An Analysis of Monthly Temperature Data from Nottingham Castle: Trends, Patterns, and Forecasts

ABSTRACT

This report examines air temperature data collected from Nottingham Castle between 1920 and 1939. Various modelling techniques were employed to identify trends, seasonal patterns, and forecast future values. The exploratory data analysis (EDA) was utilised to visualise temperature trends and seasonal patterns, followed by decomposition to separate the dataset into trend, seasonal, and residual components. The most appropriate ARIMA model was selected manually, considering the autocorrelation and partial autocorrelation functions. Diagnostic checks were conducted to ensure the adequacy of the model by evaluating residual independence and stationarity. Forecasting techniques were then applied to predict past and future temperature values, with their accuracy evaluated using relevant metrics. This research aims to gain insights into past temperature trends and provide valuable information for anticipating future climate changes, thereby supporting informed decision-making processes.

Table of Contents

ABSTRACT 2
1. INTRODUCTION 4
2. RESEARCH QUESTION 4
3. METHODS AND ASSUMPTIONS 5
4. DATA ANALYSIS 5
4.1. TREND ANALYSIS 5
4.2. SEASONAL VARIATION: 6
4.3. RESIDUALS: 7
4.4. MODEL FITTING 8
4.5. FORECASTING 11
5. RESULTS 12
6. DISCUSSION 12
7. CONCLUSION 12
REFERENCES 13

1. Introduction

The examination of historical climate data is crucial for understanding past trends and anticipating future variations in temperature. This study focuses on monthly air temperature data collected at Nottingham Castle from 1920 to 1939 Love (2018), providing insights into temperature trends and seasonal patterns during this period.

The primary research question guiding this study is: How do monthly air temperature trends at Nottingham Castle vary over the period from 1920 to 1939, what are the significant seasonal patterns observed during this time, and how accurately can past and future temperature values be forecasted?

The objective of this research is to conduct a comprehensive analysis of the temperature dataset, employing various modelling techniques to identify underlying patterns and accurately forecast past and future temperature values. Through exploratory data analysis (EDA), decomposition, and ARIMA modelling, this study aims to establish a robust framework for understanding historical temperature variations, predicting future climate trends, and evaluating the accuracy of the forecasts.

The analysis begins with EDA to visualise temperature trends and seasonal patterns. Subsequently, decomposition techniques are used to segregate the dataset into trend, seasonal, and residual components, enhancing our understanding of the underlying patterns.

Manual selection techniques are then applied to identify and fit the most appropriate ARIMA model, considering autocorrelation and partial autocorrelation functions. Diagnostic checks are conducted to ensure the adequacy of the selected model, evaluating residual independence and stationarity.

Following model fitting, forecasting techniques are employed to predict future temperature values, with forecast accuracy assessed using relevant metrics. This research aims to contribute to the understanding of past temperature trends, provide valuable insights into anticipating future climate variations, and support informed decision-making processes regarding climate-related matters.

In summary, this study seeks to provide a comprehensive analysis of Nottingham's temperature data, offering valuable insights into historical climate trends, providing a framework for forecasting future temperature variations, and evaluating the accuracy of the forecasts.

.

2. Research Question

- **Main Research Question:**
    - How do monthly air temperature trends at Nottingham Castle vary over the period from 1920 to 1939, what are the significant seasonal patterns observed during this time, and how accurately can past and future temperature values be forecasted?
- **Subsidiary Research Questions:**
    - What is the overall trend in monthly average temperatures over the observed period?
    - Are there any significant seasonal variations in temperature patterns?
    - How do these temperature patterns compare to historical averages or climatological norms?

3. Methods and Assumptions

The analysis encompasses the following methods and assumptions:

- Exploratory Data Analysis (EDA): Visualizing temperature trends and seasonal patterns through graphical representations.
- Decomposition: Utilizing decomposition techniques to segregate the time series into trend, seasonal, and residual components.
- Manual Model Selection: Employing manual techniques to identify and fit the most suitable ARIMA model, considering autocorrelation and partial autocorrelation functions.
- Model Adequacy Checks: Assessing the adequacy of the selected model through diagnostic checks for residual independence and stationarity.
- Forecasting Techniques: Applying forecasting methods to predict past and future temperature values, with forecast accuracy evaluated using relevant metrics.

Assumptions include the stationarity of the time series and the absence of significant outliers or structural breaks in the data. Additionally, it is assumed that the chosen ARIMA model adequately captures the underlying patterns in the temperature data.

4. Data Analysis

The plot below (*Figure 4*) illustrates the variation of temperatures in Fahrenheit (°F) for each month over multiple years. Each line represents a different year, allowing for the comparison of temperature patterns across months and years. The x-axis shows the months, while the y-axis indicates the temperature. The transparency of the lines helps to visualize overlapping data points.

