Postprint: Annual Precipitation Prediction for the Tianshan North Slope Economic Belt Using an EEMD-LSTM Model

Authors: Yang Qian

Date: 2021-09-14T00:00:00+00:00

Abstract

Precipitation prediction is the core and challenge of modern climate forecasting operations. Research and application of coupled models in precipitation prediction in Xinjiang are extremely limited. Therefore, this study attempts to establish a coupled model of Ensemble Empirical Mode Decomposition (EEMD) and Long Short-Term Memory Network (LSTM) to conduct predictive research on precipitation in the Northern Tianshan Economic Belt. Annual precipitation data from 1965 to 2019 (55 years) for the Northern Tianshan Economic Belt were decomposed using EEMD, transforming them into four stationary components and a trend term. Spectral analysis was employed to determine the quasi-periodic characteristics of each component, providing a foundation for subsequent LSTM model training. An LSTM network model was trained based on the components obtained from EEMD decomposition and utilized to predict precipitation in the study area. The results indicate that the EEMD-LSTM coupled model achieved an average relative error of 13.38% and a root mean square error of 38.03 mm for precipitation prediction in the Northern Tianshan Economic Belt from 2010 to 2019, suggesting that the EEMD-LSTM coupled model exhibits good prediction accuracy for precipitation in this region. The EEMD-LSTM coupled model was further employed to forecast annual precipitation in the Northern Tianshan Economic Belt for 2020–2029, revealing six years with above-normal precipitation and four years with below-normal precipitation. The year 2025 may be an extremely wet year, with precipitation exceeding normal by over 20%, whereas 2021 is projected to be an extremely dry year, with precipitation expected to be below 200 mm. This paper explores a novel method for precipitation prediction in arid regions and provides a reference basis for meteorological disaster prevention and mitigation efforts.

Full Text

Prediction of Annual Precipitation in the Northern Slope Economic Belt of Tianshan Mountains Based on an EEMD-LSTM Model

YANG Qian[1,2], **QIN Li**[2], **GAO Pei**[1,2], **ZHANG Ruibo**[2] [1] Xinjiang Branch, China Meteorological Administration Training Centre, Urumqi 830013, Xinjiang, China
[2] Institute of Desert Meteorology, China Meteorological Administration, Urumqi 830002, Xinjiang, China

Abstract

Precipitation prediction is both the core and a major challenge in modern climate forecasting operations. Research and application of coupling models for precipitation prediction in Xinjiang remain extremely limited. Therefore, this study establishes a coupling model based on Ensemble Empirical Mode Decomposition (EEMD) and Long Short-Term Memory (LSTM) neural network to predict precipitation in the Northern Slope Economic Belt of the Tianshan Mountains. Annual precipitation data from the study area spanning 55 years (1965–2019) were decomposed using EEMD into four stationary components and a trend term. Spectral analysis was applied to obtain the quasi-periodic characteristics of each component, providing a foundation for subsequent LSTM model training. Each decomposed component was then used to train an LSTM network model, which was subsequently employed for precipitation prediction. The results indicate that the EEMD-LSTM coupling model achieved an average relative error of 13.38% and a root mean square error of 38.03 mm for predicting precipitation in the Northern Slope Economic Belt of the Tianshan Mountains from 2010–2019, demonstrating good prediction accuracy. The model was further applied to forecast annual precipitation for 2020–2029; within this period, six years are predicted to have above-normal precipitation and four years below-normal precipitation. Year 2025 is forecasted to be an extremely wet year with precipitation exceeding the multi-year average by more than 20%, while 2021 is anticipated as an extremely dry year with expected precipitation below 200 mm. This study explores a new method for precipitation prediction in arid regions and provides a reference basis for meteorological disaster prevention and mitigation.

Keywords: ensemble empirical mode decomposition; long short-term memory network; annual precipitation; prediction

Introduction

Climate prediction constitutes the core and focus of meteorological operations, with precipitation prediction being particularly challenging due to the randomness and uncertainty of precipitation events. Current precipitation forecasting methods primarily include statistical prediction and numerical modeling, with statistical approaches remaining the mainstream method for interannual and decadal precipitation prediction. Numerous studies have been conducted on annual precipitation prediction in China. Traditional statistical methods such as Autoregressive Moving Average (ARMA) and wavelet analysis have been widely applied. However, these single-model approaches have limitations in accuracy and precision due to their inherent constraints, resulting in unsatisfactory prediction performance.

