Time Series Analysis of Murray Darling Basin River System

ABSTRACT

The Murray-Darling Basin (MDB) has an enormous area in the southeast region of Australia which supports around 3 million human population and is home to around 35 endangered species. MDB basin has two major rivers that are River Darling and the Murray River (Murray Darling Basin Authority [MDBA], 2022e). The Murray River provides approximately 60% of water to Adelaide city (Department for Environment and Water [DEW], 2022). More notably, European settlements in the past have led to an increase in Salinity, Blue-Green Algae, Droughts and Acid Sulfate in Soils (MDBA, 2022f). Due to this, there is a requirement to forecast the Salinity, Water Level and Water Temperature of River Murray to prevent any disastrous situation in future. To accomplish forecasting the above three parameters we need an accurate and reliable model this research project paper focuses on the comparison of two famous forecasting models Auto-Regressive Integrated Moving Average (ARIMA) and Seasonal Auto-Regressive Integrated Moving Average (SARIMA) models. The performance of ARIMA and SARIMA models was tested for Salinity, Water Level and Water Temperature in four locations specifically Biggara (Starting point of Murray River), Albury (Lies near the border of Victoria and New South Wales), Colignan (Lies near the border of Victoria and South Australia), and Murray Bridge (Ending point of Murray River). It was found that the Root Mean Square Error (RMSE) of the ARIMA model was less compared to the SARIMA model which made it a better model for Salinity, Water Level and Water Temperature. However, the ARIMA model had greater MAPE values compared to the SARIMA model which made SARIMA a better model for the above three parameters. To reach a conclusion, the difference values of the RMSE (RMSE ARIMA minus RMSE SARIMA) were compared with the difference values of the MAPE (MAPE ARIMA minus MAPE SARIMA), and this indicated that the difference values of MAPE are far greater than the difference values of RMSE. Thus, the results suggest that SARIMA is the best model.

1. INTRODUCTION

The Murray-Darling Basin (MDB) is an area made up of several intertwined lakes and rivers out of which two main rivers are the River Darling and the Murray River. Furthermore, it provides shelter to 35 threatened species along with 120 waterbird species. More importantly, a human population of approximately 2 million resides in the MDB Basin. Also, MDB Basin generates a revenue of about $11 billion annually due to tourism. A major chunk of agricultural produce comes from the MDB Basin which includes rice, grapes and dairy (Murray Darling Basin Authority [MDBA], 2022e).

Due to European settlements in the early 19th century in the MDB Basin, the industries and people grew in number which led to an increase in water consumption from the MDB River System. This abnormal use of water caused various calamities such as Drought (due to a decrease in Water Levels and River Water Flow along with an increase in Salinity, Water and Air Temperature), High Salinity, Blue-Green Algae and Acid Sulfate Soils. To tackle these issues, the Australian Parliament passed the Water Act in 2007 which made the Murray Darling Basin Authority (MDBA) in charge of maintaining the MDB Basin and making it a healthier working Basin. Moreover, a Basin Plan was also implemented so that everyone can share water in a sustainable manner (MDBA, 2022f).

Therefore, there is a requirement to do a time series analysis of the Murray Darling Basin so that we can observe whether the Water Act and the Basin Plan have improved the condition of water in terms of Salinity, Water Level and Water Temperature. More notably, the time series of Salinity, Water Level and Water Temperature needs to be forecasted to prevent any disastrous situation that is going to occur in future. For this, we would have to find a good model for predicting future values of the above variables.

2. BACKGROUND

Zitian et al. (2022) conducted a research where they examined trends of long-term channel runoff at forty-seven measuring sites present within the southern part of the MDB catchment, Australia. The objective of the research was to estimate the trends of regional channel runoff and simultaneously understand the effect of basin aspects on the spatial variation in channel runoff patterns (Zitian et al., 2022). To accomplish this, Zitian et al. (2022) utilized the Bayesian hierarchical model (BHM) on the catchment aspects and channel runoff records. The results of the study indicated that channel runoff was constantly negative, with a value of up to 2.7% annually relative to the mean flow per year (Zitian et al., 2022).

Tapas and Luke (2019) examined the temporal and spatial drivers and trends of water quality in the River Murray utilizing twenty-four gauging sites' long-term data. It was observed that the water inflow of the Darling tributary caused an increase in pH, total phosphorus and turbidity in the main Murray River (Tapas & Luke 2019). Furthermore, Tapas and Luke (2019) observed that four gauging sites, in particular, displayed a decline in electrical conductivity, colour and nutrients.

Jolly et. al (2001) did an extensive analysis of salination in surface runoff to help in forecasting the rise in the impact and magnitude of salinity in the dryland of the MDB basin. An innovative semi-parametric statistical method that utilized the Generalised Additive Model (GAM) concept was tested in the study (Jolly et. al 2001). The results of the research showed the distribution of salination in surface runoff and indicated that there are four regions of concern (Jolly et. al 2001). Jolly et. al (2001) explain that these terrestrial locations were recognised by conducting spatial analysis of surface runoff salinity patterns and basin salt balances. The region with rainfall of 500-800 mm per year had significant increasing trends and drainage basin salt export/import ratios (Jolly et. al 2001).

3. RESEARCH QUESTIONS

The Research Questions for the Time Series Analysis of the Murray Darling Basin (MDB) River System are:

1) Which model ARIMA (Auto-Regressive Integrated Moving Average) or SARIMA (Seasonal Auto-Regressive Integrated Moving Average) is better for performing a Time-Series Analysis (Forecasting) for Salinity?

2) Which model ARIMA or SARIMA is better for performing a Time-Series Analysis (Forecasting) for Water Level?

3) Which model ARIMA or SARIMA is better for performing a Time-Series Analysis (Forecasting) for Water Temperature?

4. DATA

The available data which would help in the time series analysis of the Murray Darling Basin (MDB) River System is very vast. Thus, it needs to be explained, filtered, and cleaned. Therefore, to accomplish this, the Data Section is divided into Available Data, Data Explanation, Data Parameter Selection Data Location Selection, and Data Cleaning. All these Sub-Sections of Data are elaborated below:

4.1. Available Data and Explanation

The Data which would be utilized in the Time-Series Analysis of the MDB River System is provided by the Murray Darling Basin Authority (MDBA). The Data is retrieved from a web of hydrometric observation sites present at different locations across the MDB River System which are maintained and operated by the MDBA Authority with the help of state governments. The Data which includes calculated flow rates, storage and river levels, various water quality attributes and rainfall is disseminated and recorded with the help of telemetry (Murray Darling Basin Authority [MDBA], 2022a).

The website provided by MDBA Authority for Murray River contains data of key monitoring stations for the public to view and access. The data can be downloaded in Comma Separated Values (CSV) format. The CSV file contains both historic and near-instantaneous (or near-real-time) data which can be utilized for Time Series Analysis. The data that is near-real-time is documented on a six-hourly basis. Furthermore, every time-series data has a related date and timestamp with it (MDBA, 2022a).

