DOI: 10.54254/2753-8818/51/2024CH0197

Prediction of atmospheric temperature in Sacramento area based on ARIMA and ETS models

Kairui Li

Department of Statistics & Data Science, Faculty of Science, National University of Singapore, 117546, Singapore

e1348815@u.nus.edu

Abstract. As global temperatures increased by 1.1 Celsius degrees, there has unprecedented shifts in climate systems. With the rising impact of global warming, exploring the warming trend helps to better understand and maintain the local environment and economy. This study focuses on predicting atmospheric temperature in the Sacramento area using ARIMA and ETS models. The research uses the temperature data from Sacramento Airport's Automated Surface Observing System (ASOS) and explores the prediction performance of ARIMA, ETS, and ARIMAX models in predicting daily average temperatures. The results indicate the ARIMAX (1,1,1) model is the most suitable for forecasting this temperature data with the lowest AIC and RMSE values. However, there are still challenges, in particular the dependence of the ARIMAX model on future exogenous variables, which leads to the forecast outcomes less accurate. To enable this ARIMAX models have better prediction performance, incorporating more exogenous variables is a potential solution. Therefore, the methods discussed in this paper provides ideas for the atmospheric temperature forecasting and points out the direction for further research.

Keywords: ARIMA, ARIMAX, ETS, atmospheric temperature, weather forecasting.

1. Introduction

As greenhouse gas emissions continue, the global climate is experiencing unprecedented shifts. The recently released Intergovernmental Panel on Climate Change (IPCC) Sixth Assessment Report (AR6) provides a detailed assessment of the devastating consequences from global warming, emphasizing the critical importance of understanding average temperatures [1]. As global temperatures have already increased by 1.1 Celsius degrees, there has unprecedented shifts in climate systems, including extreme weathers and rising sea level [1]. The Sacramento-San Joaquin Delta, a vital region for agriculture and water supply in California, has witnessed a notable increase in extreme weathers over the past decade, including droughts and floods [2]. Global warming and climate change lead to more intensive rainy seasons, which cause longer drought periods [3]. The region is expected to become hotter, drier, and more susceptible to droughts, flooding, and large wildfires [3]. The underlying causes involve a combination of both natural climate variability and anthropogenic factors [2]. With warmer temperatures, more extreme floods and droughts are anticipated, which poses significant challenges to the environment and economy of the Delta region [3]. The urgency of action is clear. Under such background, exploring the warming trend helps to better understand and maintain the local environment and economy. Therefore, trying to accurately predict temperature changes has become a hot topic.

© 2024 The Authors. This is an open access article distributed under the terms of the Creative Commons Attribution License 4.0 (https://creativecommons.org/licenses/by/4.0/).

DOI: 10.54254/2753-8818/51/2024CH0197

Temperature forecasting is a highly challenging mission for scientists which attracts a great deal of researchers' attention within last decades. Meteorological institutes forecast the future weather based on numerical weather prediction (NWP) models which based on current weather data. However, the predictive performance of NWP models in complex geographic areas gets challenged [4]. In response to the challenges, diverse new models are being proposed. Yi and Prybutok proposed a new model called multi-layer perceptron (MLP) [5]. And another model called deep neural network (DNN) was proposed by Jiang’s research team [6]. Pal et al. proposed two novel method, self-organizing feature map combining multi-Layer perceptron (SOFM-MLP) and feature selection MLP (FSMLP), to forecast the atmospheric temperature with better performance [7]. Radhika and Shashi used the Support Vector Machines (SVMs) to forecast atmospheric temperature, which got better prediction performance than the traditional MLP algorithm [8]. In recent year, Park and Kim successfully forecasted atmospheric temperature with defective dataset by using Long Short-Term Memory Neural Network (LSTM) [9]. Chu et al. forecasted atmospheric temperature by using recurrent neural networks only with image data [10].

