

# Time Series Prediction Method for Meteorological Data Based on the ARIMA-LSTM Model

Feifan Ji

University of Waterloo, Waterloo, Canada

---

**Abstract:** The purpose of this article is to explore and verify the effectiveness and advantages of ARIMA-LSTM hybrid model in meteorological data time series prediction. In this article, firstly, the meteorological data are pre-processed, including data cleaning, abnormal value detection and processing, and standardization operation to ensure the quality and consistency of the data. Based on this, the ARIMA-LSTM hybrid model is constructed. This model combines the ability of ARIMA (autoregressive integral moving average model) model in capturing linear relationship and short-term fluctuation, and the powerful ability of LSTM (long-term and short-term memory network) in dealing with nonlinear problems and long-term dependence. The experimental results show that the ARIMA-LSTM hybrid model has obvious advantages in forecasting meteorological data. Compared with single ARIMA model and LSTM network, the hybrid model has improved the prediction accuracy and stability. The model not only improves the prediction accuracy and stability, but also provides a reliable scientific basis for meteorological prediction and decision support.

**Keywords:** ARIMA-LSTM hybrid model, Meteorological data, Time series prediction, Prediction accuracy.

---

## 1. Introduction

As an important subject in the cross field of atmospheric science and data science, the time series prediction of meteorological data has always been widely concerned [1]. Meteorological data is not only related to people's daily life, but also plays a vital role in agricultural production, aviation safety and water resources management [2]. However, the meteorological system is a highly complex nonlinear system, which is influenced by many factors, and these factors interact with each other to form a complex and changeable meteorological phenomenon [3]. Therefore, how to accurately predict the time series of meteorological data has become a challenging problem [4].

Traditional meteorological forecasting methods mainly rely on physical models and statistical models [5]. Physical models predict meteorological changes by simulating physical processes in the atmosphere, but this method requires a lot of computing resources and accurate initial conditions, and it may be difficult to accurately simulate some complex meteorological phenomena [6]. Statistical models make predictions based on historical data, but it is often difficult to capture the nonlinear relationship and long-term dependence in the data. With the rapid development of data science and machine learning technology, especially the rise of deep learning technology, it provides new ideas and methods for meteorological data time series prediction [7].

The purpose of this study is to explore the method of forecasting meteorological data time series by using ARIMA-LSTM mixed model. By combining the traditional ARIMA model with the LSTM network in deep learning, we can give full play to their advantages in capturing linear and nonlinear relationships and improve the stability of meteorological prediction. This is not only of great significance for improving the accuracy and timeliness of meteorological forecast, but also provides more reliable meteorological data support for agricultural production, aviation safety, water resources management and other fields, and promotes scientific decision-making and intelligent development in

related fields.

## 2. Related Theoretical Basis

### 2.1. Fundamentals of time series analysis

In the field of time series analysis, understanding the law of data changing with time is the core task in many disciplines and fields [8]. Time series data widely exist in many fields, such as economy, finance, meteorology, engineering and so on. Accurate prediction of these data can provide important basis for decision-making, resource allocation and risk management.

Time series data is a collection of observation values arranged in time sequence, which usually has the characteristics of trend, seasonality and random fluctuation [9]. In order to effectively analyze time series data, researchers have developed a series of models and methods. These models are designed to capture the potential patterns in the data, thus providing reliable predictions for future data points. Time series analysis not only pays attention to the past and present of data, but also predicts its future trend by revealing the inherent laws in data.

### 2.2. ARIMA model

ARIMA is one of the classic methods in time series prediction. It captures the linear relationship and short-term fluctuation in the data by decomposing the time series data into autoregressive part, differential part and moving average part. The advantage of ARIMA model lies in its solid mathematical foundation and relatively simple model structure, which makes it perform well in dealing with linear time series data. However, when faced with nonlinear, high-dimensional or long-term dependent time series data, the prediction ability of ARIMA model may be limited.

