

­­­­DOCUMENT APPROVAL SHEET

TIMESERIES

gold boundless bar

**DOCUMENT IDENTIFICATION:**

MSBA AY 21-22/TIME SERIES

**DOCUMENT AUTHOR:**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **APPROVED BY/ ID** | **DEPT.** | **SCHOOL** | **SIGNATURE** | **DATE** |
| Teja Alluru (1150077) | MSBA/ Student | University of Washington, Tacoma, Milgard School of Business |  | 05-Dec-2021 |

*AUTHORED BY: Authored by mean of a person/entity responsible for the creation of the document.*

INDEX OF CONTENTS

gold boundless bar

[DOCUMENT APPROVAL SHEET 1](#_Toc89638960)

[INDEX OF CONTENTS 3](#_Toc89638961)

[INDEX OF FIGURES 4](#_Toc89638962)

[1 DATA GATHERING 5](#_Toc89638963)

[2 DATA STRUCTURE 7](#_Toc89638964)

[3 DATA CLEANING – REPLACING MISSING VALUES 8](#_Toc89638965)

[4 DATA CONVERSION – COLLAPSING DATA – MONTHLY AVERAGES 8](#_Toc89638966)

[5 PLOTS 9](#_Toc89638967)

[6 DECOMPOSING RAIN DATA 11](#_Toc89638968)

[6.1 Daily Rain Data 11](#_Toc89638969)

[6.2 Monthly Rain Data 12](#_Toc89638970)

[7 MODEL & FORECAST - EXPONENTIAL SMOOTHING – (HOLTWINTERS) 13](#_Toc89638971)

[7.1 Daily Data 13](#_Toc89638972)

[7.1.1 Model Output 13](#_Toc89638973)

[7.1.2 Residuals 14](#_Toc89638974)

[7.1.3 Residuals Correlograms 15](#_Toc89638975)

[7.1.4 Forecast 16](#_Toc89638976)

[7.2 Monthly Data 17](#_Toc89638977)

[7.2.1 Model Output 17](#_Toc89638978)

[7.2.2 Residuals 18](#_Toc89638979)

[7.2.3 Residuals Correlograms 19](#_Toc89638980)

[7.2.4 Forecast 20](#_Toc89638981)

[8 MODEL AND FORECAST USING AN AUTO.ARIMA MODEL 21](#_Toc89638982)

[8.1 Daily Data 21](#_Toc89638983)

[8.1.1 Model Fit 21](#_Toc89638984)

[8.1.2 Residuals 22](#_Toc89638985)

[8.1.3 Residuals Correlograms 23](#_Toc89638986)

[8.1.4 Forecast 24](#_Toc89638987)

[8.2 Monthly Data 25](#_Toc89638988)

[8.2.1 Model Fit 25](#_Toc89638989)

[8.2.2 Residuals 26](#_Toc89638990)

[8.2.3 Residuals Correlograms 27](#_Toc89638991)

[8.2.4 Forecast 28](#_Toc89638992)

[9 PARSIMONIOUS MODEL AND PREDICTION - AUTO.ARIMA MODEL 29](#_Toc89638993)

[9.1 Model Fit 29](#_Toc89638994)

[9.2 Residuals 30](#_Toc89638995)

[9.3 Prediction 31](#_Toc89638996)

[9.4 Forecast 32](#_Toc89638997)

INDEX OF FIGURES

gold boundless bar

[Figure 1‑1: Data Tools NOAA Website – Data Gathering 5](#_Toc89638998)

[Figure 1‑2: Cart Output Selection – Data Gathering 6](#_Toc89638999)

[Figure 1‑3: Cart Data Types Selection – Data Gathering 6](#_Toc89639000)

[Figure 1‑4: Figure 1‑1: Cart Data Download– Data Gathering 7](#_Toc89639001)

[Figure 2‑1: Variables Statistics 7](#_Toc89639002)

[Figure 3‑1: Imputing NAs with 0 8](#_Toc89639003)

[Figure 4‑1: Data Conversion to Time-Series 8](#_Toc89639004)

[Figure 5‑1: Daily Rain Data Plot 9](#_Toc89639005)

[Figure 5‑2: Monthly Rain Data Plot 10](#_Toc89639006)

[Figure 6‑1: Decomposed Data – Daily Rain Data 11](#_Toc89639007)

[Figure 6‑2: Decomposed Data – Monthly Rain Data 12](#_Toc89639008)

[Figure 7‑1: Holt-Winters Model – Observed/Fitted Values – Daily Data 13](#_Toc89639009)

[Figure 7‑2: Holt-Winters Model – Residual Graph – Daily Data 14](#_Toc89639010)

[Figure 7‑3: Holt-Winters Model – ACF/PACF – Correlograms – Residuals – Daily Data 15](#_Toc89639011)

[Figure 7‑4: Holt-Winters Model – Forecast – 365days – Daily Data 16](#_Toc89639012)

[Figure 7‑5: Holt-Winters Model – Observed/Fitted Values – Monthly Data 17](#_Toc89639013)

[Figure 7‑6: Holt-Winters Model – Residual Graph – Monthly Data 18](#_Toc89639014)

[Figure 7‑7: Holt-Winters Model – ACF/PACF – Correlograms – Residuals – Monthly Data 19](#_Toc89639015)

[Figure 7‑8: Holt-Winters Model – Forecast – 12 Months – Monthly Data 20](#_Toc89639016)

[Figure 8‑1: Auto.ARIMA Model – Output – Daily Data 21](#_Toc89639017)

[Figure 8‑2: Auto.ARIMA Model – Residual Graph – Daily Data 22](#_Toc89639018)

[Figure 8‑3: Auto.ARIMA Model– ACF/PACF – Correlograms – Residuals – Daily Data 23](#_Toc89639019)

[Figure 8‑4: Auto.ARIMA Model– Forecast – 365days – Daily Data 24](#_Toc89639020)

[Figure 8‑5: Auto.ARIMA Model – Output – Monthly Data 25](#_Toc89639021)

[Figure 8‑6: Auto.ARIMA Model – Residual Graph – Monthly Data 26](#_Toc89639022)

[Figure 8‑7: Auto.ARIMA Model– ACF/PACF – Correlograms – Residuals – Monthly Data 27](#_Toc89639023)

[Figure 8‑8: Auto.ARIMA Model– Forecast – 12 Months – Monthly Data 28](#_Toc89639024)

[Figure 9‑1: Auto.ARIMA Model – Output – Parsimonious Model 29](#_Toc89639025)

[Figure 9‑2: Auto.ARIMA Model – Residual Graph – Parsimonious Model 30](#_Toc89639026)

[Figure 9‑3: Auto.ARIMA Model– Predict – 12 Months – Parsimonious Model 31](#_Toc89639027)

[Figure 9‑4: Auto.ARIMA Model– Actual Future Values from NOAA – Parsimonious Model 31](#_Toc89639028)

[Figure 9‑5: Auto.ARIMA Model– Forecast – 24 Months – Parsimonious Model 32](#_Toc89639029)

# DATA GATHERING

The data for the logistic regression lab is obtained from NOAA using the following steps. The weather data was collected for Atlanta airport for the year 2020.

* Go to <https://www.ncdc.noaa.gov/cdo-web/>
* Select data tools which will direct you to the following page <https://www.ncdc.noaa.gov/cdo-web/datatools>
* Select find a station on the above page which will take you to <https://www.ncdc.noaa.gov/cdo-web/datatools/findstation>
* Enter a location data of your interest, In the dataset select “Daily Summaries”, select a data range of 01-Jan-1930 to 31-Dec-2020, and in data categories select precipitation,
* Then click on the airport icon and see if it has 100% coverage, if it has 100% coverage then add to cart and press view all items.

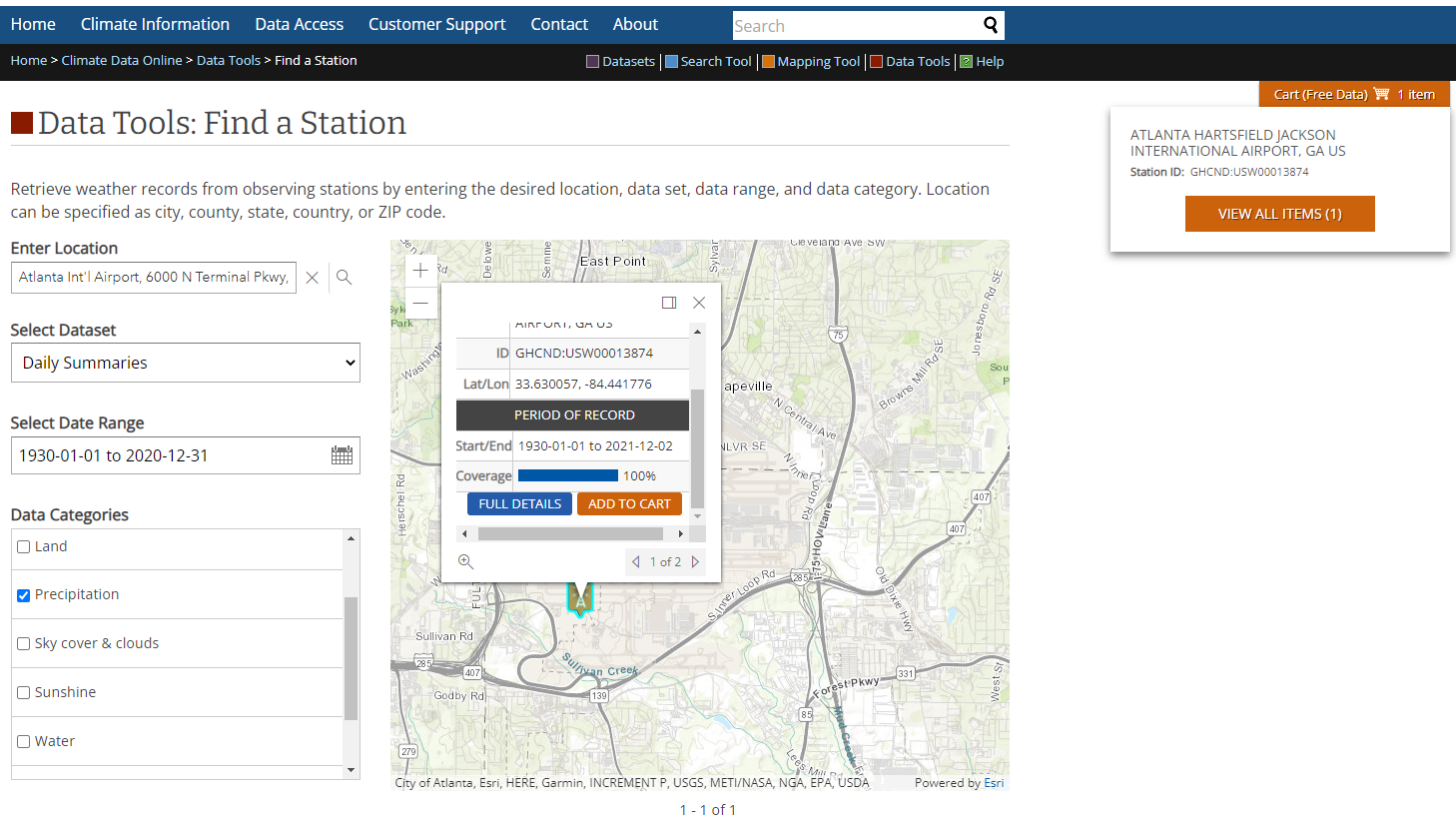


Figure 1‑1: Data Tools NOAA Website – Data Gathering

* Then in the card in the output format, select the Custom GHCN-Daily CSV option and press continue.

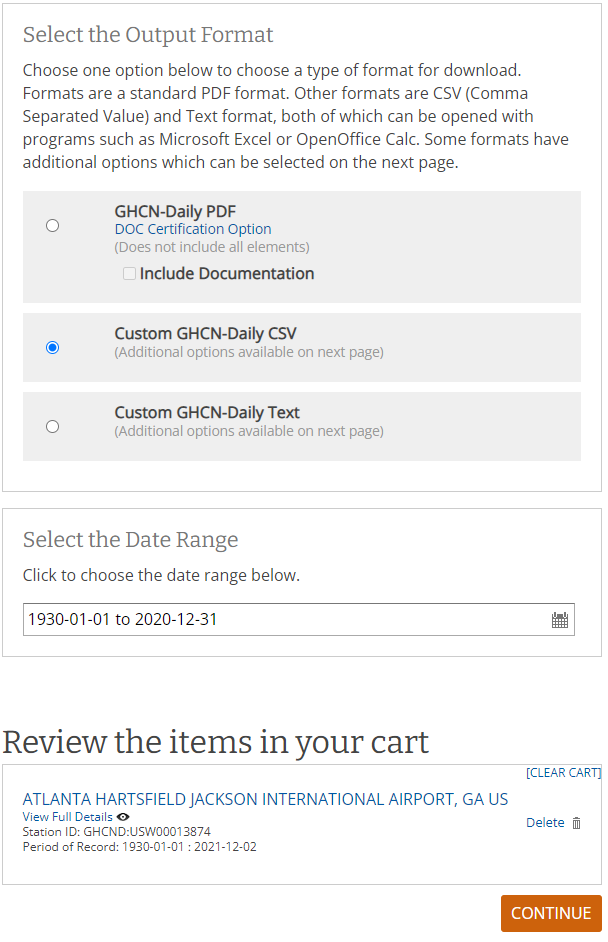


Figure 1‑2: Cart Output Selection – Data Gathering

* Select the required fields and press continue.