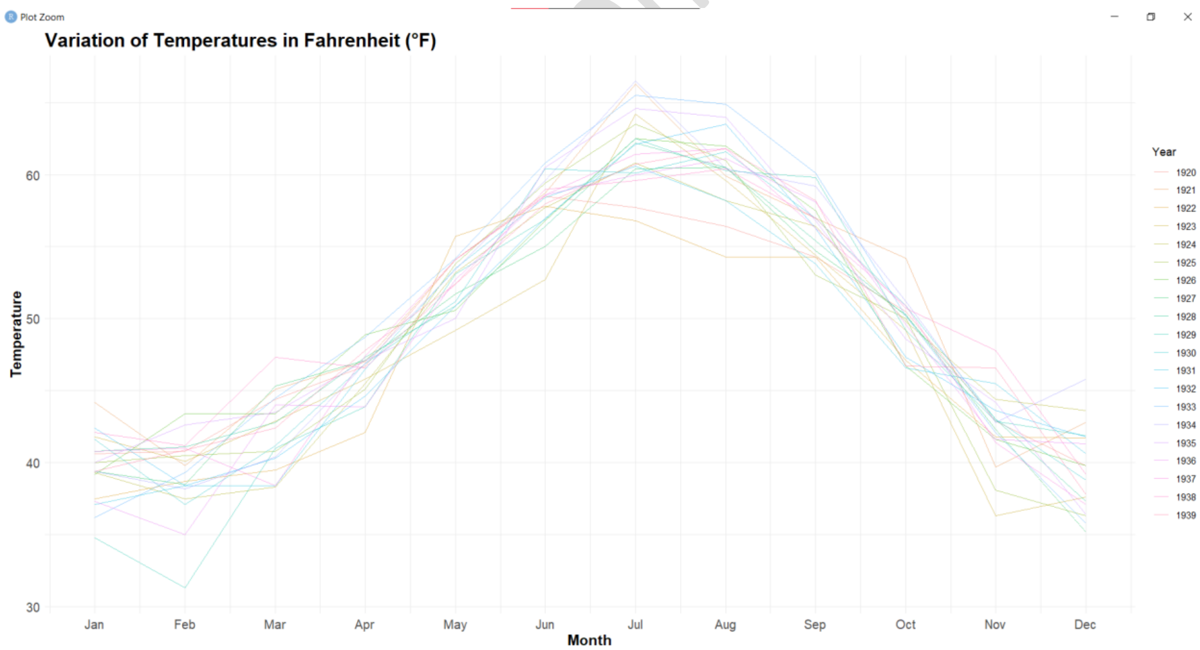


Figure 4: Plot showing the variation of temperature in °F

4.1. Trend Analysis

This plot displays the trend prediction of monthly average temperatures at Nottingham Castle over time. The green line shows the trend based on a linear regression model (lowest AIC value), indicating that the temperatures are rising over time. The blue horizontal line represents the mean temperature, providing a reference point for comparing individual temperatures to the overall average. See *Figure 4.1*

