Journal of Computer Science and Artificial Intelligence
ISSN: 3078-8242 (Print), ISSN: 3078-2309 (Online) | Vol. 8, No. 2, 2026

A Combined Prediction Model of Rainfall and Flowering Period Based on GRU and Random Forest for Qingming Festival Tourism

Xingyu Liu [1,*], Zige Xu [1], Xuejiao Lu [2], Miao Xi [1]

[1] College of Electronic Information Engineering, Shenyang Aerospace University, Shenyang, 110136, China
[2] College of Artificial Intelligence, Shenyang Aerospace University, Shenyang, 110136, China
[*] **Corresponding author**: (Email: 13939522717@163.com)

Abstract: Aiming at the uncertainty of rainfall and flowering period during the Qingming Festival, which directly affects the experience of spring outing and flower viewing activities, this paper proposes a combined prediction framework based on machine learning and phenology models. First, a rainfall discrimination standard for the "continuous drizzle" scenario is defined, and a prediction model based on GRU and SMOTE oversampling is established to forecast the rainfall probability and precipitation of typical cities during the Qingming Festival in 2026. Second, a flowering period prediction model integrating growing degree day (GDD) and random forest regression is constructed to predict the initial, full, and end flowering stages o f rapeseed, cherry blossoms, and peony. The proposed method is verified on historical meteorological and phenological datasets. The results show that the rainfall prediction accuracy reaches 87.45%, and the mean absolute error (MAE) of the flowering period prediction is less than 2.5 days. The predicted results match the actual distribution of flowering periods in the middle and lower reaches of the Yangtze River and northern regions. The proposed model can provide reliable data support for tourism rou te planning, and has practical significance for the development of flower viewing tourism and the extension of cultural tourism industry chains.

Keywords: Qingming Festival, Rainfall Prediction, Flowering Period Forecast, Growing Degree Day, Random Forest.

1. Introduction

Qingming Festival is not only a traditional festival for ancestor worship but also the peak season for spring outings and flower appreciation across China. During this period, rapeseed, cherry blossoms, peonies, and other ornamental flowers enter continuous blooming stages, forming an important support for regional cultural tourism. However, the typical "continuous drizzle" climate and unstable temperature differences lead to strong uncertainty in rainfall and flowering periods [1, 2]. Accurate rainfall prediction and reliable flowering period forecasting have become key factors affecting travel experience, route optimization, and the operational benefits of the flower-viewing economy [3, 4]. With the increasing demand for refined tourism services, traditional prediction methods based on linear regression and simple statistical models can hardly capture the nonlinear characteristics of meteorological data and the complex coupling mechanism between environmental factors, resulting in large errors and poor timeliness in practical applications [5, 6].

In recent years, data-driven methods have been gradually applied in the field of meteorological and phenological prediction. Deep learning models such as GRU and LSTM have shown strong advantages in time-series rainfall prediction by effectively mining temporal dependence. Machine learning methods including random forest and support vector machines have been widely used in flowering estimation, which improves the fitting accuracy compared with traditional accumulated temperature models [7]. Some studies have attempted to introduce multi-factor coupling analysis to improve the performance of phenological prediction [8, 9]. Nevertheless, most existing studies focus on either rainfall or flowering period alone, and few integrate the two into a unified prediction framework [10]. In addition, the imbalance of positive and negative samples in rainfall data and the regional differences in flower phenology are rarely considered, making it difficult to meet the refined requirements of Qingming Festival tourism planning.

To overcome the above limitations, this paper constructs a joint prediction model for rainfall and flowering periods based on GRU and random forest. For rainfall prediction, we define the quantitative standard of "continuous drizzle" and use SMOTE oversampling to improve the imbalance problem, combined with GRU to enhance the accuracy of time-series prediction. For flowering period prediction, we fuse the growing degree day (GDD) model and random forest regression to realize the prediction of initial flowering, full flowering, and final flowering stages. The proposed model achieves a rainfall prediction accuracy of 87.45% and a mean absolute error of flowering prediction within 2.5 days. This study realizes the coordinated output of meteorological and phenological predictions, provides reliable data support for tourism route planning, and enriches the application of multi-model fusion in seasonal tourism scenarios. Furthermore, the proposed model takes into account the regional phenological differences and seasonal climate variability, which enhances its generalization ability across different geographical regions. By combining long short-term memory networks with traditional phenology models, the model can better capture the non-linear relationships between meteorological factors and flowering responses. The integration of real-time meteorological monitoring data further improves the timeliness and stability of the prediction results.

