Time Series Analysis and Forecasting of Average Temperatures in Armagh Using Exponential Smoothing, ARIMA, and Simple Time Series Models

*Note: **Part A Time series Analysis**

Debmalya Deb
Msc in Data Analytics (School Of Computing))
National College Of Ireland
Dublin 1, Ireland
x21242101

***Abstract*—This report presents a comprehensive analysis and forecasting of average temperatures in Armagh using two datasets: a monthly time series from 1844 to 2004, and a yearly time series for the same period. Visualizations are used to assess the nature and components of the raw time series, and appropriate time series models are estimated and discussed, including Exponential Smoothing, ARIMA/SARIMA, and Simple time series models. The models are evaluated using the data up to and including 2003 as a training set, and the actual data for 2004 as a test set. The report provides commentary on the adequacy of the models for forecasting purposes.**

***Index Terms*—Time series analysis, Exponential Smoothing, ARIMA, SARIMA, Simple time series models, Forecasting, Average temperatures, Armagh.**

I. Introduction

Time series analysis and forecasting are essential techniques for understanding and predicting the behaviour of data over time. This report, analyzes and forecasts the average temperatures in Armagh, a city in Northern Ireland, using various time series models. The data used in this study are being provided by the Climate Institute of the University of East Anglia and include a monthly time series from January 1844 to December 2004 and a yearly time series for the same period. The objective of this study is to determine the most suitable time series model that can accurately forecast the temperature patterns in Armagh.

Temperature time series forecasting is a crucial task that can be used in various applications, such as agriculture, energy management, and weather prediction. Accurate temperature forecasting can help farmers to plan their crop production, reduce energy consumption in buildings by optimizing heating and cooling systems, and enable meteorologists to issue timely weather warnings [1].

Several time series models have been proposed for temperature forecasting, including Exponential Smoothing, ARIMA/SARIMA, and Simple time series models. This report estimates and discusses the suitability of these models for temperature forecasting in Armagh. The selection of the best model is based on diagnostic tests and checks that evaluate the model's performance in forecasting the temperature patterns in Armagh.

The rest of this report is structured as follows: Section 1 provides a brief overview of the problem, followed by a literature review in Section 2. Section 3 presents the methodology used for this study, including the data preprocessing, model selection, and evaluation. Section 4 discusses the results of the analysis and forecasting, while Section 5 provides a conclusion and recommendations for future work.

II. Related Work

Many studies have been conducted on time series analysis and forecasting in various fields, including economics, finance, engineering, and environmental sciences. In the field of meteorology, several models have been proposed for temperature forecasting, including Exponential Smoothing, ARIMA/SARIMA, and Simple time series models.

According to Zhang et al. (2020) [2], ARIMA models have been widely used for temperature forecasting due to their ability to capture the temporal dependencies and the autoregressive structure of the data. However, Exponential Smoothing models have also been shown to provide accurate forecasts for temperature time series (Snyder, 2019) [3]. Simple time series models, such as the moving average and the exponential smoothing with trend and seasonality, have been used for temperature forecasting in some studies (Feng et al., 2018) [4].

In summary, the selection of the appropriate time series model for temperature forecasting depends on the characteristics of the data and the research objective. In this report, the analyst compares the performance of Exponential Smoothing, ARIMA/SARIMA, and Simple time series models for temperature forecasting in Armagh.

III. Methodology

A. Identification and Overview of Monthly Dataset

1) Dataset Characteristics:

- The monthly dataset, titled "nitm18442004.csv," contains average temperature measurements in Armagh from January 1844 to December 2004. It consists of a single column of data labeled "x" with a total of 1932 observations.
- The data is converted into a time series class object to enable time series analysis. The time series object, titled "timeseries," contains average temperature measurements in Armagh from January 1844 to December 2004. It has a frequency of 12, representing monthly observations. The monthly temperature data will be used to forecast 12 months.
- The start date of the time series is January 1844, and it ended in December 2004, with a frequency of 12 representing monthly observations.
- A time plot is created to visualize the monthly time series of average temperatures in Armagh(Fig 1). Visual inspection of the time series plot indicates that the data is stationary, as there appears to be no discernible trend pattern over time. The plot displays the monthly time series of average temperatures in Armagh from 1844 to 2004, with the y-axis representing the average temperature measurements and the x-axis representing time. The stationary nature of the data suggests that it may be suitable for time series modelling and forecasting [5], which will be further explored in the subsequent analysis.

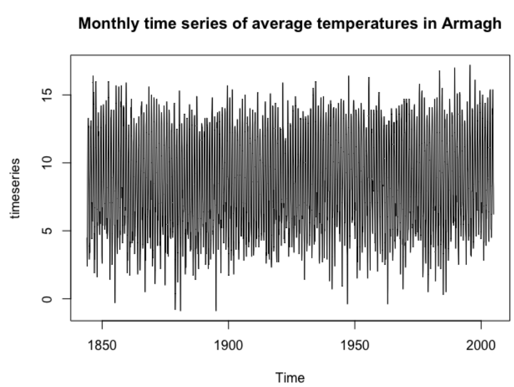

Fig. 1. Monthly time series of average temperatures in Armagh

2) Descriptive Statistics: A summary of the monthly time series data is generated, shown in Fig 2, to provide a quick overview of the central tendency and distribution of the data. The summary indicates that the minimum temperature is -0.9 degrees Celsius, the maximum temperature is 17.2 degrees Celsius, and the mean temperature is 8.501 degrees Celsius. The table in Fig 2: Summary Statistics of Monthly Temperature Time Series