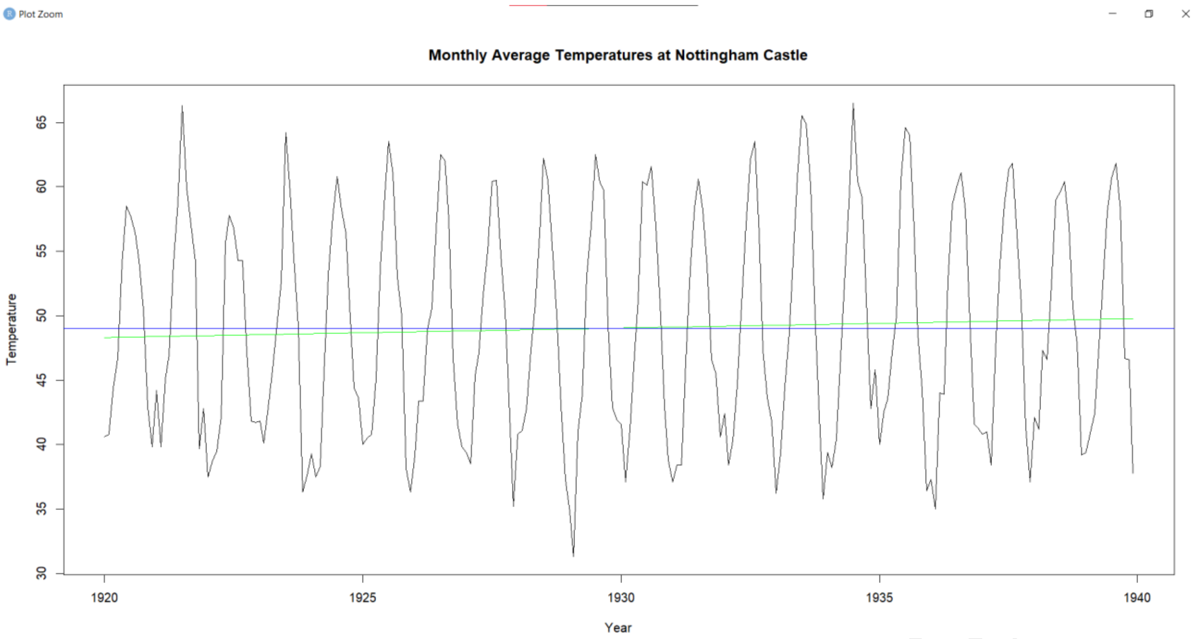


Figure 4.1: Plot showing the trend prediction of temperature in °F

4.2. Seasonal Variation:

The plot in *Figure 4.2* illustrates the seasonal variations in monthly average