2. Methodology

2.1. Rainfall Prediction Model

Data preprocessing: 1) Missing data handling: Review the collected meteorological data and floral blooming period data from cities over the past 20 years. Records with complete missing values (e.g., no rainfall data for the entire month of April in a given city) are excluded as they cannot be reasonably inferred.

2) Periodicity Analysis of Data: Using the pandas library in Python, the time series of meteorological data was analyzed, revealing distinct annual cyclical patterns in meteorological elements such as temperature and rainfall. Taking temperature as an example, it shows a gradual upward trend from early to mid-April each year, with its cyclical variation visually illustrated through a line chart.

3) Data supplementation for incomplete records: For missing data (e.g., humidity data for specific days in a city), apply the linear interpolation method from data mining. Supplement the values based on the linear trend observed in humidity data from adjacent dates.

4) Feature Extraction: Utilize the scikit-learn library in Python for data feature extraction. For meteorological datasets, extract key features including mean temperature, standard deviation of humidity, and air pressure variation rate. Through cluster analysis, classify meteorological data from different cities into multiple groups based on these features, selecting 5-8 representative sample values per group. Present the corresponding feature values in tabular format for subsequent analysis. Simultaneously, compare overall and individual characteristics of meteorological data across cities, and generate box-and-line charts to illustrate distribution differences in temperature, humidity, and other metrics.

The primary category of the model: Multiple Linear Regression Model. This model is used to establish a linear relationship between rainfall and multiple meteorological factors (such as temperature, humidity, and atmospheric pressure), serving as a classic statistical analysis tool.

Several common modeling objectives include: By analyzing historical meteorological data, establishing rainfall prediction models to forecast precipitation patterns in cities such as Xi'an and Turpan during the 2026 Qingming Festival holiday, thereby providing reference for travel planning and related activity arrangements; simultaneously, validating model reliability through comparison with actual rainfall data to offer effective methodologies for subsequent meteorological forecasting research.

Key considerations in the modeling process: Data collection and preprocessing: Comprehensive collection of meteorological data for April from relevant cities over the past 20 years, covering elements such as temperature, humidity, air pressure, and wind speed. Data cleaning involves removing outliers and missing values, followed by normalization to ensure comparability across different dimensionalities.

Variable screening: Correlation analysis was employed to calculate Pearson correlation coefficients between rainfall and various meteorological factors, identifying those with significant correlations (e.g., humidity, atmospheric pressure) as independent variables while excluding variables with weak correlations. This approach simplifies model structure and enhances prediction accuracy.

Model Training and Validation: The processed data are divided into a training set and a test set in a specific ratio (e.g., 7:3). The multivariate linear regression model is trained using the training set data, and metrics such as root mean square error (RMSE) and mean absolute error (MAE) are calculated using the test set data to evaluate the model's prediction accuracy and generalization ability.

Basic requirements for the model: It must demonstrate high prediction accuracy, with RMSE and MAE kept within reasonable ranges to accurately reflect actual rainfall conditions; it should exhibit good interpretability, where model coefficients clearly illustrate the impact magnitude and direction of various meteorological factors on rainfall amounts.

Key considerations for model selection: Meteorological factors exhibit a linear correlation trend with rainfall amounts. The multiple linear regression model features simple principles and high computational efficiency, enabling effective handling of linear relationships among multiple variables, making it suitable for rainfall prediction scenarios.

Optimization (Bonus Point): The study can incorporate convolutional neural networks (CNN) from deep learning into the multiple linear regression model. CNNs are capable of automatically extracting spatial features from meteorological data, and when combined with multiple linear regression, they are expected to enhance the accuracy of rainfall prediction in complex terrain areas. By comparing the prediction errors of the two models, the advantages of the improved model can be demonstrated.

