Dynamic Displacement Forecasting of Dashuitian Landslide in China Using Variational Mode Decomposition and Stack Long Short-Term Memory Network

Yin Xing [1] **, Jianping Yue** [1,*]**, Chuang Chen** [2]**, Kanglin Cong** [3]**, Shaolin Zhu** [1] **and Yankai Bian** [1]

[1]   College of Earth Sciences and Engineering, Hohai University, Nanjing 211100, China
[2]   College of Automation Engineering, Nanjing University of Aeronautics and Astronautics,
       Nanjing 211106, China
[3]   College of Information Science and Engineering, Shandong Agricultural University, Taian 271018, China
[*]   Correspondence: yuejping@163.com; Tel.: +86-139-1597-7693

check for
updates

Abstract: In recent decades, landslide displacement forecasting has received increasing attention due to its ability to reduce landslide hazards. To improve the forecast accuracy of landslide displacement, a dynamic forecasting model based on variational mode decomposition (VMD) and a stack long short-term memory network (SLSTM) is proposed. VMD is used to decompose landslide displacement into different displacement subsequences, and the SLSTM network is used to forecast each displacement subsequence. Then, the forecast values of landslide displacement are obtained by reconstructing the forecast values of all displacement subsequences. On the other hand, the SLSTM networks are updated by adding the forecast values into the training set, realizing the dynamic displacement forecasting. The proposed model was verified on the Dashuitian landslide in China. The results show that compared with the two advanced forecasting models, long short-term memory (LSTM) network, and empirical mode decomposition (EMD)–LSTM network, the proposed model has higher forecast accuracy.

Keywords: landslide; displacement forecasting; variational mode decomposition; stack long short-term memory network

1. Introduction

Landslides, as one of the most widespread and frequent natural hazards all over the world, not only directly threaten human life but also cause tremendous damage to the human living environment, resources, and property. To minimize the losses caused by landslide hazards, many scholars have carried out research on landslide prediction [1–5]. Among them, landslide displacement forecasting has been a contentious issue [6–9].

At present, one of the most widely used ideas for landslide displacement forecasting is to decompose the original landslide displacement, and then forecast each subsequence separately, and finally reconstruct all forecast values to obtain the forecast results. Following this idea, in [10], wavelet analysis is used to decompose the landslide displacement and a particle swarm-optimized support vector machine is used as the forecasting model. In [11], empirical mode decomposition (EMD) is matched with an extreme learning adaptive neuro-fuzzy inference system for landslide displacement forecasting. In [12], ensemble empirical mode decomposition (EEMD) is adopted to analyze the landslide displacement, which avoids the mode aliasing problem in EMD. Similarly, the authors of [13] also used the EEMD technique to decompose rainfall, reservoir level, and landslide

cumulative displacement sequences into residual sequence and finite intrinsic mode functions with frequencies ranging from high to low. Moreover, the forecasting model combined with intelligent algorithms, such as genetic algorithm-least squares support vector machine [14], genetic algorithm-back propagation neural network [15], and particle swarm-optimized support vector machine [16], is favored by researchers.

Some achievements have been made in the above studies; however, there are some shortcomings in displacement analysis and forecasting models. For displacement analysis, wavelet analysis has some difficulties in determining the basis function and estimating the wavelet order; EMD and EEMD are random and uncontrollable for the decomposition number of landslide displacement, while excessive decomposition will increase the forecasting workload. For forecasting models, neural networks have a certain dependence on the size of data samples, and the selection of the support vector machine parameters has certain blindness due to manual setting, while intelligent algorithms can easily fall into a local optimum. Therefore, how to more scientifically and reasonably analyze the original time series and establish a high-performance forecasting model are the difficulties of landslide displacement forecasting research.

To improve the forecast accuracy of landslide displacement, a dynamic forecasting model based on variational mode decomposition (VMD) and a stack long short-term memory network (SLSTM) is proposed. The main contributions of this paper are as follows:

(1) VMD with strict mathematical theory and controllable decomposition number is applied to displacement analysis, which obtains stable and fixed displacement subsequences;

(2) A SLSTM network with "3 + 1" layers is designed to model and forecast each displacement subsequence, which improves the performance of a basic long short-term memory (LSTM) network;

(3) A dynamic forecasting method is proposed, which realizes long-term landslide displacement forecasting.

The rest of this paper is organized as follows: Section 2 describes the basic principle and training method of the SLSTM network. Section 3 is devoted to describing the combined process of the VMD and the SLSTM network. The dynamic forecasting process based on the VMD-SLSTM network is also presented in this section. In Section 4, a real application case is introduced to verify the proposed forecasting model. Several comparative experiments are carried out to demonstrate the superiority of the proposed forecasting model. The last section draws conclusions.

2. Stack Long Short-Term Memory Network

2.1. LSTM Network

Recently, due to the advantages of deep learning in automatic feature extraction and high recognition rate or prediction accuracy, it has been successfully applied to speech recognition [17], action recognition [18], remaining useful life prognosis [19], traffic flow prediction [20], and other fields. As a commonly used deep learning model, the recurrent neural network (RNN) is an effective method for modeling dynamic sequences [21–28]. However, a RNN struggles to model long sequences because of gradient disappearance. To solve this problem, Hochreiter et al. proposed the long short-term memory (LSTM) network [29]. It uses multiple functionally distinct gates to control neurons and store information, thus having the ability to store important information for longer periods of time. The information is achieved by dot product $\otimes$ of activation function. Each gate state relies on a series of parameters trained by the gradient descent method.