temperatures recorded at Nottingham Castle over the years, with each point representing a specific period, likely months, or seasons.

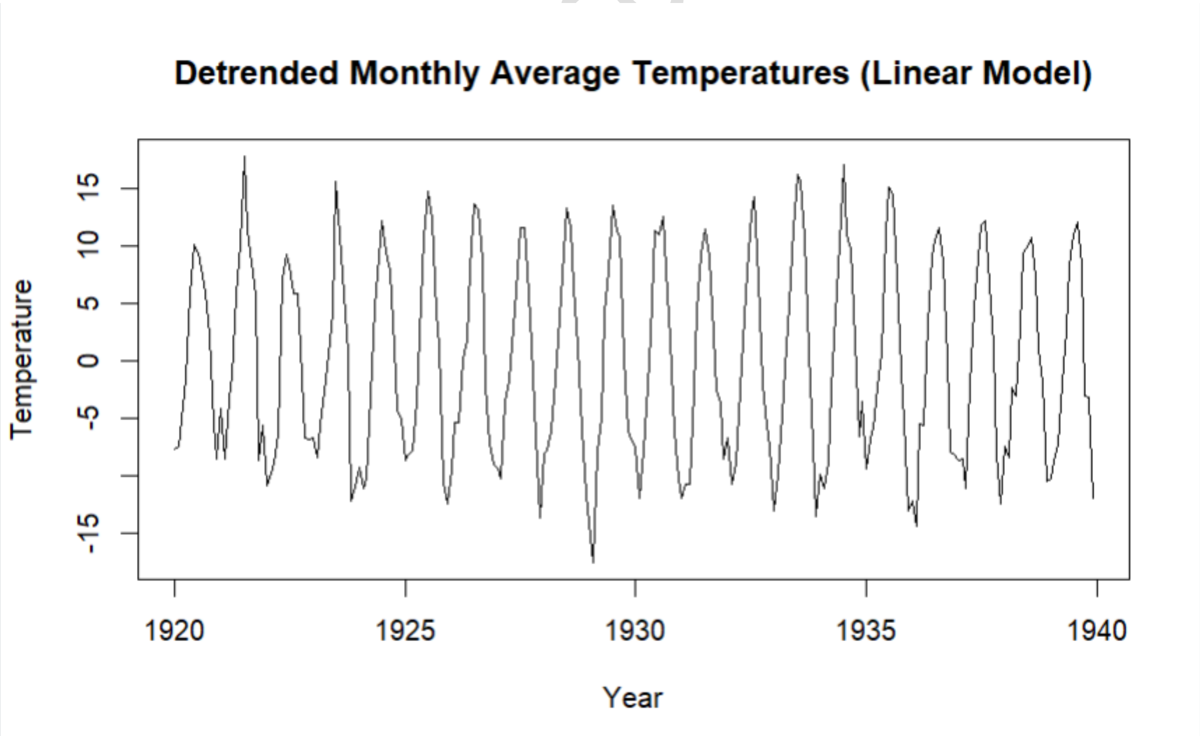
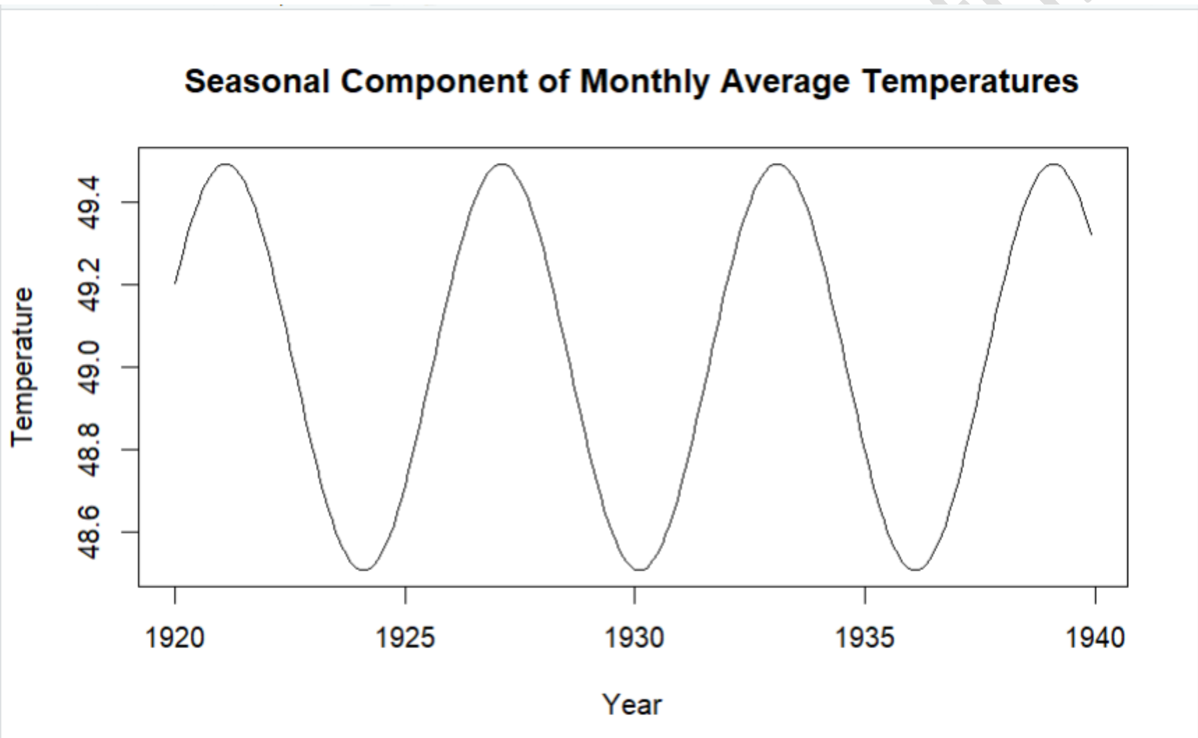