4.2. Data Parameter and Location Selection

The MDBA Authority Data explained above has three key data types or variables:

1) Electrical Conductivity (Salinity):

The Electrical Conductivity (E.C) is measured in micro-siemens per centimetre and it gives an indication of the amount of salt in the river. There is a strong correlation between the salinity of the river and the E.C. measure. The advantage of measuring salinity in form of E.C is that E.C can be documented in real-time or continuously (MDBA, 2022a).

2) Gauge Level (Water Level):

The Gauge level measure is utilized to represent the current river water level on a gauge board. These water levels are further utilized to predict the river water flow and reservoir storage volume. The Gauge Level is measured in Australian Height Datum meters (AHD). The Meters AHD is the measurement of ascent from a mean sea level (mean sea level is taken as zero) (MDBA, 2022a).

3) Water Temperature:

The Water Temperature is measured in terms of the degree Celsius, and it varies both seasonally and daily. The dependency of water temperature is simultaneously on both water discharge from reservoirs and surrounding air temperature (MDBA, 2022a).

One of the major concerns of the MDBA Authority is the Salinity of the MDB River System has increased drastically because the groundwater under the southern part of the MDB River has a high concentration of salt (MDBA, 2022b). Another concern of the MDBA Authority is the concentration of Acid Sulphates in the river water which are caused due to lower river water levels. The water near the edges of rivers and lakes can become so acidic that it might start to corrode steel (MDBA, 2022c). Lastly, another concern of the MDBA Authority is the increase in water temperature due to climatic conditions can lead to droughts (MDBA, 2022d). Therefore, the data variables that we chose for time series analysis were Salinity, Water Levels and Water Temperature.

Out of 40 monitoring stations, 4 sites namely Biggara, Albury, Colignan and Murray Bridge were selected for this research (MDBA, 2022e). The Biggara monitoring site was chosen because it is the starting point of the Murray River. Whereas Murray Bridge was selected because it is the point where Murray River ends. Furthermore, Albury was selected because it is at the border of New South Wales and Victoria. While the Colignan was chosen as it is nearer to the border of South Australia and Victoria. To conclude, we chose one ending and starting station with two intermediate sites. These sites are shown in Figure 1.

Figure 1: The four selected locations for Time Series Analysis

4.3. Data Filtering and Imputing

The Data Provided by the MDBA Authority for four locations contained data approximately ranging from the year 1900 to 2022. However, the issue was that the time series data before 2009 was very sparse as it had a lot of missing values. Moreover, because we only wanted to analyse data from the last 12 to 14 years. Therefore, we chose data from 2009 to 2022.

Another task was to impute the missing values present in the time series data from 2009 to 2022. For this Robust Linear Regression (RLR) Model was utilized because it is immune to outliers (John & Sanford, 2013). The values of the time series are fed into the RLR Model which produces a linear curve with the help of M-estimators. The missing values are replaced with the corresponding values on the linear curve. The independent variable is the time stamp whereas Water Salinity, Water Level or Water Temperature are taken as dependent variables.

The "M" in M-estimators stands for maximum likelihood function and M-estimators are of three types i.e., the least square estimator, the Huber estimator, and the Tukey bi-square estimator. The least-square estimator performs badly if the distribution of error or residual values is not normally distributed. Therefore, a combination of Huber and Tukey Bi-square estimator is utilized in RLR Model to overcome this issue (John & Sanford, 2013). The objective and weight functions of the least square estimator, the Huber estimator, and the Tukey bi-square estimator are shown in Figure 2, here "k" is the tuning constant and e is the error term or residual (John & Sanford, 2013).

The data imputation was accomplished with the help of the impute_rlm function in the simputation Library of R Programming Language (Mark, 2022).

Method	*Objective Function*	*Weight Function*
Least-Squares	$\rho_{\mathrm{LS}}(e) = e^2$	$w_{\mathrm{LS}}(e) = 1$
Huber	$\rho_{\mathrm{H}}(e) = \begin{cases} \frac{1}{2}e^2 & \text{for } \lvert e\rvert \le k \\ k\lvert e\rvert - \frac{1}{2}k^2 & \text{for } \lvert e\rvert > k \end{cases}$	$w_{\mathrm{H}}(e) = \begin{cases} 1 & \text{for } \lvert e\rvert \le k \\ k/\lvert e\rvert & \text{for } \lvert e\rvert > k \end{cases}$
Bisquare	$\rho_{\mathrm{B}}(e) = \begin{cases} \frac{k^2}{6}\left\{1 - \left[1 - \left(\frac{e}{k}\right)^2\right]^3\right\} & \text{for } \lvert e\rvert \le k \\ k^2/6 & \text{for } \lvert e\rvert > k \end{cases}$	$w_{\mathrm{B}}(e) = \begin{cases} \left[1 - \left(\frac{e}{k}\right)^2\right]^2 & \text{for } \lvert e\rvert \le k \\ 0 & \text{for } \lvert e\rvert > k \end{cases}$

Figure 2: The Objective and Weight Functions of different M-estimators

5. METHODOLOGY

For the Time Series Analysis, various factors come into play such as data stationarity, types of models used, and best model selection. Therefore, the Methodology is divided into five parts namely Time Series Data Decomposition, Box-Cox Transformation, Data Stationarity Tests, Autoregressive Integrated Moving Average (ARIMA) Model and Seasonal Autoregressive Integrated Moving Average (SARIMA) Model. These sections are elaborated below:

5.1. Data Decomposition

The Time Series Data (y_t) can be broken up or decomposed into three components specifically Seasonal (S_t), Trend-Cycle (T_t) and Remainder (R_t) components. There are two types of decomposition:

1) Additive Decomposition:

   In Additive Decomposition, the components are broken in such a way that the original time series can be formed by just adding up the Seasonal, Trend-Cycle and Remainder components. The additive decomposition is best suitable when the fluctuation around the trend cycle or seasonal pattern seems to be non-proportional to the level of the time series (Rob & George, 2018a). The formula of additive decomposition is given below in Figure 3.

$$\boxed{y_t = S_t + T_t + R_t}$$

Figure 3: The formula for Additive Decomposition

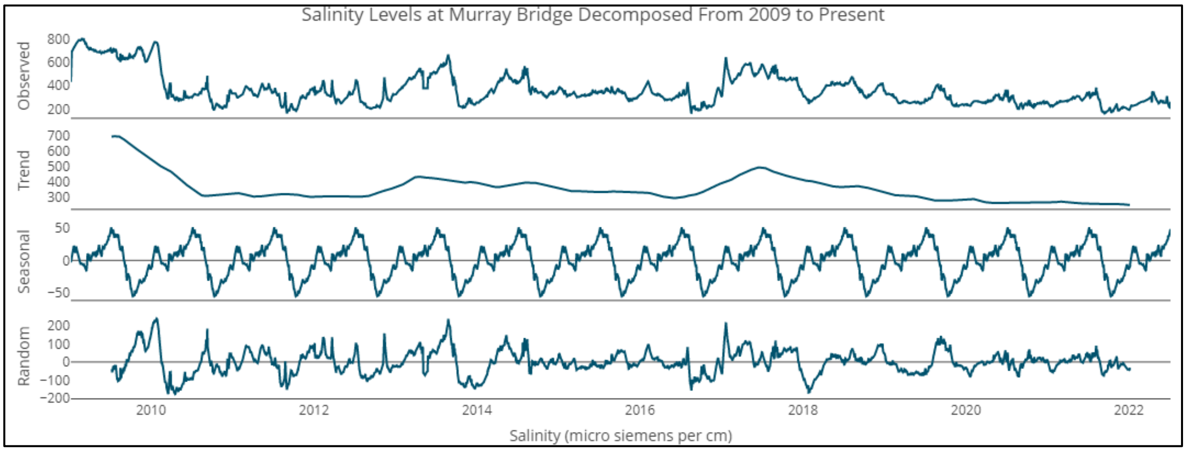

Figure 4: Decomposition of Salinity Time Series at Murray Bridge

2) Multiplicative Decomposition:

