Proceedings of the 2nd International Conference on Mathematical Physics and Computational Simulation
DOI: 10.54254/2753-8818/39/20240599

EEMD-ARIMA Model for annual precipitation forecasting to aid weather insurance decision-making research

Yuyang Xia

Nanjing University of Posts and Telecommunications, Nanjing, China

B21150329@njupt.edu.cn

Abstract. Climate risk poses significant threats to human life and property. Climate insurance can effectively mitigate and disperse these risks. This paper addresses the weakness in weather risk prediction for climate insurance formulation by combining Ensemble Empirical Mode Decomposition (EEMD) and Autoregressive Integrated Moving Average (ARIMA) models. Four models, namely EMD, EEMD, ARIMA, and EMD-ARIMA, were established for modeling and forecasting China's annual precipitation data. The results show that the EEMD-ARIMA model can suppress the modal aliasing problem in time series and has the best fit compared to other models. This model can more accurately describe the variation in annual precipitation in forecasting applications, providing significant predictive value for insurance companies and government decisions regarding insurance and climate risk management.

Keywords: Climate Insurance, Climate Risk, EMD Algorithm, EEMD-ARIMA, Precipitation Forecast

1. Introduction

From the perspective of the causes and development of climate change, climate risk arises from the excessive accumulation of greenhouse gases caused by natural and human activities, leading to global temperature rise. This subsequently triggers large-scale climate changes, ultimately posing potential losses and adverse impacts on natural systems and human socio-economic systems [1]. With global temperatures reaching record highs, people worldwide are increasingly affected by climate risks. In 2023, global natural disasters caused total losses of $380 billion, 22% higher than the average level of this century. Notably, 95,000 people worldwide lost their lives due to natural disasters, the highest number since 2010, and the number of economically significant natural disasters in 2023 set a historical record. "Human destiny is in jeopardy," warned the UN Secretary-General at the 28th United Nations Climate Change Conference (COP28) World Climate Action Summit, highlighting the common crisis humanity faces.

Strategies to address climate risk include prevention and control, as well as risk transfer and dispersion. Climate risk prevention mainly involves various measures to adapt to and mitigate climate change, while risk transfer and dispersion primarily refer to the distribution of risk costs. A report by the Intergovernmental Panel on Climate Change on managing the risks of extreme events [2] suggests that insurance can be a tool for reducing risks and restoring livelihoods. However, due to the increasing severity of losses caused by climate change, government funds can no longer meet current demands.

DOI: 10.54254/2753-8818/39/20240599

Therefore, adopting climate insurance measures to transfer and disperse risks has become an important policy direction for China [3].

With the deepening of global economic integration, China, as a developing country, is striving to innovate weather insurance, transitioning from traditional agricultural insurance to more flexible agricultural insurance [4], aiming to effectively address extreme climate risks and achieve sustainable development. Currently, the weather index insurance being piloted in China is an effective means [5] that can enhance risk resilience and reduce transaction costs, meeting the diverse needs of farmers. However, many issues still exist in reality. Insurance companies' investment decisions typically do not consider climate risk knowledge gained from underwriting, and infrastructure investment decisions often lack reflection on climate risks.

Therefore, designing a scientifically precise insurance model based on weather risk prediction holds significant practical significance. It can address the aforementioned issues, making climate insurance more feasible, creating new markets and opportunities, and playing a crucial role in helping developing country governments resist natural disasters and improve climate adaptability.

2. Research Methods

Given the ongoing intensification of global warming and its severe impact on precipitation patterns in the atmosphere, resulting in frequent extreme weather events and significant economic losses in China, it is of critical practical significance to scientifically and accurately forecast precipitation, especially during the summer flood season [6]. Currently, most annual precipitation forecasting methods are primarily meteorological, but the independent variables in these methods are difficult to precisely determine, and the models have enormous computational demands. The ARIMA method for time series analysis is widely used in precipitation research due to its advantages of low computational demand and strong resistance to interference [7,8]. However, the ARIMA method performs inadequately when dealing with nonlinear sequences, and annual precipitation sequences are predominantly nonlinear [9]. To address these issues, researchers have proposed the EEMD-based ARIMA time series model [10]. The EEMD-ARIMA model, known for its excellent adaptability, can effectively handle complex nonlinear and non-stationary time series, providing more accurate short-term runoff forecasts than traditional ARIMA models [11].