Figure 4.2: Plot of detrended monthly average temperature in °F

- **Long-Term Trends**:
  - Peaks and troughs on the plot denote periods of elevated and reduced temperatures, respectively, within each interval (approximately 5 years).
  - The cyclical nature of the fluctuations suggests recurring seasonal changes influencing temperature patterns.
- **Five-Year Intervals**:
  - Observing the plot over five-year increments allows for the identification of overarching seasonal trends.
  - Peaks and troughs occurring at regular intervals may align with seasonal changes typical of Nottingham's climate, timeanddate.com (2024).
- **Magnitude of Variation:**
  - The amplitude of the fluctuations reflects the extent of temperature variation within each five years.
  - Larger peaks and deeper troughs indicate more pronounced seasonal changes over the specified timeframe.



4.3. RESIDUALS:

After detrending the data and removing the seasonal patterns, the remaining variability is known as the residual. This residual component is critical to the analysis, as it represents the portion of the data that has not been explained by the chosen model. By examining the residual, we can gain insights into the complexities of the data that the model may have overlooked. In essence, the residual serves as an important tool for evaluating the effectiveness of the chosen analysis and the accuracy of the model' predictions.

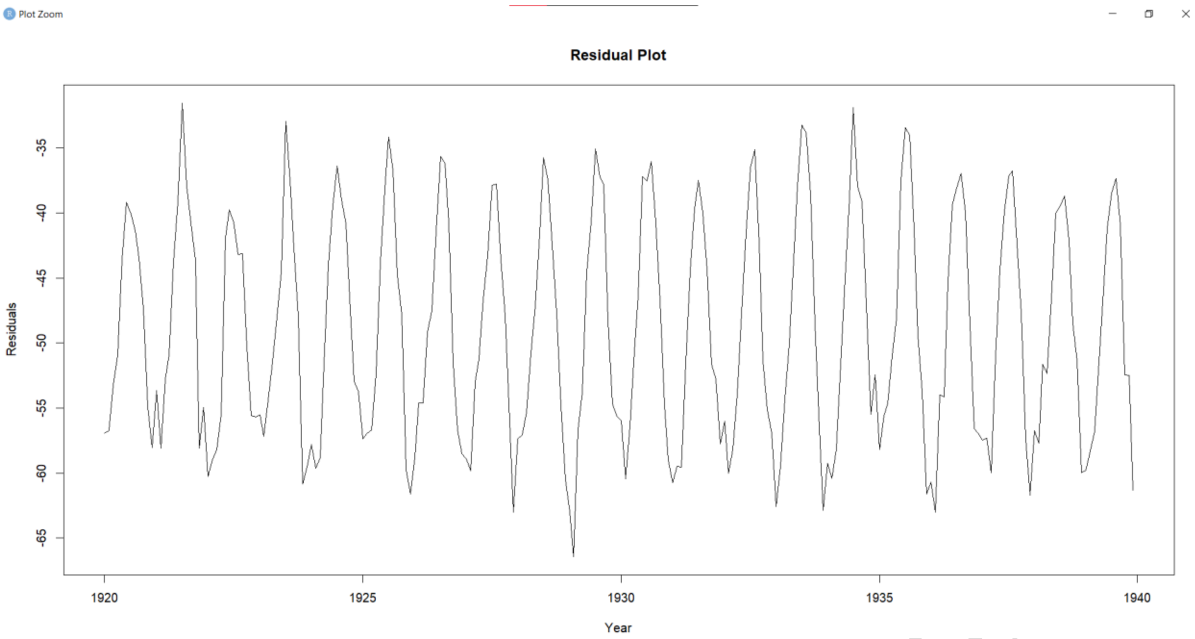


Figure 4.3: Plot of residual after removing trend and seasonality in ℉

4.4. Model Fitting

In the model fitting, an exploration into various ARIMA models was conducted to identify the most suitable one for the temperature data. The autocorrelation function (ACF) and partial autocorrelation function (PACF) of the residuals from the seasonality component were examined.

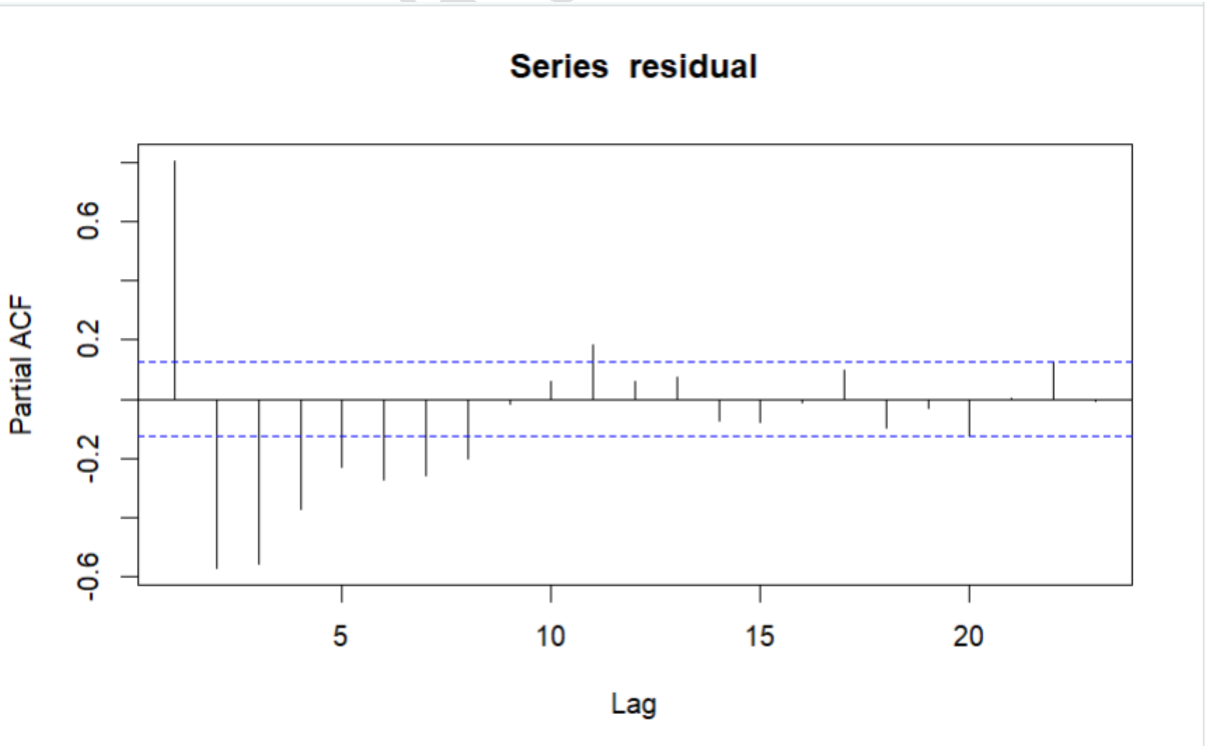


Figure 4.4 a: PACF of residual after removing trend and seasonality.

This is to identify the appropriate values for the p, d, and q parameters in the ARIMA (p, d, q) model. After thorough testing of different combinations of autoregressive (AR), differencing (I), and moving average (MA) components, the ARIMA model with the lowest AIC value was identified as the most appropriate for the dataset.

The selected ARIMA model yielded the following coefficients:

Autoregressive (AR)	
AR1	2.0125
AR2	-1.4847
AR3	0.2797

Table 4.4 a

Moving Average (MA)	
AR1	-2.7291
AR2	2.7159
AR3	-0.9835

Table 4.4 b

These coefficients, along with their respective standard errors, provide insights into the relationships between past observations and the current forecasted value. The estimated variance ($sigma^2$) of the residuals was **6.031**, derived from the model's log-likelihood of **-561.71.** The AIC value for this model was calculated to be **1135.41**, indicating its performance in balancing goodness of fit with model complexity.