### 2.3. LSTM neural network

With the rise of deep learning technology, LSTM models have begun to emerge in the field of time series prediction. As a specially designed variant of RNN (Recurrent Neural

Network), LSTM incorporates innovative structures such as memory units and forget gates, significantly overcoming the challenges of gradient decay and explosion encountered by traditional RNNs when processing long sequence data [10]. This model excels at capturing long-range correlations in time series and demonstrates excellent fitting performance when facing nonlinear problems. Therefore, when dealing with complex, multidimensional, and nonlinear time series data, LSTM not only has higher prediction accuracy compared to classical methods such as ARIMA, but also has better stability and adaptability.

#### 2.4. The concept of mixed model

In view of the advantages and limitations of ARIMA model and LSTM network, researchers began to explore a hybrid model combining the two [11]. The hybrid model aims to integrate the ability of ARIMA model in capturing linear relationship and short-term fluctuation, and the advantages of LSTM network in dealing with nonlinear problems and long-term dependence. By skillfully integrating these two methods, the hybrid model is expected to make a breakthrough in improving the accuracy and stability of time series prediction. This combination can not only give full play to the advantages of each model, but also make up for each other's shortcomings to a certain extent, bringing new research ideas to the field of time series prediction.

### 3. Time Series Prediction of Meteorological Data Based on ARIMA-LSTM Model

#### 3.1. Data preprocessing

Before the model is built, data preprocessing is a crucial step. Firstly, this article obtains multidimensional meteorological data including temperature, humidity, wind speed and precipitation from authoritative meteorological institutions. These data often have problems of missing values, abnormal values and different dimensions, so they need to be cleaned and standardized. For missing values, we use interpolation method and regression prediction method to fill in; For abnormal values, statistical methods are used to detect and eliminate them. In order to eliminate the influence of different dimensions on model training, this article standardizes the data so that all eigenvalues fall within the same range. After these preprocessing steps, we get a clean, tidy and easy-to-model meteorological data set.

#### 3.2. Model construction and training

After the data preprocessing is completed, this article begins to construct ARIMA-LSTM hybrid model. Firstly, this section uses ARIMA model to model the linear relationship and short-term fluctuation in meteorological data. In the stage of modeling, three key parameters, autoregressive order, differential order and moving average order, are carefully selected to ensure that the model can best fit the data characteristics. The ARIMA formula is as follows:

$$(1 - \phi_1 L - \phi_2 L^2 - \dots - \phi_p L^p)(1 - L)^d Y_t = c + \theta_1 L + \theta_2 L^2 + \dots + \theta_q L^q \epsilon_t \quad (1)$$

Where  $Y_t$  is the actual value of time series at time point  $t$ ;  $\phi_p$  is the coefficient of autoregressive term, and  $P$

represents the order of autoregressive term;  $L$  is the lag operator and  $d$  is the difference order, which is used to make the time series stable.  $c$  is a constant term;  $\theta_q$  is the coefficient of the moving average term, and  $q$  represents the order of the moving average;  $\epsilon_t$  is a white noise term, which represents the random error of time series.

After repeated debugging and optimization, we finally obtained the preliminary prediction results of ARIMA model, which laid a solid foundation for the subsequent hybrid model construction. However, ARIMA model has limitations in dealing with nonlinear relations, so LSTM network is further introduced in this section. With its unique network structure and memory mechanism, LSTM can easily capture long-term dependencies in data, and accurately fit complex functional relationships through nonlinear transformation of multi-layer neurons. This characteristic of LSTM network makes it show great strength and potential when dealing with nonlinear time series data. The formula of the hybrid model combining ARIMA and LSTM is as follows:

$$Y_{pred} = F_{LSTM}(F_{ARIMA}(Y_t), h_{t-1}) \quad (2)$$

Where:  $Y_{pred}$  is the final predicted value;  $h_{t-1}$  is the hidden state of the LSTM model at time  $t-1$ .

By combining ARIMA model with LSTM network, this section successfully constructs ARIMA-LSTM hybrid model. This model not only inherits the advantages of ARIMA model in capturing linear relationship and short-term fluctuation, but also fully absorbs the advantages of LSTM network in dealing with nonlinear relationship and long-term dependence.

In the stage of model training, this section adopts the method of cross-validation to select the optimal model parameters, and optimizes the weight and bias of the model through the back propagation algorithm. Through repeated iterative training, an ARIMA-LSTM hybrid model is obtained, which can not only capture linear relations but also deal with nonlinear problems.