With the development of signal processing technology and machine learning, research on annual precipitation prediction has increasingly focused on Ensemble Empirical Mode Decomposition (EEMD) and artificial neural networks. Coupling EEMD with artificial neural networks can compensate for the deficiencies of single models and is particularly suitable for predicting non-stationary, non-linear meteorological elements like precipitation. EEMD can decompose non-stationary time series into stationary Intrinsic Mode Function (IMF) components, while artificial neural networks can effectively fit non-linear relationships. The Tianshan Mountains, a unique mountain system in arid regions, intercept water vapor carried by westerly airflow, forming relatively abundant water resources in mountainous areas. The Northern Slope Economic Belt of the Tianshan Mountains, located at the northern foothills, represents the most economically developed and resource-advantaged region in Xinjiang. It also serves as a crucial hub for the Silk Road Economic Belt. However, issues such as uneven water resource distribution and frequent drought-flood disasters constrain local development. Therefore, improving precipitation prediction accuracy in this region holds significant applied value for production and livelihood.

Currently, research on applying EEMD-artificial neural network coupling models to annual precipitation prediction in Xinjiang remains scarce. Considering data availability and computational constraints, this study selects the Long Short-Term Memory (LSTM) network model, which is more suitable for time-series prediction among artificial neural networks. Referencing the EEMD-Adaboost model for commodity price prediction, this paper establishes an EEMD-LSTM coupling model for precipitation prediction in the Northern Slope Economic Belt of the Tianshan Mountains. The model is evaluated using data from 1965–2019, aiming to provide new insights for improving regional precipitation prediction models and decision-making support for water resource development and meteorological disaster prevention.

1. Data and Methods

1.1 Data Sources and Processing

Monthly precipitation data from 18 meteorological stations in the Northern Slope Economic Belt of the Tianshan Mountains were collected. After interpolation and removal of missing and anomalous values, the data were processed into annual precipitation series by averaging across stations. The 55-year annual precipitation series from 1965–2019 was then decomposed using EEMD to analyze the characteristics of each component. Meteorological data were obtained from the Xinjiang Meteorological Information Center of the China Meteorological Administration, while elevation data were derived from the General Bathymetric Chart of the Oceans (GEBCO) global ocean and land terrain model grid dataset.

1.2 Research Methods

1.2.1 EEMD Method Empirical Mode Decomposition (EMD) is a signal processing method for non-stationary time series. Ensemble Empirical Mode Decomposition (EEMD) improves upon EMD by adding Gaussian white noise sequences to effectively suppress mode mixing and scale mixing phenomena, ensuring that the decomposed Intrinsic Mode Functions (IMFs) maintain physical uniqueness. As this method requires no predetermined basis functions and is minimally affected by the amplitude of added white noise, it demonstrates good adaptability. The basic steps are as follows: First, add a white noise sequence to the original signal $x(t)$ to obtain a new signal series $X(t)$. Second, decompose $X(t)$ to obtain a set of IMFs. Third, repeat the first two steps N times, each time adding a different white noise. Finally, calculate the ensemble mean of the IMFs obtained from each decomposition, which cancels out the added white noise and yields the final result. Through these steps, IMFs at various scales and a trend term can be obtained.

1.2.2 LSTM Network The LSTM network is a variant of recurrent neural networks that effectively mines long-term dependency relationships in data transmission and mitigates the gradient vanishing problem during training. The unit structure of the LSTM network is shown in Figure 2 [Figure 2: see original paper], comprising a forget gate, input gate, output gate, and cell state information transmission components. The left and right sides represent recurrent structures of the same architecture in temporal sequence. The sigmoid and tanh functions serve as nonlinear activation functions. The combination and operation of these neurons (circular components in the unit) implement gate control functions and transmit results to the next unit structure. The nonlinear activation functions between input and output ensure the neural network can approximate any nonlinear function, while the temporal structure allows each moment' s input to be retained in the network and influence the next moment, completing long short-term memory transmission.