2.1. EEMD

The EMD method effectively converts complex nonlinear and non-stationary sequences into a series of single-frequency waves [12]. Compared to wavelet and Fourier transform decomposition, EMD offers better time-frequency resolution and adaptability and effectively resolves the issue of modal aliasing, resulting in more precise analysis and better future trend predictions. EEMD is an improvement upon EMD, capable of effectively handling original signals by incorporating noise into Intrinsic Mode Functions (IMFs) and trend items and calculating their averages to determine IMF components and residual terms, thereby better predicting signal characteristics. Additionally, EEMD provides higher accuracy and greater flexibility than EMD and can be used to analyze raw data. The specific steps to decompose the original signal using EEMD are as follows:

1) Determine the number of times to add white noise (i.e., total number of trials) n_e and the amplitude coefficient ε.

2) Add white noise to the original signal $x(t)$, where $\mathrm{w}(t)$ is the white noise sequence added for the i th time, yielding the signal $X'(t)$:

$$X'(t) = x(t) + \varepsilon * w(t) \tag{1}$$

3) Identify all local maxima and minima of the time series $X'(t)$ to be decomposed and fit all maxima and minima using the cubic spline function, forming the upper and lower envelopes $X'(t)$.

4) Let $b_1(t)$ be the average of the upper and lower envelopes. Subtract $X'(t)$ from $b_1(t)$ to obtain a new sequence:

$$X_1''(t) = X'(t) + b_1(t) \tag{2}$$

5) Determine whether $X_1''(t)$ satisfies the IMF conditions. If it does, $X_1''(t)$ is the first IMF component, IMF_1; if not, $X_1''(t)$ becomes the new original sequence and returns to steps 3 and 4 for further screening until the IMF conditions are met.

6) Subtract IMF_1 from $X'(t)$ to obtain the residual x_1, as shown in formula (3). Repeat steps 3 to 6 n times until x_1 or IMFn is less than a given value or x_n is a monotonic function, ending the decomposition process.

$$x_1 = X'(t) - IMF_1 \tag{3}$$

7) Use the standard deviation (SD) to determine whether to terminate the screening. When SD is less than the threshold α, the screening ends. The general range of α is 0.2~0.3. $X'(t)$ After processing through steps 3 to 7, different scales of IMFs and residuals are obtained.

$$SD = \frac{\sum_{t=0}^{T} (X''^{2}_{k-1}(t) - X''^{2}_{k}(t))^2}{\sum_{t=0}^{T} X''^{2}_{k-1}(t)} \tag{4}$$

8) Use the mean of the white noise spectrum, which is zero, to average the IMFs obtained from n_e decompositions, resulting in the final IMF components $c_n(t)$ and trend items $r(t)$ after EEMD decomposition:

$$C_s(t) = \frac{1}{n_e} \sum_{i=1}^{n_e} IMF_{is}(t) \tag{5}$$

$$r(t) = \frac{1}{n_e} \sum_{i=1}^{n_e} r_i(t) \tag{6}$$

9) The final decomposition result of $x(t)$ is:

$$x(t) = \sum_{s=1}^{n} C_s(t) + r(t) \tag{7}$$

2.2. ARIMA

ARIMA (p, d, q) is the most commonly used time series model [13]. AR represents the autoregressive process, with parameter p denoting the number of autoregressive terms; MA represents the moving average process, with parameter q denoting the number of moving average terms; d is the differencing order to achieve stationarity in the time series. The ARIMA (p, d, q) model structure is:

$$\begin{cases} \Phi(B)\nabla^d x_t = \Theta(B)\varepsilon_t \\ E(\varepsilon_t) = 0, Var(\varepsilon_t) = \sigma_t^2, E(\varepsilon_t \varepsilon_s) = 0, s \neq t \\ E(x_s \varepsilon_t) = 0, \forall s < t \end{cases} \tag{8}$$