#### 3.3. Experimental design and result analysis

In order to test the efficacy of ARIMA-LSTM mixed model, a set of rigorous experimental procedures was specified. In the process, the data set is divided into two parts: training and testing. The former is dedicated to model learning, and the latter is used to measure the model effect. ARIMA model, LSTM network and ARIMA-LSTM fusion model are used to predict the test set respectively, and their prediction outputs are recorded systematically. In order to comprehensively evaluate the prediction ability of the model, the experiment adopted mean square error (MSE), root mean square error (RMSE), mean absolute error (MAE) and accuracy index as evaluation criteria.

Table 1 shows the error of the model.

**Table 1.** Comparison of prediction errors

Model	MSE	RMSE	MAE
ARIMA	0.023	0.152	0.118
LSTM	0.019	0.138	0.105
ARIMA-LSTM	0.015	0.122	0.093

From the evaluation index of MSE, the MSE value of ARIMA-LSTM mixed model is 0.015, which is obviously lower than that of single ARIMA model (0.023) and LSTM network (0.019). This shows that ARIMA-LSTM hybrid model has higher accuracy in forecasting meteorological data.

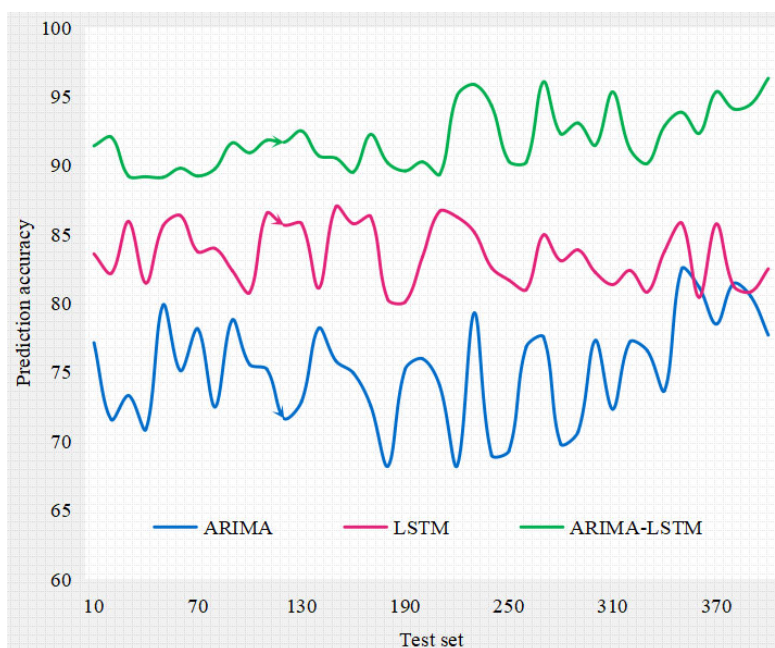
In terms of RMSE, the RMSE value of ARIMA-LSTM mixed model is 0.122, which is also lower than ARIMA model (0.152) and LSTM network (0.138). This further confirms the superiority of ARIMA-LSTM hybrid model in forecasting meteorological data.

From the perspective of MAE, the MAE value of ARIMA-LSTM mixed model is 0.093, which is also the lowest among the three models. This once again proves the effectiveness and

advantages of ARIMA-LSTM hybrid model in forecasting meteorological data time series.

To sum up, the ARIMA-LSTM hybrid model is superior to the single ARIMA model and LSTM network in MSE, RMSE and MAE. This shows that the hybrid model performs better in capturing trends and seasonal changes in meteorological data, and also shows stronger ability in dealing with nonlinear relations and long-term dependence. Therefore, this article draws the conclusion that ARIMA-LSTM hybrid model has higher prediction accuracy and stability in meteorological data time series prediction, and it is an effective prediction method.

Figure 1 shows the prediction accuracy of the model.



**Figure 1.** Prediction accuracy of the model

The prediction accuracy of ARIMA-LSTM hybrid model reaches 90% and higher. This remarkable improvement shows that the hybrid model successfully combines the advantages of ARIMA model and LSTM network, which can not only accurately capture linear relationship and short-term fluctuation, but also effectively deal with nonlinear relationship and long-term dependence.