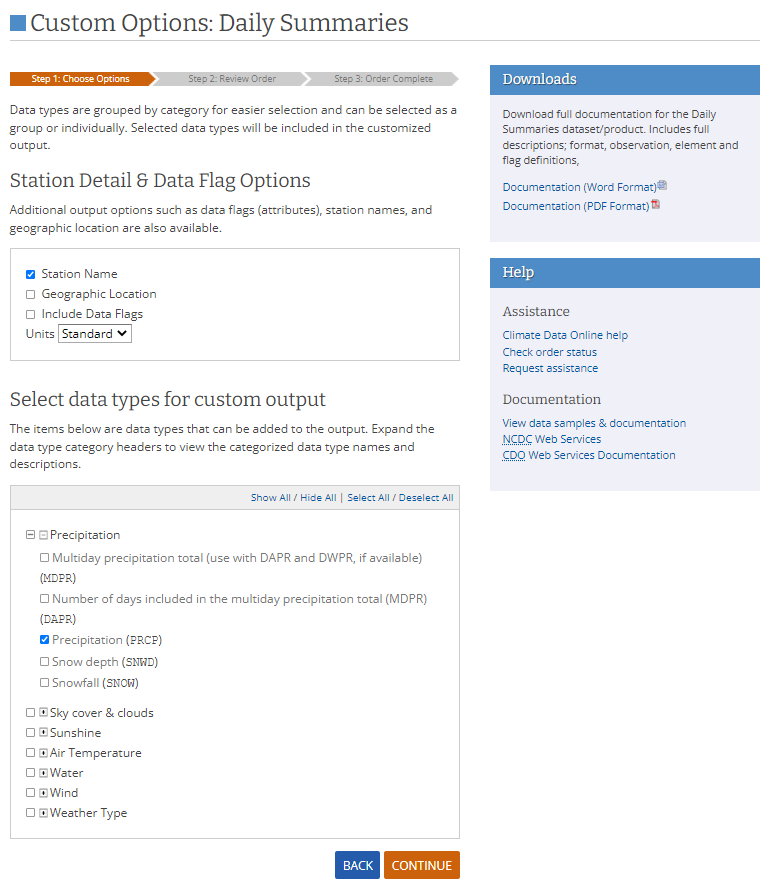


Figure 1‑3: Cart Data Types Selection – Data Gathering

* Provide the email address and submit enter and you'll get the requested data through e-mail.

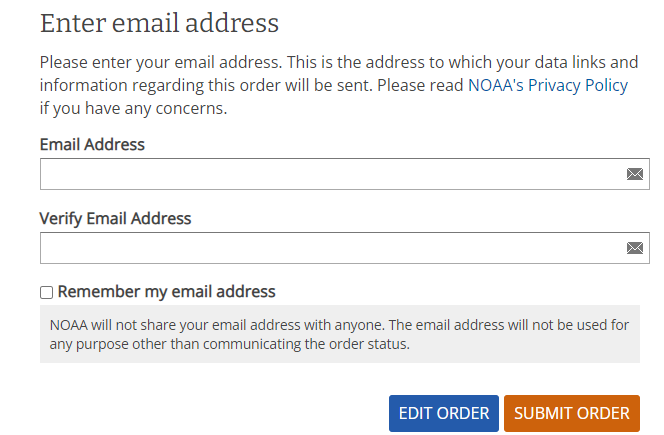


Figure 1‑4: Figure 1‑1: Cart Data Download– Data Gathering

# DATA STRUCTURE

The initial data consisted of 33237 rows and 4 columns. Figure 2‑1 gives the summary statistics of all the variables of the dataset. As seen in Figure 2‑1, it's quite evident that there are two null values in the dataset.

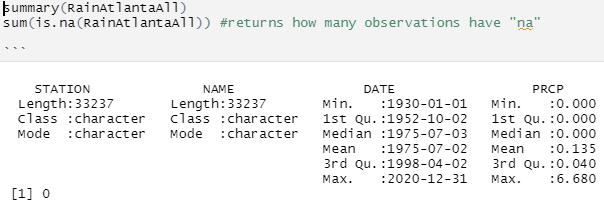


Figure 2‑1: Variables Statistics

# DATA CLEANING – REPLACING MISSING VALUES

The data is checked for null values and since only 2 null values are present the null values are replaced with 0’s. The R Code used to test for null values and the results are shown in Figure 3‑1 below.

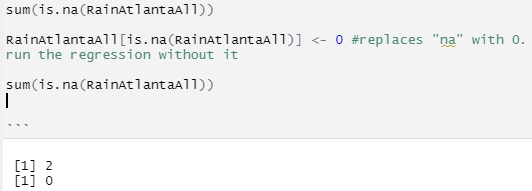


Figure 3‑1: Imputing NAs with 0

# DATA CONVERSION – COLLAPSING DATA – MONTHLY AVERAGES

The data obtained from the above section 2 is converted to time-series so that R detects it as time-series instead of normal data. The R-code used to perform the conversion of data to time-series is presented in Figure 4‑1 below.

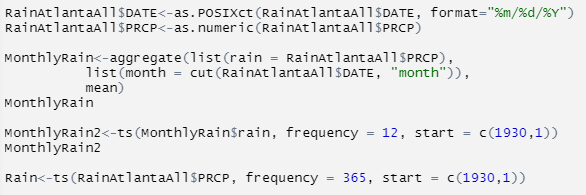


Figure 4‑1: Data Conversion to Time-Series

# PLOTS

The plots of the time series for the daily and monthly rain data are presented in Figure 5‑1 and Figure 5‑2 below.

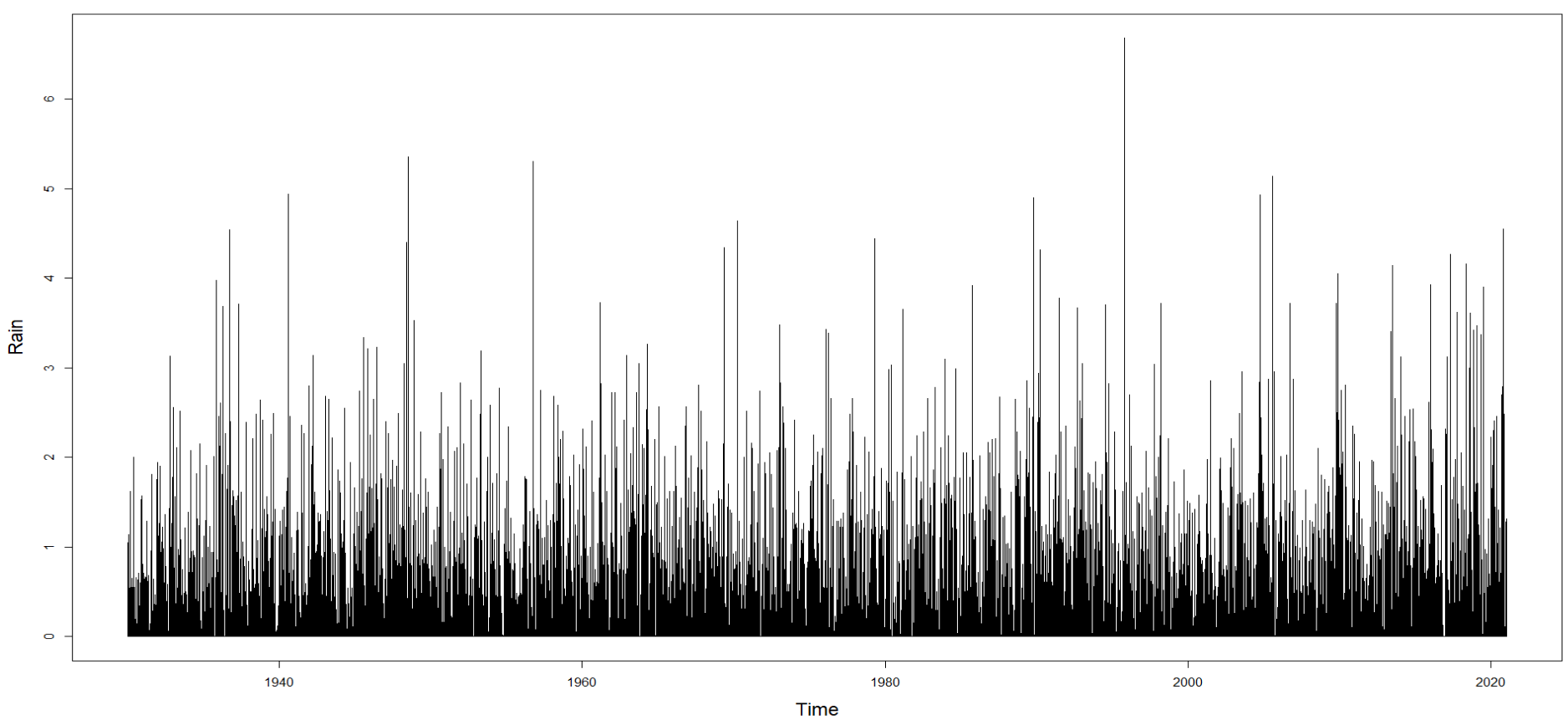


Figure 5‑1: Daily Rain Data Plot

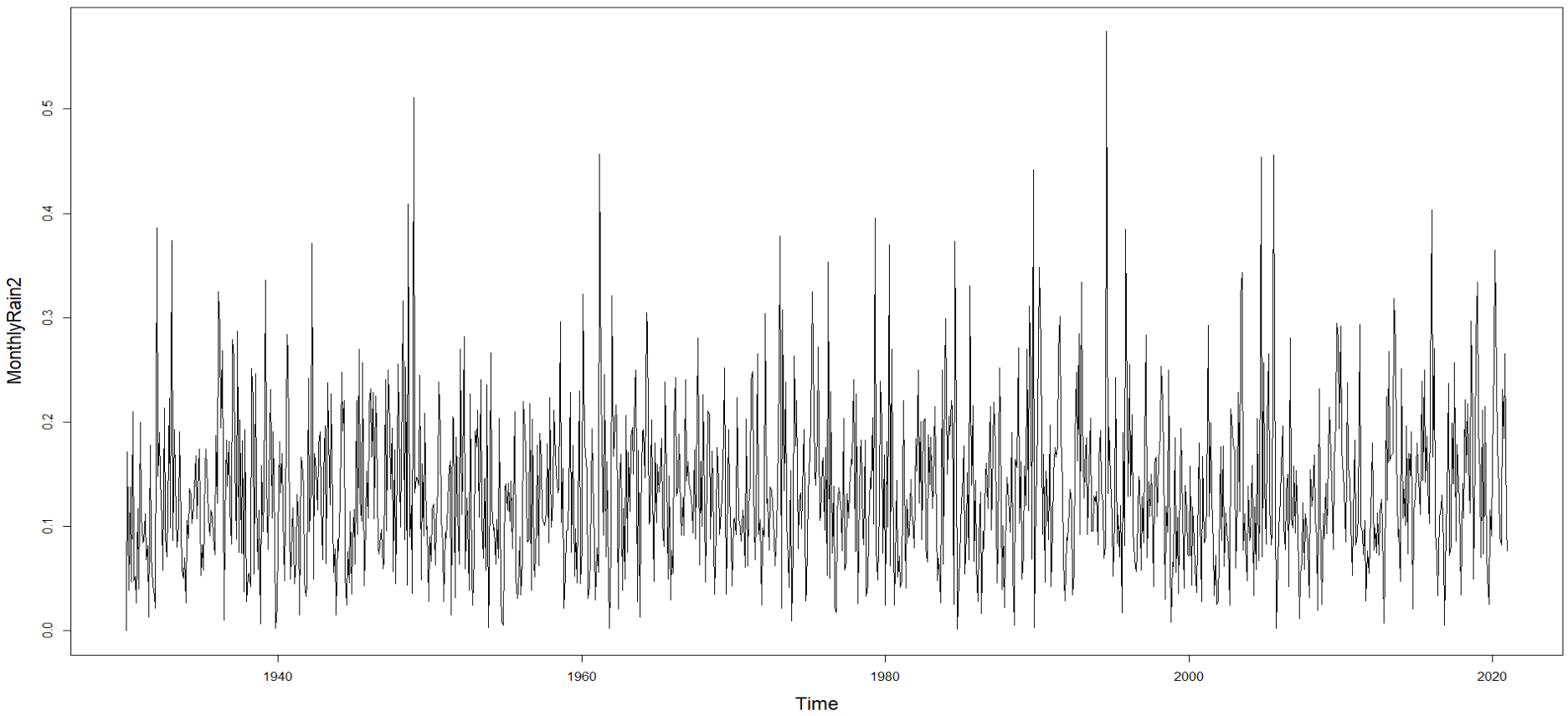


Figure 5‑2: Monthly Rain Data Plot

# DECOMPOSING RAIN DATA

## Daily Rain Data

The daily rain data was decomposed into observed, trend, seasonal and random components and is presented in Figure 6‑1. From Figure 6‑1 there is no trend but it seems that there is seasonality in the data and there is the presence of some random spikes where there was the record of higher rainfall. The differencing of the data to make data stationery/remove seasonality/trend is directly done in the ARIMA models instead of doing it manually.