   In Multiplicative Decomposition, the components are broken in such a way that the original time series can be formed by just multiplying the Seasonal, Trend-Cycle and Remainder components. The multiplicative decomposition is best suitable when the fluctuation around the trend cycle or seasonal pattern seems to vary with the level of the time series (Rob & George, 2018a). The formula of multiplicative decomposition is given below in Figure 5.

$$\boxed{y_t = S_t \times T_t \times R_t}$$

Figure 5: The formula for Multiplicative Decomposition

The decomposition of time series is utilized to make series stationery by subtracting seasonal and trend cycle components from the main time series. The stationarity of the time series will be explained in the Data Stationarity Tests Section. To implement the Additive or Multiplicative Decomposition decompose Function of stats Library in R Programming Language (Sebastien, 2021).

5.2. Box-Cox Transformation

The Box-Cox Transformation comes into use when the time series data shows fluctuations that decrease or increase with the level of series. The Box-Cox Transformation is made up of two transformations particularly Power and Logarithmic Transformations and it is dependent on the λ parameter (Box & Cox, 1964). In Logarithmic Transformation, the original observed values y_t are transformed to w_t values such that $w_t = \log(y_t)$. Furthermore, in Power Transformation the original observed values y_t are transformed to w_t values such that $w_t = (y_t)^p$ (Rob & George, 2018a). The formula of the Box-Cox Transformation is given in Figure 6.

$$\boxed{w_t = \begin{cases} \log(y_t) & \text{if } \lambda = 0; \\ (y_t^\lambda - 1)/\lambda & \text{otherwise.} \end{cases}}$$

Figure 6: Formula of the Box-Cox Transform

The logarithm being used in the Box-Cox Transformation always has a base e (i.e., natural logarithm). Therefore, if the value of $\lambda = 0$, then natural logarithm transformations are utilized and if the value of $\lambda \neq 0$, power transformations are utilized with $1/\lambda$ as the scaling factor. Moreover, if $\lambda = 1$, then the Formula in the Figure 6 becomes $w_t = y_t - 1$, thus the times series data is just shifted downwards without any change in the original time series (Rob & George, 2018a).

The Box-Cox Transformation is performed with the help of the BoxCox() method of forecast Library in R Programming Language (Hyndman et. al, 2022). The BoxCox() method utilizes a slightly modified version of the Box-Cox Transformation which is explained in Bickel & Doksum (1981), this enables the use of negative values of y_t when $\lambda > 0$. The formula is given in Figure 7.

$$w_t = \begin{cases} \log(y_t) & \text{if } \lambda = 0; \\ \text{sign}(y_t)(|y_t|^\lambda - 1)/\lambda & \text{otherwise.} \end{cases}$$

Figure 7: Formula of modified Box-Cox Transform

The Inverse of Box-Cox Transformation can also be performed to get the values of the original scale time series (Rob & George, 2018a). The formula is provided in Figure 8.

$$y_t = \begin{cases} \exp(w_t) & \text{if } \lambda = 0; \\ \text{sign}(\lambda w_t + 1)|\lambda w_t + 1|^{1/\lambda} & \text{otherwise.} \end{cases}$$

Figure 8: Formula of Inverse of Box-Cox Transform

The Box-Cox Transform reduces the variance in time series data which in turn helps in making the series more stationary.

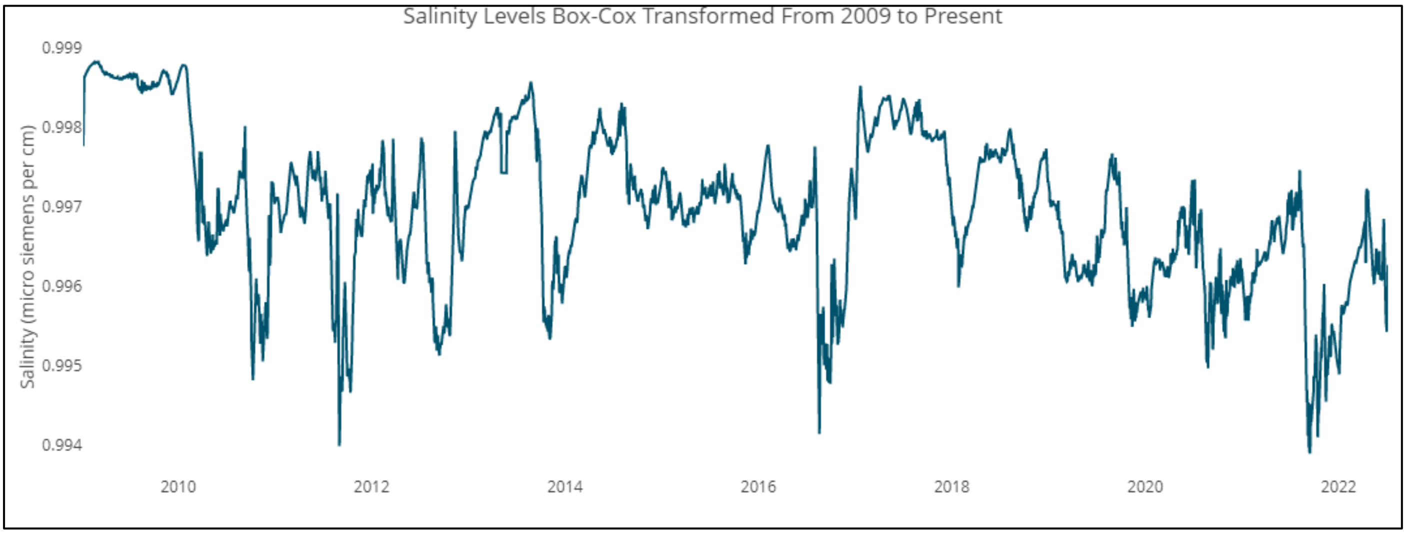

Figure 9: Box-Cox Transformation of Salinity Time-Series at Murray Bridge

5.3. Data Stationarity Tests

The time series data comes in two forms specifically Stationary and Non-Stationary. A time series whose statistical characteristics are independent of time is called a stationary time series. More specifically, if we suppose a stationary time series$\{y_t\}$, then for every s, the distribution $(y_t, \ldots, y_{t+s})$ is independent or does not dependent on t. Therefore, time series data with seasonality or trend-cycle is considered to be non-stationary time series. Whereas white noise or random time series is considered a stationary time series (Rob & George, 2021a).

