) landslide in China. Furthermore, in comparison to RVM-Cubic and RVM-Bubble, EEMD-SBLMK emerged as the most effective method, delivering superior results in the same measurement metrics.**

Keywords*—Bubble; cublic; ensemble empirical mode decomposition; landslide; Sparse Bayesian Learning

I. Introduction

Landslide, a natural geological occurrence, refers to a type of mass wasting that involves diverse ground movements [1, 2]. Essentially, it signifies a transition from a stable slope to an unstable one [3, 4]. The occurrence of this transition can be prompted by numerous internal and external factors, including vegetative cover, weather conditions, evaporation, and transpiration, either operating alone or jointly. Given the significant damage and casualties caused by landslides globally, considerable efforts are underway to establish a pr-warning system capable of predicting their occurrence. The task of landslide forecasting is not only crucial but also challenging, particularly in the context of rapidly increasing peak flows due to urbanization. To mitigate potential flood-related damages in the future, it is imperative to develop an accurate model for landslide forecasting.

It is well-established that Three Gorges Region, situated at the upstream section in Chinese Yangtze River, experiences lots of landslides, posing serious dangers to the region. These landslides, which occur almost annually, result in significant damage to both the local population and property. Given this, it is evident that the phenomenon involves numerous stochastic, interrelated components and exhibits highly nonlinear characteristics.

Currently, various methods, including artificial neural networks (ANN), fuzzy theory, chaos theory, and statistical approaches, have been extensively employed in the realm of nonlinear analysis [5-21]. A two-stage Bayesian integration framework has been effectively utilized for detecting prominent objects in light field images [5].

The resolution of nonlinear characteristics does not solely rely on a single approach; hybrid models also demonstrate their effectiveness. Methods for per-processing signals and evolutionary SVR have been developed to enhance short-term wind speed predictions [6]. Furthermore, a hybrid approach that incorporates the minimum cycle decomposition has proven effective in predicting temporary electrical load data [7]. Chen et al. proposed an innovative methodology that integrates genetic algorithm and simulated annealing algorithm with improved BPNN modeling for landslide prediction [8]. Extreme learning machines (ELM) excel in learning with superior generalization capabilities, thereby circumventing the challenges encountered by gradient-based learning methods. Lian et al. pointed out the potential applications of modified ELM in predicting landslide displacements [9, 10]. Furthermore, dynamic time series predictors leveraging echo state networks and ELM have been constructed to forecast landslide displacements [11, 12]. Functional networks (FNs) combined with hybrid methods have also been explored for landslide forecasting [13]. The paper harnessed MGGP to build a forecast method for landslide displacement without prior knowledge of the nonlinear model’s structure. Bootstrap-based generalized neural networks (Bootstrap-GRNN) have been utilized for interval prediction of displacements [14]. Kanungo et al. exhibited an integration model, combining with NN, fuzzy logic and likelihood concepts to forecast landslide occurrence [15].

Regrettably, the majority of current landslide prediction methods remain deterministic, falling short in providing meaningful insights into the uncertainty surrounding their predicted values. This significant limitation restricts the practical application of landslide forecasting in stochastic decision-making and analytical frameworks. EEMD [16] addresses the mode mixing issue by introducing finite noise, effectively eliminating it while preserving the physical uniqueness of the decomposition. On the other hand, SBL [17] leverages a parameterized prior to favor models with sparse

*Corresponding author: Ping Jiang

nonzero weights. Drawing inspiration from Yang et al.'s idea [17], we introduce a novel hybrid approach, EEMD-SBLMK, which combines EEMD and SBL. This approach generates probabilistic prediction by assessing the probabilistic distribution of weights linked to Gaussian kernel functions. Finally, the last section summarizes our findings and discusses potential avenues for future improvements.

II. THEORY

A. EEMD

EMD is a technique that exhibits great adaptability and efficiency in decomposing complex, nonlinear, and unstable signals. It leverages the HHT to accomplish this. The introduction of the IMF concept marks a pivotal innovation in EMD, as each IMF encapsulates the unique local information embedded in lots of data sheets.

Utilizing EMD allows for the decomposition of any sophisticated temporal datasets into multiple IMF components, along with a residual component that encapsulates the primary trend of data. IMFs adhere to certain criteria, which are as follows:

1) The total count of extreme points, including both peaks and valleys, should match how much zero crossings in the entire data-set, with a maximum difference of one.

2) For a specific point, the average value of the envelope formed by the local peaks and troughs should be zero.

Despite its strengths, EMD also exhibits certain limitations. A significant challenge arises from mode mixing, which occurs when signals of diverse scales coexist within a single IMF, or conversely, signals of identical scale are distributed across various IMFs. Tackling this problem, a novel method known as EEMD was introduced, which incorporates noise-assisted analysis (see Fig. 1). The EEMD approach could be summarized as:

Step 1: Augment the original signal series with white noise.

Step 2: Employ the EMD method to decompose the signal, incorporating the incorporated white noise, into its constituent IMFs.