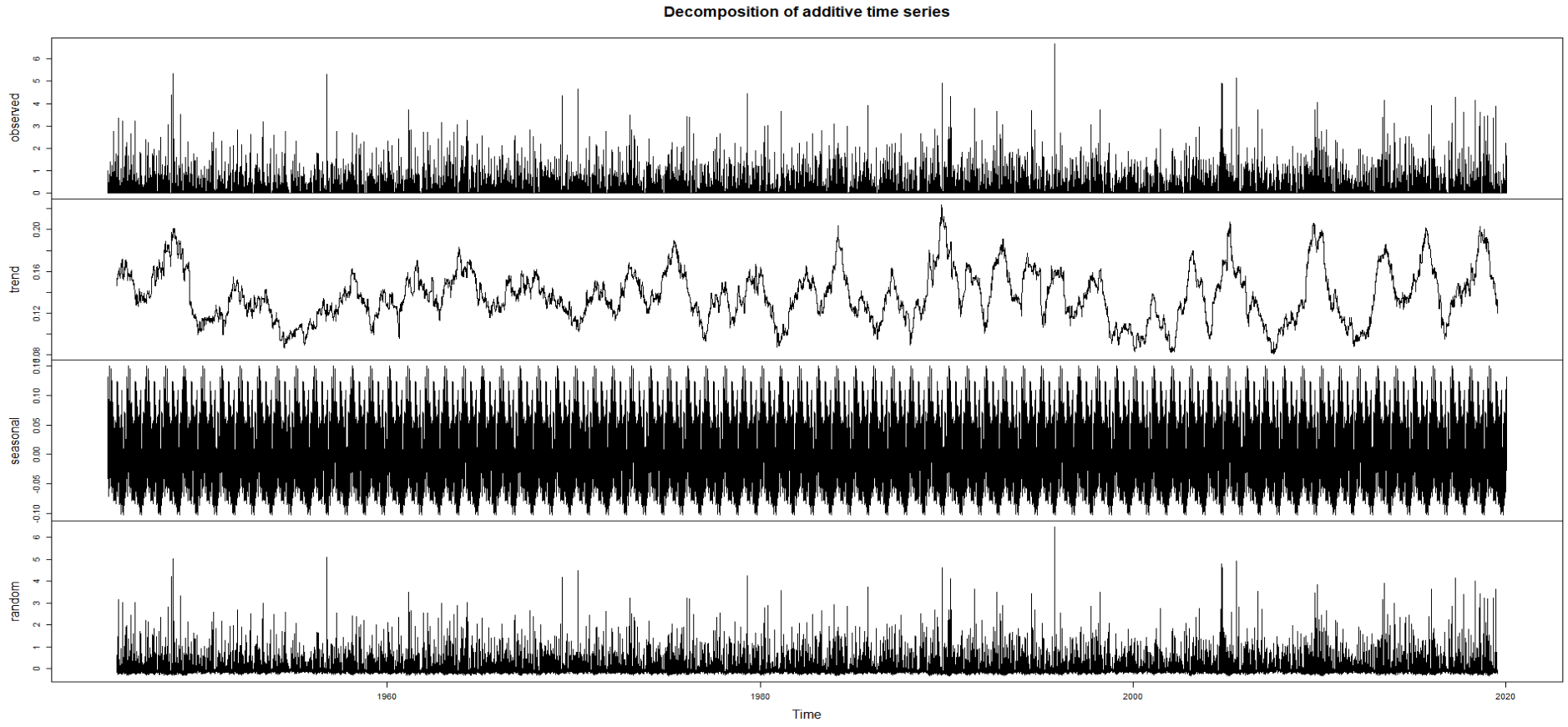


Figure 6‑1: Decomposed Data – Daily Rain Data

## Monthly Rain Data

The monthly rain data was decomposed into observed, trend, seasonal and random components and is presented in Figure 6‑2. From Figure 6‑2 there is no trend but it seems that there is seasonality in the data and there is the presence of some random spikes where there was a record of higher rainfall. The differencing of the data to make data stationery/remove seasonality/trend is directly done in the ARIMA models instead of doing it manually.

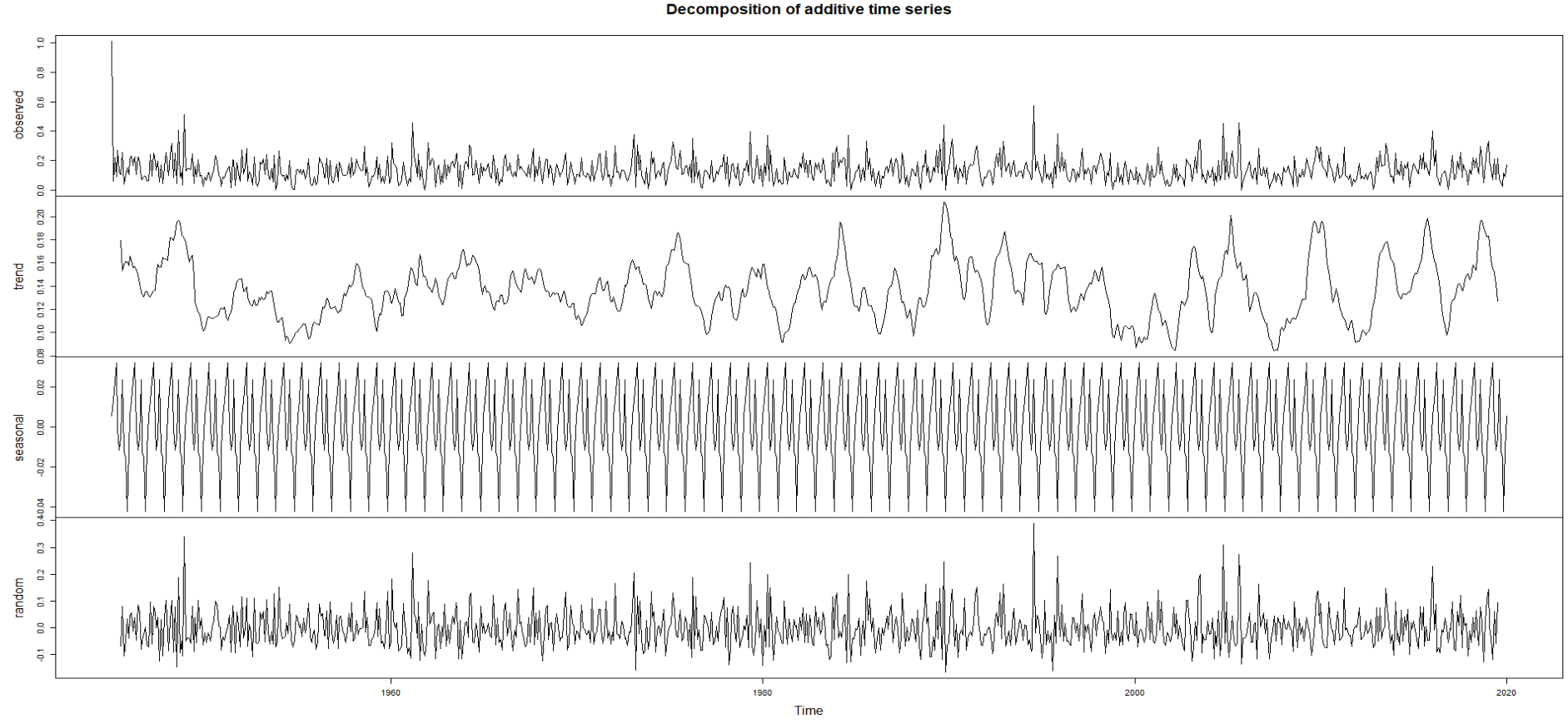


Figure 6‑2: Decomposed Data – Monthly Rain Data

# MODEL & FORECAST - EXPONENTIAL SMOOTHING – (HOLTWINTERS)

## Daily Data

### Model Output

The Holt-Winters model is run on the daily rain data and the observed/fitted graph is shown in Figure 7‑1 below. The observed values are shown in blue while fitted values are shown in purple. From Figure 7‑1 the fitted values seem ok but they were completely off in predicting spikes of rain. Overall, the data doesn’t seem to be predicted well.

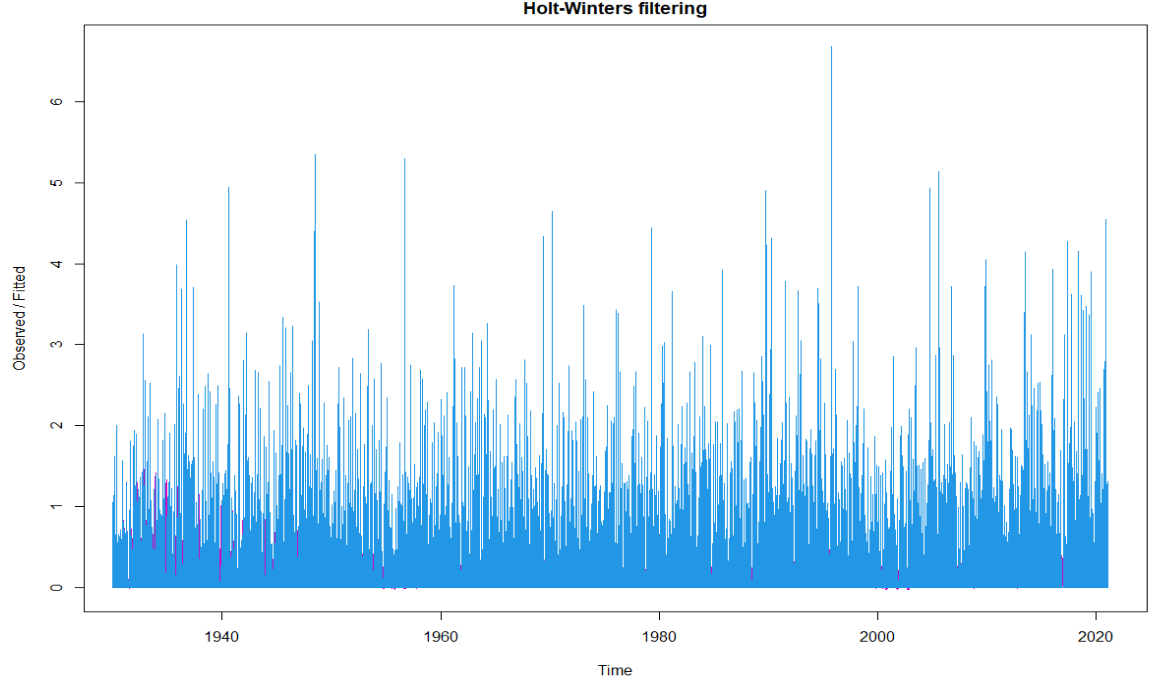


Figure 7‑1: Holt-Winters Model – Observed/Fitted Values – Daily Data

### Residuals

The residuals plot of the Holt-Winters model run on the daily rain data is presented in Figure 7‑2 below. From Figure 7‑2 the residuals suggest that the model couldn’t adequately capture information from the data as such the model is not a good fit at all. Also, from the plot, the difference between observed and fitted values seems to be great which is not an ideal scenario and this exponential smoothing is not helping to predict the rain variable accurately.