The time series can be made stationary by subtracting seasonal and trend cycle components from the original time series which is explained in detail in Data Decomposition Section. Furthermore, time series can also be made stationary by performing a Box-Cox Transformation on the original time series to make the variance constant which is explained in detail in Box-Cox Transformation Section.

However, sometimes it is difficult to know whether a time series is stationary or non-stationary just by observing the time series' pattern. To overcome this problem several Stationary Tests are utilized which are explained below:

1) Augmented Dickey-Fuller test (ADF Test):

   The Augmented Dickey-Fuller test is an Autoregressive Unit Root Test that is utilized to measure the presence of unit root in the time series data. The Null Hypothesis of the ADF Test is that the time series has a unit root. While the Alternative Hypothesis is that the time series has no unit root. Hence, if the null hypothesis is rejected then there is strong evidence that the time series is non-stationary (Statsmodels, 2022; Eric, 2014).

   The ADF test is done by utilizing the adf.test function of the aTSA Library in R Programming Language (Debin, 2015).

2) Kwiatkowski-Phillips-Schmidt-Shin (KPSS):

   The Kwiatkowski-Phillips-Schmidt-Shin test is a Stationarity Test that is utilized to check whether the time series data is stationary or not. The Null Hypothesis of the KPSS Test is that the time series has a trend stationary. While the Alternative Hypothesis is that the time series has a unit root. Hence, if the null hypothesis is rejected then there is strong evidence that the time series is non-stationary and has a unit root which is the opposite when compared to ADF (Statsmodels, 2022; Piotr & Gabriel, 2016).

   The KPSS test is done by utilizing the kpss.test function of the aTSA Library in R Programming Language (Debin, 2015).

5.4. ARIMA Model

To understand the Auto-Regressive Integrated Moving Average (ARIMA) models, we first need to understand the Auto-Regressive (AR) and Moving Average Models (MA). These Models are Explained below:

1) Auto-Regressive (AR) Models:

   In Auto-Regressive Models, we predict the future values of a variable by utilizing a linear combination of past values of the variable. The term Auto-Regressive means that the linear regression is being performed on its own values (Rob & George, 2018b). The formula of the autoregressive model is shown in Figure 10.

$$\boxed{y_t = c + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \cdots + \phi_p y_{t-p} + \varepsilon_t,}$$

Figure 10: Formula of AR Model

   In the above formula, ε_t is white noise, and the y_{t-1} to y_{t-p} are lagged values of y_t. We represent the AR Model by AR(p) (Rob & George, 2018b).

2) Moving Average (MA) Models:

   Unlike AR Model, the Moving Average Model conducts a linear regression of past forecast error values i.e., the future values are predicted with help of past error values of the forecast (Rob & George, 2018c). The formula of the moving average model is shown in Figure 11.

$$\boxed{y_t = c + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \cdots + \theta_q \varepsilon_{t-q},}$$

Figure 11: Formula of MA Model

In the above formula, ε_t is white noise, and the ε_{t-1} to ε_{t-p} are lagged values of ε_t. We represent the MA Model by MA(q) (Rob & George, 2018c).

Another major aspect of ARIMA apart from AR and MA models is differencing term which is represented as I(d). The differencing is a method in which the time-lagged series of the original time series is subtracted from the original time series to get a stationary series or differenced series (Rob & George, 2021b).

Now, if we concatenate the Auto-Regressive and Moving Average Model with a differencing term we get a nonseasonal ARIMA Model. The formula of the ARIMA Model is given below in Figure 12.

$$y'_t = c + \phi_1 y'_{t-1} + \cdots + \phi_p y'_{t-p} + \theta_1 \varepsilon_{t-1} + \cdots + \theta_q \varepsilon_{t-q} + \varepsilon_t,$$

Figure 12: Formula of ARIMA Model

Here, $(y_t)'$ represents the differenced series. Furthermore, the terms on the left-hand side of the equation include both lagged error and auto terms. The ARIMA model is represented as ARIMA(p, d, q) (Rob & George, 2021c).

The ARIMA Model is implemented by the Arima and auto.arima functions of forecast library in R Programming Language (Hyndman et. al, 2022).

Now, the selection of p, q and d terms in the ARIMA Model requires the analysis of Auto-Correlation Function (ACF) and Partial Auto-Correlation Function (PACF) plots along with the Box-Ljung test, Akaike's Information Criterion (AIC), Corrected Akaike's Information Criterion (AICc), Schwarz's Bayesian Information Criterion (BIC) and Residual plots. These all are explained below:

1) Auto-Correlation Function (ACF) plot:

   Correlation measures the amount of linear relationship between 2 variables. Similarly, autocorrelation measures the extent of the linear relationship between the original time series and its lagged version (Rob & George, 2018d). The formula of autocorrelation coefficients is given in Figure 13.

$$r_k = \frac{\sum\limits_{t=k+1}^{T} (y_t - \bar{y})(y_{t-k} - \bar{y})}{\sum\limits_{t=1}^{T} (y_t - \bar{y})^2},$$

   Figure 13: Formula of autocorrelation coefficients

   Here, r_k is the autocorrelation of k times lagged time series and T is the length of time series. The plot of these autocorrelation coefficients shows the Auto Correlation Function (ACF) (Rob & George, 2018d).

   ARIMA(p,d,0) model is selected when the ACF of the time series is sinusoidal or exponentially dampening and there is a measurable spike at lag p in the plot of PACF, however none after the lag p (Rob, 2017).

2) Partial Auto-Correlation Function (PACF) plot:

   Partial Autocorrelation is a conditional autocorrelation in which the relationship between y_t and y_{t-k} is measured by removing the effects of other time lags – 1, 2, 3, …, k-1 (Rob, 2017).