Step 3: Execute the previous two steps repeatedly, introducing a fresh white noise with each iteration.

Step 4: Compute the average of the corresponding IMFs from all decompositions to arrive at the final IMFs.

Step 5: Calculate the mean of the corresponding residue components across all decompositions to determine the final residue, as shown in Eq. (1) to Eq. (3).

$$IMF_1^{'} = \frac{imf_{11} + imf_{21} + \cdots + imf_{N1}}{N} \tag{1}$$

$$IMF_{1n}^{'} = \frac{imf_{1n} + imf_{2n} + \cdots + imf_{Nn}}{N} \tag{2}$$

$$\mathrm{Re}^{'} = \frac{\mathrm{Re}_1 + \mathrm{Re}_2 + \cdots + \mathrm{Re}_n}{N} \tag{3}$$

B. Mutual Information (MI)

Input selection serves as a crucial aspect in the development of any neural network. It holds a crucial position in ascertaining the precision of the model's forecasts. Furthermore, incorporating irrelevant inputs can significantly impact the precision and reliability of the neural network.

The Mutual Information (MI) [20, 21] between random variable X and random variable Y, is a measure that quantifies the shared information between them, as shown in Eq. (4).

$$MI = \iint \xi_{X,Y}(x,y)\log[\frac{\xi_{X,Y}(x,y)}{\xi_X(x)\xi_Y(y)}]dxdy \tag{4}$$

where, $\xi_X(x)$ and $\xi_Y(y)$ represent the the individual probability density functions of variable X and variable Y, respectively, while $\xi_{X,Y}(x,y)$ denotes their joint probability density function. Considering the restricted quantity of data accessible for this research, we employ the kth nearest neighbor approach, as described in studies [12-16], to assess MI. This evaluation method is particularly suitable for small datasets. Based on the recommendations in references [12-16], it is advisable to set k to a value between 2 and 4. Given the small size of our data sample, we have chosen to set k equal to 3 in this paper.

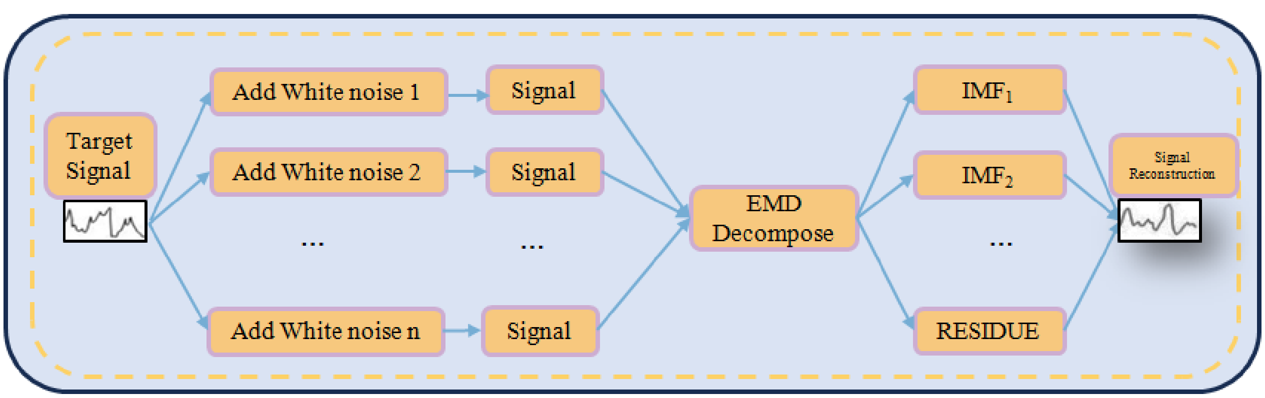

Fig. 1. EEMD.

C. Sparse Bayesian Learning (SBL)

The SBL model, alternatively known as a relevance vector machine, exhibits excellent adaptability for forecasting non-stationary random variables. This is due to its straightforward modeling of probabilistic quantity changes [17]. Fundamentally, SBL adopts a Bayesian viewpoint for kernel-based forecast models, capitalizing on a distinct prior for parameters that encourages sparsity in the prediction function.

Commonly, in a GR context, the correlation concerning the desired value t_n and the input vector x_n can be formulated as follows:

$$t_n = y(x_n;\omega) + \varepsilon_n \tag{5}$$

where, $\omega = [\omega_0, \omega_1, \omega_2, \cdots \omega_T]$ represents the weight vector that needs to be determined. On the other hand, ε_n represents the forecast error, which follows an independent and identically distributed normal distribution with a mean of N (0, σ^2). Furthermore, y_n follows a normal distribution, with its mean value designated as $f(x_n;\omega)$ and its variance designated as σ^2.

Utilizing the kernel method, $y(x_n;\omega)$ can be formally defined as:

$$y(x_n;\omega) = \omega^T \Phi(x_n) = \omega_0 + \sum_{i=1}^{M} \omega_i K(x_n, x_i) \tag{6}$$

where, $\Phi(x_n) = [1, K(x_n, x_1), \cdots, K(x_n, x_M)]^T$, $K(x_n, x_i)$ signifies the Gaussian kernel function, and M denotes the total count of such kernel functions employed. Given the inherent nonlinearity of $K(x_n, x_i)$, the model effectively captures and expresses nonlinear complexities with ease.