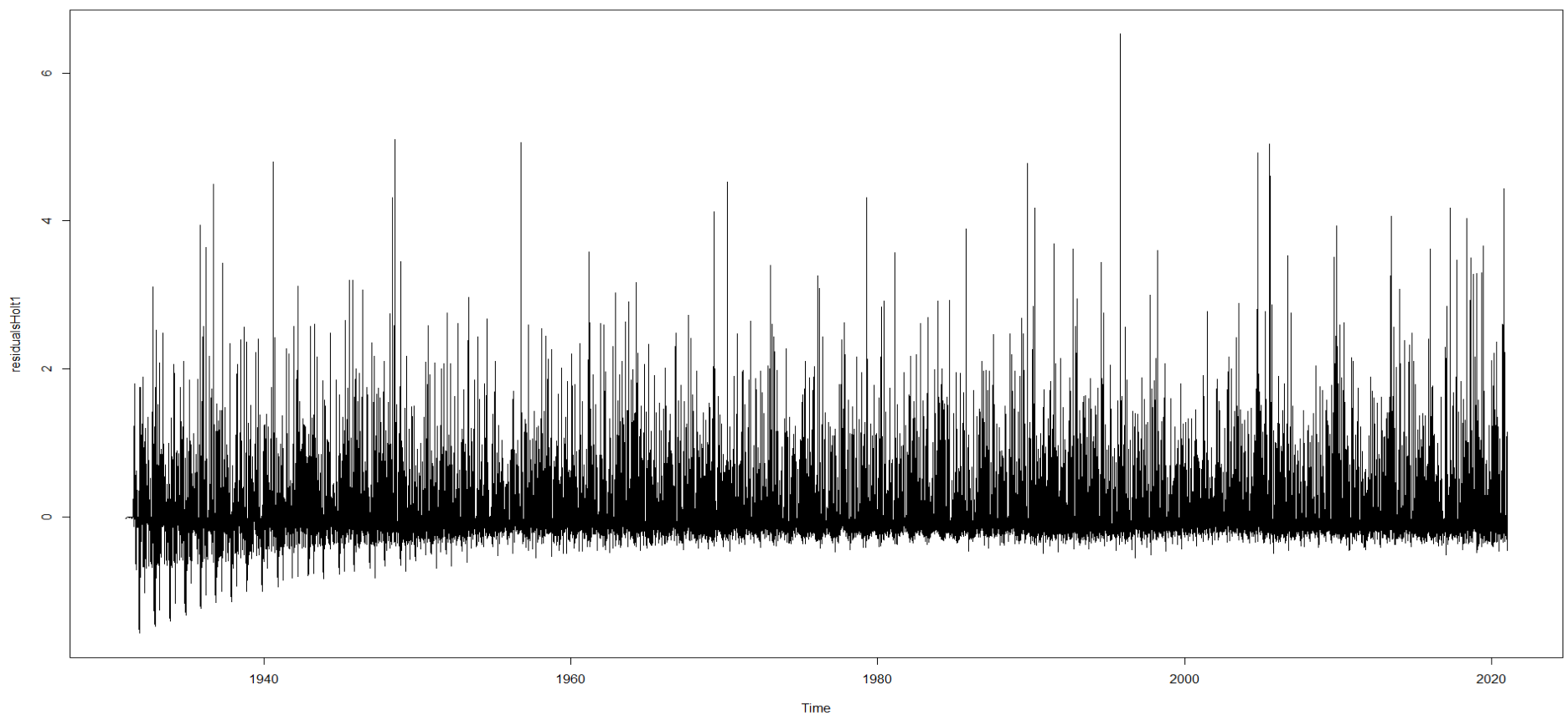


Figure 7‑2: Holt-Winters Model – Residual Graph – Daily Data

### Residuals Correlograms

The ACF and PACF plots of the residuals are presented in Figure 7‑3. From Figure 7‑3 the plots have crossed the blue lines multiple times suggesting that this is not a great model, also there seems to be lag which isn’t ideal in our case.

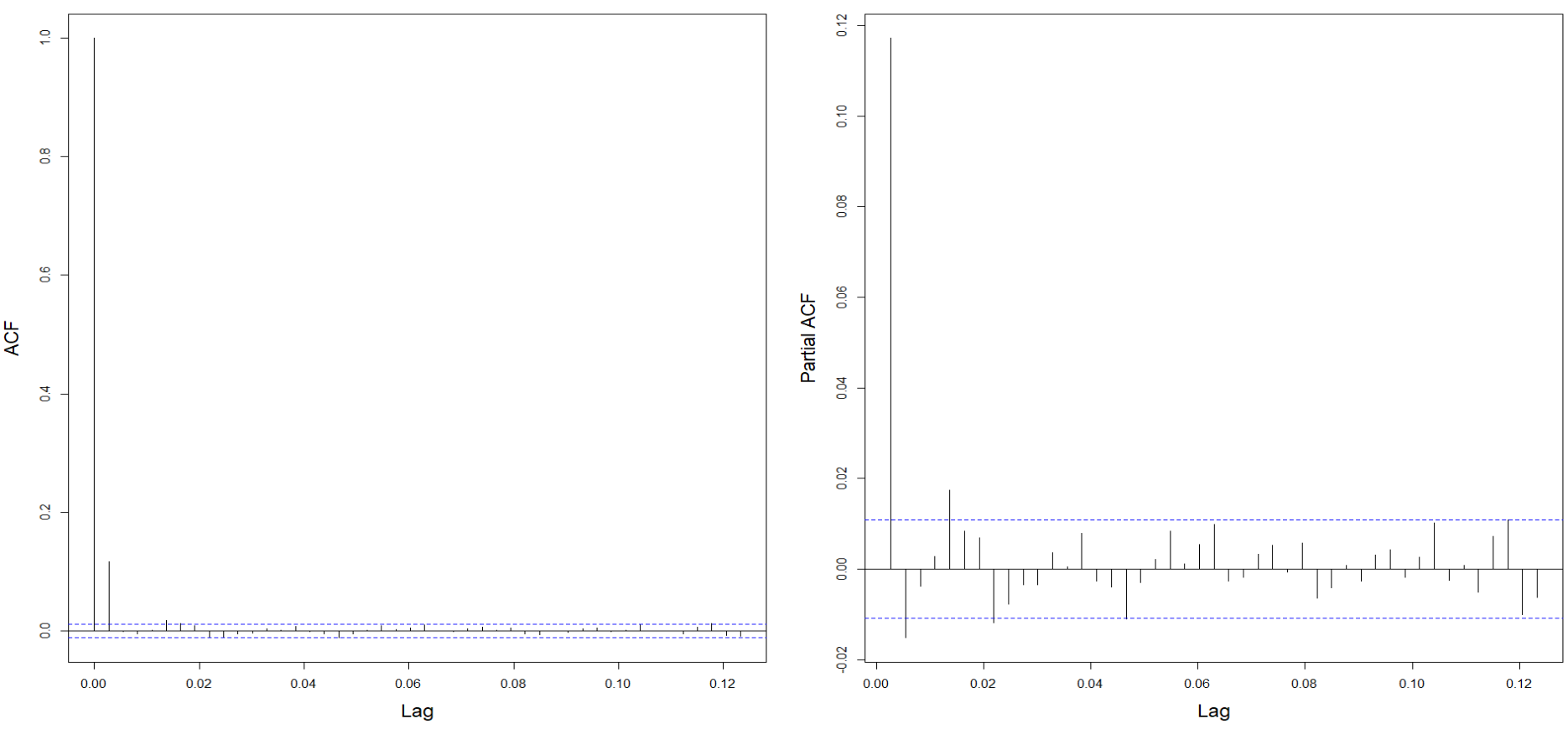


Figure 7‑3: Holt-Winters Model – ACF/PACF – Correlograms – Residuals – Daily Data

### Forecast

The holt winters exponential smoothing model is used to forecast the data for 365 days into the future and the forecasting plot is presented in Figure 7‑4. From Figure 7‑4 the forecast doesn’t seem to be great for this model.

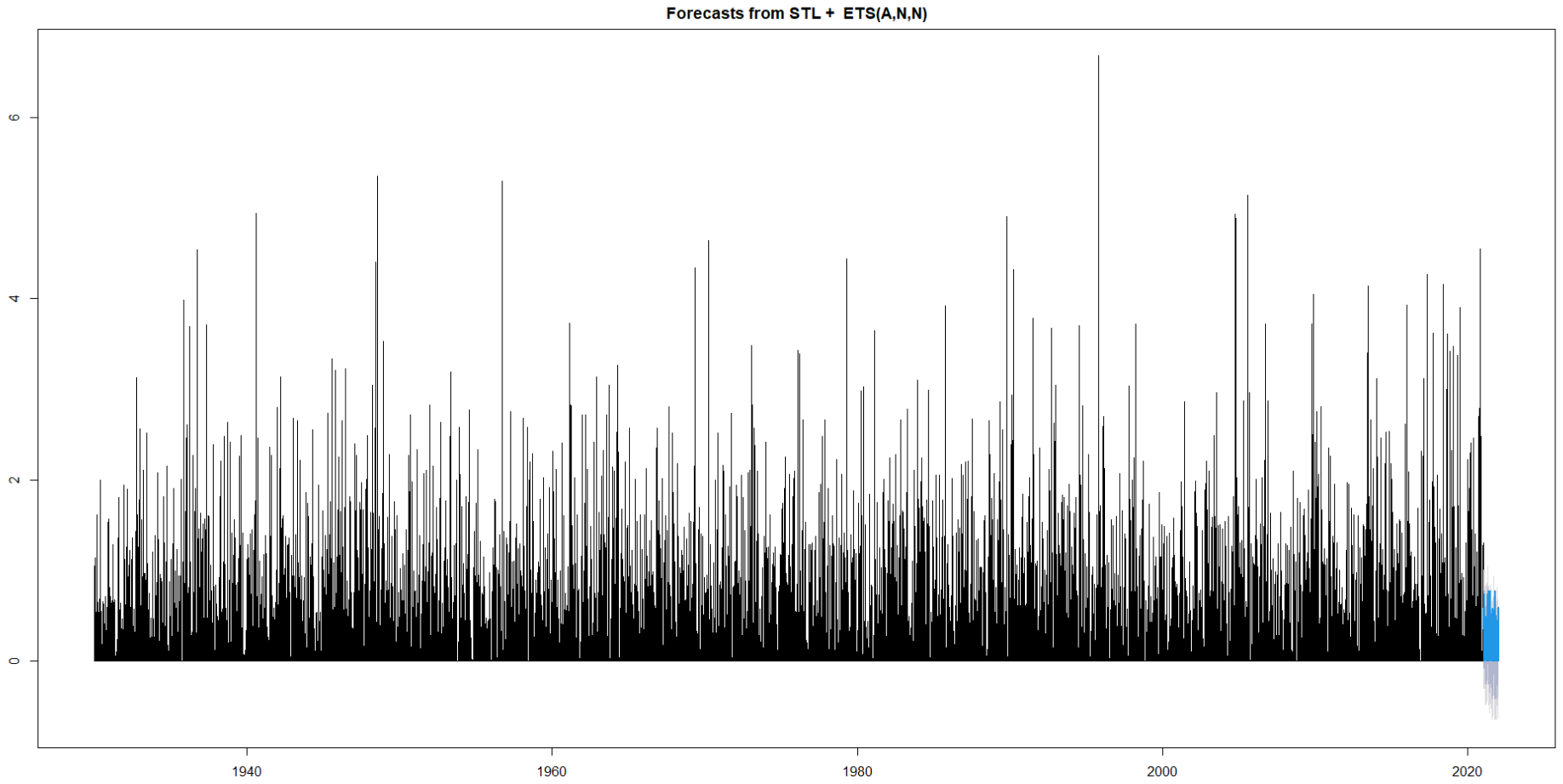


Figure 7‑4: Holt-Winters Model – Forecast – 365days – Daily Data

## Monthly Data

### Model Output

The Holt-Winters model is run on the monthly rain data and the observed/fitted graph is shown in Figure 7‑5 below. The observed values are shown in blue while fitted values are shown in purple. From Figure 7‑5 the fitted values seem better but they were completely off in predicting spikes of rain. But as observed this seems to be a better prediction than daily data as we have already done smoothing by converting to monthly data here.