Where $\nabla^d = (1-B)^d; \Phi(B) = 1 - \phi_1 B - ... - \phi_q B^q$ is the autoregressive polynomial coefficient of the stationary invertible ARMA (p, q) model; $\Theta(B) = 1 - \theta_1 B - ... - \theta_q B^q$ is the moving average polynomial coefficient of the stationary invertible ARMA (p, q) model, and $\{\varepsilon_t\}$ is a white noise sequence with zero mean. The ARMA (p, q) model is denoted as $\nabla^d x_t = \frac{\Theta(B)}{\Phi(B)} \varepsilon_t$.

DOI: 10.54254/2753-8818/39/20240599

2.3. EEMD-ARIMA Model

Precipitation variation is a complex phenomenon. Therefore, a stable method must be employed for better prediction. This study establishes an EEMD-ARIMA prediction model based on EEMD decomposition. The modeling method is shown in Figure 1.

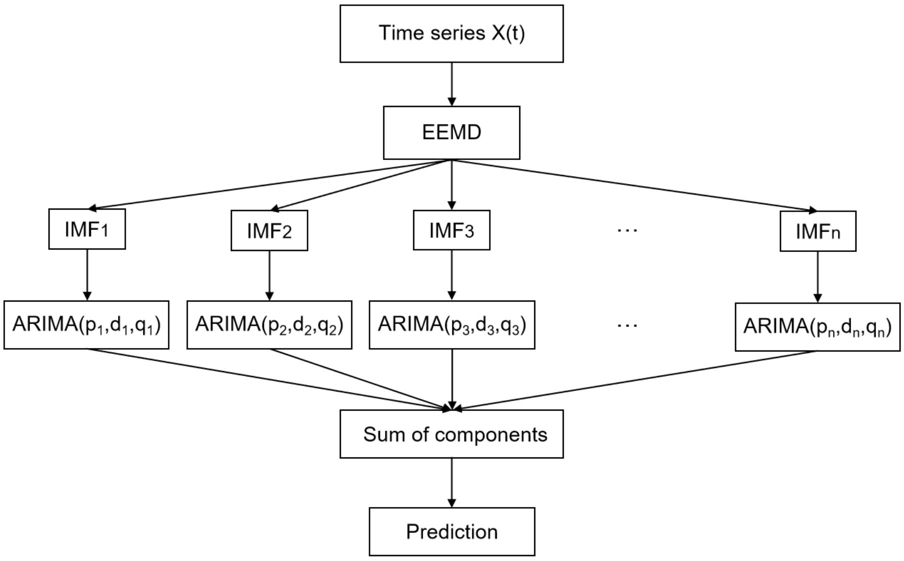

Figure 1. Flow chart of EEMD-ARIMA prediction model

The implementation of the EEMD-ARIMA prediction model involves two steps:

1. First, convert the original signal into several independent IMFs and residuals using EEMD, and analyze each IMF and residual to obtain their forecast values.
2. Then, use the ARIMA method to reconstruct these forecast values, achieving more accurate predictions.
3. Finally, calculate the forecast values of each component to derive the most accurate precipitation forecast.

3. Model Results and Analysis

This section uses the EEMD-ARIMA model to model and analyze China's annual precipitation data. To ensure the model's stability and reliability, given that time series prediction accuracy decreases over time, annual precipitation data from 1953 to 2007 were used as the training sample, while data from 2008 to 2012 were used as the test set.

3.1. EEMD Decomposition

Following the EEMD algorithm process, which involves adding noise, EMD decomposition, and calculating the ensemble, the 55-year annual precipitation series was decomposed into four IMF component series (IMF1-IMF4) and one residual component IMF5, as shown in Figure 2. The horizontal axis represents time, and the vertical axis represents precipitation.