Recalling our earlier discussion, the joint distribution of target values $t = [t_1, t_2, \cdots t_N]$, pertaining to N independent groups of sampling data, can be formulated based on the distribution of t as follows:

$$p(t|\omega,\sigma^2) = (2\pi\sigma^2)^{-\frac{N}{2}} \exp\left\{-\frac{1}{2\sigma^2}\|t - \Phi\omega\|\right\} \tag{7}$$

where, $\Phi = [\varphi(x_1), \varphi(x_2), \cdots \varphi(x_N)]^T$.

Employ the process of maximizing the likelihood function, which signifies the likelihood of observing the provided data given the assumed model, to estimate ω_i and σ^2, but it may have over fitting phenomenon. Then to avoid it, we use the mandatory additional prerequisites to some parameters, based on Bayesian theory then define ω_i function, normal distribution:

$$p(\alpha) = \prod_{i=0}^{N} Gamma(\alpha_i | a, b) \tag{8}$$

$$p(\beta) = Gamma(\beta | c, d) \tag{9}$$

where, $\beta = \sigma^2$.

$$Gamma(\alpha | a, b) = \Gamma(a)^{-1} b^a \alpha^{a-1} e^{ba} \tag{10}$$

Then, $\Gamma(a) = \int_0^\infty t^{a-1} e^{-t} dt$, parameters a , b, c, d have no prior knowledge, values are small, a=b=c=d=10-4. Then, it can obtain uniform hyper parameters a=b=c=d=0.

Under bias framework, the prediction is based on the training data ω_i , α , σ^2 posterior distribution. According to Bayesian formula:

$$p(\omega,\alpha,\sigma^2 | t) = \frac{p(t|\omega,\alpha,\sigma^2)\, p(\omega,\alpha,\sigma^2)}{p(t)} \tag{11}$$

But, the above formula ride is hard to solve up, the left formula can be decomposed into:

$$p(\omega,\alpha,\sigma^2 | t) = p(\omega | t,\alpha,\sigma^2)\, p(\alpha,\sigma^2 | t) \tag{12}$$

Through the above analysis, the original problem is decomposed into two steps to solve:

1) Compute α , σ^2 under t posterior distribution.

2) Compute ω under α , σ^2 , t posterior distribution.

In practice, to simplify the calculation, Dirac distribution $\delta(\alpha_{MP}, \sigma^2_{MP})$ as ω under α , σ^2 , t posterior distribution:

$$\int p(t|\alpha,\sigma^2) = \int p(t|\omega,\sigma^2) p(\omega|\alpha) d\omega \tag{13}$$

$$p(\omega | t,\alpha,\sigma^2) = \frac{p(t|\omega,\sigma^2) p(\omega|\alpha)}{p(t|\alpha,\sigma^2)} \tag{14}$$

After the model parameters are obtained by training data, new input vectors x*, target value t* distribution density:

$$p(t_* | t) = \int p(t_* | \omega,\alpha,\sigma^2) p(|\omega,\alpha,\sigma^2) d\omega d\alpha d\sigma^2 \tag{15}$$

RVM regression model σ^2_*:

$$\sigma^2_* = \sigma^2_{MP} + \Phi^T(x_*) F \Phi(x_*) \tag{16}$$

Finally, the main problem α_{MP} and σ^2_{MP} , the maximum likelihood estimation method.

III. Forecast Model and Analysis

In the RVM [18] model training, it assumed that there exist no errors in the historical data of each sample. And it used eight kernel functions respectively in Eq. (17) to Eq. (22).

1) Gaussian

$$G(x, x_i) = \exp(-\frac{\|x - x_i\|^2}{\sigma^2}) \quad (17)$$

2) Cauchy

$$Ca(x, x_i) = \frac{1}{1 + \sigma^2 . \|x - x_i\|^2} \quad (18)$$

3) Cublic

$$C(x, x_i) = (\sigma^2 . \|x - x_i\|^2)^{\frac{3}{2}} \quad (19)$$

4) Bubble

$$B(x, x_i) = -\frac{\|x - x_i\|^2}{\sigma^2} \quad (20)$$

5) Laplace

$$L(x, x_2) = \exp(-\frac{\|x - x_i\|}{\sigma}) \quad (21)$$

6) R-distance

$$R(x, x_i) = \exp(-\frac{\|x - x_i\|}{\sigma}) \quad (22)$$

The landslide data, presented as a time series, typically exhibit nonlinear and non-stationary characteristics. To address this, we adopt an approach that combines decomposition and ensemble techniques. Specifically, we utilize the ensemble EEMD method to decompose three distinct types of landslide data. Three sets of sequences are obtained, the correlation between three groups of sub sequences and landslide displacement was calculated, and the best correlation group was selected as SBL parameters. Then, different kernel functions with each selected parameters are used to compute. Using distinct kernel functions in a mixed kernel model for landslide prediction offers benefits in terms of enhanced model flexibility, improved feature representation, enhanced prediction accuracy, robustness and generalization, as well as increased interpretability and understanding of model decisions. Based on the minimum number of computed RVM rules, it can obtain one selected parameter corresponds to one kernel function. Moreover, EEMD-SBLMK used selected kernels functions with corresponded input parameters to gain the final predicted results by assembling.