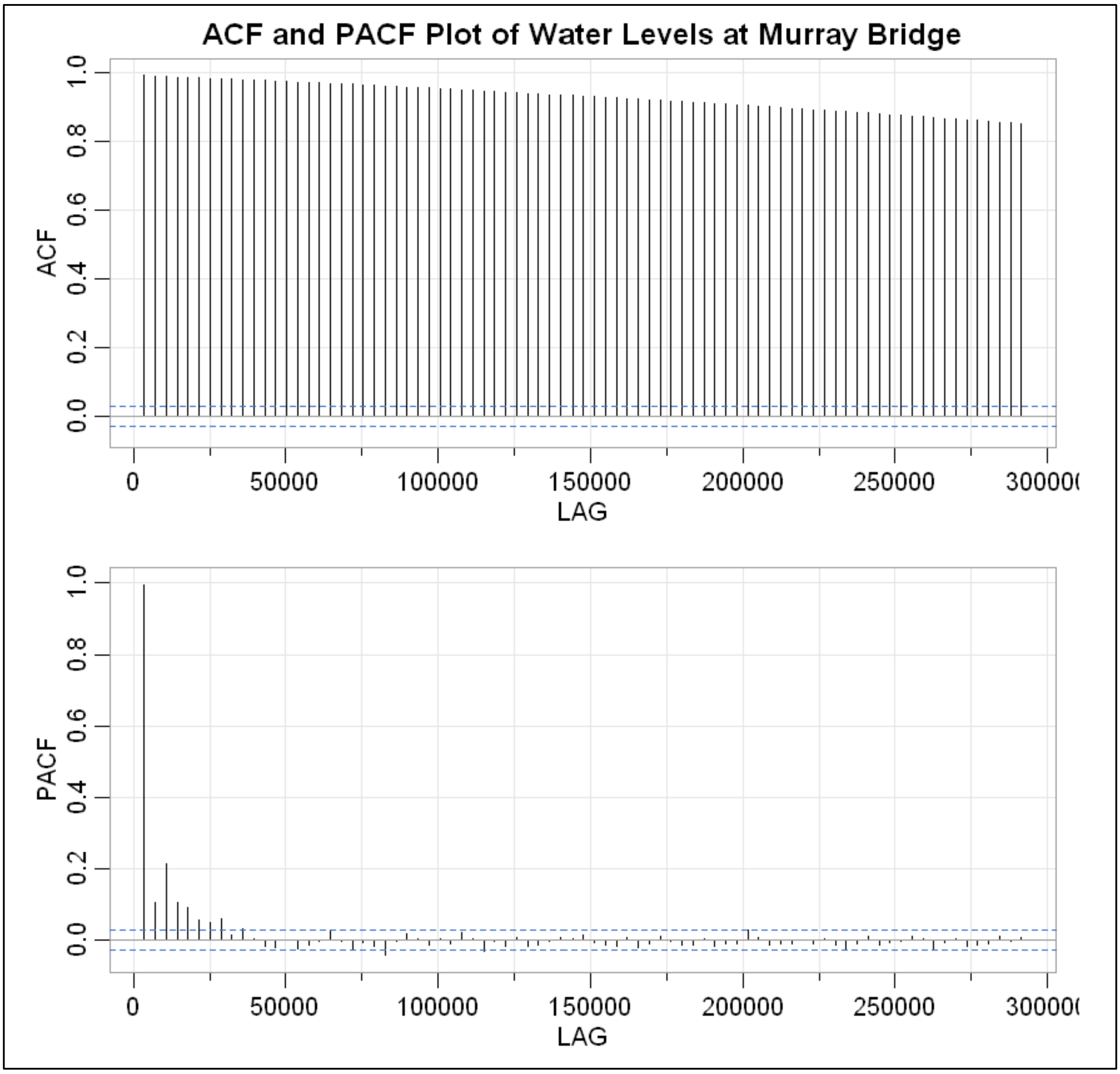

Figure 14: ACF and PCF plot of Water Level at Murray Bridge

ARIMA(0,d,q) model is selected when the PACF of the time series is sinusoidal or exponentially dampening and there is a measurable spike at lag q in the plot of ACF, however none after the lag q (Rob, 2017).

3) Box-Ljung test:

Box-Ljung is a Portmanteau test where a more professional test for autocorrelation is conducted to check whether the residuals of a model have any autocorrelation or not (Rob & George, 2021e). The formula of the Box-Ljung is given in Figure 15.

$$Q^* = T(T+2)\sum_{k=1}^{\ell}(T-k)^{-1}r_k^2.$$

Figure 15: Formula of Box-Ljung Test

```
        Ljung-Box test

data:  Residuals from ARIMA(3,0,1) with zero mean
Q* = 765.4, df = 726, p-value = 0.1507

Model df: 4.   Total lags used: 730
```

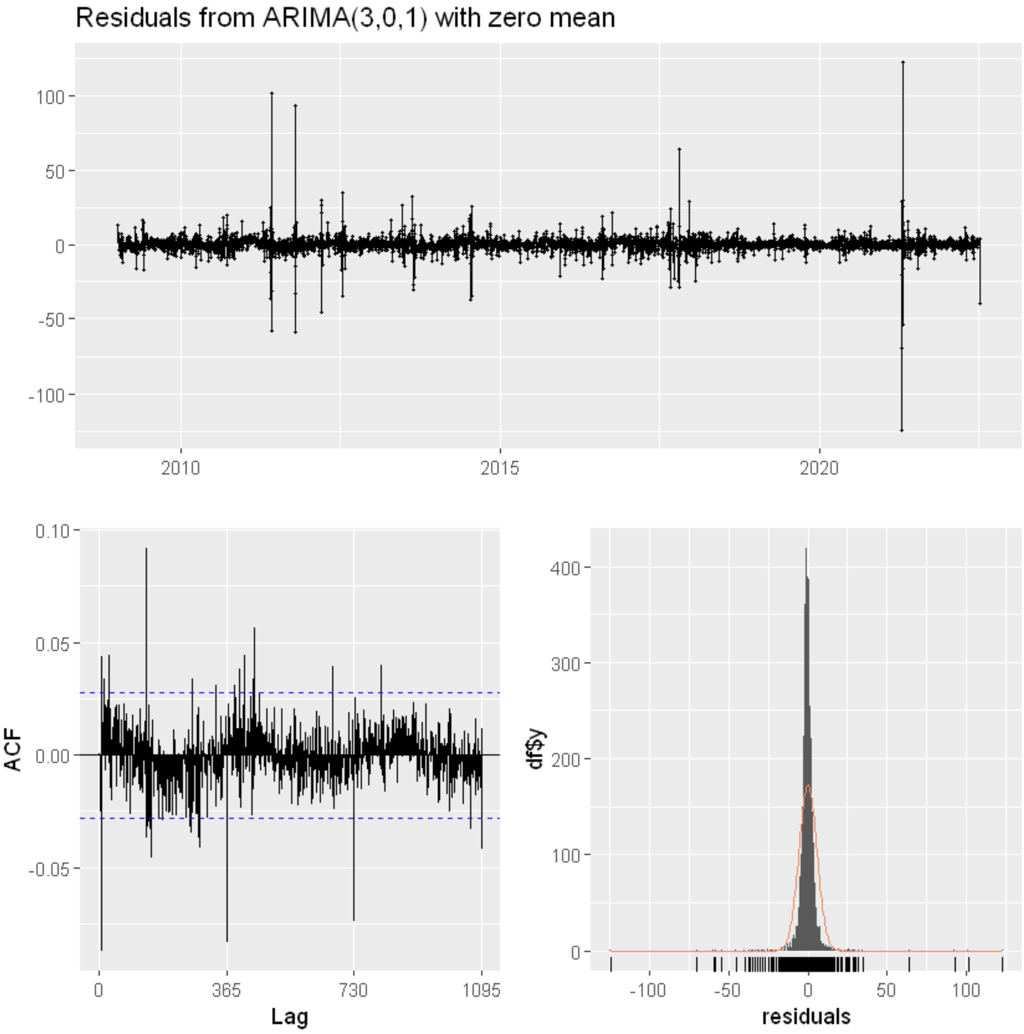

Figure 16: Ljung-Box test with Residuals for Salinity Levels at Colignan

Here, l is the largest lag considered for autocorrelations, r_k is the autocorrelation for lag k and T is the number of observations. Furthermore, a large value of Q^* suggests that the autocorrelations are just a White Noise Time-Series. A Large value here means that the Q^* will have a χ^2 distribution (Rob & George, 2021e).