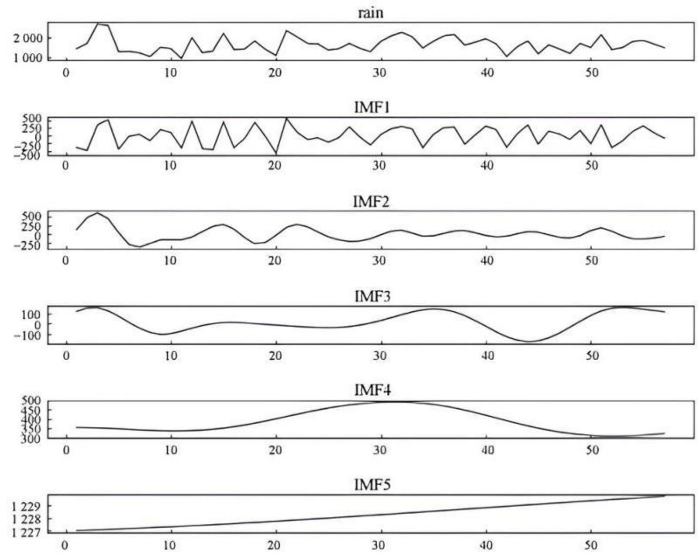

Figure 2. EEMD Decomposition of Annual Precipitation Series

It can be observed that IMF1 has the highest frequency, followed by IMF2, IMF3, and IMF4, while IMF5 is a residual component and a monotonically decreasing function. Among these variables, IMF1, representing short-term random variations, usually serves noise processing. However, directly removing noise may reduce fitting and prediction accuracy, so random fluctuations fitting is also used to reduce model errors.

To quantitatively evaluate the denoising results of EMD and EEMD, the Root Mean Square Error (RMSE) metric is used to better reflect the distribution characteristics of the predicted values. RMSE measures the degree of difference between the estimated values and the actual values, providing an overall reflection of the prediction value dispersion. Hence, the smaller the RMSE, the higher the reliability of the estimator and the better the denoising effect. The formula is as follows:

$$RMSE=\sqrt{\frac{1}{n}\sum_{t=1}^{n}(x(t)-x'(t))^2} \tag{9}$$

Where $x(t)$ is the original time series, and $x'(t)$ is the denoised time series.

The RMSE results are shown in Table 1. The RMSE of EEMD is smaller, indicating its superior denoising effect compared to EMD.

Table 1. RMSE calculation value

Method	RMSE
EMD	308.11
EEMD	294.90

3.2. Model Fitting

In annual precipitation forecasting, most researchers choose a single ARIMA time series model. With technological advancements, the ARIMA model alone no longer meets the current needs for annual precipitation forecasting. Therefore, this study uses the EEMD method to decompose complex series into multiple simple series and a residual series to suppress noise and multiple frequency characteristics. To avoid the modal aliasing problem brought by EMD decomposition, the improved EEMD algorithm is used to determine the non-stationary sequences IMF1, IMF2, IMF3, IMF4, and IMF5, and then fit them using the ARIMA model.

To enhance fitting, ARMA (2,0), ARMA (4,2), ARMA (2,4), ARIMA (5,0,1), and ARIMA (1,1,1) models were chosen for IMF1, IMF2, IMF3, IMF4, and IMF5, respectively. The fitting results are shown in Figure 3.