1.2.3 Evaluation Metrics Three evaluation metrics were selected: Root Mean Square Error (RMSE), Mean Relative Error (MRE), and anomaly sign consistency rate. The formulas are as follows:

$$\mathrm{RMSE} = \sqrt{\frac{1}{m}\sum_{t=1}^{m}(y_t - \hat{y}_t)^2}$$

$$\mathrm{MRE} = \frac{1}{m}\sum_{t=1}^{m}\left|\frac{y_t - \hat{y}_t}{y_t}\right| \times 100\%$$

where y_t represents observed values at time t, and $\hat{y}_t$ represents predicted values at time t.

2. Results

2.1 EEMD-LSTM Coupling Model Construction

The modeling procedure includes: (1) EEMD decomposition of annual precipitation data (1965–2019) into components and trend terms. (2) Dataset partitioning: component data from 1965–2009 as training set, 2010–2014 as validation set, and 2015–2019 as test set. (3) Normalization of each component using min-max scaling. (4) LSTM network training to determine network structure, parameters, and hyperparameters. The network structure is determined by input layer, hidden layer, and output layer. Input and output layer structures are determined by feature dimensions and time steps. Basic parameters are generated through supervised learning, while hidden layer neuron count, iteration times, and Dropout rate are determined through trial-and-error experiments. (5) Model performance evaluation using test set data. (6) Prediction of annual precipitation in the study area using the trained model and comparison with original data to assess model performance.

2.2 Model Testing and Assessment

2.2.1 EEMD Decomposition of Precipitation The 55-year annual precipitation series (1965–2019) was decomposed into four IMFs and one trend term. Each component contains information at different time scales, with the sum of components in any time period equaling the original series. The physical significance of each component can be judged through significance testing. As shown in Figure 4 [Figure 4: see original paper], only IMF3 is above the 95% confidence line, indicating it contains the most significant information, while other components contain relatively less information but still contribute to variance calculation, maintaining total signal energy.

The quasi-periods were calculated through spectral analysis, and variance contribution rates indicate the influence degree of each component' s frequency and amplitude on original data. As shown in Table 1 , IMF1 has a quasi-period of 2.9 years with a variance contribution rate of 54.54%, representing the most significant oscillation signal. IMF2 has a quasi-period of 7.5 years with 15.70% variance contribution, showing two relatively stable periods during 1970–1985 and 2000–2010. IMF3 has a quasi-period of 18.5 years with 7.75% variance contribution, with amplitude gradually decreasing over time. IMF4 has a quasi-period of 33 years with 20.15% variance contribution, fluctuating within about

20 units. The trend term has a variance contribution rate of 0.45%, indicating an upward trend with gradually slowing growth rate.

Xinjiang' s water vapor originates primarily from the Mediterranean Sea, North Atlantic, and Arctic Ocean, transported by westerly winds to the northern slope of Tianshan Mountains. This process is mainly influenced by the Arctic Oscillation (AO) and North Atlantic Oscillation (NAO). Therefore, precipitation quasi-periods in the study area should be consistent with AO/NAO cycles. Previous research has confirmed that precipitation in northern Xinjiang corresponds well with AO index at decadal scales, which aligns with our quasi-period results.

2.2.2 LSTM Network Training The LSTM network structure is determined by variable count, input time steps, and prediction time steps. With precipitation as the only variable, the variable count is 1. Input time steps should be greater than component periods but less than original data length. Based on this requirement and reference to previous studies, input time steps were set to 10. The activation function was ReLU, optimization function was Adam, and loss function targeted RMSE. Dropout was used to prevent overfitting. Hyperparameters were determined through trial-and-error experiments: fixing two hyperparameters while testing the third, selecting values minimizing RMSE on validation data. Final hyperparameters are shown in Table 2 .

2.2.3 Model Testing and Evaluation The EEMD-LSTM coupling model was used to predict annual precipitation from 2010–2019. Figure 5 [Figure 5: see original paper] compares predicted and observed values for each component. RMSE values for IMF1-IMF4 and trend term were 1.85 mm, 2.18 mm, 7.68 mm, 28.82 mm, and 3.89 mm respectively. Summing component predictions reconstructed annual precipitation predictions for 2010–2019. Absolute errors, relative errors, and anomaly signs were calculated (Table 3). The average relative error was 13.38%, RMSE was 38.03 mm, and anomaly sign consistency rate was 81.58%.