Figure 1 shows the structure of the LSTM network in detail. Each gate in the LSTM network has a specific and unique function. Forget gate f decides which information should be discarded from the previous state h_{t-1}. Input x_t and previous state h_{t-1} after update gate u operation, together with the revised forget gate f, determine how much weight the candidate state $\widetilde{h_t}$ should use to update state h_t. To generate output ..., a non-linear function g_2 is used to filter its current state, and then it is returned after output gate operation. Herein, the returned partial state y_t is taken as the next input

y_{t-1}. Each gate depends on the current external input x_t and the previous output y_{t-1}. The updated process of the LSTM network is given by Equation (1).
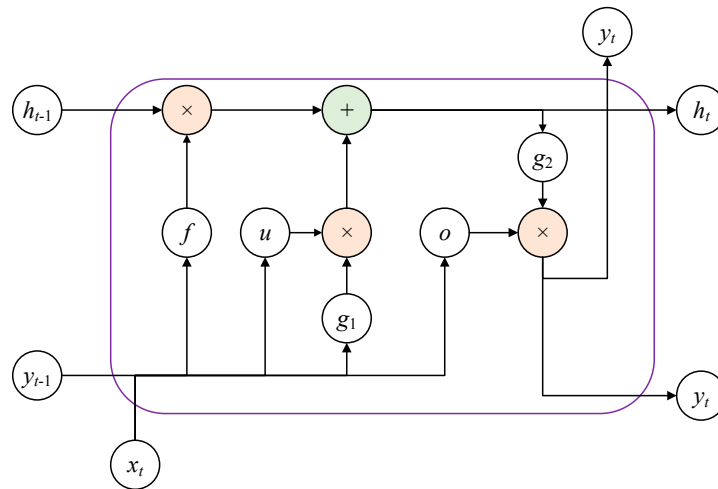

Figure 1. Structure of the long short-term memory (LSTM) network.

$$\begin{cases} forget\ gate : f_t = \sigma(W_f x_t + R_f y_{t-1} + b_f) \\ candidate\ state : \widetilde{h}_t = g_1(W_f x_t + R_f y_{t-1} + b_f) \\ update\ state : u_t = \sigma(W_u x_t + R_u y_{t-1} + b_u) \\ cell\ state : h_t = u_t \otimes \widetilde{h}_t + f_t \otimes h_{t-1} \\ output\ gate :_t = \sigma(W x_t + R y_{t-1} + b) \\ output : y_t =_t \otimes g_2(h_t) \end{cases} \tag{1}$$

where x_t is the input vector at t time; W_f, W_h, W_u, and W are the weight matrices associated with the input unit; R_f, R_h, R_u, and R are the weight matrices for hidden layer connections; b_f, b_h, b_u, and b are bias vectors; σ is the sigmoid or tanh function; g_1 and g_2 are the non-linear activation functions; and $\otimes$ denotes the dot product operation.

2.2. Training Method of the SLSTM Network

The architecture of the SLSTM network is shown in Figure 2. It consists of n LSTM layers and one fully connected layer. To reduce the model dimensions, the nth LSTM layer outputs a one-dimensional vector. The SLSTM network consists of two processes: forward computation and back propagation.

Similar to an artificial neural network (ANN), the forward calculation of the SLSTM network is given by

$$Y_t = V f(U X_t + W f(U X_{t-1} + W f(U X_{t-2} + \cdots))) \tag{2}$$

where U is the weight matrix of input X, W is the weight matrix of this input from the previous state h_{t-1}, f is the activation function, and V is the weight matrix of output layer.

On the other hand, the SLSTM network uses the back propagation through time (BPTT) algorithm [30]. The main idea of the BPTT algorithm is to train the SLSTM network using a back propagation algorithm after it is unfolded. Specifically, the error between the actual output and the expected output is calculated, and then the weight matrix is adjusted by the minimum error. Finally, the weight gradient is obtained and updated iteratively.
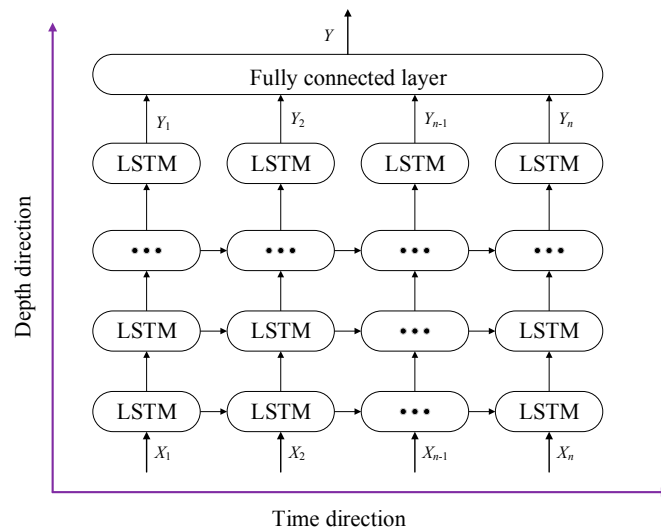
Figure 2. Architecture of the stack long short-term memory (SLSTM) network.