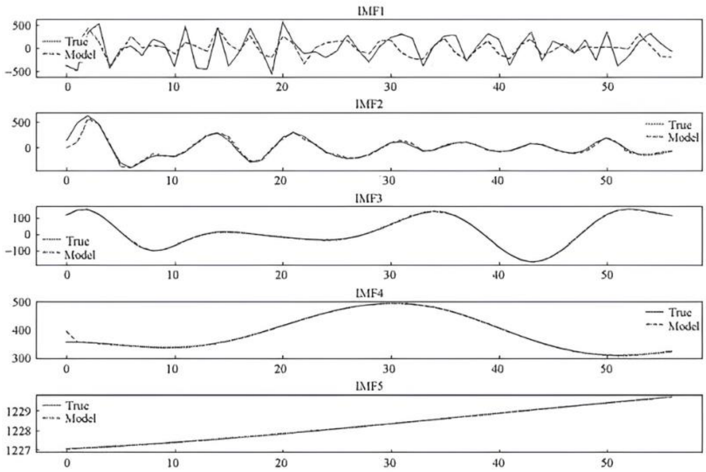

Figure 3. Component Fitting Results

According to the ARMA prediction results, the fitting ability of IMF1 still needs improvement. However, compared to discarding high-frequency signals, the fitting ability of IMF2, IMF3, IMF4, and IMF5 is more outstanding, significantly reducing the bias in fitting results. EEMD provides complete information, so the original signal can be reconstructed by summing its parts. This reconstruction results in a new sequence very close to the original sequence, with errors approaching zero. From the model training results, it is found that the fitting effect of low-frequency elements is better than that of high-frequency elements.

3.3. Model Results

Finally, after parameter tuning, the ARIMA (1,0,1) model was set, and the EMD-ARIMA, EEMD-ARIMA, and ARIMA models were used to model the annual precipitation series. The prediction results are shown in Figure 4, which indicates that the EEMD-ARIMA model's prediction is closest to the original values and has the highest fitting degree.

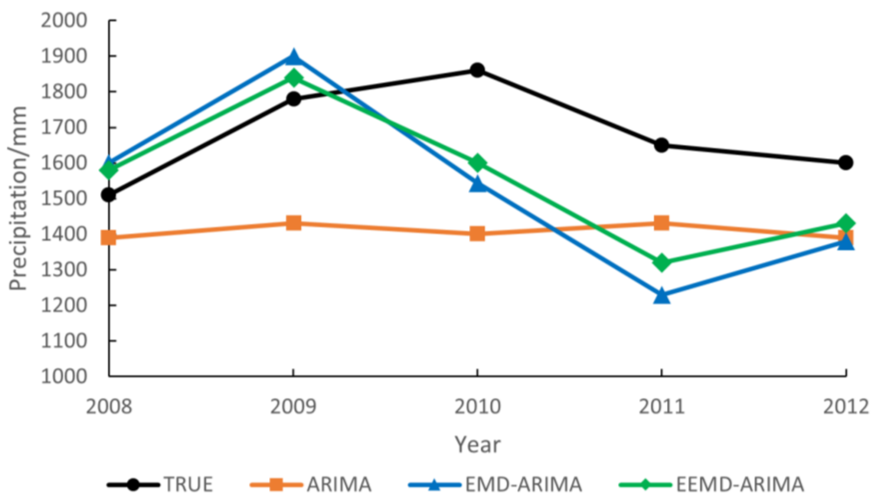

Figure 4. Comparison of Model Results

The prediction effects of the three models were evaluated using the Mean Relative Error (MRE). MRE reflects the model's error magnitude and prediction accuracy. The smaller the MRE, the smaller the deviation between the predicted and original values, and the higher the prediction model's accuracy. The formula is as follows:

$$MRE = \frac{1}{n}\sum_{t=1}^{n}\left|\frac{x(t)-x'(t)}{x(t)}\right| \tag{10}$$

Where $x(t)$ is the original value, and $x'(t)$ is the predicted value.

The MRE calculation results for the three models are shown in Table 2. The EEMD-ARIMA prediction model has the smallest MRE, followed by the EMD-ARIMA model, and the ARIMA prediction model has the largest MRE, indicating that the EEMD-ARIMA prediction model has the best prediction effect.

Table 2. Calculated value of mean relative error(MRE)

Method	MRE
ARIMA	0.157
EMD-ARIMA	0.138
EEMD-ARIMA	0.105

4. Conclusion

This paper aims to explore an efficient precipitation forecasting method to aid climate insurance decision-making. Given the importance of climate change factors to the insurance industry, the proposed model holds significant relevance for climate insurance decision-making. By utilizing the EEMD decomposition algorithm and ARIMA time series analysis model, this study models annual precipitation time series, allowing for accurate predictions of annual precipitation trends. This provides a scientific basis for insurance companies to assess risks, set insurance rates, and optimize product designs. Moreover, the model assists governments and regulatory agencies in formulating more effective climate risk management policies, thereby enhancing overall societal climate resilience.