In summary, the research on the global and local atmospheric temperature has attracted numerous researchers. This research paper will use the Autoregressive Integrated Moving Average (ARIMA) and Error-Trend-Seasonality (ETS) model to predict and analyse average daily atmospheric temperature in the Sacramento area. Besides, this article will also explore the prediction performance of ARIMA and ETS models with non-stationary datasets.

2. Methodology

2.1. Data source

This research paper mainly uses the temperature data got from Sacramento Airport Automated Surface Observing System (ASOS). The data is downloaded from National Oceanic and Atmospheric Administration (NOAA). This dataset contains daily information on average temperatures, maximum temperatures, minimum temperatures, and precipitation. It includes a total of 3865 observations from 2014/1/1 to 2024/7/31.

2.2. Variable selection

The daily average temperature is an important indicator of trends in climate change. As global warming occurs, the frequency of extreme weathers increases (Figure 1).

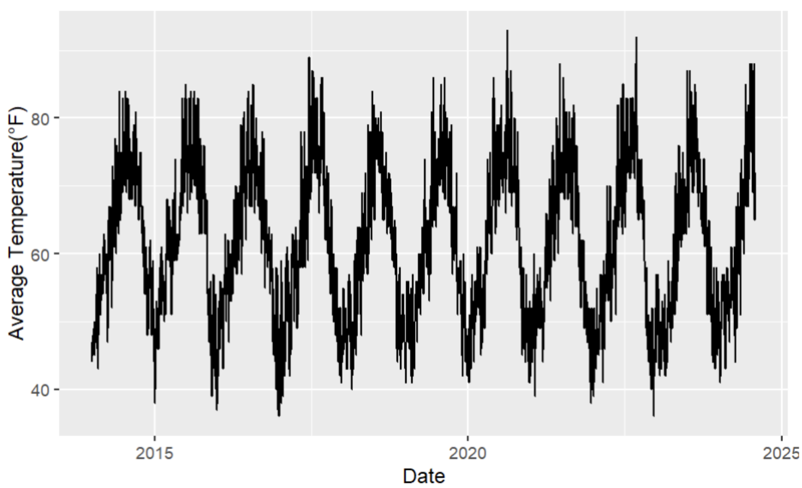

Figure 1. Average Temperature(°F) from 2014/1/1 to 2024/7/31 in Sacramento Area

The irregular and unpredictable nature of these extreme weathers makes meteorological data change frequently and hard to anticipate. The average daily temperature mitigates the effects of extreme

DOI: 10.54254/2753-8818/51/2024CH0197

weathers over short periods of time. Thus, the study of daily average temperature helps in studying and predicting the long-term climate change. Figure 1 shows the daily average temperature from 2014/1/1 to 2024/7/31 in Sacramento Area. Figure 1 reveals that the average daily temperatures show a strong periodicity, with a one-year cyclic period.

2.3. Model selection

This research paper will use the ARIMA and ETS models to predict and analyse average daily atmospheric temperature in the Sacramento area. In working with time series data, ARIMA model is a widely used model which consists of three parts: autoregression (AR), integration (I) and moving average (MA). The AR part regresses the lagged values of the variables, which takes into account the influence of previous values on the present values. The I component is designed to smooth the non-stationary time series by using differencing. The MA section is designed to deal with the effect of past prediction errors on the present values.

ETS model is another approach for time series forecasting. The three main components of the ETS model are error (E), trend (T), and seasonality (S). The error represents the difference between actual values and predicted values. The trend represents changes in the time series. Trends can be linear, nonlinear, or constant. The seasonality represents the seasonal variations and periodicity in the time series.

3. Results and discussion

3.1. Data preprocessing

The time series utilised in the ARIMA model requires stability. It is evident from the ACF plot in Figure 2 that the average temperature data have significant autocorrelation, which means the time series is non-stationary. Therefore, differencing is required to smooth the data.