3. Dynamic Forecasting Model Based on VMD-SLSTM Network

3.1. Variational Mode Decomposition

Unlike EMD, the intrinsic mode function (IMF) in VMD is redefined as an amplitude modulated–frequency modulated signal $u_k(t)$, as follows [31]:

$$u_k(t) = A_k(t) \cos(\phi_k(t)) \tag{3}$$

where $A_k(t)$ and $\phi_k(t)$ are the instantaneous amplitude and phase.

In VMD, to obtain IMF components, the signal stripping method used in EMD is discarded. Instead, after setting the parameters such as the mode number K, the penalty parameter a, and the rising step τ, the signal is decomposed by finding the optimal solution of the model shown below in the variational framework:

$$\left.\begin{array}{l} \min\limits_{\{u_k\},\{\omega_k\}} \left\{ \sum\limits_{k=1}^{K} \|\partial_t\left[(\sigma(t) + \frac{j}{\pi t})u_k(t)\right]e^{-j\omega_k t}\|_2^2 \right\} \\ s.t. \quad \sum\limits_{k=1}^{K} u_k = f(t) \end{array}\right\} \tag{4}$$

where $\{u_k\}$ is the component that is finally obtained by decomposition; $\{\omega_k\} = \{\omega_1, \ldots, \omega_K\}$ is the actual center frequency of each IMF component, $[\sigma(t) + j/(\pi t)]u_k(t)$ is the analytical signal of each IMF component, $e^{-j\omega_k t}$ is the estimated center frequency of each analytical signal, and $f(t)$ is the original signal.

To solve for the optimal solution of the constrained variational model, Equation (4) needs to be transformed into an unconstrained model, as follows:

$$\begin{aligned} &L(\{u_k\}, \{u_k\}, \lambda) = \\ &a\sum\limits_{k} \|\partial_t\left[(\sigma(t) + \frac{j}{\pi t})u_k(t)\right]e^{-j\omega_k t}\|_2^2 + \|f(t) - \sum\limits_{k} u_k(t)\|_2^2 + \left\langle \lambda(t), f(t) - \sum\limits_{k} u_k(t) \right\rangle \end{aligned} \tag{5}$$

where λ is the Lagrange multiplier.

The saddle point of the unconstrained model is obtained by using the alternating direction multiplier algorithm, which is the optimal solution of the constrained variational model. Hence, the original signal $f(t)$ is decomposed into K IMF components.

3.2. VMD-SLSTM Forecasting Model

The landslide displacement-time curve is a non-stationary time series. If a forecasting model is applied directly to the landslide displacement, it is very difficult to learn the intrinsic law of displacement change. A better idea is to decompose the original displacement into several stationary subsequences that are more suitable for modeling.

Figure 3 shows the landslide displacement forecasting flow chart based on the VMD-SLSTM network. The landslide displacement is decomposed into *K* components by the VMD algorithm. On the other hand, the SLSTM network is used to forecast each component. Hence, the forecast values of landslide displacement are obtained by reconstructing the forecast values of all components.
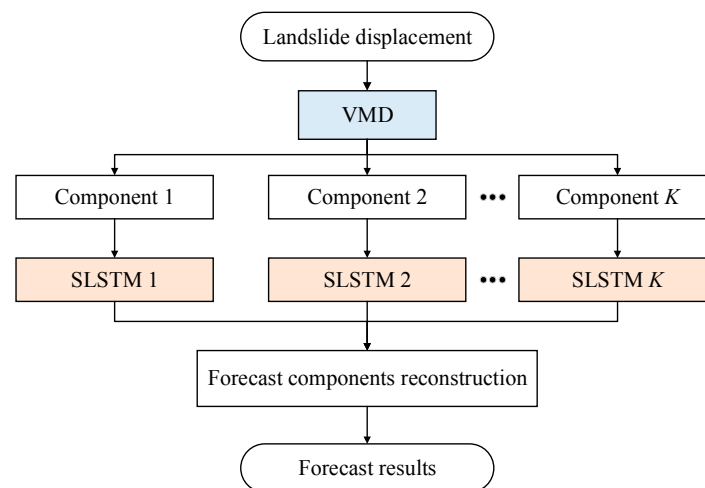

Figure 3. Landslide displacement forecasting flow chart based on the variational mode decomposition (VMD)-SLSTM network.

3.3. Dynamic Forecasting Process

For landslide displacement $[y_1, y_2, \cdots, y_{t-1}]$, to forecast the displacement value at $t + q$ time, Figure 4 shows the dynamic forecasting process based on VMD-SLSTM network. The main steps are as follows:
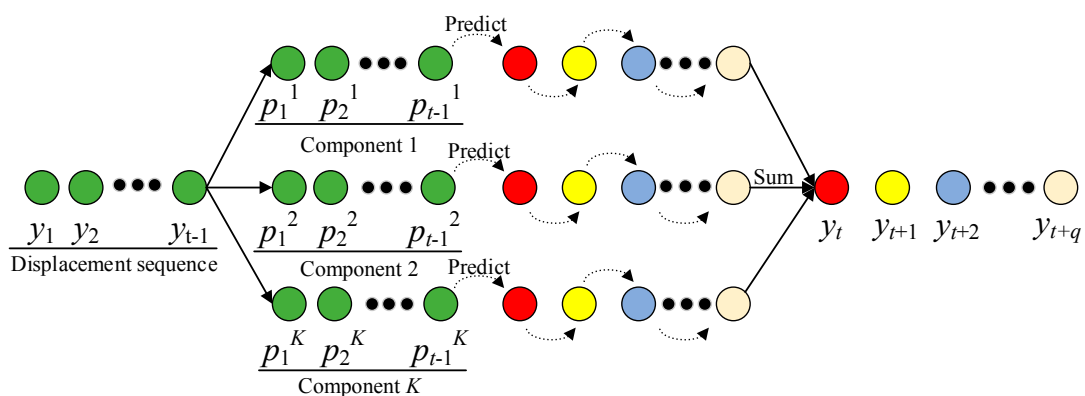

Figure 4. Dynamic forecasting process based on the VMD-SLSTM network.