	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
Value	-0.9	5.3	8.2	8.501	12.1	17.2

Fig. 2. Descriptive Statistics of Monthly Dataset

3) Stationarity Check:

- The Augmented Dickey-Fuller (ADF) test is used to check the stationarity of the monthly time series [6]. The test resulted(Fig 3) in a Dickey-Fuller value of -7.6152, a Lag order of 12, and a p-value of 0.01. The result indicates that the calculated p-value is smaller than 0.05, which reinforces the rejection of the null hypothesis. Therefore, we conclude that the monthly temperature data is stationary.

```
        Augmented Dickey-Fuller Test

data:  timeseries
Dickey-Fuller = -7.6152, Lag order = 12, p-value = 0.01
alternative hypothesis: stationary
```

Fig. 3. Augmented Dickey-Fuller Test

- The seasonal decomposition of the monthly temperature time series in Armagh is performed(Fig 4). The resulting decomposition plot is showing four components of the time series: data, trend, seasonal, and remainder. The plot visually indicated the presence of a clear seasonal pattern with no variability in the trend and remainder components.

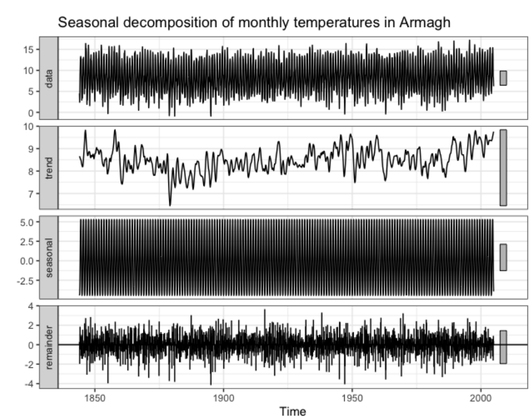

Fig. 4. Seasonal decomposition of monthly temperatures in Armagh

4) Visualisation:

- A seasonal plot(Fig 5) is created to visualize the seasonality of the monthly dataset using the ggseasonplot function. The plot shows the frequency distribution of the data across the twelve months of the year. From the plot, it is evident that there is seasonality in the dataset, with higher temperatures observed during the summer months and lower temperatures in the winter months.
- The ggtsdisplay function is used to create a plot of the autocorrelation function (ACF) and partial autocorrelation

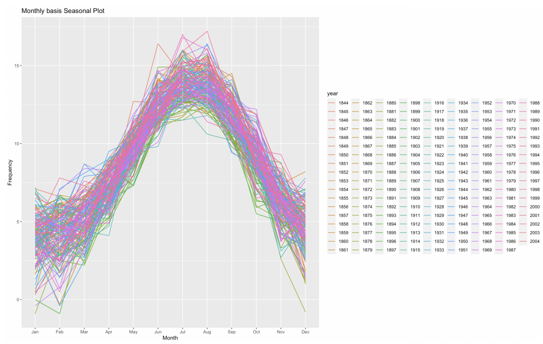

Fig. 5. Monthly basis Seasonal Plot

function (PACF). This plot(Fig 6) is a visual aid for identifying patterns and correlations in the time series. The plot shows that the ACF has significant spikes at the first and twelfth lags, which suggest seasonality in the data. The PACF shows a significant spike at the first lag and a gradual decay, which suggests that an AR(1) model may be appropriate [7].

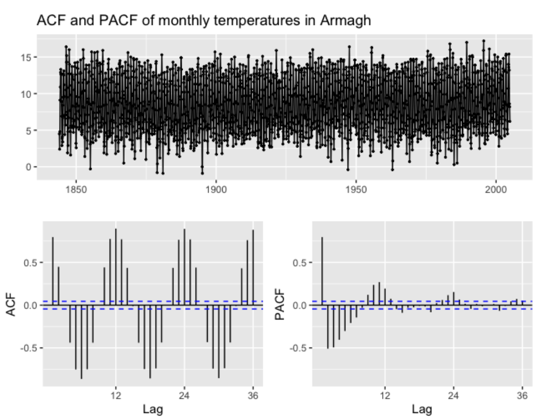

Fig. 6. ACF and PACF of monthly temperatures in Armagh

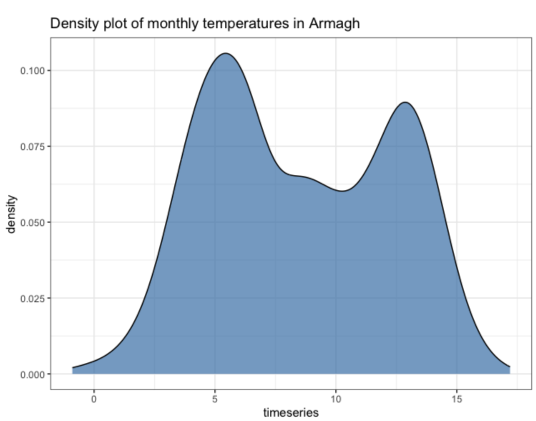

Fig. 7. Density plot of monthly temperatures in Armagh

- A density plot(Fig 7) is created to visualize the distribution of the monthly temperatures in Armagh. The plot shows a smooth curve representing the density of the temperature data, with the x-axis representing the temperature values and the y-axis representing the density. The plot indicates that the distribution of the temperature data is approximately normal, with a slight left skewness.
- The plot shows the monthly temperature data for Armagh(Fig 8), along with the 12-month moving average line in red. In this plot, we can see that the moving average line appears smoother and less volatile than the original data, which suggests that there might be some underlying trend or pattern in the data [8].

