Proceedings of the 2nd International Conference on Mathematical Physics and Computational Simulation
DOI: 10.54254/2753-8818/42/2024CH0223

Research on the climate change and prediction of the temperature in Beijing

Xuefei Liu

School of Mathematical Sciences, University of Nottingham, Nottingham, NG7 2RD, United Kingdom

smyxl10@nottingham.ac.uk

Abstract. Over the past few years, global warming has become a growing threat to the human living environment. Thus, preventing extreme temperature changes is a crucial area of study for the global community. However, temperature changes have diverse causes, making it hard to accurately calculate the data that impacts subsequent temperature forecasts. The ARIMA model is widely favoured by researchers as a typical time series prediction model. Additionally, it holds significant research value in forecasting dynamic models. This paper will use the ARIMA model to forecast temperature changes in Beijing for the next month, starting on 21 July. The findings show that the temperature remains relatively stable in the upcoming month, with minimal fluctuations. Based on historical data, this is the peak temperature in Beijing for the year. So, if the temperatures are not extremely high now, they might not get higher later. This forecast can be used as a guide for managing extreme temperatures in Beijing in the future.

Keywords: Prediction, ARIMA model, temperature changing.

1. Introduction

Global climate change has already become one of the most emergency environmental challenges which confronted by the world. Meteorological data, from 1976 to 2006, showed that global surface temperatures have been rising at an average rate of 0.2°C per decade [1]. It posed an unprecedented threat to the human well-being, economies, and the ecosystem [2]. It showed the effectiveness of Beijing's defenses against temperature fluctuations to a certain degree [3]. Therefore, the trend and cause of Beijing's rising temperature have a crucial impact on relevance and academic value.

Multiple researchers and scholars have developed various methods for predicting temperature changes and the corresponding prediction models. Several factors, such as air temperature, air pressure, relative temperature, PM2.5 concentration, and surface temperature, play a significant role in the model of temperature predictions [4, 5]. Wan et al. developed a spatial and temporal model with varying coefficients, which can indirectly impact temperature predictions while forecasting PM2.5 concentrations in Beijing, accounting for the influence of climatic conditions [6]. This approach also enables people to comprehend the temperature alteration. Because of the correlation between air temperature and pavement temperature, this is also a factor in making predictions about air temperature. In addition, Chao et al. collected ground temperature data and utilized the partial least squares method to create a regression model that predicts asphalt pavement temperature using environmental data [7]. In their study, Dong et al. developed three statistical models using the

Characteristics of Road Surface Temperature in Beijing Winter and its statistical forecasting models based on the Rapid-refresh Multi-scale Analysis and Prediction System (RMAPS) product. The model incorporating additional factors improved accuracy by 25% compared to the basic data model, while keeping the mean absolute error within 2.5°C [8]. The model with more factors was used as a model with a higher accuracy than the simple data model. By analyzing data from 1960 to 2004, Yang compared the accuracy of PSO_LSSVM (The particle swarm optimization (PSO) is added to a least squares support vector machine (LSSVM) prediction model) and Elman neural network models in predicting the average temperature of Beijing for the next four years [9]. Both models showed a high feasibility in simulating the overall temperature, yet they could not forecast the low temperature. Among these two models, the PSO_LSSVM model has higher accuracy and is better for temperature prediction in the real production industry. As seen above, including more factors and utilizing suitable and adaptable models are two methods to boost prediction accuracy. Deep learning has also shown that it is possible to predict temperatures by examining various prediction examples. Qi et al. investigated the possibilities of machine learning in accurately expecting local temperatures and organizing the Winter Olympic event [10]. By utilizing a Long Short Term Memory (LSTM) model, they employed historical temperature data to forecast the upcoming hour's temperature in a bustling megacity in North China. The model showed strong potential in accurately predicting both warming and cooling patterns, with Beijing having the most accurate predictions among all cities.

In conclusion, the forecast of Beijing's temperature has captured attending a multitude of scholars. This article will utilize the ARIMA model to predict Beijing's temperature between 2014 and 2024. Following the forecast, a corresponding interpretation and analysis will be presented.