Step 1 The displacement is decomposed into *K* components by the VMD: $[p_1^i, p_2^i, \cdots, p_{t-1}^i]$ with $i = 1, 2, \cdots, K$.

Step 2 Each component $\hat{p}_t^i$ is forecasted using the SLSTM network, respectively.

Step 3 The sum of *K* forecast components denotes the final forecast displacement $\hat{y}_t$.

Step 4 Each forecast component $\hat{p}_t^i$ is added to the input, and the SLSTM network is updated to forecast the next value.

Step 5 Steps 2–4 are repeated q times, and the sum of each forecast component $\hat{p}_{t+q}^i$ denotes the final forecast displacement $\hat{y}_{t+q}$.

To evaluate the forecast accuracy of the VMD-SLSTM network, mean absolute error (MAE) and root mean square error (RMSE) are adopted, which are defined as follows:

$$RMSE = \sqrt{\frac{1}{q+1}\sum_{i=t}^{t+q}(\hat{y}_i - y_i)^2} \tag{6}$$

$$MAE = \frac{1}{q+1}\sum_{i=t}^{t+q}|\hat{y}_i - y_i| \tag{7}$$

4. A Real Application Case

4.1. Dashuitian Landslide

The Dashuitian landslide is located in Baiquan Village, Longjing Township, Wushan County, Chongqing City, China, as shown in Figure 5. In this paper, the landslide is taken as the research object. It is located at an elevation of 340.8–482.5 m and has a relative height of 141.7 m. The topographic cutting depth is generally about 110 m, and the largest cutting depth is 150 m. The landslide is located at the waist of the hillside, lying north to south tilt. The slope of the landslide ranges from 15° to 35°. The hillside is mainly planted with crops. The slope foot of the landslide is steeper, while the slope body and top are relatively gentle. The landslide tends to the north with a width of 400 m and a length of 380 m. The thickness of the overlying loose layer is generally 1.0–6.4 m.
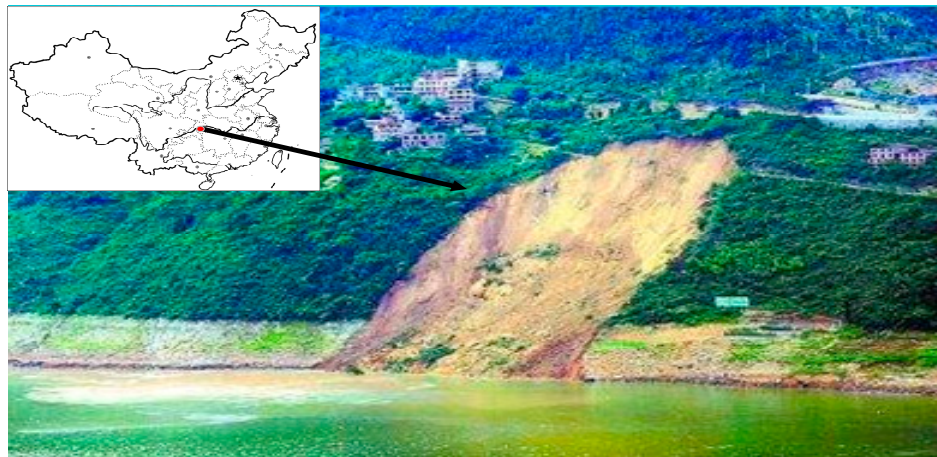

Figure 5. Location of the Dashuitian landslide in China.

To ensure the safety of people's lives and property, the Dashuitian landslide has been professionally monitored since 1 March 2009. Three GPS displacement detectors were installed to measure the deformation of the landslide. Monitoring point A1 was installed at the trailing edge of the landslide, where many cracks had developed. In this paper, the displacements measured at monitoring point A1 were used to establish a forecast model. The displacement was monitored every 10 days, and 36 measured values were obtained when the deformation monitoring ended on 1 March 2010. The landslide has an undulating oscillation characteristic due to non-periodic factors such as rainfall and human engineering activities during the evolution process, as shown in Figure 6.
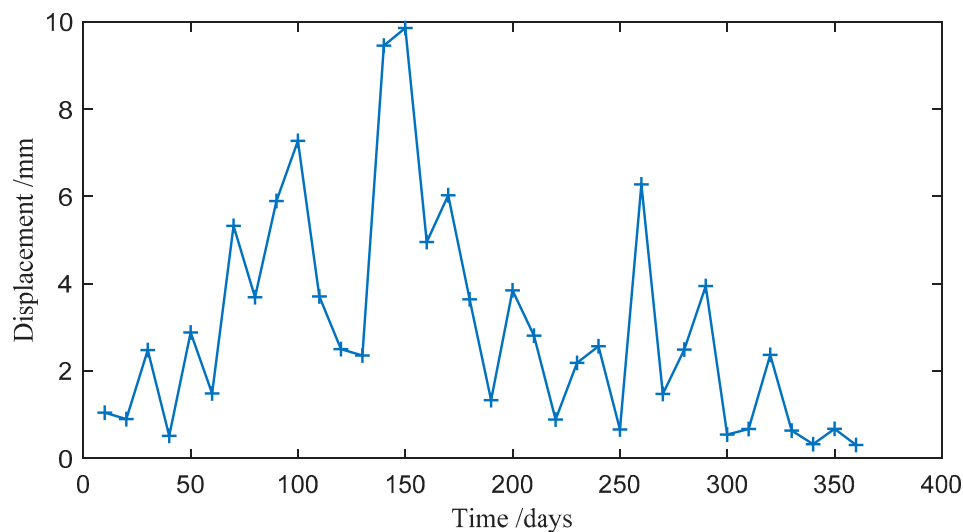
Figure 6. Displacement-time curve of the Dashuitian landslide.