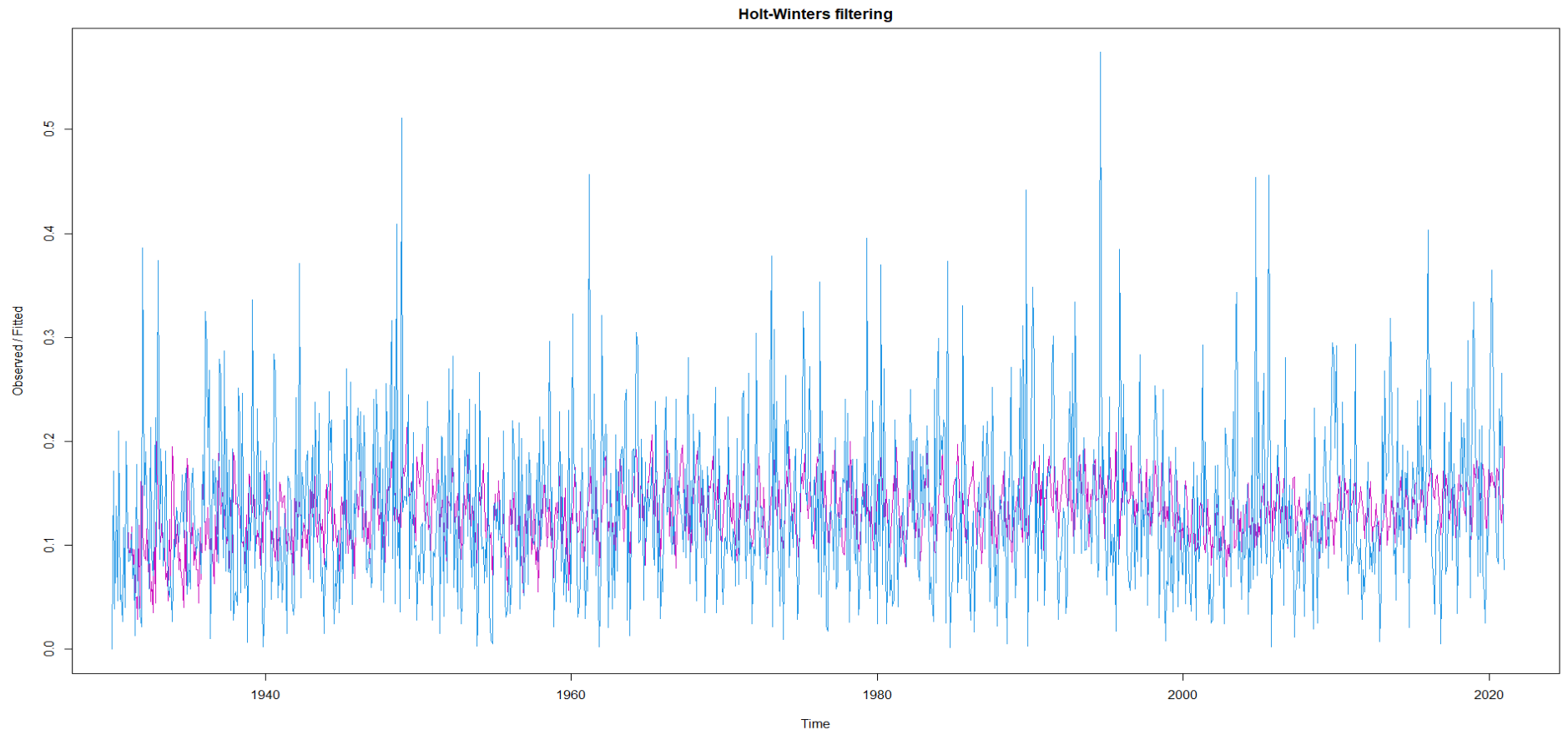


Figure 7‑5: Holt-Winters Model – Observed/Fitted Values – Monthly Data

### Residuals

The residuals plot of the Holt-Winters model run on the daily rain data is presented in Figure 7‑6 below. From Figure 7‑6 the residuals suggest that the model couldn’t adequately capture information from the data as such the model is not a good fit at all. Also, from the plot, the difference between observed and fitted values seems to be more which is not ideal but this is a lot better than the daily data as the data is already smoothened due to consideration of monthly data.

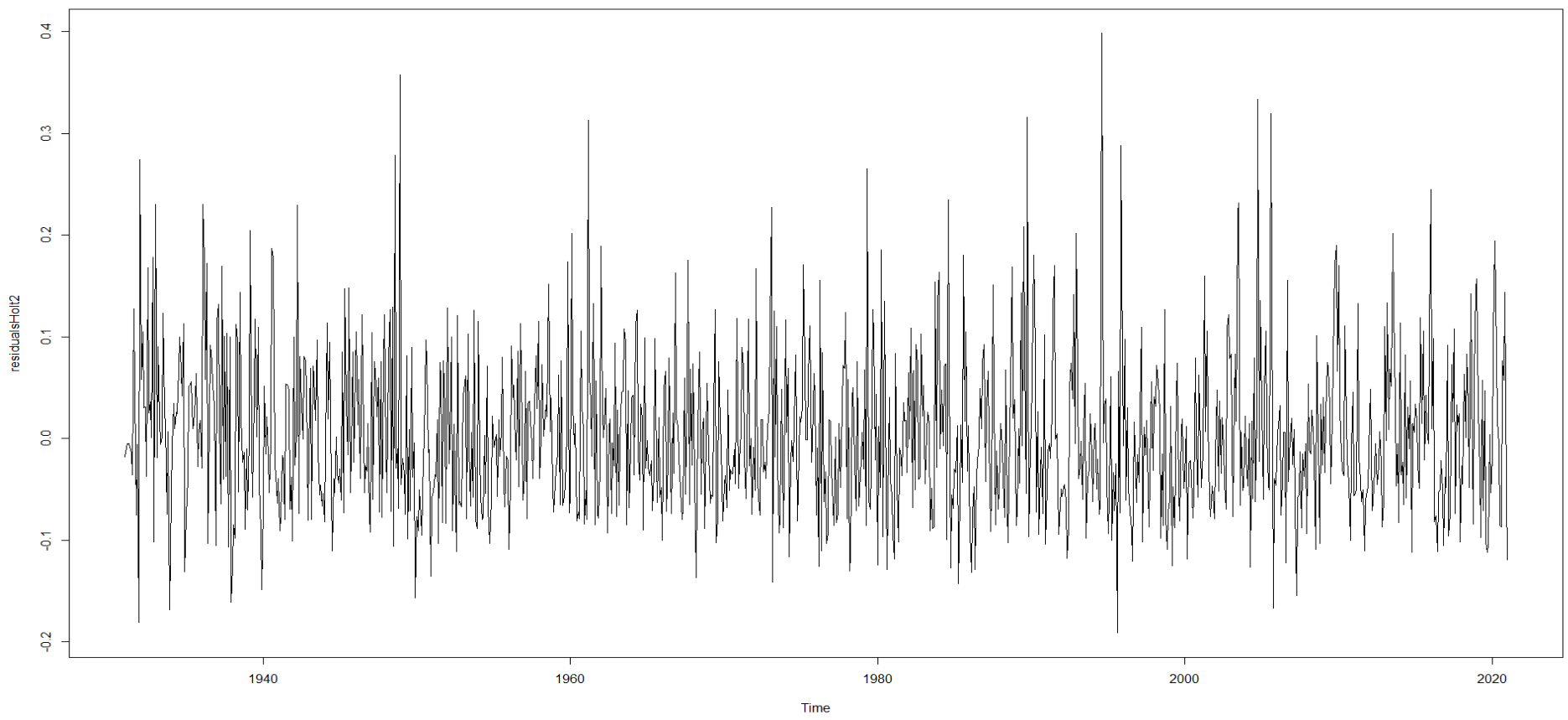


Figure 7‑6: Holt-Winters Model – Residual Graph – Monthly Data

### Residuals Correlograms

The ACF and PACF plots of the residuals are presented in Figure 7‑7. From Figure 7‑7 the plots have crossed the blue lines multiple times suggesting that this is not a great model, also there seems to be lag which isn’t ideal in our case. Again, there seems to be quite an improvement from daily data due to smoothing due to consideration of monthly data.

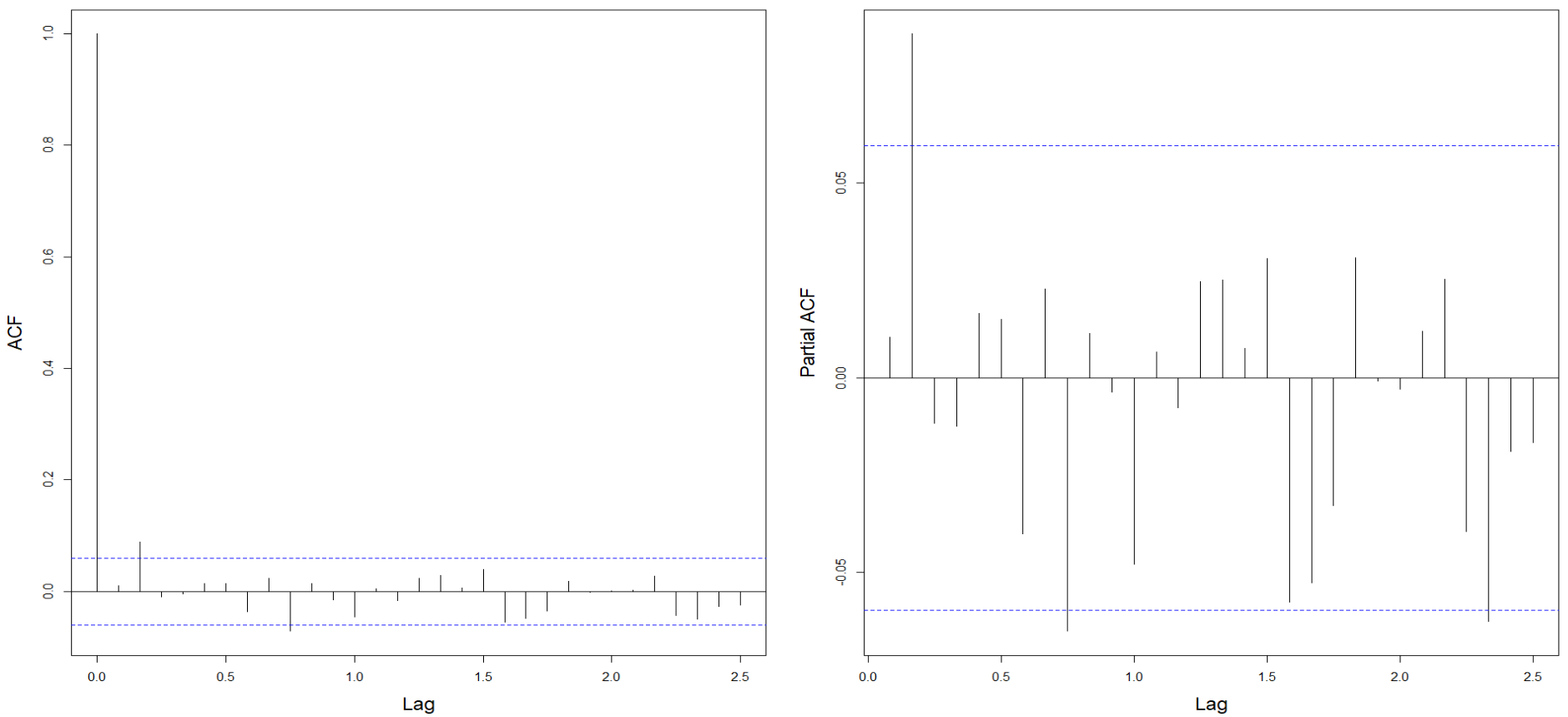


Figure 7‑7: Holt-Winters Model – ACF/PACF – Correlograms – Residuals – Monthly Data

### Forecast

The holt winters exponential smoothing model is used to forecast the data for 12 months into the future and the forecasting plot is presented in Figure 7‑8. From Figure 7‑8 the forecast seems a little better compared to the daily data one.