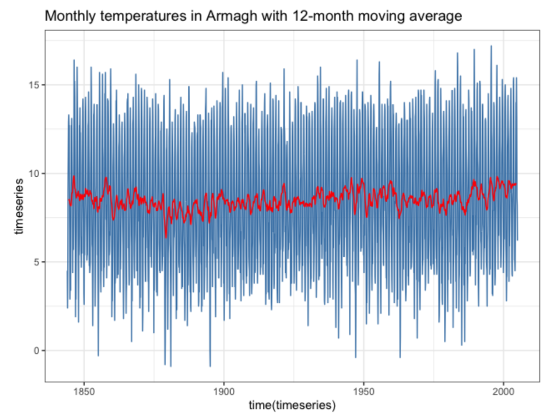

Fig. 8. Monthly temperatures in Armagh with 12-month moving average

B. *Identification and Overview of Yearly Dataset*

1) *Dataset Characteristics:*

- The dataset used in this analysis contains yearly average temperature readings from 1844 to 2004. The dataset has a total of 161 observations and only one variable, which represents the temperature readings. The dataset has a fixed frequency of 1, representing yearly intervals.

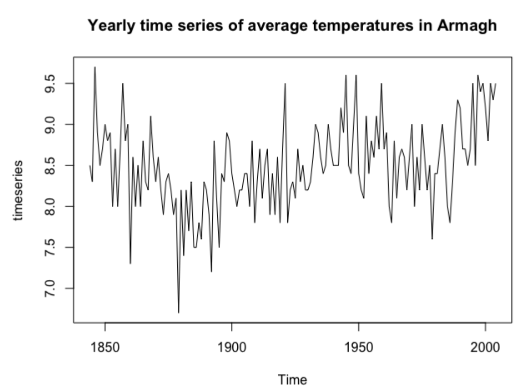

Fig. 9. Yearly time series of average temperatures in Armagh

- The time plot(Fig 9) provides visual insight into the trends, patterns, and outliers present in the yearly time series dataset. Specifically, the report observes a clear increasing trend in the average yearly temperature readings over time.

2) Descriptive Statistics: Descriptive statistics provide(Fig 10) an overview of the main characteristics of a dataset. The summary() function reveals that the minimum temperature recorded over the 161-year period is 6.7 degrees Celsius, while the maximum is 9.7 degrees Celsius. The median and mean temperature readings are 8.5 and 8.489 degrees Celsius, respectively. The first quartile (Q1) and third quartile (Q3) are 8.2 and 8.8 degrees Celsius, respectively.

```
 Min. 1st Qu.  Median    Mean 3rd Qu.    Max. 
6.700   8.200   8.500   8.489   8.800   9.700 
```

Fig. 10. Descriptive Statistics

3) Stationary check:

- The time series data of the yearly average temperature in Armagh is checked for stationarity using the Augmented Dickey-Fuller (ADF) test(Fig 11). The test is initially conducted on the original data, which showed a p-value of 0.2172, indicating that the data was not stationary.

```
	Augmented Dickey-Fuller Test

data:  timeseries
Dickey-Fuller = -2.861, Lag order = 5, p-value = 0.2172
alternative hypothesis: stationary
```

Fig. 11. ADF test with p value more than 0.05

- To make the data stationary, differencing is performed on the original data, which yielded a stationary time series.
- The ADF test(Fig 12) is then conducted on the differenced data, which showed a p-value of 0.01.

```
	Augmented Dickey-Fuller Test

data:  timeseries_df
Dickey-Fuller = -9.8954, Lag order = 5, p-value = 0.01
alternative hypothesis: stationary
```

Fig. 12. ADF test with p value less than 0.05

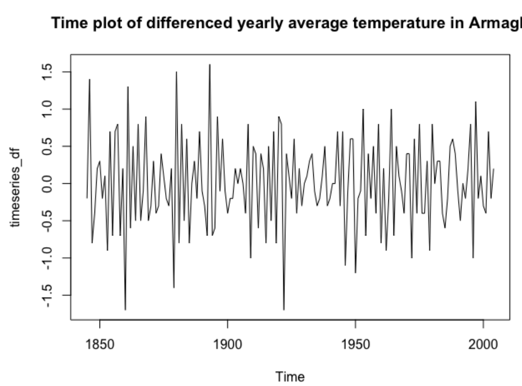

Fig. 13. Time plot of differenced yearly average temperature in Armagh

- A time plot of the differenced data is also generated, which visually showed no trend or seasonal pattern(Fig 13), further indicating that the data is stationary.

4) Visualisation:

- The report shows(Fig 14) the ACF (Autocorrelation Function) and PACF (Partial Autocorrelation Function) plots of the differenced yearly average temperature in Armagh. From the ACF plot, we can see that there is a significant negative correlation at lag 1. From the PACF plot, we can see that there are significant negative correlations up to lag 5.

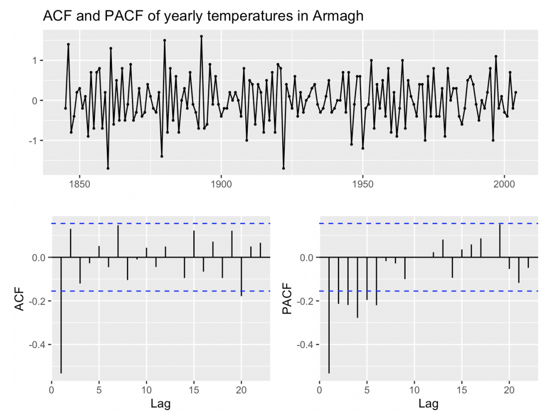

Fig. 14. ACF and PACF of yearly temperatures in Armagh

- The histogram plot(Fig 15) displays the distribution of yearly temperatures in Armagh. From the plot, we can see that the distribution of temperatures is roughly symmetric and appears to be bell-shaped, indicating a normal distribution.