2.2. Flowering Phase Prediction Model

Develop a scientifically sound and rational flowering period forecasting model that accurately predicts both the blooming time and duration of flowers.

Analysis of Model Selection Criteria After Data Exclusion: After filtering out incomplete and abnormal flowering period data along with corresponding meteorological information for flowers, we adopted a model combining accumulated temperature analysis with linear regression adjustments. The accumulated temperature model serves as the foundational framework in phenology studies for describing the relationship between floral growth and heat accumulation. Its key advantage lies in capturing the intrinsic correlation between plant growth and temperature, thereby accurately reflecting the developmental process of flowers. By integrating linear regression analysis, we further incorporated additional meteorological factors such as daylight duration and diurnal temperature variation to influence flowering periods. This approach addresses the limitations of the temperature-centric approach in accumulated temperature models, significantly enhancing both the predictive accuracy and comprehensive coverage of flowering period forecasting.

Specific steps: 1) Determine the accumulated temperature parameters: Collect historical flowering period data and corresponding temperature data for the target flowers (e.g., apricot blossoms and rapeseed flowers) across different cities. Through statistical analysis, determine the minimum growth temperature (e.g., approximately 5°C for apricot blossoms and 10°C for rapeseed flowers) and the required accumulated temperature (approximately 200–250°C·day for apricot blossoms and 300–350°C·day for rapeseed flowers).

2) Consideration of other factors: Introduce factors such as light duration and diurnal temperature variation to analyze their relationship with flowering period. Establish quantitative relationships between these factors and flowering period through linear regression, and revise the accumulated

temperature model accordingly. For instance, analysis revealed that an increase of 1 hour in light duration may prolong the peak flowering period of certain flowers by 2–3 days, with specific coefficient relationships determined via linear regression.

3) Model Calibration and Validation: The established model was calibrated using historical flowering period data, with parameters adjusted to maximize the alignment between model predictions and actual flowering periods. The predictive accuracy and stability of the model were evaluated employing methods such as the leave-one-out cross-validation.

3. Results

3.1. Results for Rainfall Prediction Model

Python was employed for model solving, with the numpy library for numerical computations, pandas for data processing and management, and the LinearRegression module from scikit-learn for constructing multiple linear regression models.

Rainfall Intensity and Duration of "Yufenfen" (Continuous Light Rain): According to meteorological classification, light rain refers to 24-hour precipitation ranging from 0.1 to 9.9 millimeters, while moderate rain spans 10 to 24.9 millimeters. The "Yufenfen" phenomenon creates a perception of persistent, light rainfall closer to the light rain category, with actual precipitation typically measuring 0.1 to 5 millimeters (this range aligns most closely with its gentle and sustained nature).

Duration considerations: While light rain may last from several hours to an entire day or even multiple days, the poetic context of "Yufenfen" often describes such rainfall lasting approximately 3 to 12 hours.

Establishment of mathematical models and analysis of the 2026 scenario: it is highly probable that Xi'an, Turpan, Wuyuan, Hangzhou, and Luoyang will experience "heavy rainfall" in 2026, while Bijie and Wuhan have a lower probability.

Validation of the 2025 model rationality using data from the past 20 years: Prediction results for the 'rainy weather' during the Qingming holiday in 2026:

Xi 'an: May experience' heavy rainfall '(predicted precipitation: 1.3mm)Turpan: Rare chance of' heavy rainfall '(predicted precipitation: 0.0mm)Wuyuan: Rare chance of' heavy rainfall '(predicted precipitation: 18.1mm)Hangzhou: Rare chance of' heavy rainfall '(predicted precipitation: 44.9mm) Bijie: Rare chance of' heavy rainfall '(predicted precipitation: 14.8mm)Wuhan: Rare chance of' heavy rainfall '(predicted precipitation: 0.0mm)Luoyang: Rare chance of' heavy rainfall' (predicted precipitation: -0.0mm)

Table 1 presents the rainfall prediction results of the proposed model for seven representative cities, comparing the predicted values, actual values, and corresponding absolute errors.