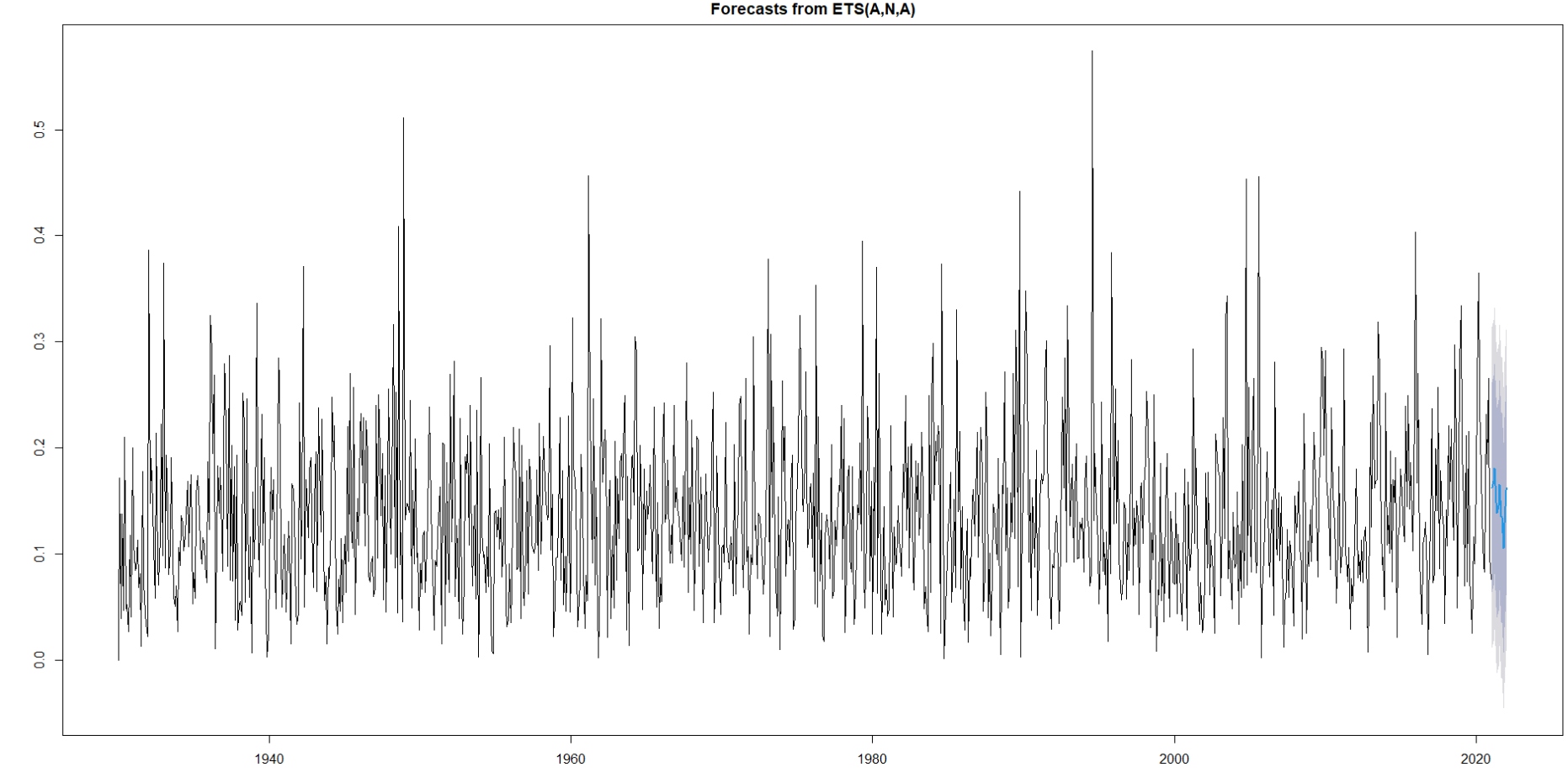


Figure 7‑8: Holt-Winters Model – Forecast – 12 Months – Monthly Data

# MODEL AND FORECAST USING AN AUTO.ARIMA MODEL

## Daily Data

### Model Fit

The Auro.ARIMA model is run on the daily rain data and the output is shown in Figure 8‑1 below. The AIC value is high. From Figure 8‑1 the p,d & q values are 2, 0 and 0 respectively. It seems like there is no differencing done and the ma value also seems to be 0. It’s quite a basic model and I don’t think so it’ll have a good fit for the data.

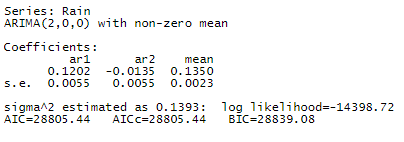


Figure 8‑1: Auto.ARIMA Model – Output – Daily Data

### Residuals

The residuals plot of the Auto-ARIMA model run on the daily rain data is presented in Figure 8‑2 below. From Figure 8‑2 the residuals suggest that the model couldn’t adequately capture information from the data as such the model is not a good fit at all. Also, from the plot, the difference between observed and fitted values seems to be great which is not an ideal scenario.

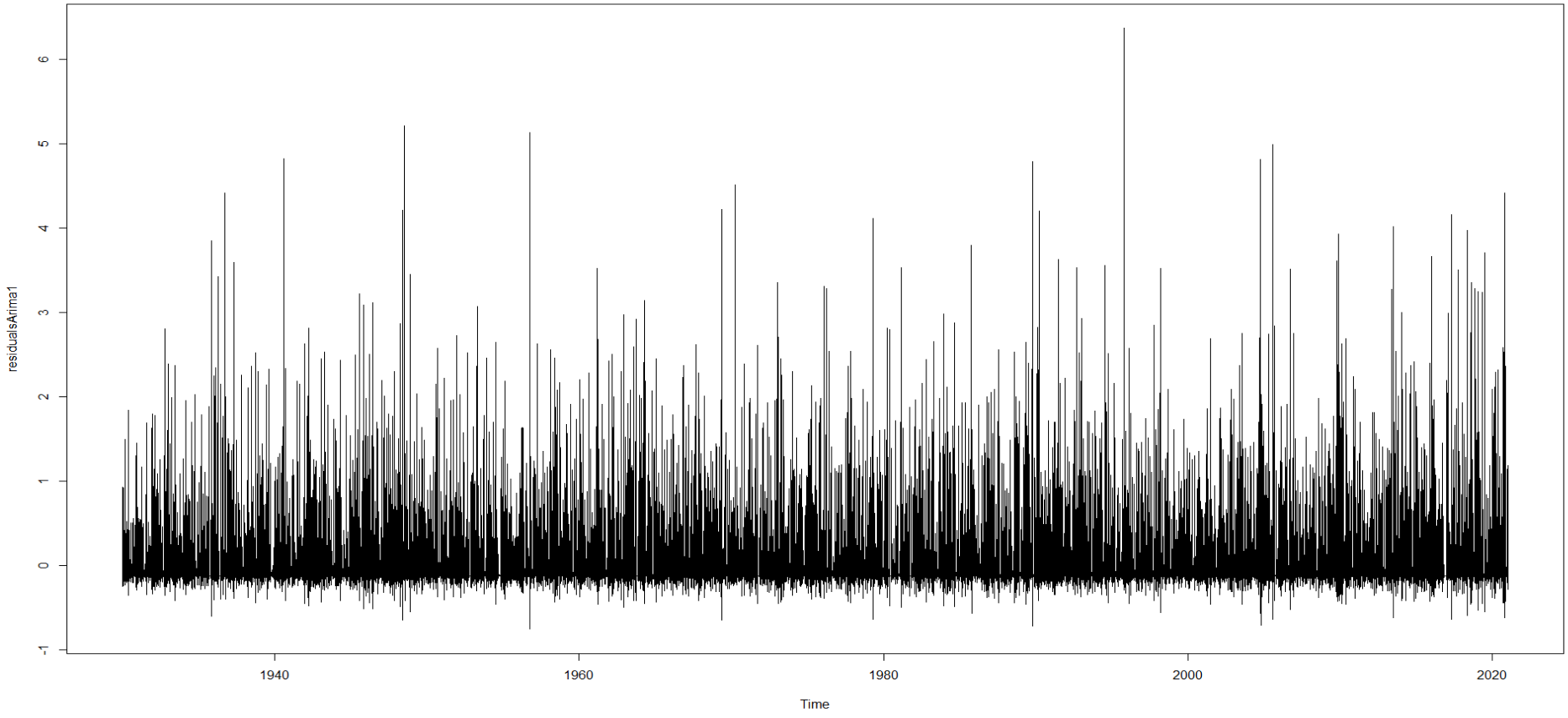


Figure 8‑2: Auto.ARIMA Model – Residual Graph – Daily Data

### Residuals Correlograms

The ACF and PACF plots of the residuals are presented in Figure 8‑3. From Figure 8‑3 the plots have crossed the blue lines multiple times suggesting that this is not a great model.

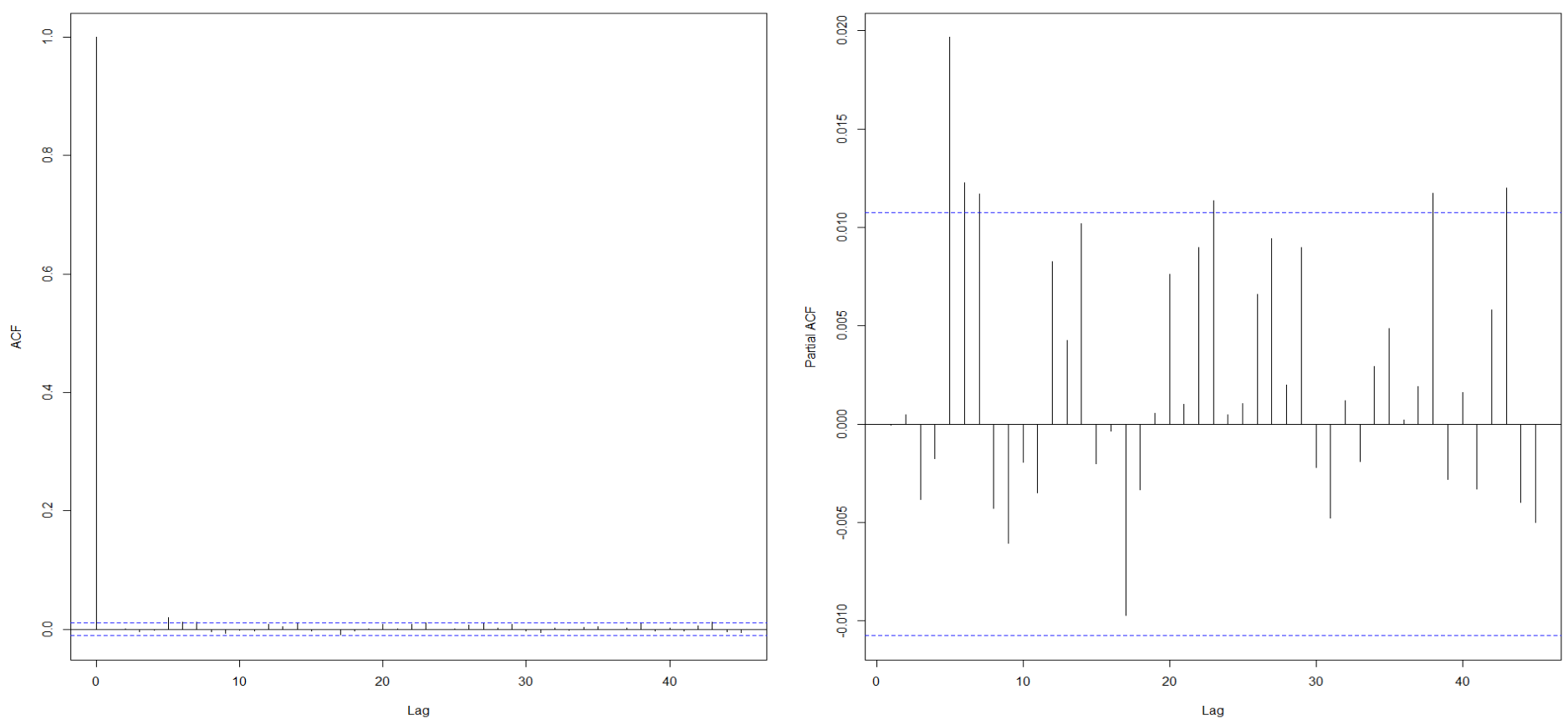


Figure 8‑3: Auto.ARIMA Model– ACF/PACF – Correlograms – Residuals – Daily Data

### Forecast

The Auto.ARIMA model is used to forecast the data for 365 days into the future and the forecasting plot is presented in Figure 8‑4. From Figure 8‑4 the forecast doesn’t seem to be good for this model.