There several steps for EEMD-SBLMK:

1) All data (including displacement reservoir level and rainfall) are decomposed using EEMD into n IMFs and one residual function Residue (t) (see Fig. 2).

2) Use MI method to choose strong correlation between the IMFs component and displacement, and then it can decide the input parameters of EEMD-SBLMK (see Fig. 3).

3) Each selected IMFs component to be trained by different kernels functions, which can be predefined based on domain knowledge or determined through a data-driven approach, where different kernels are tested to find the optimal combination.

4) According to the minimum number of computed RVM rules, it gets some computed rules between kernel functions and selected parameter.

5) The final predicted result presents the sum of each subsequence prediction value of IMF and residual function Residue (t) with corresponded kernel function.

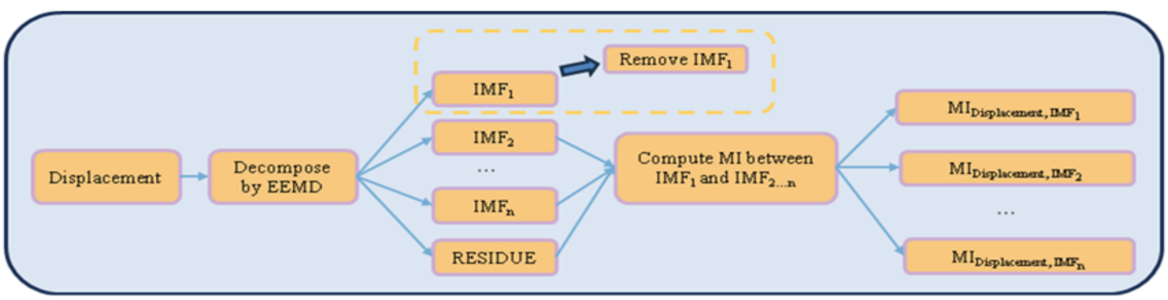

Fig. 2. Decomposed by EEMD.

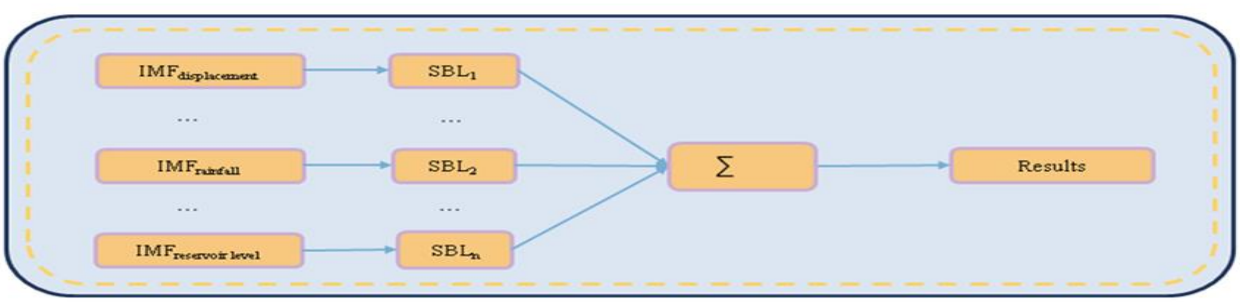

Fig. 3. EEMD-SBLMK.

IV. Application of EMMD-SBLMK on Landslide Prediction: A Case Study

A. *Dataset*

In this paper, we endeavor to introduce the EMMD-SBLMK approach for elucidating significant nonlinear relationships among diverse parameters pertaining to a practical geotechnical problem. All experiments conducted in this study were executed on the MATLAB 2013 platform. Given the uncertainty, instability, and intricate nature of landslides, their formation remains highly elusive. This complexity encompasses factors such as loose loess material susceptible to sliding, variations in reservoir levels, rainfall patterns, intricate geological formations, precipitation, and anthropogenic engineering activities, among others.

The landslide is very complicate, and some data about landslide are extremely difficult to collect or measure. So, we cannot analyze all collected data. All data have internal relations, not a single existence. Actually, scholars devote to study landslide based on two sides. Some scholars pay some interest in inter factor like mechanics, the other scholars are pay attention to numerical value. Then, the data of displacement, reservoir level and rainfall were collected to study landslide like [12-14]. Given the computational intensity of the EMD-SBLMK algorithm, a practical application was conducted by selecting the LSJ landslide at monitoring point 24 in the Three Gorges Reservoir area of China as a test case (see Fig. 4).The inclusion of mixed kernel functions in EEMD-SBLMK enables the model to effectively capture diverse patterns and features in landslide displacement data, enhancing generalization, robustness, interpretability, and overall modeling performance. Monitoring data about displacement and reservoir water level (see Fig. 5) and (see Fig. 6) are date from April 6, 2009 to May 25, 2011 at time interval six days. Monitoring about rainfall data (see Fig. 6) are date from April 6, 2009 to June 16, 2010 at time interval six days. The left data about rainfall data are recorded 0.