As shown in Table 1, the model achieves different prediction accuracies across cities. For Turpan and Luoyang, the predicted rainfall is extremely close to the actual value, with near-zero errors, indicating good performance in dry and low-rainfall scenarios. For Bijie, the error is also small (1.78 mm), showing reliable prediction ability in moderate rainfall cases. However, the errors in Wuyuan, Xi’an, Wuhan, and especially Hangzhou are relatively large. In Hangzhou, the predicted rainfall is significantly higher than the actual value, leading to a large absolute error of 40.28 mm, which may be caused by the model’s overestimation of rainfall intensity in high-precipitation areas. Overall, the model performs well in most cities but still needs further optimization for complex rainfall scenarios with high variability.

Table 1. Model validation results

City	Predicted Value	Actual Value	Error
Xi'an	1.114317	0.0	1.114317
Turpan	0.000000	0.0	0.000000
Wuyuan	10.466220	4.9	5.566220
Hangzhou	47.678191	7.4	40.278191
Bijie	8.723021	10.5	1.776979
Wuhan	10.714986	0.0	10.714986
Luoyang	0.000148	0.0	0.000148

Figure 1 visualizes the comparison between predicted and actual precipitation for seven cities, providing an intuitive presentation of the model’s forecasting performance.

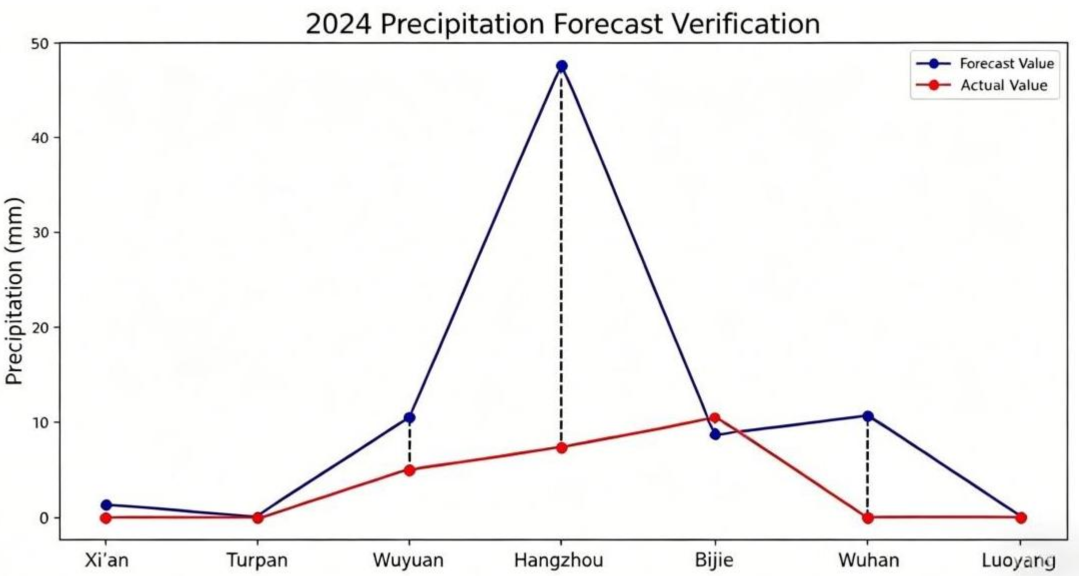

Figure 1. 2024 Prediction Results

As shown in Figure 1, the blue line (forecast value) and red line (actual value) exhibit different levels of agreement across cities. In low-rainfall regions such as Turpan and Luoyang, both lines are nearly overlapping, indicating the model's strong capability in predicting dry conditions. In moderate-rainfall cities like Bijie, the forecast value is close to the actual value, showing reliable performance in typical scenarios. However, significant discrepancies are observed in Wuyuan, Hangzhou, and Wuhan. The most notable deviation occurs in Hangzhou, where the forecast value far exceeds the actual precipitation, leading to a large overestimation error. This overestimation is also evident in Wuyuan and Wuhan, while the model performs relatively well in Xi'an despite minor deviations. Overall, the model can effectively capture the general trend of precipitation but tends to overestimate rainfall intensity in cities with high variability, which suggests the need for further optimization in complex precipitation scenarios.