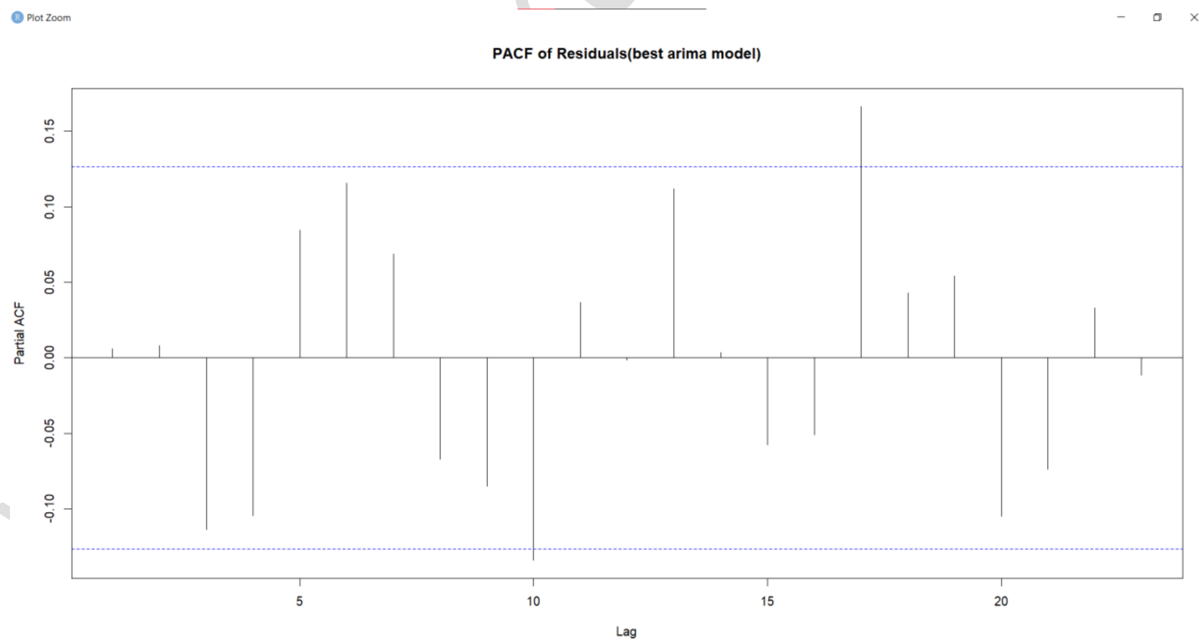


Figure 4.4 b: PACF plot of residuals from the model

- **PACF:** The PACF measures the correlation between observations at different lags while removing the effects of correlations at shorter lags as seen in *Figure 4.4 b*. A value of around **0.20** suggests a weak correlation at that lag.
- **ACF:** The ACF measures the correlation between observations at different lags without removing the effects of correlations at shorter lags. Similarly, a value of around **0.20** indicates a weak correlation.

Since the values are small and centred around **0**, it implies that the residuals are relatively independent and do not exhibit strong autocorrelation patterns. This is a desirable characteristic for residuals in a time series model, indicating that the model has adequately captured the temporal dependencies in the data.

Shapiro-Wilk normality test	
residuals(best_arima_model)	
W	0.99432
p-value	0.5059

Table 4.4 c

In addition, the Shapiro-Wilk normality test was conducted on the residuals of the ARIMA model. The test yielded a W statistic of **0.99432** and a corresponding p-value of **0.5059**. With a p-value above the conventional significance level of **0.05**, there is insufficient evidence to reject the null hypothesis of normality. Therefore, it is reasonable to assume that the residuals follow a normal distribution.