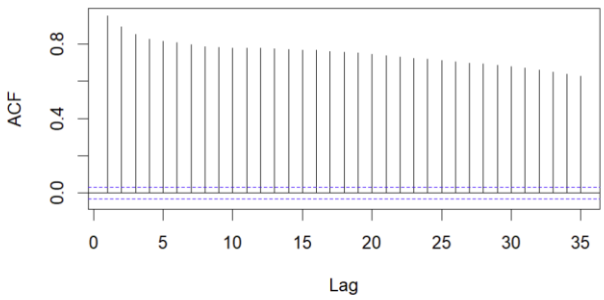

Figure 2. ACF results for Average Temperature dataset (d = 0)

Then, by differentiating the average temperature data once, the corresponding ACF plot and PACF plot are showed below in Figure 3 and Figure 4. Based on the results of the first differentiation, both of its ACF plot and PACF plot are satisfied. Therefore, the data obtained after first differentiation can be viewed as stationary.

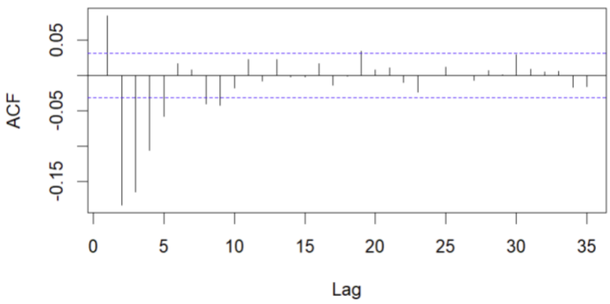

Figure 3. ACF results for Average Temperature dataset after first differentiation (d = 1)

DOI: 10.54254/2753-8818/51/2024CH0197

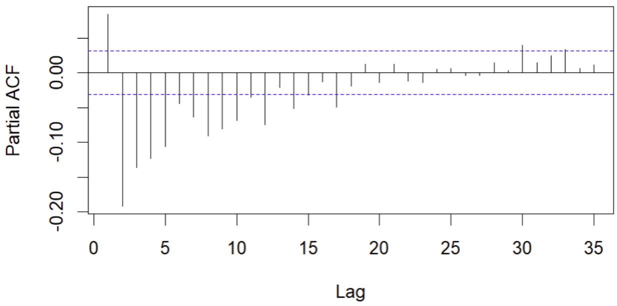

Figure 4. PACF results for Average Temperature dataset after first differentiation (d = 1)

The stability is also confirmed by the Augmented Dickey–Fuller (ADF) test with corresponding P-value = 0.01, which is less than 0.05 (Table 1).

Table 1. P-value result from ADF test after first differentiation (d = 1)

Dickey-Fuller	Lag order	p-value
-22.256	15	0.01

3.2. Model evaluation

The predicting performance of the model is reflected by the variance between the fitted and real values. The less the variance, the more accurate the model. The article has already determined the I component as d = 1 in the previous section. The next sections are designed to identify the AR and MR components.

This research mainly uses Akaike information criterion (AIC) to evaluate the predicting performance of models. The root mean square error (RMSE) helps to check the model accuracy. AIC is widely used to select best model and avoids overfitting problems. RMSE measures the variance between predicted and actual values. Both of the indicators prefer a smaller value.

Table 2. RMSE and AIC with different models

Model	RMSE	AIC
ARIMA (0,1,5)	3.337	16235.72
ARIMA (2,1,1)	3.338	16233.85
ARIMA (3,1,4)	3.319	16205.90
ETS (A,N,N)	3.541	32674.11
ETS (A,A,N)	3.541	32678.43
ETS (M,N,N)	3.541	32731.44
ARIMAX (1,1,1)	1.591	11660.94

The first model in Table 2, ARIMA (0,1,5), is selected based on the ACF plot and PACF plots. The last significant autocorrelation occurs at lag = 5 and partial autocorrelation decreases to 0 quickly. Hence, the paper uses ARIMA (0,1,5) to fit the data. The second model, ARIMA (2,1,1), is selected by auto fitting with Hyndman’s technique. ARIMA (3,1,4) is selected by finding a model with the best AIC value. The next four ETS model is selected by using different variables. The last ARIMAX model is an extended version of ARIMA model, which incorporates exogenous variables such as maximum temperatures, minimum temperatures, and precipitation.