Fig. 4. LSJ landslide.

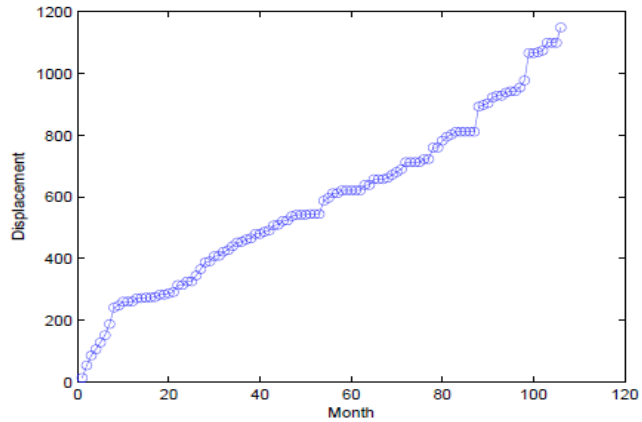

Fig. 5. LSJ displacement.

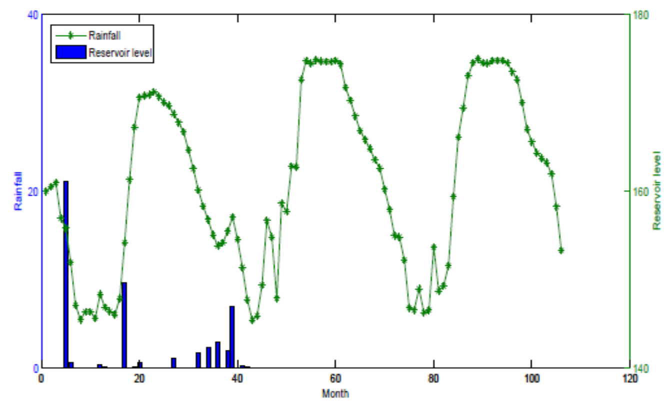

Fig. 6. LSJ rainfall and reservoir level.

The SBL method departs from ANNs in its requirement for equal-length training and prediction datasets, focusing on maintaining a balance between capturing complex patterns for model expressiveness and ensuring good generalization to unseen data. Consequently, we divide the entire dataset evenly into two parts to establish our prediction model. The dataset is bifurcated for analysis, with 50% allocated to the first group for model construction and the remaining 50% reserved for landslide displacement predictions. Additionally, we restrict the minimum number of time delays for input parameters to 10. Our EEMD integration totaled 100 iterations, augmented with 0.2 of white noise. This technique facilitates the decomposition of initial landslide displacement, reservoir water level, and rainfall time series. Specifically, displacement and reservoir water level series are broken down into five finite subsequences (IMF) and a residual function, while rainfall series yield four IMF subsequences and a residual function. The decomposition outcomes are graphically represented in Fig. 7, 8, and 9.

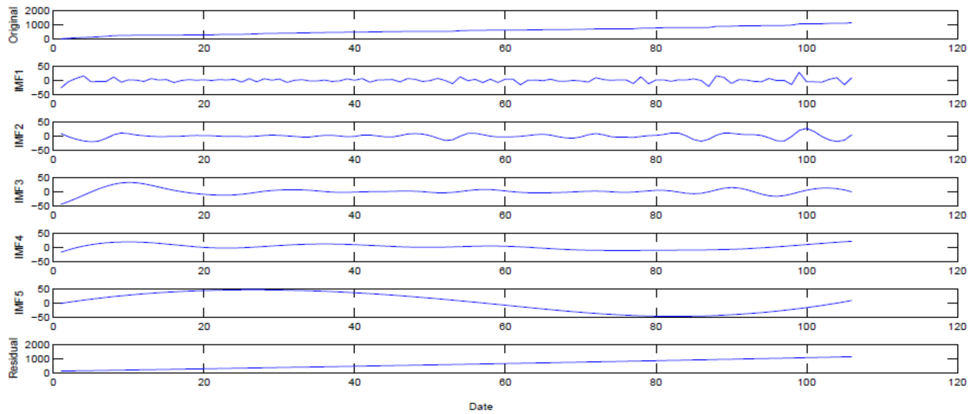

Fig. 7. EEMD decomposition of displacement.