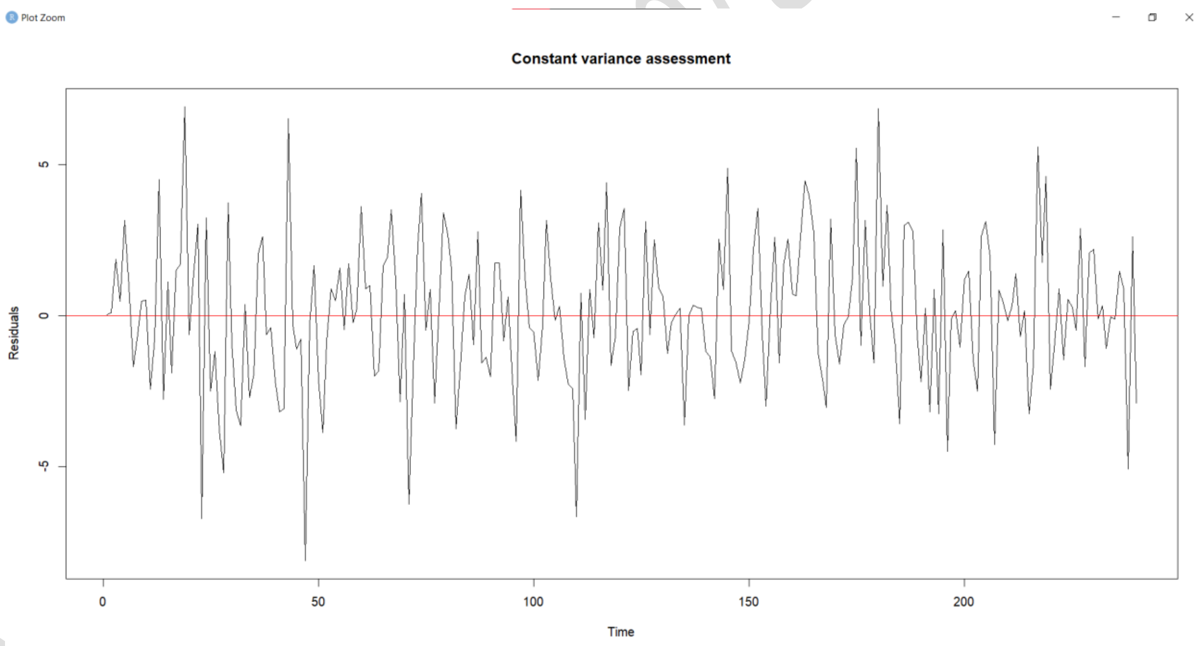


Figure 4.4 c: Plot of residuals from the best model against time

In conclusion, a plot of residuals against time was reviewed to assess constant variance. A consistent spread around the horizontal axis was observed, as depicted in *Figure 4.4(c),* indicating that residual variance remains stable across time. This supports the assumption of constant variance in the model's residuals, ensuring the reliability of predictions.

4.5. FORECASTING

In this report, the series was forecasted for both past and future using the ARIMA model previously selected.

Firstly, the series was forecasted for future values beyond the existing dataset. A forecast was made for 1940, 1941 and 1942.

Additionally, the series was forecasted for past values, covering the year 1937 to 1939, allowing for an assessment of the model's performance in reproducing known data points. See *Figure 4.5a*

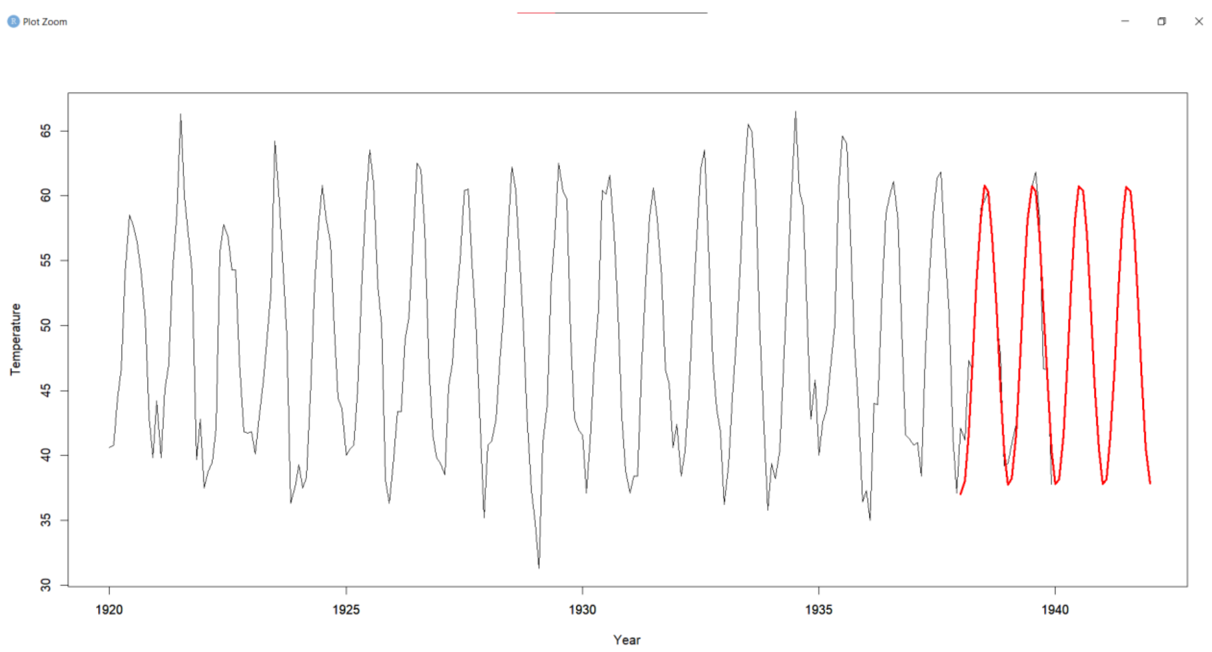


Figure 4.5 a: Prediction plot from 1937 to 1942

The forecast accuracy was evaluated using two key metrics:

1. **Mean Absolute Percentage Error (MAPE):** This metric measures the average absolute percentage difference between the forecasted and actual values. A lower MAPE indicates better accuracy in the forecasts.
2. **Root Mean Squared Error (RMSE):** This metric measures the square root of the average squared differences between the forecasted and actual values. Like MAPE, a lower RMSE value signifies better accuracy in the forecasts.