2. Methodology

2.1. Data sources

The data comes from the U.S. National Oceanic and Atmospheric Administration's (NOAA) temperature data, which counts Beijing's daily average temperatures from January 2023 to July 2024, with a total of 563 data.

2.2. Variable selection

Changes in Beijing's temperature is closely related to global warming, and changes in Beijing's temperature reflect the trend of global temperature change, so it is important to study changes in Beijing's temperature. Temperatures are influenced by topography, climate, human activities and other factors, which are unpredictable and it is difficult to identify the trend of variations, as illustrated in Figure 1:

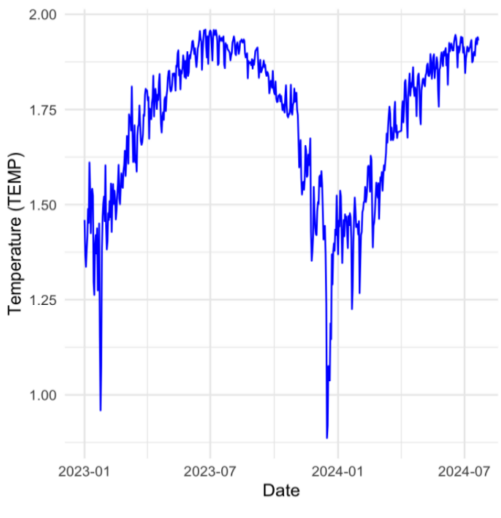

Figure 1. Temperature in Beijing

DOI: 10.54254/2753-8818/42/2024CH0223

In 2023, the data can be concluded from the figure above. The lowest temperatures marked both the start and end of the year, and from the first one, volatility gradually rose to its highest point in July 2023. Afterward, it is a drop after the highest point and then a decline in fluctuation which experiences approximately 5 months. It can also be deduced from the graph that there is a periodicity in the variation of temperature within a period of about one year. Additionally, it has a cause for concern that the start of 2024 is hotter than the start of 2023.

2.3. Model selection

This article uses the Autoregressive Integrated Moving Average (ARIMA) model to predicate the variation of daily temperature in Beijing from 2023 to 2024. ARIMA model, widely applies in time series data, contains three parameters to analyse which are autoregressive (AR), integrated(I), and moving average (MA). AR is the autoregressive component used to predict forward values using a linear combination of historical data, denoted by p, showing the use of the first p lag terms. In addition, this paper indicates the difference, which is used to render the time series more stable, and is denoted by d, which denotes the number of times the difference is conducted. MA is the moving average part of the model, which represents the relationship between the current value and a linear combination of past random error values, denoted by the parameter q, indicating the use of the first q error values. Due to a value of p greater than 0.05, it requires careful consideration of the difference.

Firstly, definition of the formula for the p-order autoregressive process: $y_t = \mu + \sum_{i=1}^{p} \gamma_{t-i} + \epsilon_t$, y_t is current value, μ is constant term, p is the order, γ_i is the autocorrelation coefficient, ϵ_t is the error. For P-order, the first order represents today and yesterday, and the second order represents today and the day before. Secondly, the moving average model is concerned with the accumulation of error terms in the autoregressive model. The q-order autoregressive process is defined by the formula: $y_t = \mu + \epsilon_t + \sum_{i=1}^{q} \theta_i \epsilon_{t-i}$, The method of moving average is conducive in eliminating random fluctuations in forecasting. Finally, the formula for combining autoregressive and moving average:$y_t = \mu + \gamma_i Y_{t-i} + \epsilon_t + \sum_{i=1}^{q} \theta_i \epsilon_{t-i}$, p, q need to specified, γ_i and θ_i need to solved [10].