4.2. Forecast Results Using the VMD-SLSTM Network

In the experiment, the measured data from the first 260 days were used as a training set, and the remaining data were used as a testing set. To ensure the fidelity after displacement decomposition, the penalty parameter $a = 0.5$ and the rising step $\tau = 0.1$ were finally determined through multiple trial calculations. After several tests, it was considered that $K = 3$ has better decomposition effect. Figure 7 shows the decomposition results of original displacement for the training set.
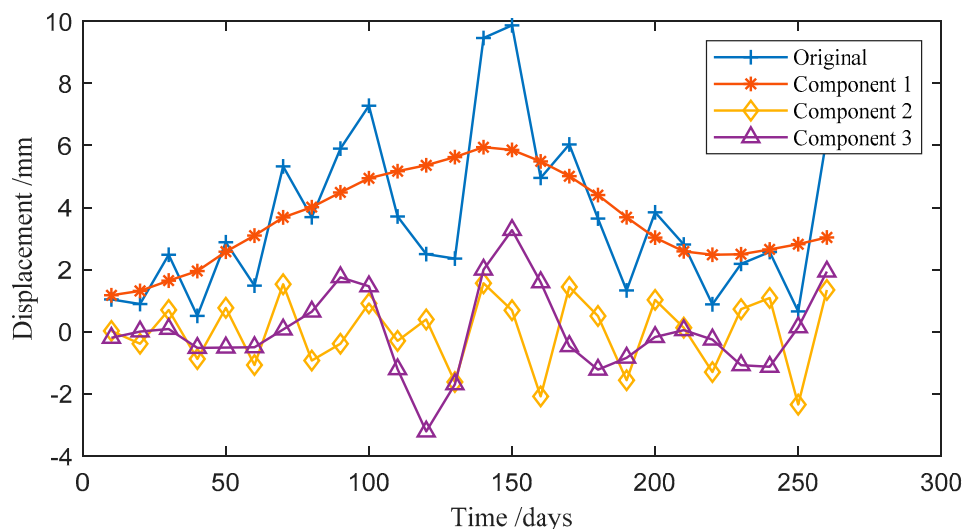

Figure 7. Decomposition results of original displacement for training set.

Next, the SLSTM network with "3 + 1" layers (3 LSTM layers and 1 fully connected layer) was built to learn the training set. To improve the convergence speed of the network, the data was normalized to have the same mean 0 and standard deviation 1 by Z-score. Figure 8 shows the forecast results of displacement subsequences using SLSTM networks. It can be seen that the forecast curve of each component conforms intuitively to the changing trend of the corresponding observed displacement subsequence.
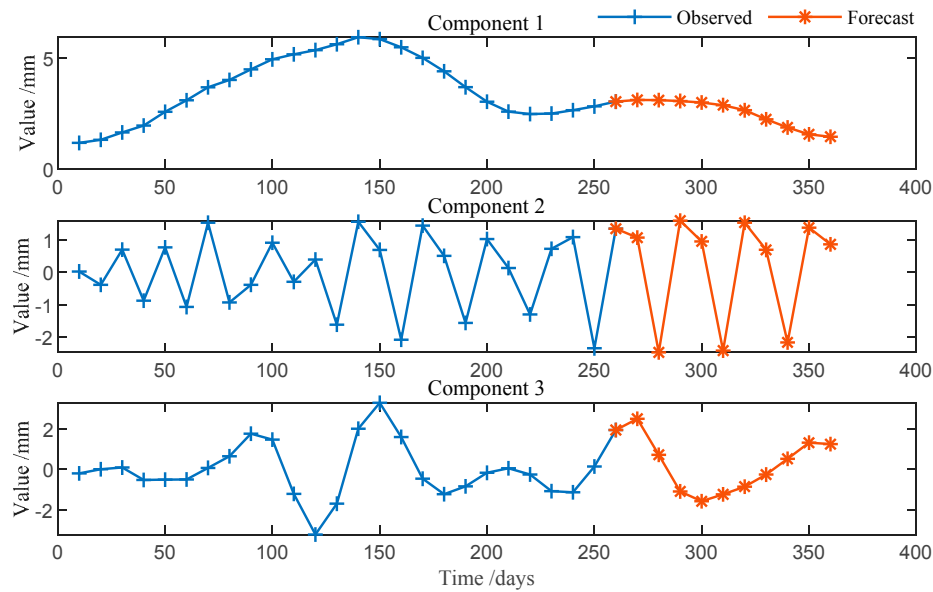
Figure 8. Forecast results of displacement subsequences using SLSTM networks.