4) Akaike's Information Criterion (AIC):

The Akaike's Information Criterion is a measure of predictive accuracy to check which model might be the best (Rob & George, 2021d). The formula of AIC is given in Figure 17.

$$\mathrm{AIC} = T \log\left(\frac{\mathrm{SSE}}{T}\right) + 2(k+2),$$

Figure 17: Formula of AIC

Here SSE is the minimum sum of squared errors, T is the number of observations for prediction, and the k is the coefficients for the estimators. The main idea behind AIC is to chastise the fit of the model with the number of parameters that are required to be predicted. The model with the minimum value of AIC is often considered the best (Rob & George, 2021d).

5) Corrected Akaike's Information Criterion (AICc):

   The Corrected Akaike's Information Criterion is utilized for small values of T because AIC selects a lot of estimators than required. Therefore, AICc is the bias-corrected version of AIC. Also, the model with a minimum value of AICc is the best model (Rob & George, 2021d). The Formula is given in Figure 18.

$$\text{AIC}_c = \text{AIC} + \frac{2(k+2)(k+3)}{T-k-3}.$$

Figure 18: Formula of AICc

6) Schwarz's Bayesian Information Criterion (BIC):

   Schwarz's Bayesian Information Criterion is similar to AIC. The only difference is that it chastises the fit of the model with the number of parameters that are required to be predicted more heavily. Moreover, the best model is the one with a minimum value of BIC (Rob & George, 2021d). The formula of BIC is given in Figure 19.

$$\text{BIC} = T\log\left(\frac{\text{SSE}}{T}\right) + (k+2)\log(T).$$

Figure 19: Formula of BIC

All the above parameters specifically ACF, PCF, Box-Ljung Test, AIC, AICc and BIC are implemented by the Acf, checkresiduals and glance methods of forecast Library in R Programming Language (Hyndman et. al, 2022).

5.5. SARIMA Model

A Seasonal Autoregressive Integrated Moving Average (SARIMA) is an ARIMA Model that accepts seasonal data. Furthermore, SARIMA is created by adding certain seasonal terms in ARIMA Model (Rob & George, 2018e). The representation of the SARIMA Model is shown in Figure 20.

$$\text{ARIMA} \quad \underbrace{(p,d,q)}_{\substack{\uparrow \\ \text{Non-seasonal part} \\ \text{of the model}}} \quad \underbrace{(P,D,Q)_m}_{\substack{\uparrow \\ \text{Seasonal part of} \\ \text{of the model}}}$$

Figure 20: Representation of SARIMA Model

Here, m is the number of observations annually. The terms present in the seasonal part shown in the Figure 20 are similar to the Non-Seasonal Terms, but the former terms perform the backshift of the seasonal period (Rob & George, 2018e). For instance, let us consider an ARIMA(1,1,1)$(1,1,1)_4$ model for quarterly data is shown in Figure 21.

$$(1-\phi_1 B)\ (1-\Phi_1 B^4)(1-B)(1-B^4)y_t = (1+\theta_1 B)\ (1+\Theta_1 B^4)\varepsilon_t.$$

Figure 21: ARIMA(1,1,1)(1,1,1)$_4$ model

Here, B is a Backshift Operator which lags a variable depending on its power; for example, $B^2\ x_t = x_{t-2}$ (PennState, 2022).

The SARIMA is implemented by Sarima and auto.sarima methods in bayesforecast Library in R Programming Language (Matamoros et al.,2021). The ACF, PCF, Box-Ljung Test, AIC, AICc and BIC explained in the ARIMA Model Section are applied in the SARIMA Model in the exact same way as in the ARIMA Model.

An ARIMA(0,0,0)(1,0,0)$_{12}$ will be utilized when the seasonal lags of the ACF plots dampen exponentially, and there is a single measurable spike at the 12th time lag in the plot of PACF (Rob & George, 2018e). An ARIMA(0,0,0)(0,0,1)$_{12}$ will be utilized when the seasonal lags of the PACF plots dampen exponentially, and there is a measurable spike at the 12th time lag in the plot of ACF (Rob & George, 2018e).

6. EXPLANATORY DATA ANALYSIS

For Explanatory Data Analysis different ARIMA and SARIMA Models were trained and tested to get the best models for time series analysis (forecasting). All the tested and trained ARIMA and SARIMA models are given in Tables 9 to 32 in the Appendix Section of this Report. Moreover, the Forecast of ARIMA along with SARIMA, ACF and PACF plots are also given in the Appendix Section in Figures 22 to 57.

The Explanatory Data Analysis Section has two parts i.e., Results and Discussion. In the Results Section, we have tables displaying the accuracy of the best ARIMA and SARIMA Models. Whereas, in the Discussion Section we will discuss the results we have received. These Sections are explained below in detail:

6.1. Results

The Root Mean Square Error (RMSE) and Mean Absolute Percentage Error (MAPE) values of the best ARIMA Models for Salinity, Water Level and Water Temperature at Biggara, Albury, Colignan and Murray Bridge are given below:

Table 1: RMSE and MAPE Score of Best Models for Salinity, Water Level and Water Temperature at Biggara

Location	Parameter	Best ARIMA Model	Model Accuracy (RMSE)	Model Accuracy (MAPE)
Biggara	Salinity	(4,0,1)	27.34155	157.7315
	Water Level	(3,0,2)	0.09440942	157.7211
	Water Temperature	(1,0,4)	0.9167563	515.5143

Table 2: RMSE and MAPE Score of Best Models for Salinity, Water Level and Water Temperature at Albury

Location	Parameter	Best ARIMA Model	Model Accuracy (RMSE)	Model Accuracy (MAPE)
Albury	Salinity	(2,0,3)	2.908709	192.4262
	Water Level	(5,0,0)	0.07970485	84.67822
	Water Temperature	(1,0,4)	0.4027536	147.1829

Table 3: RMSE and MAPE Score of Best Models for Salinity, Water Level and Water Temperature at Colignan

Location	Parameter	Best ARIMA Model	Model Accuracy (RMSE)	Model Accuracy (MAPE)
Colignan	Salinity	(3,0,1)	5.624947	90.96274
	Water Level	(1,0,4)	0.06132232	38.26355
	Water Temperature	(3,0,1)	0.4024886	218.4029

Table 4: RMSE and MAPE Score of Best Models for Salinity, Water Level and Water Temperature at Murray Bridge

Location	Parameter	Best ARIMA Model	Model Accuracy (RMSE)	Model Accuracy (MAPE)
Murray Bridge	Salinity	(3,0,2)	9.154455	225.77
	Water Level	(1,0,2)	0.04572144	254.9606
	Water Temperature	(3,0,2)	0.3442569	109.1723

The Root Mean Square Error (RMSE) and Mean Absolute Percentage Error (MAPE) values of the best SARIMA Models for Salinity, Water Level and Water Temperature at Biggara, Albury, Colignan and Murray Bridge are given below:

Table 5: RMSE and MAPE Score of Best Models for Salinity, Water Level and Water Temperature at Biggara