3. Results and discussion

3.1. Data processing

The premise of using the ARIMA model is that the data is smooth. If the data is not smooth, it is necessary to subject the data to a specific process in order to achieve smoothness. The analysed ACF graph shows irregularities in the data (Figure 2):

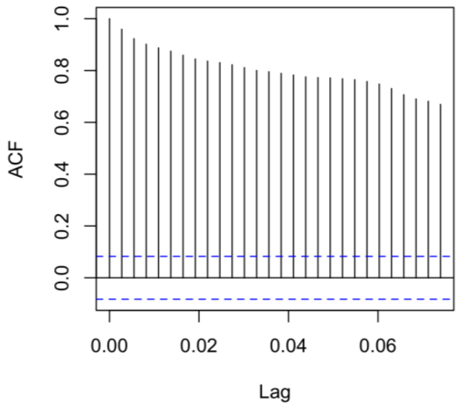

Figure 2. The original plot of ACF

Since the data is not smooth, this paper needs to perform differencing. This paper first performs the first order differencing process and then the author can get the differenced ACF (Figure 3) and PACF (Figure 4) plots. The table 1 below presents the results got from the ADF test, showing a p-value of

DOI: 10.54254/2753-8818/42/2024CH0223

0.01. Therefore, this paper can deduce that the data becomes stationary when first-order differencing is applied.

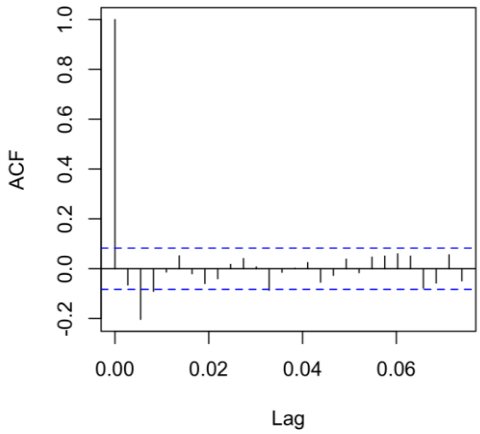

Figure 3. The ACF plot

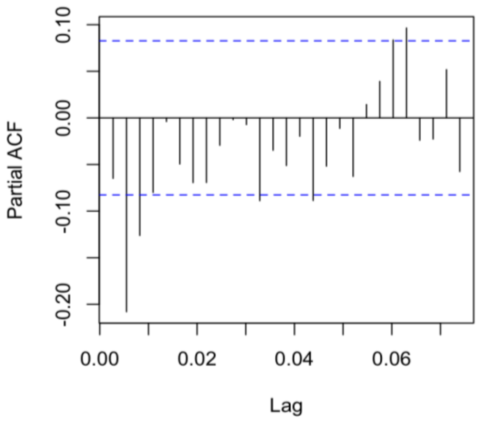

Figure 4. The PACF plot

Table 1. The ADF test

Dickey-Fuller	Lag order	p-value
-10.111	8	0.01

3.2. Model evaluation

To improve the fit of the ARIMA model, adjusting the values of p, d, and q can help select a more accurate model for predicting the data. A smaller error between the fitted and the actual values indicates a better fit for the model. This paper uses a comparison of the magnitude of the values of Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC) and Root Mean Square Error (RMSE), in Table 2 to determine the parameters of the ARIMA model.

AIC and BIC are utilised to assess the statistical model, measuring its goodness and complexity. As a result, the smaller their values and the more accurately the model represent balance, effectively fitting the data without undue complications. BIC has a preference for more straightforward models. The Root Mean Square Error (RMSE) is a metric utilised to evaluate the predictive capability of the model. It is calculated as the square root of the average of the squared differences.

Table 2. Model evaluation

ARIMA Model	RMSE	AIC	BIC
(1,1,1)	0.054	-1675.899	-1662.904
(1,1,2)	0.054	-1680.856	-1663.53
(2,1,1)	0.054	-1679.783	-1662.457
(2,1,2)	0.054	-1679.169	-1657.512

From the table 2, the results show that ARIMA (1, 1, 2) and ARIMA (2, 1, 2) have the smallest values of RMSE under comparison. Then, by comparing the AIC and BIC, it is found that the value of ARIMA (1, 1, 2) is smaller. Therefore, the subsequent predictions will be performed using this model in the following section.

3.3. Forecasting results

The below ARIMA (1, 1, 2) model will be applied to predict the weather changes in the following 30 days (Figure 5).