Finally, the forecast results of the landslide displacement were obtained by reconstructing the forecast results of all displacement subsequences. Figure 9 shows the forecast results and errors of landslide displacement. It can be seen that the forecast curve can follow the observed curve well, indicating that the proposed model is valid. It is noted that after 320 days, the forecast results are not ideal. This is because although the forecast values are retrained, the accumulated errors will continue to impair the network performance.
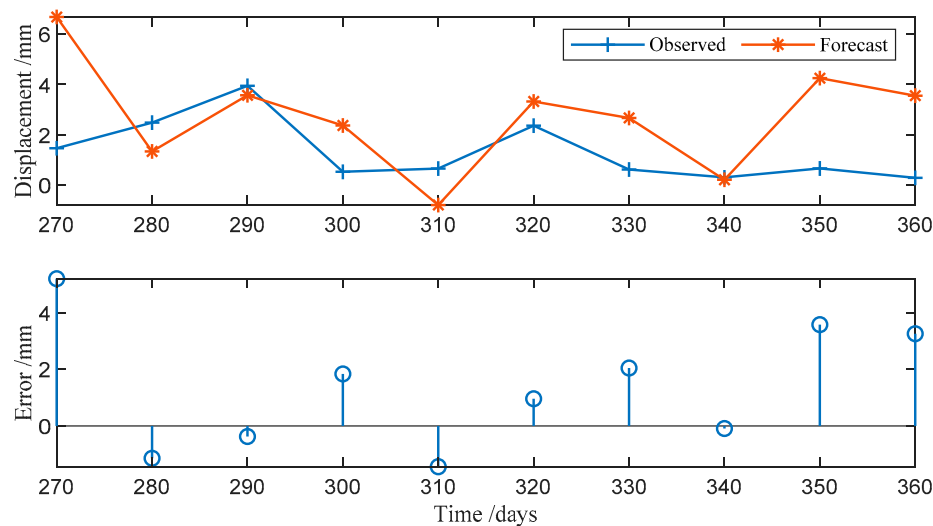

Figure 9. Forecast results and errors of landslide displacement.

4.3. Comparison with Other Forecasting Models

To demonstrate the superiority of the proposed VMD-SLSTM network, two advanced forecasting models—the EMD-LSTM network [21] and the LSTM network [32]—are used for comparison. Figure 10 shows the forecast results of landslide displacement using the three forecasting models. It can be seen that the forecast results of the LSTM network are the worst. This is because the original landslide displacement-time curve is a non-stationary oscillation curve, and it is difficult to learn the law of

displacement change if the LSTM network is used directly. The forecast curve of EMD-LSTM network fluctuates regularly, while that of VMD-SLSTM network can follow the observed curve well.
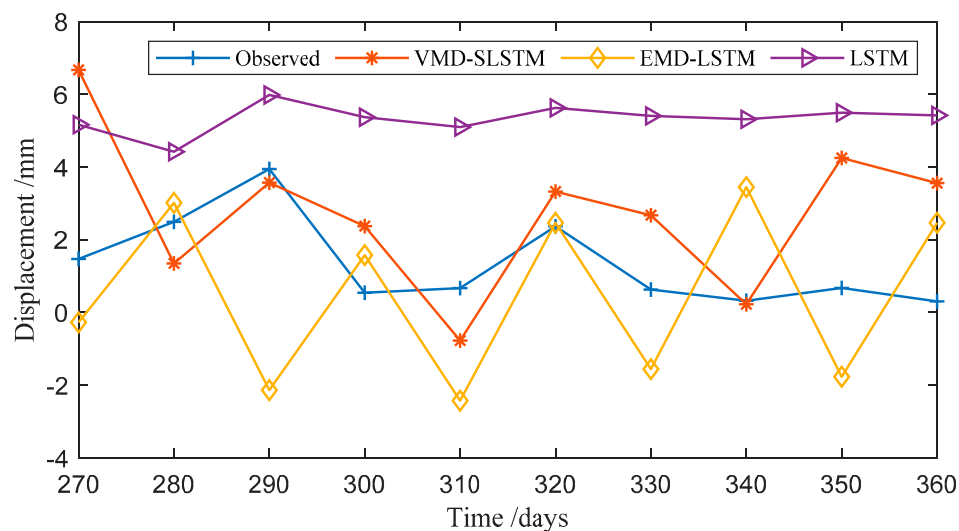

Figure 10. Forecast results of landslide displacement using three forecasting models: VMD-SLSTM, empirical mode decomposition (EMD)-SLSTM, and LSTM.

Table 1 reports the mean absolute error (MAE) and root mean square error (RMSE) of the forecast results using the three forecasting models. It can be seen that the MAE of the VMD-SLSTM network is 1.99 mm, lower than the 3.99 mm of the LSTM network and 2.25 mm of the EMD-LSTM network. On the other hand, the RMSE of the EMD-SLSTM network is 2.50 mm, lower than the 4.15 mm of the LSTM network and 2.76 mm of the EMD-SLSTM. These results show that the VMD-SLSTM network outperforms the LSTM network and EMD-LSTM network.