The model showed relatively poor prediction performance for 2015 and 2016, mainly due to large precipitation variations in these years. However, according to precision grading standards, average relative error between 10%-20% indicates good prediction performance. For comparison, ARIMA model was applied, yielding average absolute error of 84.76 mm, RMSE of 102.97 mm, and average relative error of 38.69%—all substantially larger than EEMD-LSTM, demonstrating the coupling model' s superiority.

A paired-sample T-test was conducted to verify significant differences between predicted and observed values. Results showed significant correlation ($r=0.82$, $p<0.01$). The 95% confidence interval for mean differences was [-12.34, 5.86] with two-tailed probability $p=0.47>0.05$, indicating no significant difference between predictions and observations (Table 4).

2.2.4 Annual Precipitation Prediction for 2020–2029 The EEMD-LSTM model was applied to predict annual precipitation for 2020–2029 (Table 5). Six years are predicted to have above-normal precipitation and four years below-normal. Year 2025 shows maximum precipitation, exceeding the multi-year average by 18.87%, while 2021 shows minimum precipitation below 200 mm, indicating an extremely dry year.

3 Discussion

Interannual and decadal precipitation variations on the northern slope of Tianshan Mountains are primarily influenced by mid-high latitude circulation in the Northern Hemisphere, showing synchronous response to AO/NAO disturbances. Research has found that atmospheric teleconnection patterns correspond well with IMF components of precipitation series in frequency domain, demonstrating that EEMD can adaptively extract multi-timescale information from precipitation signals with physical significance. EEMD transforms non-stationary time series into multiple stationary components plus a trend term, providing a solid foundation for subsequent modeling. Neural network models can fit arbitrary nonlinear relationships, making them suitable for precipitation modeling. However, their generalization ability is relatively low, with long-term prediction accuracy decreasing significantly when data patterns change substantially. Therefore, data stationarity positively impacts neural network prediction accuracy, highlighting the necessity of combining EEMD with LSTM. While the established coupling model shows good performance, improvements can be made by incorporating reanalysis grid data and atmospheric circulation factors as input variables in future work.

4 Conclusion

This study established an EEMD-LSTM coupling model to predict precipitation in the Northern Slope Economic Belt of Tianshan Mountains. The 55-year precipitation data (1965–2019) were decomposed into four stationary IMFs and one trend term. Spectral analysis revealed quasi-periods of 2.9a, 7.5a, 18.5a, and 33a. LSTM networks were trained for each component, achieving average relative error of 13.38% and RMSE of 38.03 mm for 2010–2019 predictions—superior to traditional ARIMA models. The model predicts six years with above-normal and four years with below-normal precipitation during 2020–2029, with 2025 as an extremely wet year and 2021 as an extremely dry year. This study provides a new approach for precipitation prediction in arid regions.

References

[1] Liu Ying, Ren Hongli, Zhang Peiqun, et al. Application of the hybrid statistical downscaling model in summer precipitation prediction in China [J]. Climatic and Environmental Research, 2020, 25(2): 163-171.

[2] Ren Ran. The Research about the Summer Precipitation Prediction in China Based on EEMD Method[D]. Nanjing: Nanjing University of Information Science & Technology, 2014.

[3] Chi Daocai, Zhang Tenan, Wu Xiuming, et al. Application of ARMA model in the annual precipitation forecast of Taizi River[J]. Journal of Shenyang Agriculture University, 2012, 43(3): 607-610.

[4] Chen Husheng, Zhou Yuliang, Zhou Ping, et al. Analysis and prediction of annual precipitation in Huangshan City based on wavelet and ARIMA[J]. South North Water Transfers and Water Science & Technology, 2019, 17(5): 50-55.

[5] Li Jiaxiu, Du Chunli, Du Shifei, et al. Temporal spatial variation and trend prediction of extreme precipitation events in Xinjiang[J]. Arid Zone Research, 2015, 32(6): 60-69.

[6] Xue Chunfang, Hou Wei, Zhao Junhu, et al. The application of ensemble empirical mode decomposition method in multiscale analysis of region precipitation and its response to the climate change[J]. Acta Physica Sinica, 2013, 62(10): 1-7.

[7] Fang Chuanglin. Strategic thinking and spatial layout for the sustainable development of urban agglomeration in northern slope of Tianshan Mountains[J]. Arid Land Geography, 2019, 42(1): 1-11.