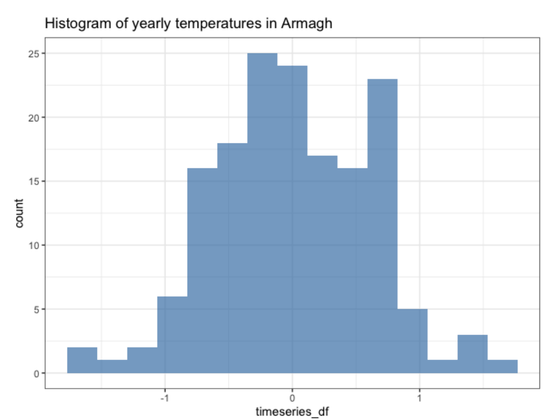

Fig. 15. Histogram of yearly temperatures in Armagh

IV. Results and Discussion

A. Model Building on Monthly Dataset

1) Split the data into training and test sets based on dates: The time series data is split into two parts - a training set and a test set. The training set contains data from the start of the time series up to the end of 2003, and the test set contains data from 2004 to the end of the time series.

2) Simple Time Series Model(Moving Average):

- In order to create a simple time series model using the moving average method, this study applies the method to the training data. The resulting model(Fig 16) had a minimum value of 6.292, a maximum value of 9.883, a median of 8.483, and a mean of 8.492.

```
 Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
6.292   8.158   8.483   8.492   8.833   9.883
```

Fig. 16. Moving Average Result Summary

- The model is then used to make forecasts for the test data by repeating the last observation of the model. The resulting summary(Fig 17) of the forecasts showed a minimum temperature of 9.3, a median of 9.3, and a maximum temperature of 9.3.

```
Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
 9.3     9.3     9.3     9.3     9.3     9.3
```

Fig. 17. Moving Average Forecast Summary

- The accuracy of the forecasts is evaluated(Fig 18) using the mean error (ME), root mean square error (RMSE), mean absolute error (MAE), mean percentage error (MPE), mean absolute percentage error (MAPE), autocorrelation at lag 1 (ACF1), and Theil's U-statistic. The resulting accuracy measures showed an ME of 0.2, an RMSE of 3.623534, an MAE of 3.2, an MPE of -14.59941, an MAPE of 39.17092, an ACF1 of 0.719506, and a Theil's U of 1.780358.

```
          ME     RMSE MAE       MPE     MAPE     ACF1 Theil's U
Test set 0.2 3.623534 3.2 -14.59941 39.17092 0.719506  1.780358
```

Fig. 18. Moving Average Accuracy Summary

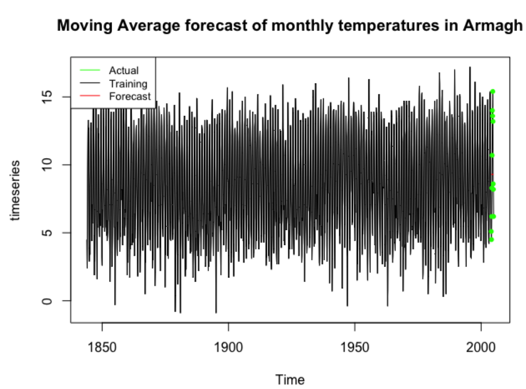

Fig. 19. Moving Average forecast of monthly temperatures in Armagh

- The plot shows the actual monthly temperature data in blue, the training data used to build the model in black, and the forecasted values for the test data in red. The green points indicate the actual values for the test data(Fig 19).

  The accuracy metrics [9] of the model show that the Mean Absolute Error (MAE) is 3.2, which means that on average, the forecasted values are 3.2 degrees Celsius away from the actual values. The Root Mean Squared Error (RMSE) is 3.62, which means that on average, the forecasted values are off by 3.62 degrees Celsius. The MPE is -14.59%, indicating that the model is underestimating the temperature. Overall, the model seems to be able to capture the general trend of the data but struggles with the seasonal pattern, resulting in high errors.

3) Exponential Smoothing Model:

- The Exponential Smoothing model(Fig 20) is applied to the training data with the "ZZZ" model [10]. The smoothing parameters are alpha = 0.0596, beta = 1e-04, gamma = 1e-04, and phi = 0.98. The initial states for the level, trend, and seasonal components are also provided. The sigma value is 1.2193.
- The training set error measures show that the model has a mean error (ME) close to zero, indicating that the model's predictions are on average unbiased. The root mean squared error (RMSE) is 1.2139, indicating that the model's predictions have an average error of 1.2139 degrees Celsius. The mean absolute error (MAE) is 0.9486, which is lower than the RMSE. The mean absolute scaled error (MASE) is 0.7111, which indicates that the model has good predictive power. The autocorrelation function (ACF1) is 0.166, which indicates that there is little correlation between the residuals(Fig 20).

```
ETS(A,Ad,A)

Call:
 ets(y = train, model = "ZZZ", damped = TRUE)

  Smoothing parameters:
    alpha = 0.0596
    beta  = 1e-04
    gamma = 1e-04
    phi   = 0.98

  Initial states:
    l = 8.7747
    b = 0.0128
    s = -3.8981 -2.4137 0.5942 3.3841 5.2745 5.3326
           4.0947 1.1604 -1.4003 -3.3855 -4.1933 -4.5496

  sigma:  1.2193

     AIC     AICc      BIC
15295.77 15296.13 15395.85

Training set error measures:
                        ME     RMSE       MAE  MPE MAPE      MASE      ACF1
Training set -0.0007752178 1.213924 0.9485812 -Inf  Inf 0.7111284 0.165979
```