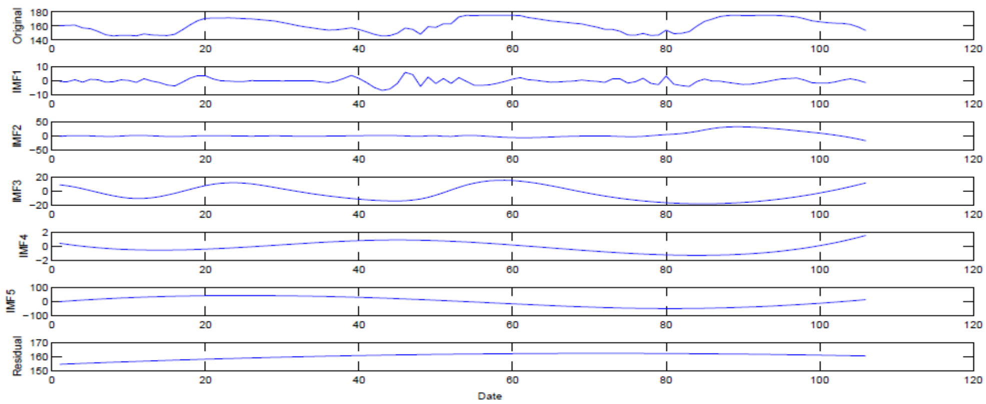

Fig. 8. EEMD decomposition of reservoir level.

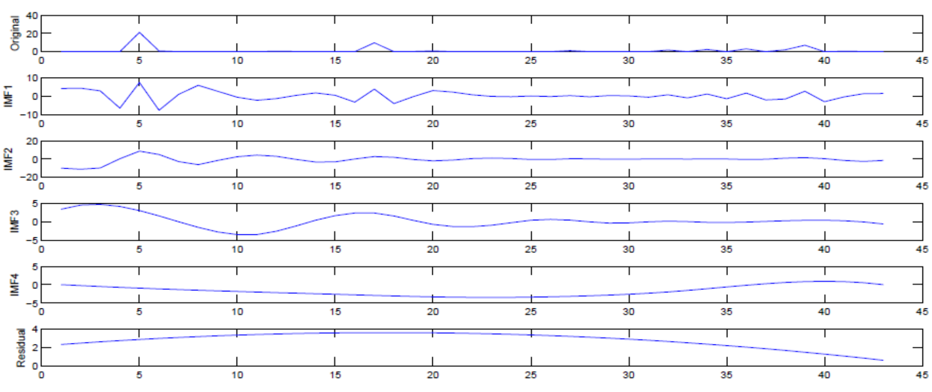

Fig. 9. EEMD decomposition of rainfall.

The choice of input parameters is critical to the outcome of the prediction, where we compute the correlation between each subsequence and the original displacement by MI (see Table I). In Table I, C represents for category, DD represents for displacement, RL represents for Reservoir Level, RR represents for Rainfall. According to the value MI between original displacement and component decomposition in Table I, we chose seven values as input to build model, such as displacement decomposition IMF3, IMF4, IMF5, residual, reservoir water level IMF4, IMF5, and Residual. The value MI between original displacement and decomposition of Rainfall are the same, also is lowest among three values. So the rainfall is not as inputs in the paper. The process of selecting sub-series for forecast model construction involves segmenting the data-set based on relevant criteria to represent key patterns and features, ensuring a balanced representation of training and testing sub-series.

TABLE I. CALCULATE THE MI VALUES

C	IMF1	IMF2	IMF3	IMF4	IMF5	RESIDUE
DD	0.0739	0.1360	0.5523	1.1587	1.5003	2.7303
RL	0.1114	0.1168	0.2243	0.3214	0.3795	0.4800
RF	0.1163	0.1163	0.1163	0.1163	0.1163	0.1163

B. Analysis and Results

Then we choose eight kernel functions to train each input separately, and compute the number of RVM. Each input parameter use eight kernel functions to compute. According to the minimum number of computed RVM rules, each input parameter can choose best kernel function. That is mean each input parameter have own kernel function to compute. The model uses many different kernel functions to build. In Table II, D, E, F, G, H, I, J stand for displacement decomposition (IMF3, IMF4, IMF5, Residual), reservoir water level (IMF4, IMF5, Residual). 0 cannot be computed by kernel functions, other number means that can be computed by kernel functions and the number of kernel functions. Each variable selects different kernel functions as much as possible base on least number of using RVM.

The symbols A through G correspond to various kernel functions: A represents the Gauss, B the Laplace, C the R, D the Spline, E the Cubic kernel function (chosen twice), F the Cauchy kernel function, and G the Thin-plate spline (TPS) kernel function. In Table II, all data set can be computed only by two kernel functions. One is Cublic, the other is Bubble. Because the prerequisite of SBL is that the array of Hessian should be positive definite. Then it can be decomposed by Cholesky. Then, in this paper, we use hybrid kernel models, Cublic kernel model and Cholesky kernel model to build our model.