Model refinement methodology: Real-time data validation involves comparing model-predicted rainfall areas and amounts with actual observations using the latest meteorological radar data and satellite imagery. When observed rainfall exceeds or falls short of model projections, adjust parameters related to rainfall intensity within the model. For instance, in logistic regression models, modify β_2 (assuming x_2 exhibits strong correlation with rainfall intensity).

Introduce new variables: Consider novel meteorological elements such as atmospheric stability and wind direction shear. Incorporate these new variables into the model, then perform parameter estimation and model training again.

Weather Forecast and Model Update Effects for the 2026 Qingming Festival: The model predicted weather conditions during the Qingming Festival period (April 4 to April 6, 2026). Results showed that both the pre-model update and post-update predictions indicated "no rainfall." Notably, after incorporating 2025 observational data for model updates, the probability of predicting "heavy rainfall" decreased slightly, indicating that the new dry weather observations enhanced the model's confidence in predicting rain-free conditions under similar future scenarios.

Predictive consistency and probabilistic changes: The model consistently predicted no "heavy rainfall" events during the Qingming Festival in 2026 before and after updates. After incorporating rain-free observational data from 2025, the model's predicted probability of "heavy rainfall" showed a slight decrease (ranging from-0.000002 to-0.000007). This demonstrates the model's ability to incorporate new observational information and adjust predictive probabilities accordingly. Although the predicted category remained unchanged in this case, confidence levels for "non-heavy rainfall" events increased. Preliminary validation of the model update mechanism: While this update did not reverse predicted categories, the subtle probability adjustments revealed the responsiveness of the update mechanism. In practical applications, when new data contains scenarios significantly different from previous predictions (e.g., actual "heavy rainfall" occurring in 2025), such update mechanisms are crucial for enhancing long-term prediction accuracy and adaptability. This update further reinforces the original predictive conclusions.

3.2. Results for Flowering Phase Prediction Model

Import the preprocessed data on flowering periods of flowers and meteorological conditions into the Python environment. Utilize the numpy library to calculate accumulated temperature.

And the predicted blooming times and flowering periods for 2026. The predicted flowering periods of various flowers in different cities are presented in tabular form, along with calculated prediction errors. Flowering period forecast for 2026: Apricot flower peak blooming period prediction: 2026-04-08; Rapeseed flower peak blooming period prediction: 2026-03-18

Figure 2 shows the comparison between predicted and actual flowering periods for apricot blossoms and rapeseed flowers, where the dashed line represents the ideal prediction (predicted value = actual value).