Location	Parameter	Best SARIMA Model	Model Accuracy (RMSE)	Model Accuracy (MAPE)
Biggara	Salinity	(1,0,2)(0,1,0)[365]	40.91339	17.86406
	Water Level	(2,1,2)(0,1,0)[365]	0.1357823	8.165705
	Water Temperature	(1,0,2)(0,1,0)[365]	1.155269	5.935066

Table 6: RMSE and MAPE Score of Best Models for Salinity, Water Level and Water Temperature at Albury

Location	Parameter	Best SARIMA Model	Model Accuracy (RMSE)	Model Accuracy (MAPE)
Albury	Salinity	(1,1,2)(0,1,0)[365]	3.9286	3.922016
	Water Level	(2,1,2)(0,1,0)[365]	0.1122539	4.528721
	Water Temperature	(2,0,2)(0,1,0)[365]	0.5586905	2.935275

Table 7: RMSE and MAPE Score of Best Models for Salinity, Water Level and Water Temperature at Colignan

Location	Parameter	Best SARIMA Model	Model Accuracy (RMSE)	Model Accuracy (MAPE)
Colignan	Salinity	(0,1,1)(0,1,0)[365]	8.327019	3.441338
	Water Level	(3,1,5)(0,1,0)[365]	0.06660298	Infinite
	Water Temperature	(1,0,5)(0,1,0)[365]	0.532371	1.922765

Table 8: RMSE and MAPE Score of Best Models for Salinity, Water Level and Water Temperature at Murray Bridge

Location	Parameter	Best SARIMA Model	Model Accuracy (RMSE)	Model Accuracy (MAPE)
Murray Bridge	Salinity	(2,1,2)(0,1,0)[365]	12.3331	2.282056
	Water Level	(1,1,3)(0,1,0)[365]	0.06121044	7.000789
	Water Temperature	(2,0,2)(0,1,0)[365]	0.4468404	1.173231

6.2. Discussion

There were seven research questions asked which we would be answering by analysing Tables 1 to 8 above. Each of the Research Questions is answered and discussed below in detail:

1) Research Question 1 (Salinity Forecast):

   The RMSE values of the Salinity ARIMA Model of four locations (Biggara: 27.34, Albury: 2.91, Colignan: 5.62, Murray Bridge: 9.15) were less than the RMSE Values of the Salinity SARIMA Model of four locations (Biggara: 40.91, Albury: 3.92, Colignan: 8.32, Murray Bridge: 12.33) which indicates that ARIMA Model has outperformed SARIMA Model.

   However, when comparing MAPE values of the Salinity ARIMA Model of four locations (Biggara: 157.73, Albury: 192.43, Colignan: 90.96, Murray Bridge: 225.77) to MAPE values of the Salinity SARIMA Model of four locations (Biggara: 17.86, Albury: 3.92, Colignan: 3.44, Murray Bridge: 2.28) the MAPE values of SARIMA are less compared to MAPE values of ARIMA. Thus, the SARIMA Model performs better than ARIMA Model.

   Now, if we compare the difference in RMSE values between ARIMA and SARIMA Models with the difference in MAPE values between ARIMA and SARIMA Models, we observe that difference in MAPE values is far greater than that of RMSE. Therefore, the best Model is SARIMA Model.

2) Research Question 2 (Water Level Forecast):

   The RMSE values of the Water Level in the ARIMA Model of four locations (Biggara: 0.09, Albury: 0.07, Colignan: 0.06, Murray Bridge: 0.04) were less than the RMSE Values of the Water Level SARIMA Model of four locations (Biggara: 0.14, Albury: 0.11, Colignan: 0.07, Murray Bridge: 0.06) which indicates that ARIMA Model has outperformed SARIMA Model.

   However, when comparing MAPE values of the Water Level ARIMA Model of four locations (Biggara: 157.72, Albury: 84.67, Colignan: 38.26, Murray Bridge: 254.96) to MAPE values of the Water Level SARIMA Model of four locations (Biggara: 8.16, Albury: 4.53, Colignan: Infinite or Not a Number, Murray Bridge: 7.00) the MAPE values of SARIMA are less compared to MAPE values of ARIMA. Thus, the SARIMA Model performs better than ARIMA Model for Salinity.

   Now, if we compare the difference in RMSE values between ARIMA and SARIMA Models with the difference in MAPE values between ARIMA and SARIMA Models, we observe that difference in MAPE values is far greater than that of RMSE. Therefore, the best Model is SARIMA Model for Water Level.

3) Research Question 3 (Water Temperature Forecast):

   The RMSE values of the Water Temperature ARIMA Model of four locations (Biggara: 0.92, Albury: 0.40, Colignan: 0.40, Murray Bridge: 0.34) were less than the RMSE Values of the Water Temperature

SARIMA Model of four locations (Biggara: 1.15, Albury: 0.55, Colignan: 0.53, Murray Bridge: 0.44) which indicates that ARIMA Model has outperformed SARIMA Model.

However, when comparing MAPE values of the Water Temperature ARIMA Model of four locations (Biggara: 515.51, Albury: 147.18, Colignan: 218.40, Murray Bridge: 109.17) to MAPE values of the Water Temperature SARIMA Model of four locations (Biggara: 5.93, Albury: 2.93, Colignan: 1.92, Murray Bridge: 1.17) the MAPE values of SARIMA are less compared to MAPE values of ARIMA. Thus, the SARIMA Model performs better than ARIMA Model.

Now, if we compare the difference in RMSE values between ARIMA and SARIMA Models with the difference in MAPE values between ARIMA and SARIMA Models, we observe that difference in MAPE values is far greater than that of RMSE. Therefore, the best Model is SARIMA Model for Water Temperature.

Lastly, we got to know that SARIMA is the best model for all three parameters i.e., Salinity, Water Level and Water Temperature. However, mostly the values of MAPE for both ARIMA and SARIMA were out of the acceptable range which is generally between 0.1 and 0.25 (Swanson, 2015). Therefore, there is a need for a better more accurate model than SARIMA and ARIMA.

7. CONCLUSION

The research project presented a comparison between two statistical models i.e., ARIMA and SARIMA Models for Salinity, Water Level and Water Temperature time series data provided by MDBA Authority for four locations specifically Biggara, Albury, Colignan and Murray Bridge. The synopsis of key findings is provided below:

1) The ARIMA Model had lower RMSE values compared to SARIMA Model for Salinity. Hence, making ARIMA a better Model. However, MAPE values of ARIMA were greater than MAPE values of SARIMA. Thus, making SARIMA a better Model. Nevertheless, if compare the difference values of RMSE (i.e., RMSE ARIMA – RMSE SARIMA) and the difference values of MAPE (i.e., MAPE ARIMA – MAPE SARIMA), the difference between MAPE is far greater than RMSE due to which SARIMA is overall best Model for Salinity.

2) The ARIMA Model had lower RMSE values compared to SARIMA Model for Water Level. Hence, making ARIMA a better Model. However, MAPE values of ARIMA were greater than MAPE values of SARIMA. Thus, making SARIMA a better Model. Nevertheless, if compare the difference values of RMSE (i.e., RMSE ARIMA – RMSE SARIMA) and the difference values of MAPE (i.e., MAPE ARIMA – MAPE SARIMA), the difference between MAPE is far greater than RMSE due to which SARIMA is overall best Model for Water Level.