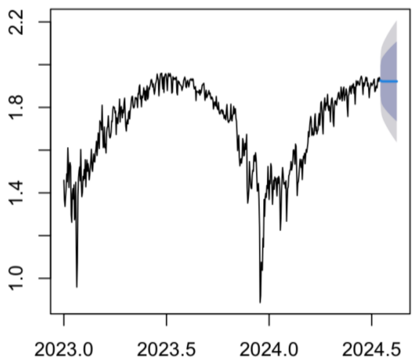

Figure 5. Forecasting Plot

The prediction results and upper and lower confidence intervals for the first 10 values in the forecast are shown below (Table 3). On the left side of the forecast chart, you can see the actual observations, which exhibit a more pronounced pattern of fluctuations in the overall data, including some cyclical variations. The predicted values of the model are shown on the smoother part of the right-hand side of the graph, with the blue line showing the predicted point estimate. Around the blue portion of the forecast, the grey and dark blue areas show 80 percent and 95 percent confidence intervals. The confidence intervals become progressively wider over time, showing that uncertainty increases with extension.

Table 3. Forecast data

Date	Point Forecast	Lo 80	Hi 80	Lo 95	Hi 95
2024/7/22	1.926	1.857	1.995	1.820	2.032
2024/7/23	1.923	1.832	2.016	1.783	2.064
2024/7/24	1.923	1.820	2.025	1.767	2.079
2024/7/25	1.922	1.814	2.031	1.756	2.088
2024/7/26	1.922	1.809	2.036	1.748	2.096
2024/7/27	1.922	1.800	2.040	1.742	2.102
2024/728	1.922	1.800	2.043	1.736	2.108
2024/7/29	1.922	1.797	2.047	1.730	2.113
2024/7/30	1.922	1.793	2.050	1.725	2.119
2024/7/31	1.922	1.790	2.054	1.720	2.124

3.4. Residuals check

In the forthcoming section, the aim is to evaluate whether the residual makes up white noise, in order to ensure the model's effective prediction of the data.

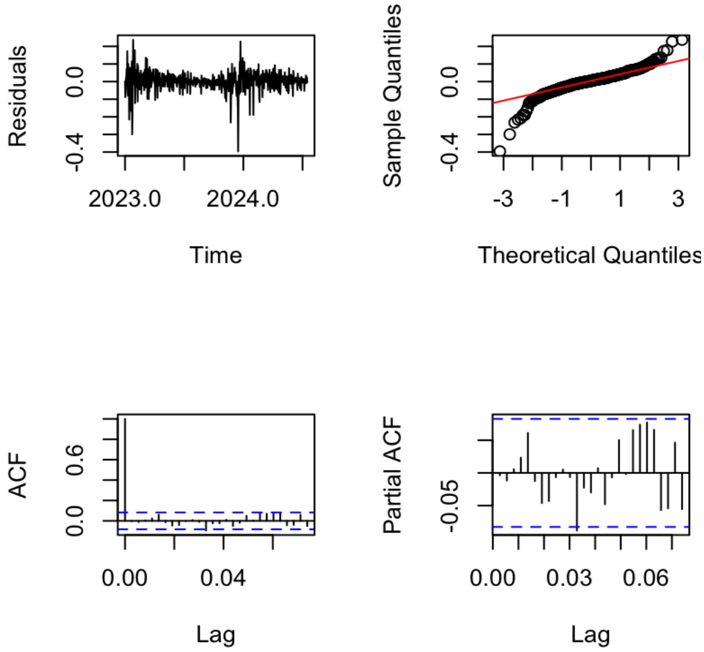

Figure 6. Residuals plot

In the Figure 6 above, the top left corner depicts the residual time series, which lacks any noticeable trend or periodicity. This shows a relatively good fit for the model. The QQ normality plot in the top right corner tests compliance with the normal distribution. Most of the points in this graph are distributed on the diagonal, showing that the residuals are in line with the normal distribution. Although there are extreme points present at the ends, their impact on the overall validity of the model is negligible. To evaluate white noise in the residuals, this paper utilizes the following pair of graphs. In the ACF graph on the left, aside from the initial lag, all other lags fall within the confidence interval, showing no autocorrelation. Similarly, the residuals in the right PACF graph mostly lie within the confidence interval, suggesting the absence of significant autocorrelation. This implies that the model's fit is relatively satisfactory.