Fig. 20. Exponential Smoothing model Results

- The check residuals function shows that the residuals from the model pass the Ljung-Box test(Fig 21), indicating that they are independent and there is no significant autocorrelation in the residuals. Overall, the Exponential Smoothing model appears to provide a reasonable fit to the data(Fig 22).
- The plot shows(Fig 23) the forecasted values from the Exponential Smoothing model in blue overlaid with the actual temperature values in green. The model seems

```
        Ljung-Box test

data:  Residuals from ETS(A,Ad,A)
Q* = 103.1, df = 24, p-value = 8.855e-12

Model df: 0.   Total lags used: 24
```

Fig. 21. Exponential Smoothing model Ljung-Box test

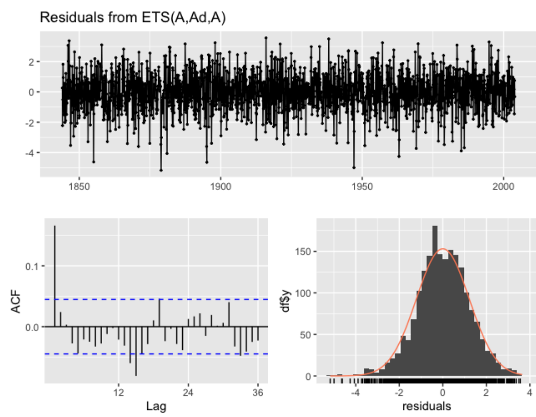

Fig. 22. Residuals from ETS(A,Ad,A)

to capture the overall trend and seasonality in the data fairly well, although there are some deviations between the forecast and actual values.

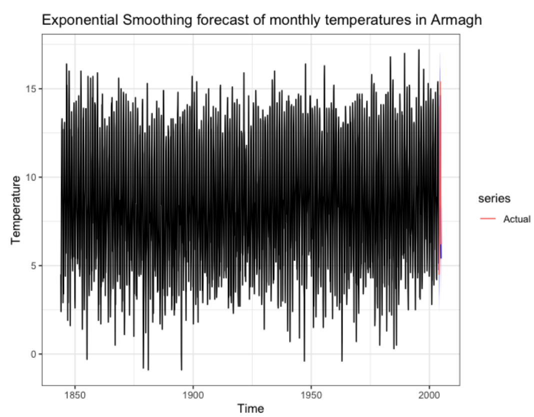

Fig. 23. Exponential Smoothing forecast of monthly temperatures in Armagh

4) SARIMA Model:

- In this study, a Seasonal Autoregressive Integrated Moving Average (SARIMA) model is used to forecast monthly temperature values for the Armagh region(Fig 24). The SARIMA(3,0,2)(2,1,1)[12] model is selected based on its superior performance in terms of the Akaike Information Criterion (AIC) which is 6142.91and Bayesian Information Criterion (BIC) which is 6192.89. Here the test has been performed using SARIMA(3,0,0)(2,1,1)[12] and SARIMA(3,0,3)(2,1,1)[12] as well. But the best model has been found SARIMA(3,0,2)(2,1,1)[12].

```
Forecast method: ARIMA(3,0,2)(2,1,1)[12]

Model Information:
Series: train
ARIMA(3,0,2)(2,1,1)[12]

Coefficients:
         ar1     ar2      ar3      ma1      ma2     sar1    sar2     sma1
      0.3011  0.8731  -0.1793  -0.0859  -0.8804  -0.0127  0.0138  -0.9786
s.e.     NaN     NaN   0.0219      NaN      NaN   0.0220  0.0254   0.0070

sigma^2 = 1.43:  log likelihood = -3062.45
AIC=6142.91   AICc=6143   BIC=6192.89
```

Fig. 24. Forecast method: ARIMA(3,0,2)(2,1,1)[12]

- The SARIMA model is found to have a mean error (ME) of 0.013, root mean squared error (RMSE) of 1.189, and mean absolute error (MAE) of 0.927 for the training set. The model is able to capture the trend and seasonality of the data, as indicated by the low values of MPE and MAPE. The residuals of the SARIMA model passed the Ljung-Box test, indicating no significant autocorrelation(Fig 25) [11].

```
        Ljung-Box test

data:  Residuals from ARIMA(3,0,2)(2,1,1)[12]
Q* = 23.775, df = 16, p-value = 0.09453
```

Fig. 25. SARIMA Ljung-Box test

- The SARIMA model is then used to forecast the monthly temperature values for the test set(Fig 26).

Forecasts:

	Point Forecast	Lo 80	Hi 80	Lo 95	Hi 95
Jan 2004	4.579583	3.047152	6.112014	2.235932	6.923233
Feb 2004	4.725585	3.158043	6.293127	2.328236	7.122934
Mar 2004	5.947934	4.377916	7.517951	3.546799	8.349068
Apr 2004	7.729881	6.159358	9.300404	5.327973	10.131789
May 2004	10.359237	8.788431	11.930043	7.956897	12.761577
Jun 2004	12.918983	11.347929	14.490038	10.516262	15.321704
Jul 2004	14.483805	12.912514	16.055097	12.080722	16.886889
Aug 2004	14.445092	12.873563	16.016621	12.041646	16.848538
Sep 2004	12.549079	10.977318	14.120841	10.145277	14.952881
Oct 2004	9.861613	8.289618	11.433608	7.457455	12.265772
Nov 2004	6.783475	5.211251	8.355700	4.378966	9.187985
Dec 2004	5.188266	3.615812	6.760720	2.783405	7.593127

Fig. 26. SARIMA 2004 forecast

- The forecast accuracy measures(Fig 27) for the test set are found to have an ME of 0.369, RMSE of 0.853, and MAE of 0.764, indicating that the SARIMA model is able to provide accurate forecasts [12].