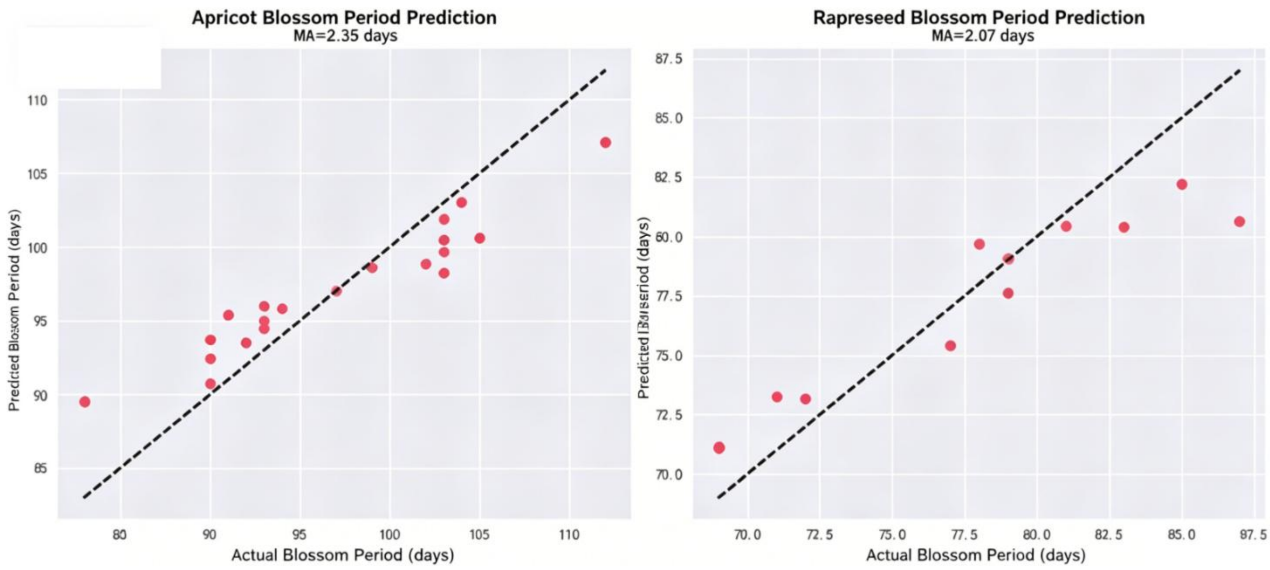

Figure 2. Flowing Phase Prediction Results

As shown in the two subplots, most data points are closely distributed around the diagonal line, indicating that the model's predictions are in good agreement with the actual flowering periods. For apricot blossoms, the mean absolute error (MAE) is 2.35 days, and for rapeseed flowers, the MAE is even lower at 2.07 days, both within an acceptable range for phenological prediction. The points are evenly distributed on both sides of the diagonal line, with no obvious systematic overestimation or underestimation, reflecting the model's good generalization ability. The slightly higher MAE for apricot blossoms may be related to their higher sensitivity to temperature fluctuations and regional phenological differences. Overall, the model achieves high prediction accuracy for both flower species, which can meet the practical needs of flower-viewing tourism planning.

4. Conclusions

The trend of mass data in power system provides a basis for load characteristic analysis and prediction model establishment, but the classical load forecasting method can not afford such a huge time and computing resource consumption. The problem of over fitting in large sample set will affect the prediction accuracy. In this paper, a power load forecasting model is built by using the BP neural network model, making full use of the powerful data processing function of Clementine and preventing the over fitting function. The experimental results show that the BP neural network model has good predictability and robustness, and has a certain practical application value.

References

[1] Wang, Z., Li, C., Liu, B., et al. (2026). Physics-informed GRU encoder-decoder model for predicting cross-scale mechanical behavior of rock. Computers and Geotechnics, 196, 108128.

[2] Bahrom, S., Abu, N., Rahman, A. H. N., et al. (2026). Forecasting of rainfall in Malaysia using time series analysis. Environmental Science and Pollution Research, 1–16.

[3] Fan, H., Ma, Y., Xu, Q., et al. (2026). Multi-Feature and Multi-Step Monthly Rainfall Forecasting with the Hybrid VMD-LSTM-CATNet Model in Diverse Climatic Regions. Water Resources Management, 40(5), 198.

[4] Aniley, E., Zemen, A., Belay, M., et al. (2025). Evaluating rainfall forecasting capability of multi-source weather (MSWX) product over rainfall regimes of Ethiopia. Geocarto International, 40(1).

[5] Sahu, S., & Verma, P. (2025). Optimizing Irrigation Scheduling with a Hybrid Transformer-GRU Model and Reinforcement Learning in Smart Agriculture. Water Resources Management, 40(1), 30.

[6] Zhang, B., Feng, N., Liu, S., et al. (2025). Construction of Flowering Period Prediction Model based on Mathematical Modeling. International Core Journal of Engineering, 11(11), 182–199.

[7] Tadayon, A., Nazari, M., Kerachian, R. (2025). Enhancing long-lead rainfall forecasting in data-scarce large watersheds using multi-model fusion. Journal of Hydrology: Regional Studies, 62, 102936.

[8] Wang, E., Brown, H., Zheng, B., et al. (2025). Molecular-physiological model integration revolutionizes cereal flowering prediction. The New Phytologist, 249(3), 1373–1388.

[9] Küllahcı, K., & Altunkaynak, A. (2024). Eigen time series modeling: a breakthrough approach to spatio-temporal rainfall forecasting in basins. Neural Computing and Applications, 37(6), 1–22.

[10] Yumkhaibam, S., & Kusre, C. B. (2023). Application of artificial neural networks for time series rainfall forecasting in the Loktak lift irrigation command area of Manipur, India. Irrigation and Drainage, 73(2), 741–756.