Combined with the previous analysis of model errors (MSE, RMSE, MAE), we can further confirm the advantages of ARIMA-LSTM hybrid model in prediction accuracy and stability. The hybrid model not only improves the prediction accuracy, but also makes remarkable achievements in reducing the prediction error. These results fully prove the effectiveness and superiority of ARIMA-LSTM hybrid model in forecasting meteorological data time series.

## 4. Conclusions

This study is devoted to exploring and verifying the effectiveness of ARIMA-LSTM hybrid model in forecasting meteorological data time series. By combining the ability of ARIMA model in capturing linear relationship and short-term fluctuation, and the strong ability of LSTM network in dealing with nonlinear problems and long-term dependence, a hybrid forecasting model is constructed in this article, which can not only accurately reflect the linear characteristics of meteorological data time series, but also effectively capture

its nonlinear changes. The experimental results show that compared with single ARIMA model and LSTM network, ARIMA-LSTM hybrid model has obvious advantages in prediction accuracy and stability. This discovery not only provides new ideas and methods for time series prediction of meteorological data, but also provides useful reference for time series prediction in other fields.

In the stage of model construction and training, this article deeply analyzes the characteristics and laws of meteorological data, and comprehensively preprocesses the original data to ensure the quality and consistency of the data. At the same time, through repeated experiments and parameter optimization, the ARIMA-LSTM hybrid model with superior performance is obtained. When forecasting meteorological data, the model can accurately capture the trend, seasonality and random fluctuation in the data, which provides a reliable scientific basis for meteorological forecasting and decision support.

## References

- [1] Işık E, Inallı M. Artificial neural networks and adaptive neuro-fuzzy inference systems approaches to forecast the meteorological data for HVAC: the case of cities for Turkey[J]. Energy, 2018, 154: 7-16.
- [2] Liao S, Wang H, Liu B, et al. Runoff Forecast Model Based on an EEMD-ANN and Meteorological Factors Using a Multicore

- Parallel Algorithm[J]. *Water Resources Management*, 2023, 37:1539-1555.
- [3] Huang T, Jiao F. Study on data transfer in meteorological forecast of small and medium-sized cities and its application in zhaqing city[J]. *The Computer Journal*, 2020, 63(7): 1076-1083.
- [4] Zarei A R, Mahmoudi M R, Pourbagheri A. Meteorological Drought Prediction Based on Evaluating the Efficacy of Several Prediction Models[J]. *Water resources management*, 2024, 38(7):2601-2625.
- [5] Gupta A, Chug A, Singh A P. Meteorological Factor-Based Tomato Early Blight Prediction Using Hyperparameter Tuning of Intelligent Classifiers[J]. *Agricultural Research*, 2024, 13(2):232-242.
- [6] Kimura Y. Numerical Weather Prediction Model in Japan Meteorological Agency[J]. *The Journal of The Institute of Electrical Engineers of Japan*, 2023, 143(5):267-270.
- [7] Zhai B, Wang Y, Wu B. An ensemble learning method for low visibility prediction on freeway using meteorological data[J]. *IET intelligent transport systems*, 2023, 17(11):2237-2250.
- [8] Singh T, Uppaluri R V S. Feed-forward ANN and traditional machine learning-based prediction of biogas generation rate from meteorological and organic waste parameters[J]. *Journal of supercomputing*, 2024, 80(2):2538-2571.
- [9] Cho S, Youn Y, Kim S, et al. A Comparative Evaluation of Multiple Meteorological Datasets for the Rice Yield Prediction at the County Level in South Korea[J]. *Journal of remote sensing*, 2021, 37:337-357.
- [10] Wen H, Dang Y, Li L. Short-Term PM<sub>2.5</sub> concentration prediction by combining GNSS and meteorological factors[J]. *IEEE Access*, 2020, 8: 115202-115216.
- [11] Liang H, Zhang M, Wang H. A neural network model for wildfire scale prediction using meteorological factors[J]. *IEEE Access*, 2019, 7: 176746-176755.