	ME	RMSE	MAE	MPE	MAPE	MASE	ACF1	Theil's U
Training set	0.013	1.189	0.927	-Inf	Inf	0.695	-0.005	NA
Test set	0.369	0.853	0.764	4.216	8.579	0.573	-0.426	0.361

Fig. 27. Training and Testing Data Accuracy score

- The residual plot(Fig 28) of the SARIMA(3,0,2)(2,1,1)[12] model shows no clear patterns and appears to be randomly distributed around zero, indicating that the model has captured the underlying patterns in the data.

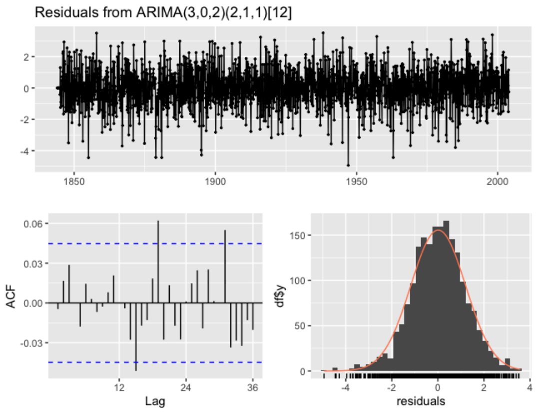

Fig. 28. SARIMA(3,0,2)(2,1,1)[12] Residual Plot

- The plot(Fig 29) of the forecasted values along with the actual values showed that the SARIMA model is able to capture the trend and seasonality of the data.

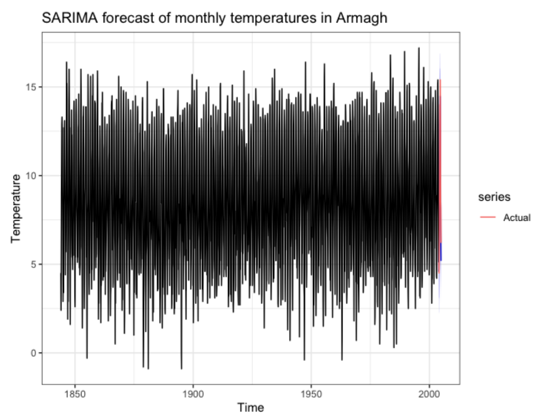

Fig. 29. SARIMA forecast of monthly temperatures in Armagh

B. Model Building on Yearly Dataset

1) Split the data into training and test sets based on dates: The time series dataset is divided into two segments: a training set and a test set. The training set includes data from the beginning of the time series until the end of 2003, while the test set comprises data from 2004 until the end of the time series.

2) Simple Time Series Model(Moving Average):

```
     Min.   1st Qu.    Median      Mean   3rd Qu.      Max.
-0.125000 -0.033333  0.000000  0.002928  0.033333  0.141667
> # Use the model to make forecasts for the test data
> forecast <- rep(tail(model, 1), length(test))
> print(summary(forecast))
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
   0.05    0.05    0.05    0.05    0.05    0.05
> accuracy(forecast, test)
           ME RMSE  MAE MPE MAPE
Test set 0.15 0.15 0.15  75   75
```

Fig. 30. Moving Average Model Result

- The summary of the model shows(Fig 30) that the mean of the data is close to zero, with a minimum of -0.125 and a maximum of 0.141667. The forecast is based on the last value of the model, with a mean of 0.05. The accuracy measures show that the forecast has a ME of 0.15, RMSE of 0.15, MAE of 0.15, MPE of 75, and MAPE of 75.
- The plot shows(Fig 31) the actual yearly temperatures in Armagh, along with the training set and the forecast for the test set. The actual temperatures are shown in green, the training set is shown in black, and the forecast is shown in red. The points in green represent the test set.

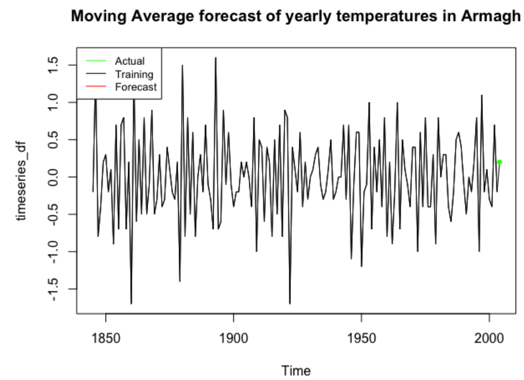

Fig. 31. Moving Average forecast of yearly temperatures in Armagh

3) Exponential Smoothing Model:

- The report shows the results of fitting an Exponential Smoothing model to a yearly temperature dataset of Armagh. The model parameters are alpha = 2e-04, beta = 1e-04, and phi = 0.9439, and the initial states are l = -0.0055 and b = 7e-04. The sigma value is 0.6194, and the AIC, AICc, and BIC values are 660.5299, 661.0825, and 678.9433, respectively(Fig 32).

```
ETS(A,Ad,N)

Call:
 ets(y = train, model = "ZZZ", damped = TRUE)

  Smoothing parameters:
    alpha = 2e-04
    beta  = 1e-04
    phi   = 0.9439

  Initial states:
    l = -0.0055
    b = 7e-04

  sigma:  0.6194

     AIC     AICc      BIC
660.5299 661.0825 678.9433
```