TABLE II. 8 KERNEL FUNCTIONS FOR EACH COMPONENT

Category	A	B	C	D	E	F	G
Rvm-Gauss	5	7	2	7	7	3	0
Rvm-Cauchy	0	15	0	52	45	2	0
Rvm-Cublic	5	7	8	6	5	2	6
Rvm-Bubble	5	52	52	52	29	28	23
Rvm-Laplace	18	5	45	29	6	48	0
Rvm-R	18	49	2	28	7	5	0
Rvm-Spline	3	7	0	2	0	7	15
Rvm-Tps	2	44	4	49	7	7	0

Measuring the quality of algorithms involves various commonly employed methods, including the Relative Error (RE), Root Mean Square Error (RMSE), Mean Absolute Error (MAE), and Correlation Coefficient (R), as shown in Eq. (23) to Eq. (26).

$$RE = \frac{\left|\hat{Y}_i - Y_i\right|}{Y_i} \tag{23}$$

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(x_{texti} - x_{reali})^2} \tag{24}$$

$$MAE = \frac{1}{n}\sum_{i=1}^{n}\left|x_{testi} - x_{reali}\right| \tag{25}$$

$$R = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2 \sum_{i=1}^{n}(y_i - \bar{y})^2}} \tag{26}$$

The findings pertaining to three distinct kernel functions are presented in Fig. 10, Fig. 11, and Table III. Fig. 10 illustrates that the predicted values deviate slightly from the actual values. Notably, the first 38 data points utilizing the hybrid kernel function align most closely with the original data, followed by the Cubic kernel function for the remaining data. While the Bubble kernel function exhibits a similar trend to the hybrid kernel, its performance is inferior. In Fig. 11, the relative values of these three methods mirror the patterns observed in Fig. 10. Notably, the hybrid kernel function averages the best prediction results among the three methods.

In addition to these metrics, we computed four additional values for the three kernel functions: MAE, RMSE, R, and the number of RVM. The evaluation criteria for MAE, RMSE, and the number of RVM variables favor lower values, whereas a higher value is preferred for R. The hybrid kernel function achieved the minimum values for MAE, RMSE, and the number of RVM, while attaining the maximum value for R. According to the current evaluation standards, the hybrid kernel function demonstrates superior predictive performance. The hybrid approach involves selecting the most appropriate kernel calculation for each variable, thereby leveraging the unique characteristics of each kernel.

TABLE III. COMPARISON OF THREE METHODS

Method	MAE	RMSE	R	RVM
Cublic	266.8843	273.5266	0.9873	42
Bubble	280.1885	286.6568	0.9593	204
Hybrid	244.6038	247.7012	0.9710	38

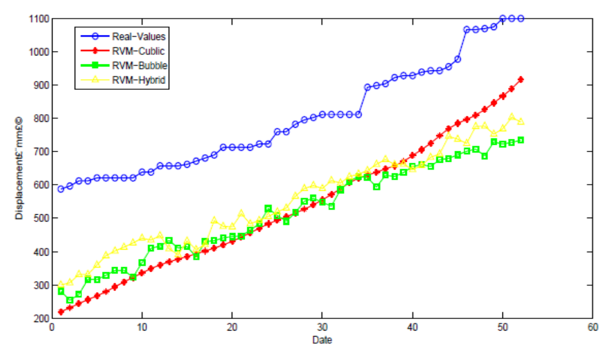

Fig. 10. Three methods predictive values.

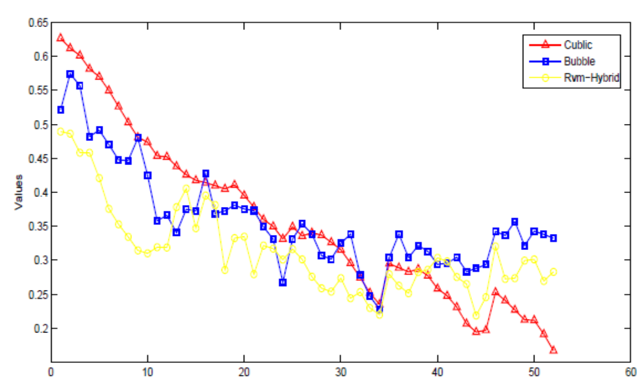

Fig. 11. Three methods relative values.

V. Conclusion

Employing the principles of decomposition and ensemble, we commence by decomposing three distinct types of landslide data using EEMD methods. This decomposition results in three separate groups, each containing multiple subseries. Subsequently, we utilize mutual information (MI) to assess the correlation between each subseries and landslide displacement, enabling us to identify potential input variables for our forecast model. Next, we select specific subseries to construct forecast models using support vector regression with mixed kernels. Ultimately, the results of these predictive models are combined to reconstitute the initial landslide displacement sequence. To showcase the potency of our model across varying kernels, we provide a case study centered on the LSJ landslide monitoring site ZJG24 in the vicinity of China Three Gorges. The EEMD-SBLMK method we introduce is notably beneficial due to its suitability for single-step-ahead (SS) forecasts in real-world situations. Additionally, it possesses the capability for precise multi-step-ahead (MS) forecasts down the line.

Acknowledgment

The work was supported by the Natural Science Foundation of China under Grants 62273136 and 61976085.

References

[1] S. Q. Qin, J. J. Jiao, S. J. Wang, “The predictable time scale of landslides,” *Bulletin of Engineering Geology and the Environment*. Vol. 59, No. 4, pp. 307-312, 2001.