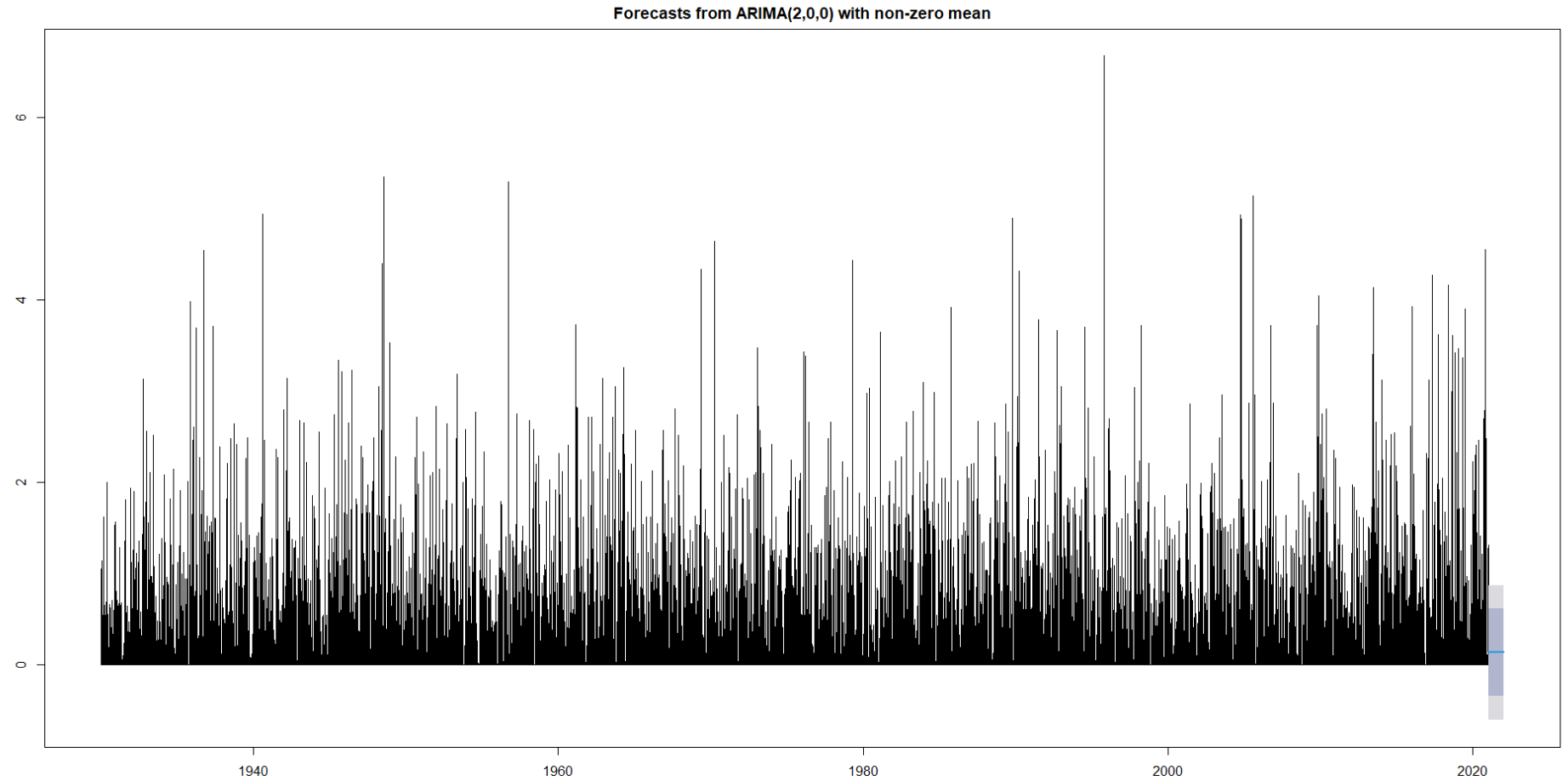


Figure 8‑4: Auto.ARIMA Model– Forecast – 365days – Daily Data

## Monthly Data

### Model Fit

The Auro.ARIMA model is run on the daily rain data and the output is shown in Figure 8‑5 below. The AIC value is high. From Figure 8‑5 the p,d & q values are 1, 0, and 3 respectively. It seems like there is no differencing done. It’s quite a basic model and I don’t think so it’ll have a good fit for the data. But it looks to be a better fit than the daily rain data model.

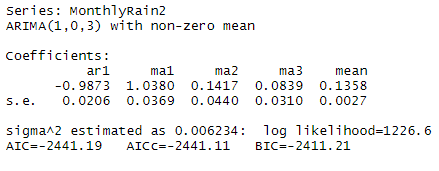


Figure 8‑5: Auto.ARIMA Model – Output – Monthly Data

### Residuals

The residuals plot of the Auto-ARIMA model run on the daily rain data is presented in Figure 8‑6 below. From Figure 8‑6 the residuals suggest that the model could better capture information from the data as such the model is a better fit than the previous model which was run on daily data. Also, from the plot, the difference between observed and fitted values seems to be a little better than the model run on daily data. This might be because there is smoothening happening due to converting the data from daily to monthly.

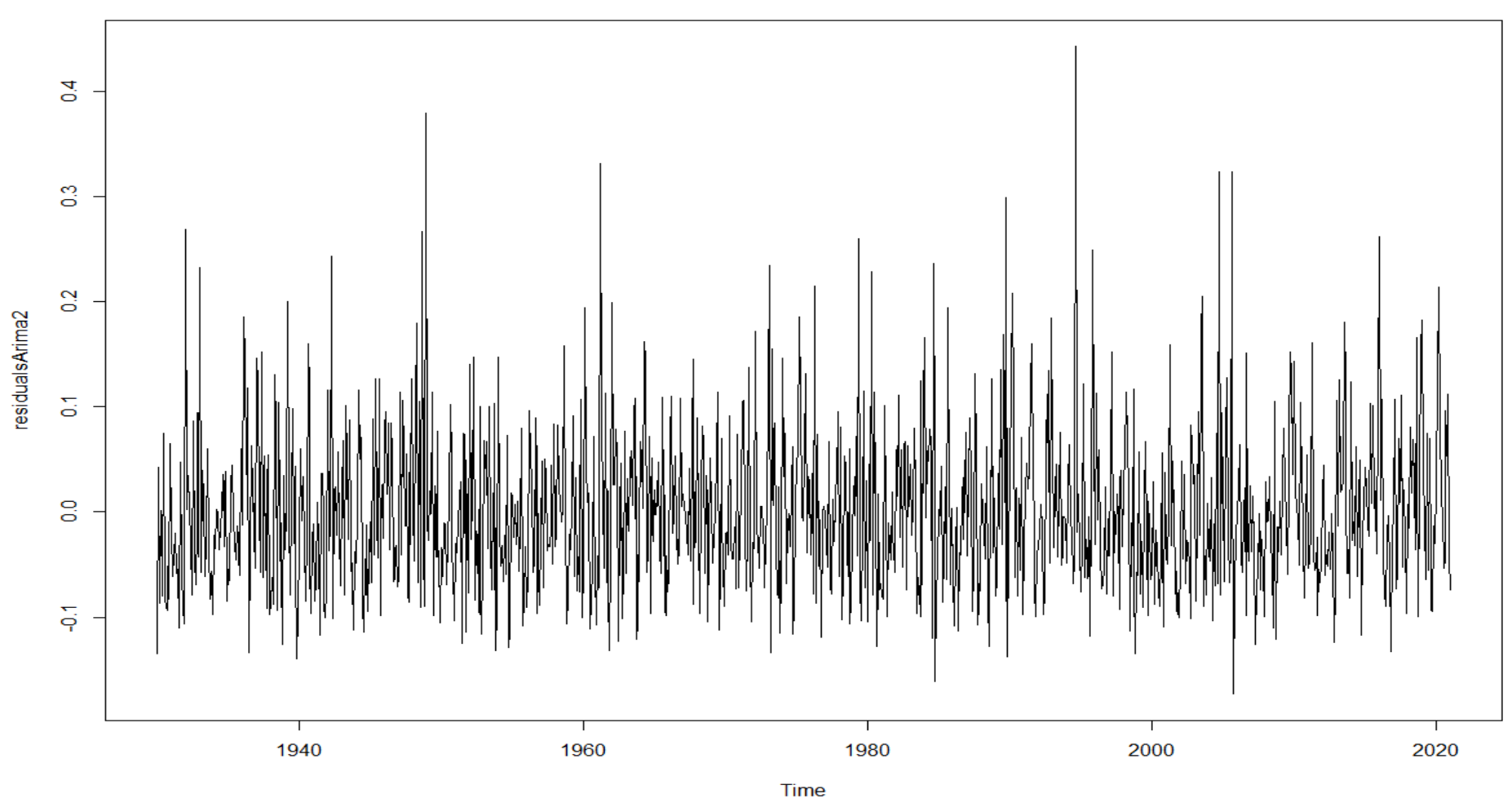


Figure 8‑6: Auto.ARIMA Model – Residual Graph – Monthly Data

### Residuals Correlograms

The ACF and PACF plots of the residuals are presented in Figure 8‑7. From Figure 8‑7 the plots have crossed the blue lines two times suggesting that this is a better model compared to the previous model.

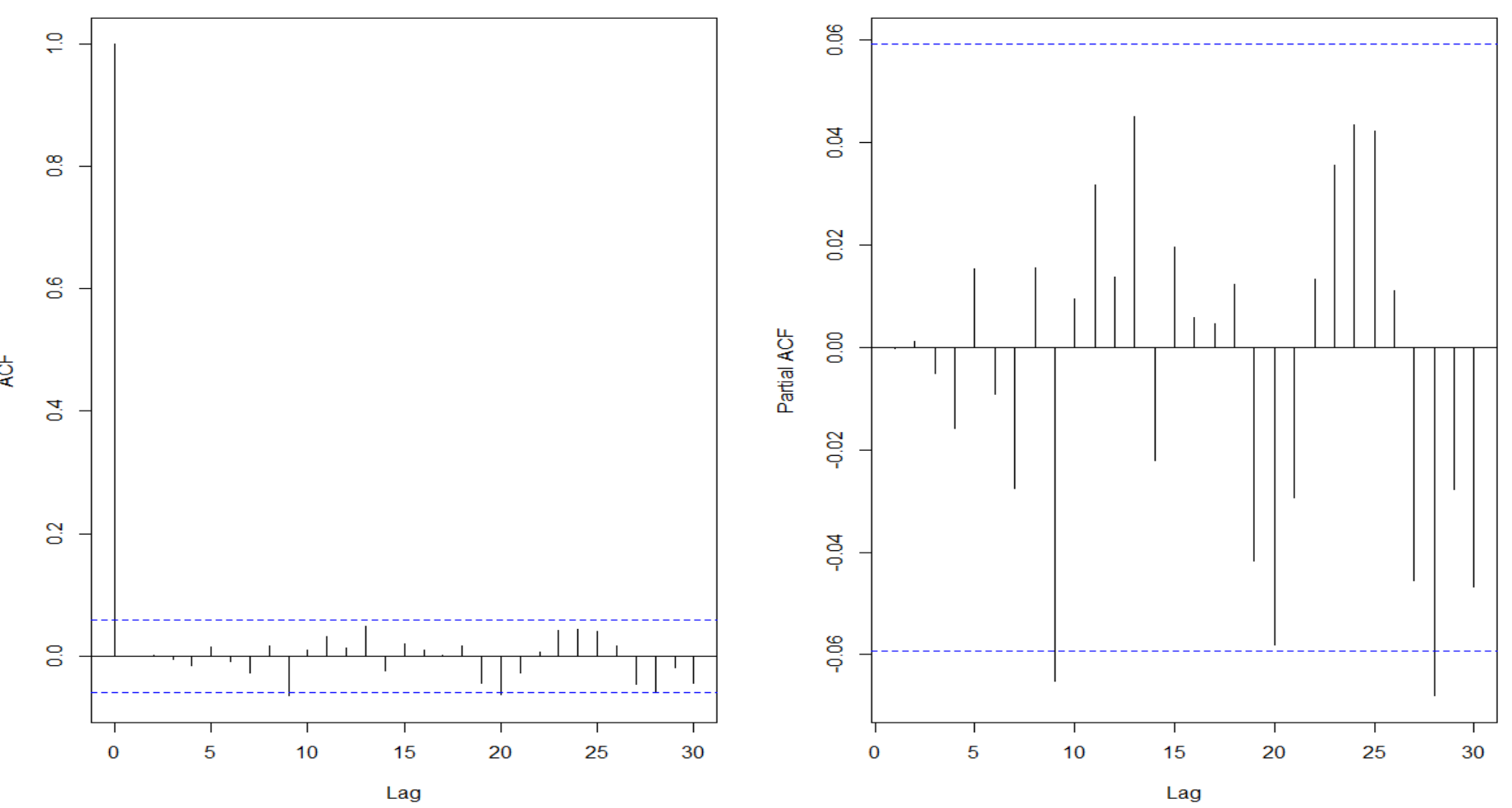


Figure 8‑7: Auto.ARIMA Model– ACF/PACF – Correlograms – Residuals – Monthly Data

### Forecast

The Auto.ARIMA model is used to forecast the data for 12 months into the future and the forecasting plot is presented in Figure 8‑8. From Figure 8‑8 the forecast seems a little better than the previous model.