Table 1. Forecast errors of landslide displacement using three forecasting models.

Model	MAE/mm	RMSE/mm
LSTM network	3.99	4.15
EMD-LSTM network	2.25	2.76
VMD-SLSTM network	1.99	2.50

5. Conclusions

In this work, a dynamic forecasting model based on variational mode decomposition (VMD) and stack long short-term memory network (SLSTM) is proposed. The proposed model is successfully applied to the Dashuitian landslide in China. Compared with the two advanced forecasting models—the LSTM network and the EMD-LSTM network—the proposed model has higher forecast accuracy.

The use of the VMD method, which is backed by mathematical theory, is proposed to solve the problem of the incomplete or excessive decomposition problems caused by traditional methods such as EMD and EEMD. The landslide displacement is decomposed into stable and fixed components by the VMD method, which lays the foundation for each component forecasting task.

The use of the SLSTM network, which is a novel predictor that can remember historical information, is proposed to conduct each component forecasting task. As a deep learning method, the SLSTM network improves the network forecasting performance through stacking LSTM layers. Finally, the forecast values of landslide displacement are obtained by reconstructing the forecast values of all components.

Considering that the proposed forecasting model only achieves slightly better performance than the EMD-LSTM network when the dataset is small, the VMD-SLSTM network is promising if the dataset is expanded in future work. In addition, the complexity of the SLSTM network makes parameter

training more difficult. If better network parameters are trained, the prediction accuracy can be further improved. The residual compensation for the forecasting model should also be paid more attention.

Nomenclature

Acronyms		Notations	
ANN	artificial neural network	a	penalty parameter
BPTT	back propagation through time	A_k	instantaneous amplitude
EEMD	ensemble empirical mode decomposition	b_f, b_h, b_u, b	bias vectors
EMD	empirical mode decomposition	g_1, g_2	non-linear activation functions
GPS	global positioning system	h_t	state of cell at t time
IMF	intrinsic mode function	$\tilde{h}_t$	candidate state of cell
LSTM	long short-term memory	K	mode number
MAE	mean absolute error	R_f, R_h, R_u, R	weight matrices of hidden layer connections
RMSE	root mean square error	U	weight matrix of input
RNN	recurrent neural network	u_k	amplitude modulated-frequency modulated signal
SLSTM	stack long short-term memory network	V	weight matrix of output
VMD	variational mode decomposition	W_f, W_h, W_u, W	weight matrices associated with the input unit
		X	input matrix of network
		Y	output matrix of network
		τ	rising step
		λ	Lagrange multiplier
		ω_k	actual center frequency
		ϕ_k	instantaneous phase

References

1. Bui, D.T.; Pradhan, B.; Lofman, O.; Revhaug, I.; Dick, O.B. Spatial prediction of landslide hazards in Hoa Binh province (Vietnam): A comparative assessment of the efficacy of evidential belief functions and fuzzy logic models. *Catena* **2012**, *96*, 28–40.
2. Blahut, J.; Westen, C.J.V.; Sterlacchini, S. Analysis of landslide inventories for accurate prediction of debris-flow source areas. *Geomorphology* **2010**, *119*, 36–51. [CrossRef]
3. Ruette, J.V.; Papritz, A.; Lehmann, P.; Rickli, C.; Or, D. Spatial statistical modeling of shallow landslides—Validating predictions for different landslide inventories and rainfall events. *Geomorphology* **2011**, *133*, 11–22. [CrossRef]
4. Hong, H.; Pradhan, B.; Jebur, M.N.; Bui, D.T.; Xu, C.; Akgun, A. Spatial prediction of landslide hazard at the Luxi area (China) using support vector machines. *Environ. Earth Sci.* **2016**, *75*, 40. [CrossRef]
5. Chen, W.; Pourghasemi, H.R.; Panahi, M.; Kornejady, A.; Wang, J.; Xie, X.; Cao, S. Spatial prediction of landslide susceptibility using an adaptive neuro-fuzzy inference system combined with frequency ratio, generalized additive model, and support vector machine techniques. *Geomorphology* **2017**, *297*, 69–85. [CrossRef]
6. He, K.; Zhao, M.; Zhang, Y.; Zhang, J. Unload–load displacement response ratio parameter and its application in prediction of debris landslide induced by rainfall. *Environ. Earth Sci.* **2017**, *76*, 55. [CrossRef]
7. Lian, C.; Zeng, Z.; Yao, W.; Tang, H. Multiple neural networks switched prediction for landslide displacement. *Engl. Geol.* **2015**, *186*, 91–99. [CrossRef]
8. Huang, F.; Yin, K.; Zhang, G.; Gui, L.; Yang, B.; Liu, L. Landslide displacement prediction using discrete wavelet transform and extreme learning machine based on chaos theory. *Environ. Earth. Sci.* **2016**, *75*, 1376. [CrossRef]
9. Du, J.; Yin, K.; Lacasse, S. Displacement prediction in colluvial landslides, Three Gorges Reservoir, China. *Landslides* **2013**, *10*, 203–218. [CrossRef]

10. Ren, F.; Wu, X.; Zhang, K.; Niu, R. Application of wavelet analysis and a particle swarm-optimized support vector machine to predict the displacement of the Shuping landslide in the Three Gorges, China. *Env. Earth Sci.* **2015**, *73*, 4791–4804. [CrossRef]

11. Shihabudheen, K.V.; Pillal, G.N.; Peethambaran, B. Prediction of landslide displacement with controlling factors using extreme learning adaptive neuro-fuzzy inference system (Elanfis). *Appl. Soft Comput.* **2017**, *61*, 892–904.