[2] C. T. Chen, M. L. Lin, K. L. Wang, “Landslide seismic signal recognition and mobility for an earthquake-induced rockslide in Tsaoling, Taiwan,” *Engineering Geology*, Vol. 171, pp. 31-44, 2014.

[3] D. P. Kanungo, S. Sarkar, S. Sharma, “Combining neural network with fuzzy, certainty factor and likelihood ratio concepts for spatial prediction of landslides,” *Nat Hazards Rev*, Vol. 59, No. 3, pp. 1491-1512, 2011.

[4] R. N. Bi, D. Ehret, W. Xiang, “Landslide reliability analysis based on transfer coefficient method: a case study from Three Gorges Reservoir,” *Journal of Earth Science*, Vol. 23, No. 2, pp. 187-198, 2012.

[5] A. Z. Wang, M. H.Wang, X. Y. Li, Z. T. Mi, H. Zhou, “A two-stage bayesian integration framework for salient object detection on light field,” *Neural Processing Letters*, Vol. 46, No. 3, pp. 1083-1094. 2017.

[6] Wang, J. J., Li, Y. N. “Short-Term wind speed prediction using signal preprocessing technique and evolutionary support vector regression,” *Neural Processing Letters*, Vol. 46, No. 3, pp. 1-19, 2017.

[7] Z. S. He, M. H. Wang, C. H. Li, Y. L. Shen, A. P. He, “A Hybrid model equipped with the minimum cycle decomposition concept for short-term forecasting of electrical load time series,” *Neural Processing Letters*, Vol. 46, No. 3, pp. 1059-1081, 2017.

[8] H. Q. Chen, Z. G. Zeng, “Deformation prediction of landslide based on improved back-propagation neural network,” *Cognitive Comput*, Vol. 5, No. 1, pp. 56-62, 2013.

[9] C. Lian, Z. G. Zeng, W. Yao, H. M. Tang, “Displacement prediction model of landslide based on a modified ensemble empirical mode decomposition and extreme learning machine,” *Nat Hazards*, Vol. 66, pp. 759-771, 2013.

[10] C. Lian, Z. G. Zeng, W. Yao, H. M. Tang, “Extreme learning machine for the displacement prediction of landslide under rainfall and reservoir level,” *Stoch Enviorn Res Risk Assess*, Vol. 28, No. 8, pp. 1957-1972, 2014.

[11] F. Grasso, A. Luchetta, S. Manetti, “A multi-valued neuron based complex elm neural network,” *Neural Processing Letters*, 2017.

[12] P. J. Chang, J. S. Zhang, J. Y. Hu, Z. J. Song, “A deep neural network based on elm for semi-supervised learning of image classification,” *Neural Processing Letters*. 2017.

[13] J. J. Chen, Z. G. Zeng, P. Jiang, H. M. Tang, “Deformation prediction of landslide based on functional network,” *Neurocomputing*, Vol. 149, pp. 151-157, 2015.

[14] J. J. Chen, Z. G. Zeng, P. Jiang, H. M. Tang, “Application of multi-gene genetic programming based on separable functional network for landslide displacement prediction,” *Neural Computing & Application*, Vol. 27, pp. 1771-1784, 2016.

[15] H. A. Nefelioglu, Gokceoglu, C., Sonmez, H. “An assessment on the use of logistic regression and artificial neural networks with different sampling strategies for the preparation of landslide susceptibility maps,” *Engineering Geology*, Vol. 97, pp. 171-191, 2008.

[16] Z. H. Guo, W. G. Zhao, H. Y. Lu, J. Z. Wang, “Multi-step forecasting for wind speed using a modified EMD-based artificial neural network model,” *Renewable Energy*, Vol. 37, pp. 241-249, 2012.

[17] M. Yang, S. Fan, W. J. Lee, “Probabilistic short-term wind power forecast componential sparse bayesian learning,” *IEEE Transaction on Industry Application*, Vol. 49, No. 6, pp. 2783-2792, 2013.

[18] M. E. Tipping, A. C. Faul, “Fast marginal likelihood maximization for sparse bayesian model,” *In Proceedings of the Ninth International Workshop on Artificial Intelligence and Statistics*, 2003.

[19] G. J. Wang, C. Xie, S. Chen, J. J. Yang, M. Y. Yang, “Random matrix theory analysis of cross-correlations in the US stock market: Evidence from pearson’s correlation coefficient and detrended cross-correlation coefficient,” *Physica a Statistical Mechanics & Its Applications*, Vol. 392, No. 17, pp. 3715-3730, 2013.

[20] A. Kraskov, H. Stogbauer, P. Grassberger, “Estimating mutual information,” *Physical Review E*, Vol. 69, No. 6, pp. 066138, 2004.

[21] S. Frenzel, B. Pompe, “Partial mutual information for coupling analysis of multivariate time series,” *Physical Review Letters*, Vol. 99, No. 20, pp. 204101, 2007.