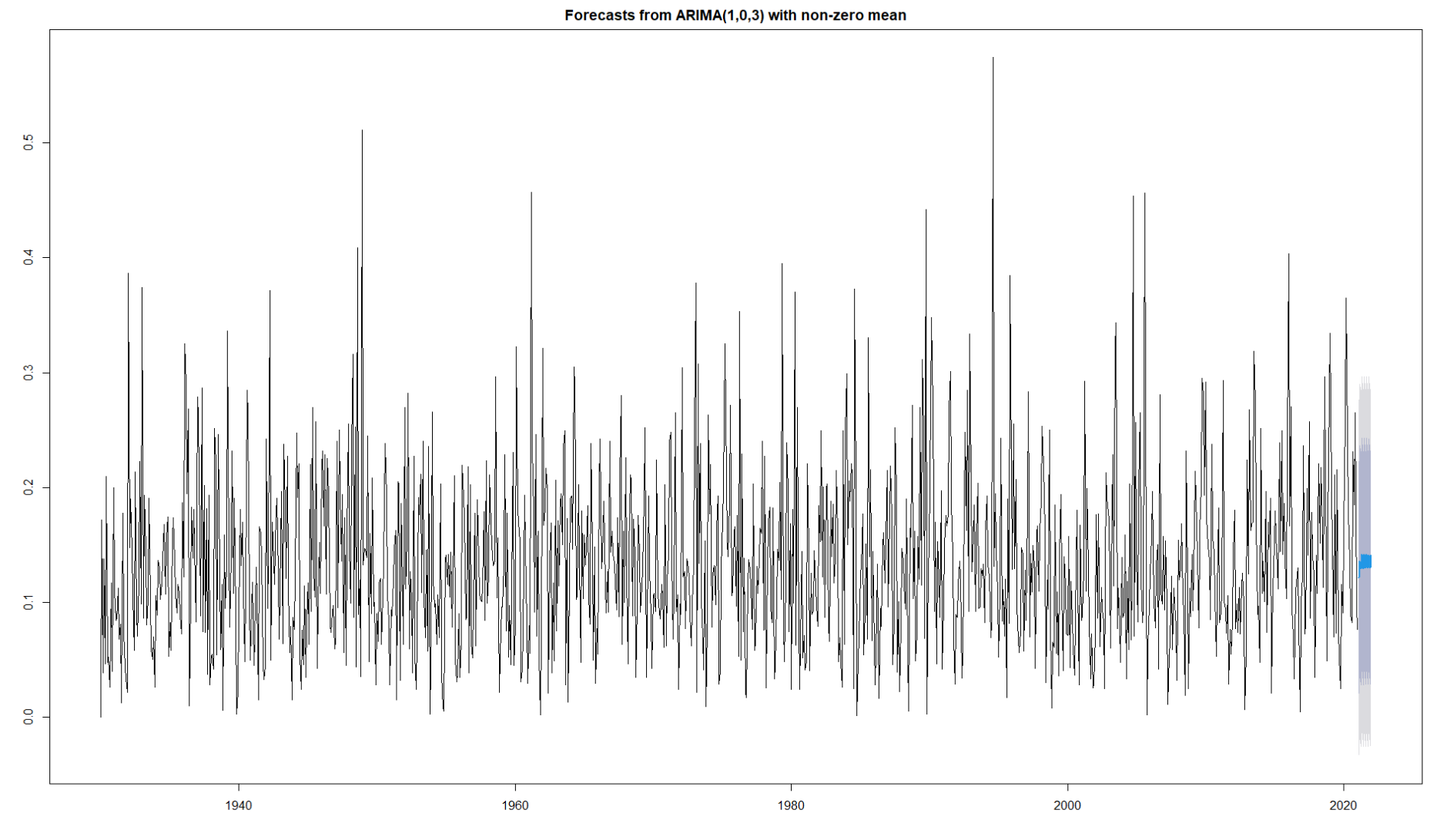


Figure 8‑8: Auto.ARIMA Model– Forecast – 12 Months – Monthly Data

# PARSIMONIOUS MODEL AND PREDICTION - AUTO.ARIMA MODEL

## Model Fit

Several Auro.ARIMA models are run on the monthly data and the input/output for the most parsimonious model is shown in Figure 9‑1 below. The AIC value is high. From Figure 9‑1 the p,d & q values are 5, 2, and 3, and the P, D & Q values are 4, 2, and 0 respectively. It seems like the difference is improving the model as it is making the model stationary and removing seasonality. This model has an AIC of 1701 which improved from the previous one of 2441.



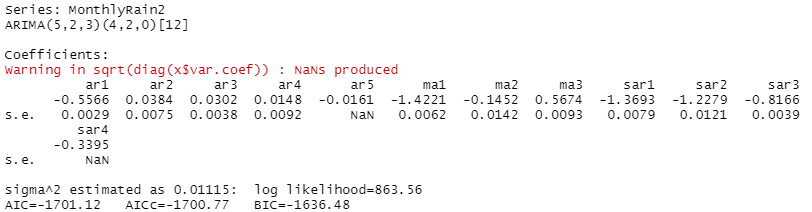


Figure 9‑1: Auto.ARIMA Model – Output – Parsimonious Model

## Residuals

The residuals plot of the Auto-ARIMA model run on the monthly data is presented in Figure 9‑2 below. From Figure 9‑2 the residuals suggest that the model was a better fit than all previous models. Also, from the plot, the difference between observed and fitted values seems to be comparatively very small.

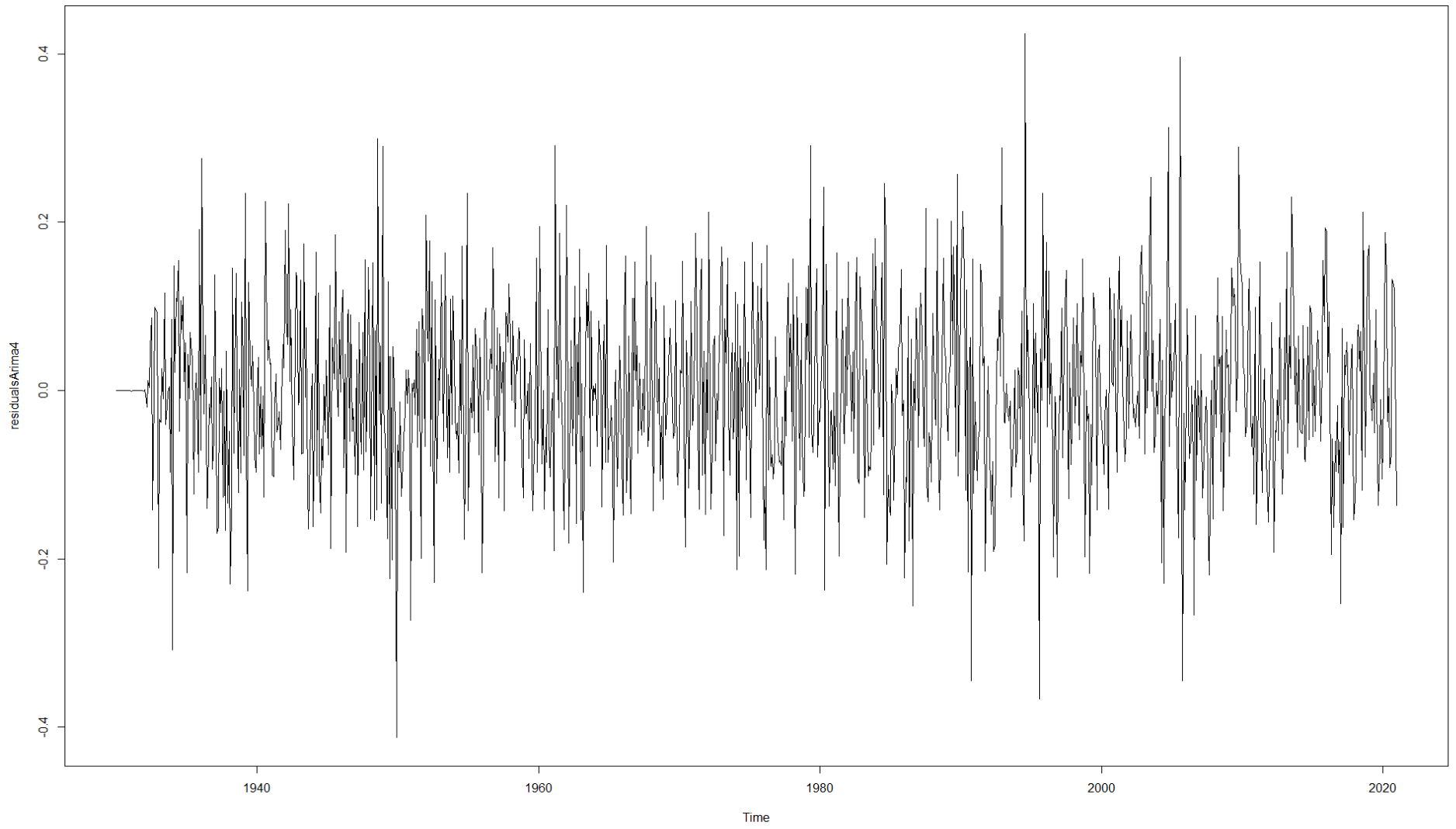


Figure 9‑2: Auto.ARIMA Model – Residual Graph – Parsimonious Model

## Prediction

The Auto.ARIMA model is used to predict the data for 10 months into the future and the predicted values are presented in Figure 9‑3. Figure 9‑4 shows the actual values recorded by NOAA. Therefore from Figure 9‑3 and Figure 9‑4, the predicted values look closer to the actual values.



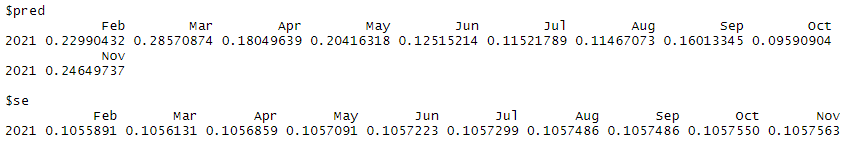


Figure 9‑3: Auto.ARIMA Model– Predict – 12 Months – Parsimonious Model

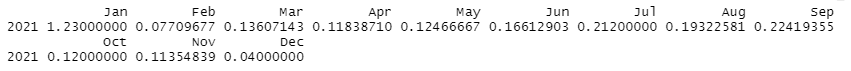


Figure 9‑4: Auto.ARIMA Model– Actual Future Values from NOAA – Parsimonious Model

## Forecast

The Auto.ARIMA model is used to forecast the data for 24 months into the future and the forecasting plot is presented in Figure 9‑5. From Figure 9‑5 the forecast seems a very good fit with nice confidence intervals.

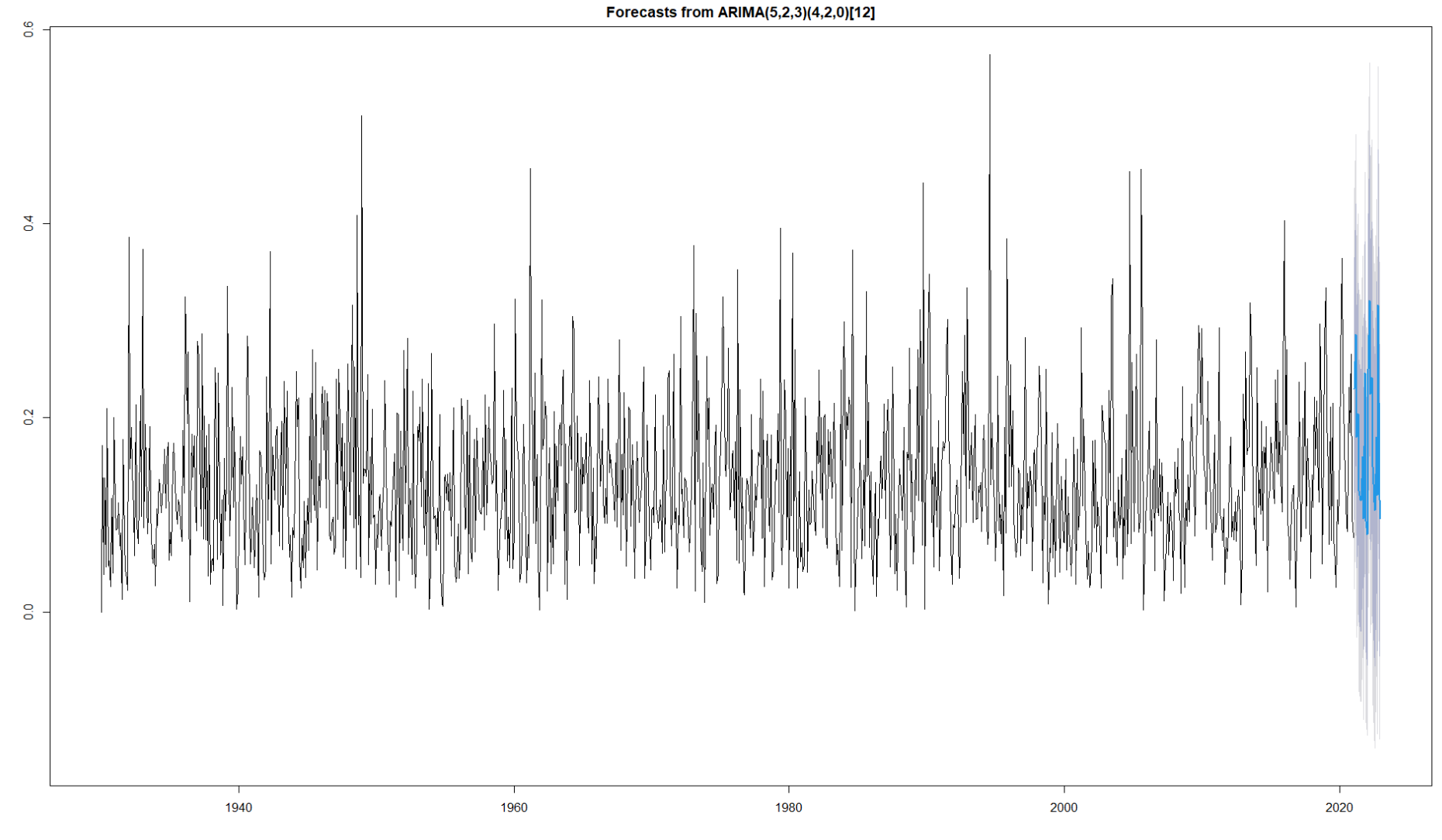


Figure 9‑5: Auto.ARIMA Model– Forecast – 24 Months – Parsimonious Model