12. Deng, D.; Liang, Y.; Wang, L.; Wang, C.; Sun, Z.; Wang, C.; Dong, M. PSO-SVR prediction method for landslide displacement based on reconstruction of time series by EEMD: A case study of landslides in Three Gorges Reservoir area. *Rock Soil Mech.* **2017**, *38*, 1001–1009.

13. Lian, C.; Zeng, Z.; Yao, W.; Tang, H. Extreme learning machine for the displacement prediction of landslide under rainfall and reservoir level. *Stoch. Environ. Res. Risk Assess.* **2014**, *28*, 1957–1972. [CrossRef]

14. Cai, Z.; Xu, W.; Meng, Y.; Shi, C.; Wang, R. Prediction of landslide displacement based on GA-LSSVM with multiple factors. *Bull. Engl. Geol. Environ.* **2016**, *75*, 637–646. [CrossRef]

15. Ran, Y.; Xiong, G.; Li, S.; Ye, L. Study on deformation prediction of landslide based on genetic algorithm and improved BP neural network. *Kybernetes* **2010**, *39*, 1245–1254. [CrossRef]

16. Zhou, C.; Yin, K.; Cao, Y.; Ahmed, B. Application of time series analysis and PSO-SVM model in predicting the Bazimen landslide in the Three Gorges Reservoir, China. *Engl. Geol.* **2016**, *204*, 108–120. [CrossRef]

17. Sak, H.; Senior, A.; Beaufays, F. Long short-term memory based recurrent neural network architectures for large vocabulary speech recognition. *Comput. Sci.* **2014**, *1128*, 338–342.

18. Du, Y.; Wang, W.; Wang, L. Hierarchical recurrent neural network for skeleton based action recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Boston, MA, USA, 7–12 July 2015; pp. 1110–1118.

19. Wang, C.; Lu, N.; Wang, S.; Cheng, Y.; Jiang, B. Dynamic long short-term memory neural-network- based indirect remaining-useful-life prognosis for satellite lithium-ion battery. *Appl. Sci.* **2018**, *8*, 2078. [CrossRef]

20. Ma, X.; Tao, Z.; Wang, Y.; Yu, H.; Wang, Y. Long short-term memory neural network for traffic speed prediction using remote microwave sensor data. *Transp. Res. Part C Emerg. Technol.* **2015**, *54*, 187–197. [CrossRef]

21. Xu, S.; Niu, R. Displacement prediction of Baijiabao landslide based on empirical mode decomposition and long short-term memory neural network in Three Gorges area, China. *Comput. Geosci.* **2018**, *111*, 87–96. [CrossRef]

22. Wang, H.; Wang, L. Modeling temporal dynamics and spatial configurations of actions using two-stream recurrent neural networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 499–508.

23. Krause, B.; Kahembwe, E.; Murray, I.; Renals, S. Dynamic evaluation of neural sequence models. *arXiv* **2017**, arXiv:1709.07432.

24. Chung, J.; Kastner, K.; Dinh, L.; Goel, K.; Courville, A.C.; Bengio, Y. A recurrent latent variable model for sequential data. In *Advances in Neural Information Processing Systems*; MIT Press: Cambridge, MA, USA, 2015; pp. 2980–2988.

25. Strobelt, H.; Gehrmann, S.; Pfister, H.; Rush, A.M. Lstmvis: A tool for visual analysis of hidden state dynamics in recurrent neural networks. *IEEE Trans. Vis. Comput. Graph.* **2017**, *24*, 667–676. [CrossRef] [PubMed]

26. Xue, W.; Nachum, I.B.; Pandey, S.; Warrington, J.; Leung, S.; Li, S. Direct estimation of regional wall thicknesses via residual recurrent neural network. In *International Conference on Information Processing in Medical Imaging*; Springer: Cham, Switzerland; Boone, NC, USA, 2017; pp. 505–516.

27. Bashivan, P.; Rish, I.; Yeasin, M.; Codella, N. Learning representations from EEG with deep recurrent-convolutional neural networks. *arXiv* **2015**, arXiv:1511.06448.

28. Coulibaly, P.; Baldwin, C.K. Nonstationary hydrological time series forecasting using nonlinear dynamic methods. *J. Hydrol.* **2005**, *307*, 164–174. [CrossRef]

29. Hochreiter, S.; Schmidhuber, J. Long short-term memory. *Neural Comput.* **1997**, *9*, 1735–1780. [CrossRef] [PubMed]

30. Bengio, Y.; Simard, P.; Frasconi, P. Learning long-term dependencies with gradient descent is difficult. *IEEE Trans. Neural Netw.* **1994**, *5*, 157–166. [CrossRef]

31. Dragomiretskiy, K.; Zosso, D. Variational mode decomposition. *IEEE Trans. Signal Process.* **2014**, *62*, 531–544. [CrossRef]

32. Zhao, Z.; Chen, W.; Wu, X.; Chen, P.C.Y.; Liu, J. LSTM network: A deep learning approach for short-term traffic forecast. *IET Intell. Transp. Syst.* **2017**, *11*, 68–75. [CrossRef]