The Ljung-Box test is used to test for autocorrelation in the residuals (Table 4). When the p-value is greater than 0.05, the residuals are not considered being auto-correlated and the model is selectable. The table below shows the magnitude of the p-value and the relevant parameters.

Table 4. Ljung-Box test

Q*	df	p-value
17.102	20	0.65

3.5. Critical thinking

Despite the conclusions and the corresponding validation in this paper, the predicted parts have a certain feasibility and goodness of fit. Moreover, the data used is available for prediction and analysis. However, there are still several issues that are unresolved in the whole process. First, as seen by the residual time series plot, there are some regions that still have large fluctuations. This indicates that the model's accuracy is lacking in specific regions and a more complex model might be necessary. It's important to consider the existence of outliers and their potential impact on the findings. In order to account for the substantial volatility in the original data, logarithmic processing was performed to create smoother data and enhance model suitability. However, there are still some extreme values that can affect the model fit. This paper examines the cyclical temperature change in Beijing, which is influenced by its semi-humid and semi-arid monsoon climate and four distinct seasons. Thus, using a dataset that spans approximately 1.5 years, this paper observes a single cycle to decrease periodicity and enhance the model's feasibility. However, the temperature projection in this article is subject to limitations given the influence of global warming and extreme weather.

4. Conclusion

This paper concludes that temperature can be predicted to some extent, but because of global warming, monsoon climate, and extreme weather, temperature comparisons are more flexible and unpredictable. Hence, it is important to utilize advanced models and precise measurement tools for temperature predictions in different areas. The present study employs a time series model to forecast temperature, presenting a potential avenue for weather researchers.

Through weather prediction, the government can issue timely warnings to the public and adapt the schedule and location of human activities in response to extreme temperatures. The ability to determine temperature changes allows individuals to decide regarding clothing additions or subtractions, as well as the timing of their outings. This study holds great importance because of its direct relevance to the well-being and livelihood of people.

References

[1] Hansen J, Makiko S, et al. 2006 Global temperature change. Proceedings of the National Academy of Sciences, 103, 14288-14293.
[2] Zhang L, Xu Z X and Ruan B Q 2006 Effect of Urban Heat Island on the Air Temperature and Precipitation in Beijing Region. JOURNAL OF NATURAL RESOURCES, 21(5), 746-755
[3] Wang Y, et al. 2014 The Impact of Urbanization on the Annual Average Temperature of the Past 60 Years in Beijing. Advances in Meteorology, 374987.
[4] Hansen J, Ruedy R, et al. 2001 A closer look at United States and global surface temperature change. Journal of Geophysical Research, 106, 23947-23963.
[5] Xiong Y J 2015 Application of KNN Data Mining Algorithm to Haze Grade Forecasting in Beijing. Meteorological Monthly.
[6] Wan Y, Xu M Y, et al. 2020 A spatio - temporal model for the analysis and prediction of fine particulate matter concentration in Beijing. Environmetrics, 32.
[7] Jing C and Zhang J X 2018 Prediction Model for Asphalt Pavement Temperature in High-Temperature Season in Beijing. Advances in Civil Engineering.
[8] Dong Y, Guo W L, et al. 2019 Characteristics of Road Surface Temperature in Beijing Winter and its statistic forecasting models Based on the RMAPS Product. IOP Conference Series: Earth and Environmental Science, 384.
[9] Yang J P 2013 Comparison of Beijing Temperature Prediction Effect Based on PSO_LSSVM and Elman Neural Network Model. Journal of Henan Agricultural Sciences.
[10] Qi Y J and Guo C W 2020 Deep Learning-based Hourly Temperature Prediction: A Case Study of Mega-cites in North China. Proceedings of the 3rd International Conference on Big Data Technologies.