The Table 2 above indicates that ARIMA models have smaller AIC and RMSE values than ETS models for this dataset. ARIMAX (1,1,1) has the lowest AIC and RMSE values. Therefore, the ARIMAX (1,1,1) model is considered to be the most suitable for this research on average temperature. Only this model will be used for forecasting and analyses in the following sections.

DOI: 10.54254/2753-8818/51/2024CH0197

3.3. Model fitting and prediction

This research uses ARIMAX (1,1,1) model to fit and forecast the training data and the test data for average temperature. The corresponding results are shown below in Figure 5 and Figure 6:

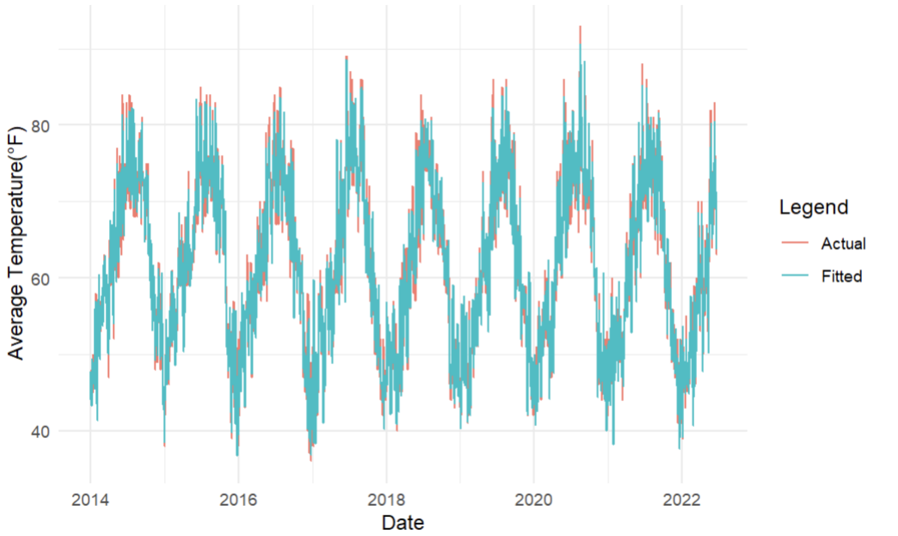

Figure 5. Fitted model from ARIMAX (1,1,1)

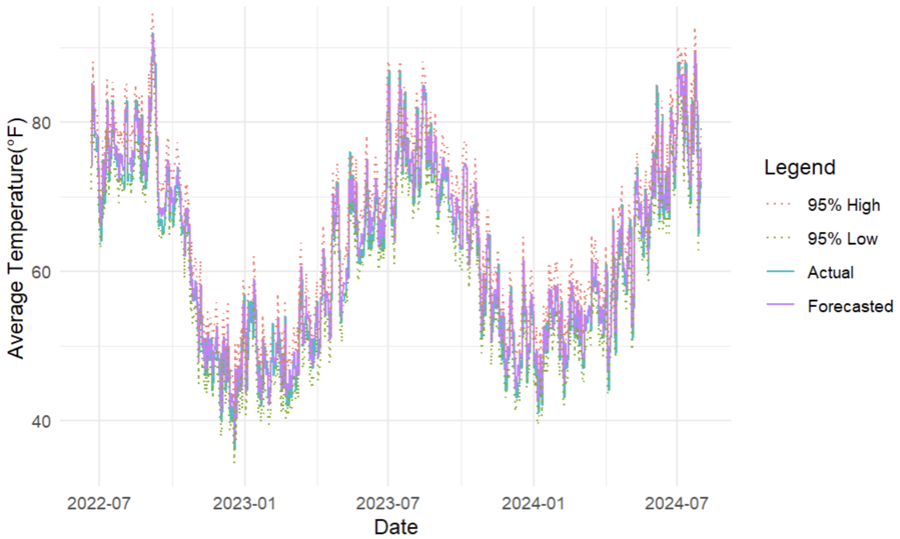

Figure 6. Prediction Results from ARIMAX (1,1,1)