[8] Liu Tianhu, Liu Tianlong. Regional features of precipitation variation trends over Xinjiang in China by the Ensemble Empirical Mode Decomposition method[J]. Desert and Oasis Meteorology, 2015, 9(4): 17-24.

[9] Di Hao, Zhao Xuejun, Zhang Zili. Commodity price forecasting based on EEMD-Adaboost[J]. Statistics & Decision, 2018, 34(13): 72-76.

[10] Luo Nana, Bake Batur, Wu Yanfeng. Precipitation multi-scale characteristics by Ensemble Empirical Mode Decomposition in Northern Xinjiang[J]. Research of Soil and Water Conservation, 2017, 24(4): 362-367.

[11] Wu Z, Huang N E. Ensemble empirical mode decomposition: a noise-assisted data analysis method[J]. Advances in Adaptive Data Analysis, 2009, 1(1): 1-41.

[12] Zheng Yi, Sun Xiaofeng, Chen Jian, et al. Extracting pulse signals in measurement while drilling using optimum denoising methods based on the ensemble empirical mode decomposition[J]. Petroleum Exploration and Development, 2012, 39(6): 113-116.

[13] Dang Chiheng, Zhang Hongbo, Chen Keyu, et al. Application of the long short-term memory neural network for rainfall-runoff simulation in seasonal snowmelt basin[J]. Journal of North China University of Water Resources and Electric Power (Natural Science Edition), 2020, 41(5): 10-18, 33.

[14] Hochreiter S, Schmidhuber J. Long short-term memory[J]. Neural Computation, 1997, 9(8): 1735-1780.

[15] Graves A. Long Short-Term Memory: Supervised Sequence Labeling with Recurrent Neural Networks[M]. Berlin, Heidelberg: Springer, 2012: 1735-1780.

[16] Wu Lin, Feng Lihua. Prediction of precipitation in Yiwu based on BP neural networks[J]. Zhejiang Hydrotechnics, 2014, 42(2): 52-54.

[17] Shen Yan, Yang Chunlei, Zhang Qingguo, et al. Prediction of precipitation series based on RBF neural network in Chizhou City[J]. Journal of Anhui Agricultural University, 2012, 39(3): 451-455.

[18] Zhang Feilian, Liu Yanping. Application of Empirical Mode Decomposition and Neural Network in Precipitation Prediction[C]// Proceedings of the 18th Annual Meeting of Systems Engineering Society of China. Hefei: Systems Engineering Society of China, 2014.

[19] Zhu Lingzi. Study of Flood Season Precipitation Prediction Based on Statistical Model: Taking Yiwu City as an Example[D]. Jinhua: Zhejiang Normal University, 2015.

[20] Bai Hui, Gao Hui, Liu Changzheng, et al. Assessment of multi-model downscaling ensemble prediction system for monthly temperature and precipitation prediction in Guizhou[J]. Desert and Oasis Meteorology, 2016, 10(5): 58-63.

[21] Ning Like. Study on the Influence of Topography and Geomorphology on Precipitation over Tianshan Mountains[D]. Shihezi: Shihezi University, 2013.

[22] Yang Lianmei, Liu Jing. Some advances of water vapor research in Xinjiang[J]. Journal of Natural Disasters, 2018, 27(2): 1-13.

[23] Dai Xingang, Wang Ping, Zhang Kaijing. A study on precipitation trend and fluctuation mechanism in northwestern China over the past 60 years[J]. Acta Physica Sinica, 2013, 62(12): 527-537.

[24] Hu Yuewen, Yang Xiaoyi. Low frequency variability of AO and NAO[J]. Journal of the Meteorological Sciences, 2007, 99(3): 84-90.

[25] Guan Xuefeng, Sun Weiguo, Li Minjiao, et al. Climate change in north Xinjiang and its response to Arctic Oscillation during the period of 1965-2012[J]. Arid Zone Research, 2016, 33(4): 681-689.

[26] Liu Dehao, Zhou Haidong, Rong Li, et al. Predicting water demands of cities by GNNM(1,1)[J]. Journal of Shandong Jianzhu University, 2006, 21(4): 335-337.

[27] Wang Tingting, Qian Xiaodong. Survey of non-linear trend forecast and application of time series[J]. Computer Engineering and Design, 2010, 31(7): 1545-1549.

Note: Figure translations are in progress. See original paper for figures.