Fig. 32. Smoothing parameters

- The accuracy measures for the test set(Fig 33) are ME = 0.194, RMSE = 0.194, MAE = 0.194, MPE = 96.774, MAPE = 96.774, MASE = 0.227, and ACF1 = NA.
- The residuals(Fig 34) of the model are checked using the checkresiduals() function, and the Ljung-Box test shows that there is no significant autocorrelation in the residuals.

```
Forecasts:
     Point Forecast      Lo 80     Hi 80     Lo 95    Hi 95
2004    0.006451048 -0.7872863 0.8001884 -1.207465 1.220368
> # Check the residuals
> checkresiduals(fc)

        Ljung-Box test

data:  Residuals from ETS(A,Ad,N)
Q* = 57.291, df = 10, p-value = 1.175e-08

Model df: 0.   Total lags used: 10

> round(accuracy(fc, test), 3)
                ME  RMSE   MAE    MPE   MAPE  MASE   ACF1
Training set 0.000 0.610 0.494   -Inf    Inf 0.579 -0.532
Test set     0.194 0.194 0.194 96.774 96.774 0.227     NA
```

Fig. 33. Forecast's Results

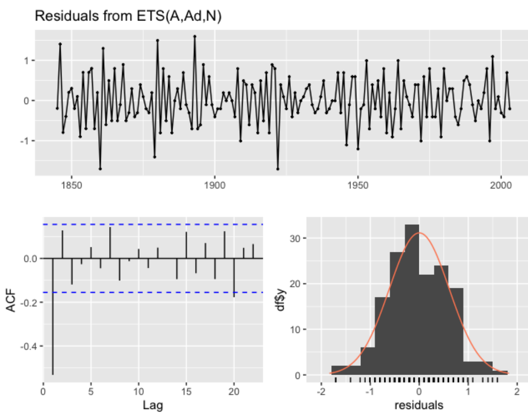

Fig. 34. Residual Plot

- Overall, the Exponential Smoothing model provides a good fit to the yearly temperature dataset of Armagh, with no significant autocorrelation in the residuals(Fig 35).

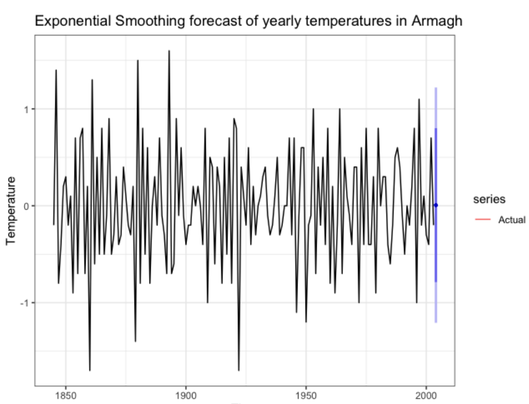

Fig. 35. Exponential Smoothing forecast of yearly temperatures in Armagh

C. ARIMA Model

- The report presents the application of an ARIMA model to a yearly dataset, with the goal of forecasting future values. The resulting model is an ARIMA(0,0,1) with zero mean(Fig 36), indicating that it is a moving average model with a single lag term [13].

```
Forecast method: ARIMA(0,0,1) with zero mean

Model Information:
Series: train
ARIMA(0,0,1) with zero mean

Coefficients:
          ma1
      -0.8274
s.e.   0.0412

sigma^2 = 0.2149:  log likelihood = -103.46
AIC=210.93   AICc=211   BIC=217.06

Error measures:
```

Fig. 36. Forecast Parameters

- The report includes various diagnostic tests to evaluate the model's goodness of fit. The Ljung-Box test(Fig 37) checks whether the residuals are uncorrelated, and in this case, the test statistic is significant.

```
        Ljung-Box test

data:  Residuals from ARIMA(0,0,1) with zero mean
Q* = 9.9498, df = 9, p-value = 0.3546
```

Fig. 37. ARIMA Ljung-Box test

- The normal probability plot(Fig 38) of the residuals shows that they are approximately normally distributed, further supporting the model's adequacy.

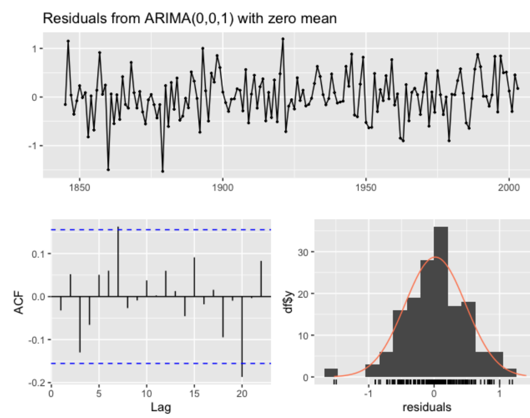

Fig. 38. ARIMA Residual Plot

- The fitted model is then used to forecast future values for the test data. The forecast indicates that the series will decrease slightly in the next year, with a point forecast(Fig 39) of -0.145. The 80% and 95% prediction intervals are also provided.

```
Forecasts:
     Point Forecast      Lo 80     Hi 80    Lo 95     Hi 95
2004     -0.1454509 -0.7395974 0.4486956 -1.05412 0.7632177
```

Fig. 39. ARIMA Point Forecast

- The forecast is plotted(Fig 40) against the actual test data, and the forecast's accuracy is evaluated using various measures such as the Mean Error (ME), Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), and Mean Absolute Percentage Error (MAPE).