The selection of simple models for the high-frequency components after EEMD decomposition in this study could be improved in the future by using neural network models, which may provide better fitting for high-frequency components. Although the EEMD algorithm used in this paper addresses the modal aliasing problem, it still faces the issue of endpoint effects during decomposition. Future research could explore more effective decomposition algorithms to improve accuracy. Additionally, given the high complexity of the climate system, any predictive model will have certain limitations and uncertainties. This study only selected the more prominent annual precipitation series as the model fitting object in climate change. Therefore, when applying this model to climate insurance decision-making, it is necessary to fully consider various influencing factors and integrate other climate information and professional knowledge for comprehensive judgment.

It is important to note that while climate insurance has the advantages of prevention, dispersion, and market-based fundraising, it also faces challenges, such as the inability to fully adhere to the law of large numbers and moral hazards, which may lead investors to make incorrect decisions, thus reducing their long-term investment returns. To promote the development of climate insurance in China, this paper proposes the following measures: 1. Strengthen the construction of meteorological infrastructure and data, develop natural disaster risk maps for different regions, encourage insurance companies to offer diverse and multi-level climate insurance products, and increase public willingness to purchase insurance to avoid adverse selection. 2. Enhance moral hazard warnings to prevent negative impacts and take effective measures to improve the social environment, thereby promoting the sustainable development of climate insurance in China.

References

[1] Wu, S. H., Gao, J. B., Deng, H. Y., Liu, L. L., & Pan, T. (2018). Climate change risks and quantitative assessment methods. Progress in Geography, 37(1), 28-35.
[2] Intergovernmental Panel on Climate Change. Managing the Risks of Extreme Events and Disasters to Advance Climate Change Adaptation. Field, C.B. et al. (eds.). Cambridge University Press, Cambridge, UK (2012).
[3] XU Guang-Qing, CHEN Xiao-Yu, LIU Hai-Bo, HUANG Dan-Ze, ZHANG You-Yi. The conception and theory of climate insurance and its enlightenment to China. Advances in Climate Change Research[J], 2020, 16(3): 373-382 doi:10.12006/j.issn.1673-1719.2019.137
[4] Wang L P. A comparative study on innovative weather index insurance products [J]. Insurance Studies, 2016, 10: 81-88. DOI: 10.13497/j.cnki.is.2016.10.008 (in Chinese)
[5] Yao Q H. Research on meteorological disaster and weather index insurance [J]. Shanghai Insurance Monthly, 2015 (1): 7-11, 20 (in Chinese)
[6] Shen, H. J., Luo, Y., Zhao, Z. C., & Wang, H. J. (2020). Research on summer precipitation prediction in China based on LSTM network. Advances in Climate Change Research.
[7] Cheng, M., Zhang, Y. W., Jiang, J. Y., Ren, J., & Zhao, Z. H. (2019). Rainfall prediction analysis based on time series models. Water Science and Engineering Technology, (1), 5.
[8] Liu, J. (2018). Application of ARIMA model in precipitation trend analysis and prediction. Water Conservancy Science and Technology and Economy, 24(9), 3.
[9] Xu, S. M. (2018). Study on nonlinear disturbance response of precipitation runoff based on SLM model. Applied Technology of Soil and Water Conservation, (3), 2.
[10] Guo, J. P., She, Y. L., & Wen, B. Y. (2018). Regional monthly load forecasting based on EEMD-ARIMA model. Power Engineering Technology, 37(6), 6.
[11] Liu, Y., Yang, Y., Nie, L., & Song, Q. Y. (2017). EEMD-ARIMA prediction of runoff at the exit of the Manas River. Research of Soil and Water Conservation, (6), 9.
[12] Huang NE, Shen Z, Long SR, et al. The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis. The Royal Society, 1998,454:903-995.
[13] Mai, H. K., Xiao, J. H., Wu, X. C., & Chen, C. (2015). Parallel research on load forecasting ARIMA model based on R language. Power Grid Technology, 39(11), 5.