In Figure 5, the red line shows the actual daily average temperature in the training dataset. And the blue line shows the fitted value from ARIMAX (1,1,1). In Figure 6, the blue line shows actual daily average temperature in the test dataset. The purple line shows the forecasted value from ARIMAX (1,1,1). The red and green dash line represent the 95% confidence interval. ARIMAX model performs exceptionally well, with a low AIC value indicating a well-balanced model, and a low RMSE value showing minimal prediction errors. This indicates that ARIMAX model is highly effective in predicting daily average temperature.

DOI: 10.54254/2753-8818/51/2024CH0197

3.4. Check residuals

To make sure the ARIMAX (1,1,1) model is effective in explaining information from the data, this section is designed to determine whether the residual errors of the model is a white noise. The residual check results are shown below in Figure 7:

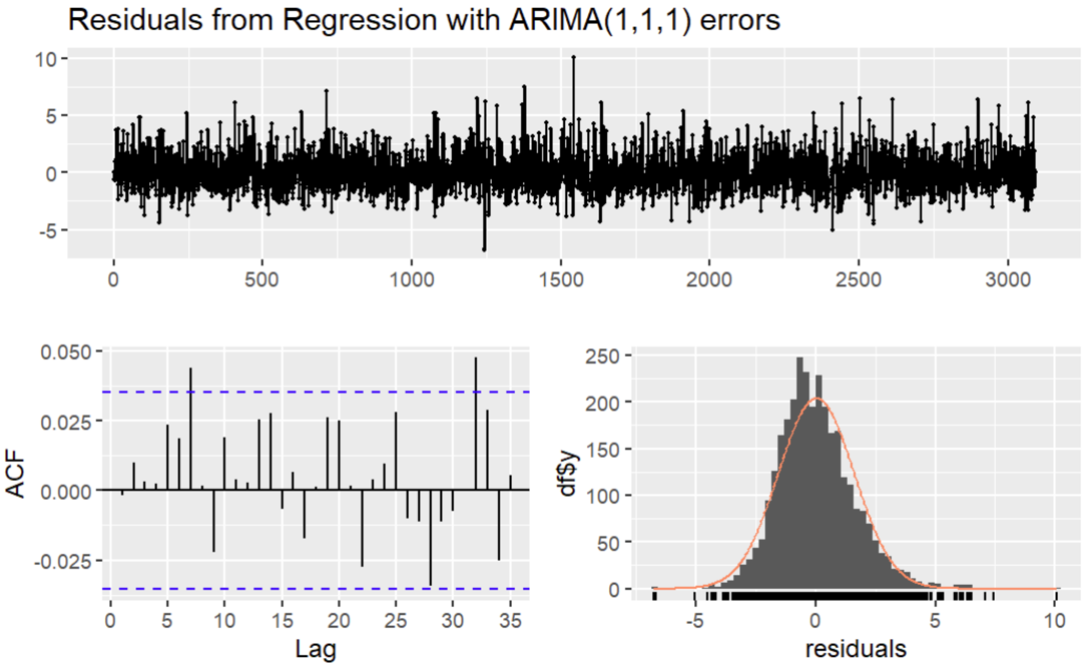

Figure 7. Residual plots of ARIMAX (1,1,1)

The upper distribution graph shows the residuals fluctuate around 0. In the ACF plot of residuals, there exists no significant autocorrelations in the data. And most of the autocorrelations are within the 95% confidence intervals. In the last residuals distribution graph, the residuals are normally distributed. Therefore, the Figure 7 confirms the distribution of residuals is a white noise.