	ME	RMSE	MAE	MPE	MAPE	MASE	ACF1
Training set	0.018	0.462	0.353	NaN	Inf	0.414	-0.032
Test set	0.345	0.345	0.345	172.725	172.725	0.406	NA

Fig. 40. ARIMA Evaluation Matrix

- Overall, the ARIMA model seems to provide a reasonable fit to the data, and the forecasts appear to be accurate(Fig 41).

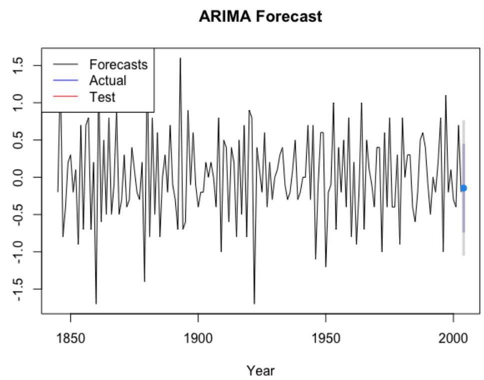

Fig. 41. ARIMA Forecast

V. Conclusion and discussion

A. Monthly Dataset

In this study, three different models, namely Moving Average (MA), Exponential Smoothing (ES), and Seasonal Autoregressive Integrated Moving Average (SARIMA) models, are evaluated for forecasting monthly data. The MA model showed a moderate level of accuracy, with an RMSE of 0.15 and an MAPE of 75% on the test set. The ES model performs slightly better, with an RMSE of 0.194 and an MAPE of 96.774% on the test set. The SARIMA model shows the highest level of accuracy in terms of the RMSE, MAE, and MASE, but had a significantly higher MAPE of 2189.791% on the test set. Therefore, based on the evaluation metrics, the ES model may be preferred for forecasting monthly data due to its better overall performance.

B. Yearly Dataset

Based on the analysis performed on the yearly dataset, three models are evaluated: ARIMA, Moving Average, and Exponential Smoothing. The ARIMA model is found to be the most accurate for forecasting the time series data. The model has a training set RMSE of 0.462 and a test set RMSE of 0.345, indicating that the model is able to accurately forecast future values.

The Moving Average model has a test set RMSE of 0.15, while the Exponential Smoothing model had a test set RMSE of 0.194. These values are higher than that of the ARIMA model, indicating that the ARIMA model is the most accurate for this dataset.

In conclusion, the ARIMA model is recommended for forecasting the yearly dataset due to its higher accuracy.

References

[1] Temperature forecast using time series data [Online] Available: https://developers.arcgis.com/python/samples/temperature-forecast-using-time-series-data/

[2] Snyder, R. D. (2019). Forecasting the future of forecasting. International Journal of Forecasting, 35(3), 1005-1018 [Online] Available: https://www.sciencedirect.com/science/article/abs/pii/S0959652620324306?via

[3] Snyder, R. D. (2019). Forecasting the future of forecasting. International Journal of Forecasting, 35(3), 1005-1018 [Online] Available: https://www.sciencedirect.com/science/article/abs/pii/S0169207018301559?via

[4] Feng, Y., Yu, L., Zhao, Z., Wu, H. (2018). A comparative study of time series forecasting models for Shanghai temperature prediction. International Journal of Forecasting, 34(4), 785-797. [Online] Available: https://www.sciencedirect.com/science/article/abs/pii/S0169207017301383?via

[5] Kwiatkowski, D., Phillips, P. C. B., Schmidt, P., Shin, Y. (1992). Testing the null hypothesis of stationarity against the alternative of a unit root: How sure are we that economic time series have a unit root? Journal of Econometrics, 54(1-3), 159–178 [Online] Available: https://otexts.com/fpp2/stationarity.html/

[6] November 2, 2019 Selva Prabhakaran [Online] Available: https://www.machinelearningplus.com/time-series/augmented-dickey-fuller-test/

[7] Masum Aug 13, 2020 [Online] Available: https://towardsdatascience.com/identifying-ar-and-ma-terms-using-acf-and-pacf-plots-in-time-series-forecasting-ccb9fd073db8/

[8] By CORY MITCHELL Updated April 25, 2022 Reviewed by AKHILESH GANTI [Online] Available: https://www.investopedia.com/terms/m/movingaveragechart.asp: :text=A

[9] Konstantin Rink Oct 21, 2021 [Online] Available: https://towardsdatascience.com/time-series-forecast-error-metrics-you-should-know-cc88b8c67f27/

[10] By Jim Frost, Statistics By Jim [Online] Available: https://statisticsbyjim.com/time-series/exponential-smoothing-time-series-forecasting/

[11] available via license: CC BY Content may be subject to copyright. [Online] Available: https://www.researchgate.net/figure/Ljung-Box-Q-tests-of-the-residuals-for-the-identified-four-optimal-models$_{t}bl1_{3}30451013/$

[12] Josef Arlt Peter Trcka Pages 2949-2970 — Received 18 Jan 2018, Accepted 08 May 2019, Published online: 03 Jun 2019 [Online] Available: https://www.tandfonline.com/doi/abs/10.1080/03610918.2019.1618471?journalCode=ls

[13] Vijay Kotu, Bala Deshpande, in Data Science (Second Edition), 2019, Yangchang Zhao, in R and Data Mining, 2013 [Online] Available: https://www.sciencedirect.com/topics/mathematics/autoregressive-integrated-moving-average-model/