For the forecasted series:

- Mean Absolute Percentage Error (MAPE): 7.73%
- Root Mean Squared Error (RMSE): 3.71

These accuracy metrics provide insights into the performance of the ARIMA model in forecasting temperature variations. The low MAPE and RMSE values suggest that the model provides relatively accurate forecasts, with errors averaging around 7.73% and 3.71 units, respectively. This indicates that the ARIMA model is effective in capturing the underlying patterns in the temperature data and producing reliable forecasts for both existing and future values as discussed by Hyndman and Athanasopoulos (2018)

5. RESULTS

The analysis of monthly air temperature data collected at Nottingham Castle between 1920 and 1939 revealed several key findings:

Temperature Trends: Through exploratory data analysis (EDA), it was observed that the temperatures exhibited a rising trend over the specified period. The trend analysis, based on a linear regression model with the lowest AIC value, indicated a consistent increase in temperatures over time.

Seasonal Patterns: Significant seasonal variations in temperature patterns were identified. The seasonal variation analysis illustrated recurring fluctuations in temperature, with peaks and troughs corresponding to seasonal changes. These variations were observed both in short-term intervals and over longer five-year increments, indicating the cyclical nature of temperature patterns.

ARIMA Model Selection: Manual selection techniques were applied to identify and fit the most appropriate ARIMA model to the temperature data. By examining the autocorrelation and partial autocorrelation functions of the residuals, the ARIMA model with the lowest AIC value was determined as the most suitable for the dataset. The selected ARIMA model provided coefficients for autoregressive (AR) and moving average (MA) components, along with their respective standard errors.

6. DISCUSSION

The ARIMA model chosen provides a satisfactory performance in capturing temperature patterns, which is a crucial component of accurate temperature forecasting. However, evaluating its predictive ability through accuracy metrics is equally important. The Mean Absolute Percentage Error (MAPE) of 7.73% indicates that, on average, forecasts deviated by approximately 7.73% from actual temperature values, while the Root Mean Squared Error (RMSE) of 3.71 suggests an average error magnitude of 3.71 degrees Fahrenheit.

Interpreting these metrics within the context of the dataset and forecasting objectives is essential. While lower MAPE and RMSE values indicate better accuracy, it is vital to ensure they align with the desired precision levels. The MAPE value of 7.73% and the RMSE value of 3.71 degrees Fahrenheit suggest relatively accurate forecasts since no precision level was given. However, it is crucial to conduct further analysis to understand the sources of forecast errors and potential enhancements.

Furthermore, although the present model performs well within the observed period, its reliability for future forecasts may vary due to external factors or climate dynamics shifts. Consequently, continuous evaluation and refinement are necessary for reliable forecasting and informed decision-making.

In summary, the current ARIMA model offers valuable insights into temperature trends and variations. However, ongoing assessment and improvement are necessary for robust forecasting, particularly when considering the potential impact of external factors on future forecasts.

7. CONCLUSION

In conclusion, this research contributes to the comprehension of past temperature trends and seasonal patterns at Nottingham Castle between 1920 and 1939. The model selection was based on manual selection. Conducting a comprehensive analysis, valuable insights have been gained into historical climate variations. The selected ARIMA model provides a framework for forecasting future temperature values with reasonable accuracy, supporting informed decision-making processes regarding climate-related matters.

Overall, this study underscores the importance of manual model selection methods in analysing historical climate data and offers a valuable approach to understanding and anticipating climate variations in the future.

REFERENCES

Hyndman, R.J. and Athanasopoulos, G. (2018) Forecasting : Principles and Practice. 2nd edn. Heathmont, Vic.: Otexts. Available at: https://otexts.com/fpp2/ (Accessed: 9 April 2024).

Love, J. (2018) Nottingham Castle Dataset. Available at: https://github.com/jamovi/r-datasets/blob/master/data/nottem.csv (Accessed: 25 March 2024).

OpenAI (2024) Explanations of Terms, Grammer Correction and Paragraph Reduction, Provided by ChatGPT. Available at: https://chat.openai.com/ (Accessed: 6 April 2024).

timeanddate.com (2024) Climate & Weather Averages in Nottingham, England, United Kingdom. Available at: https://www.timeanddate.com/weather/uk/nottingham/climate (Accessed: 14 April 2024).