3.5. Further discussion

From the above study, ARIMAX model performs well in daily average temperature forecasting, but there are still certain problems for application. The most significant problem is the ARIMAX model used in this research can only forecast past results. Without knowing the maximum temperatures, minimum temperatures, and precipitation for the next day, this ARIMAX model becomes ARIMA model, which leads to the forecast outcomes less accurate. To make this ARIMAX models more accurate, incorporating more exogenous variables will help improve prediction performance.

Another problem with this research is this time series is too frequent. In the research, the frequency of the temperature data is 365, which exceed the maximum frequency for ETS model and ARIMA model. This means the research cannot use seasonal ETS model and seasonal ARIMA (SARIMA) model. However, this dataset shows a strong periodicity and seasonality. Hence, the forecasting results from ARIMA model and ETS model will only perform a strong linearity and the forecasting results are not very well.

4. Conclusion

Based on the previous results, it is evident that ARIMA models perform better than ETS models in predicting daily average temperatures for the Sacramento area. Among the various models tested, the ARIMAX (1,1,1) model demonstrated the highest accuracy, and lowest AIC and RMSE values. This indicates that ARIMAX (1,1,1) is highly effective in capturing the underlying patterns in the temperature data and providing reliable forecasts.

DOI: 10.54254/2753-8818/51/2024CH0197

However, one thing worth to note is that the ARIMAX model is problematic with its applicability. Without access to future exogenous variables such as maximum and minimum temperatures and precipitation, the ARIMAX model becomes ARIMA model, which can reduce its prediction performance. To address the issue, the integration of LSTM is a considerable solution. Additionally, incorporating more exogenous variables will help improve prediction performance.

In conclusion, while the ARIMAX (1,1,1) model currently offers the best performance for this dataset, future advancements in model could yield more accurate and robust temperature predictions. This study underscores the potential of advanced time series models in climate research and their critical role in understanding and mitigating the impacts of climate change on regional environments.

References

[1] Calvin K, Dasgupta D, et al. 2023 Climate Change 2023: Synthesis Report. Contribution of Working Groups I, II and III to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change. Intergovernmental Panel on Climate Change (IPCC).

[2] Fernandez-Bou A S, Ortiz-Partida J P, Pells C, et al. 2021 California's Fourth Climate Change Assessment: San Joaquin Valley region report. California Governors.

[3] He M 2022 Assessing Changes in 21st Century Mean and Extreme Climate of the Sacramento–San Joaquin Delta in California. Climate, 10(2), 16.

[4] De Giorgi M G, Ficarella A and Tarantino M 2011 Assessment of the benefits of numerical weather predictions in wind power forecasting based on statistical methods. Energy, 36(7), 3968-3978.

[5] Yi J and Prybutok V R 1996 A neural network model forecasting for prediction of daily maximum ozone concentration in an industrialized urban area. Environmental Pollution, 92(3), 349-357.

[6] Jiang D, Zhang Y, Hu X, Zeng Y, Tan J and Shao D 2004 Progress in developing an ANN model for air pollution index forecast. Atmospheric Environment, 38(40), 7055-7064.

[7] Pal N R, Pal S, Das J and Majumdar K 2003 Sofm-mlp: a hybrid neural network for atmospheric temperature prediction. IEEE Transactions on Geoscience and Remote Sensing, 41(12), 278-2791.

[8] Radhika Y and Shashi M 2009 Atmospheric Temperature Prediction using Support Vector Machines. International Journal of Computer Theory and Engineering, 55-58.

[9] Park I, Kim H S, Lee J, Kim J H, Song C H and Kim H K 2019 Temperature Prediction Using the Missing Data Refinement Model Based on a Long Short-Term Memory Neural Network. Atmosphere, 10(11), 718.

[10] Chu W T, Ho K C and Borji A 2018 Visual Weather Temperature Prediction. 2018 IEEE Winter Conference on Applications of Computer Vision (WACV).