3) The ARIMA Model had lower RMSE values compared to SARIMA Model for Water Temperature. Hence, making ARIMA a better Model. However, MAPE values of ARIMA were greater than MAPE values of SARIMA. Thus, making SARIMA a better Model. Nevertheless, if compare the difference values of RMSE (i.e., RMSE ARIMA minus RMSE SARIMA) and the difference values of MAPE (i.e., MAPE ARIMA minus MAPE SARIMA), the difference between MAPE is far greater than RMSE due to which SARIMA is overall best Model for Water Temperature.

Overall, when observing the RMSE and MAPE values of ARIMA and SARIMA, it was found that SARIMA is the best model for the above three parameters. Finding the best model was necessary to create a better Forecasting Model in Future to prevent disastrous situations in the MDB Basin which has an enormous area and is a significant area for a lot of endangered species for their survival.

8. FUTURE WORK

In Part B of the Master of Data Science Research Project following new models and datasets will be explored:

1) The Auto-Regressive Fractionally Integrated Moving Average (ARFIMA) model will be utilized as it gives better efficiency than ARIMA in terms of forecasting accuracy and it can also handle non-stationary time series (Sitohang et al., 2017).

2) The Gated Recurrent Unit (GRU) and Long-Term Short Memory (LSTM) Deep learning models can be utilized as they outperform ARIMA and SARIMA (ArunKumar et al.,2022).

3) The United States Geological Survey (USGS) Current Water Data for the Nation will also be utilized to test different time series models such as ARIMA, SARIMA, ARFIMA, GRU and LSTM (United States Geological Survey [USGS], 2022).

REFERENCES

ArunKumar, K. E., Kalaga, D. V., Kumar, C. M. S., Kawaji, M., & Brenza, T. M. (2022). Comparative analysis of Gated Recurrent Units (GRU), long Short-Term memory (LSTM) cells, autoregressive Integrated moving average (ARIMA), seasonal autoregressive Integrated moving average (SARIMA) for forecasting COVID-19 trends. *Alexandria Engineering Journal*, *61*(10), 7585-7603.

Bickel, P. J., & Doksum, K. A. (1981). An analysis of transformations revisited. *Journal of the american statistical association*, *76*(374), 296-311.

Biswas, T. K., & Mosley, L. M. (2019). From mountain ranges to sweeping plains, in droughts and flooding rains; River Murray water quality over the last four decades. *Water Resources Management*, *33*(3), 1087-1101.

Box, G. E., & Cox, D. R. (1964). An analysis of transformations. *Journal of the Royal Statistical Society: Series B (Methodological)*, 26(2), 211-243.

Department for Environment and Water. (2022). *Taking action on Salinity*. Department for Environment and Water. Retrieved August 19, 2022, from https://www.environment.sa.gov.au/topics/river-murray/improving-river-health/water-quality-and-salinity/taking-action-on-salinity

Fox, J., & Weisberg, S. (2013, October 8). *Robust regression - School of Statistics*. Robust regression. Retrieved August 15, 2022, from http://users.stat.umn.edu/~sandy/courses/8053/handouts/robust.pdf

Gao, Z., Guo, D., Peel, M. C., & Stewardson, M. J. (2022). Understanding regional streamflow trend magnitudes in the Southern Murray–Darling basin, Australia. *Australasian Journal of Water Resources*, 1-14.

Hyndman, R. J. (2017). *Forecasting: Principles and practice (UWA)*. 2.4 Non-seasonal ARIMA models. Retrieved August 16, 2022, from https://robjhyndman.com/uwa2017/2-4-NonseasonalARIMA.pdf

Hyndman, R. J., & Athanasopoulos, G. (2018a). *Forecasting: Principles and practice (2nd ed)*. 6.1 Time series components. Retrieved August 15, 2022, from https://otexts.com/fpp2/components.html

Hyndman, R. J., & Athanasopoulos, G. (2018b). *Forecasting: Principles and practice (2nd ed)*. 8.3 Autoregressive models. Retrieved August 16, 2022, from https://otexts.com/fpp2/AR.html

Hyndman, R. J., & Athanasopoulos, G. (2018c). *Forecasting: Principles and practice (2nd ed)*. 8.4 Moving average models. Retrieved August 16, 2022, from https://otexts.com/fpp2/MA.html

Hyndman, R. J., & Athanasopoulos, G. (2018d). *Forecasting: Principles and practice (2nd ed)*. 2.8 Autocorrelation. Retrieved August 15, 2022, from https://otexts.com/fpp2/autocorrelation.html

Hyndman, R. J., & Athanasopoulos, G. (2018e). *Forecasting: Principles and practice (2nd ed)*. 8.9 Seasonal ARIMA models. Retrieved August 15, 2022, from https://otexts.com/fpp2/seasonal-arima.html

Hyndman, R. J., & Athanasopoulos, G. (2021a). *Forecasting: Principles and practice (3rd ed)*. 9.1 Stationarity and differencing. Retrieved August 16, 2022, from https://otexts.com/fpp3/stationarity.html#fn15

Hyndman, R. J., & Athanasopoulos, G. (2021b). *Forecasting: Principles and practice (3rd ed)*. 9.1 Stationarity and differencing. Retrieved August 16, 2022, from https://otexts.com/fpp3/stationarity.html

Hyndman, R. J., & Athanasopoulos, G. (2021c). *Forecasting: Principles and practice (3rd ed)*. 9.5 Non-seasonal ARIMA models. Retrieved August 16, 2022, from https://otexts.com/fpp3/non-seasonal-arima.html

Hyndman, R. J., & Athanasopoulos, G. (2021d). *Forecasting: Principles and practice (3rd ed)*. 7.5 Selecting predictors. Retrieved August 16, 2022, from https://otexts.com/fpp3/selecting-predictors.html

Hyndman, R. J., & Athanasopoulos, G. (2021e). *Forecasting: Principles and practice (3rd ed)*. 5.4 Residual diagnostics. Retrieved August 16, 2022, from https://otexts.com/fpp3/diagnostics.html#diagnostics

Hyndman, R., Athanasopoulos, G., Bergmeir, C., Caceres, G., Chhay, L., Kuroptev, K., O'Hara-Wild, M., & Petropoulos, F. (2022, July 25). *Forecasting functions for time series and linear models [R package forecast version 8.17.0]*. The Comprehensive R Archive Network. Retrieved August 16, 2022, from https://cran.r-project.org/web/packages/forecast/index.html

Jolly, I. D., Williamson, D. R., Gilfedder, M., Walker, G. R., Morton, R., Robinson, G., ... & McNeill, V. (2001). Historical stream salinity trends and catchment salt balances in the Murray–Darling Basin, Australia. *Marine and Freshwater Research*, *52*(1), 53-63.

Kokoszka, P., & Young, G. (2016). *KPSS test for Functional Time Series*. Colorado State University. Retrieved August 16, 2022, from https://www.stat.colostate.edu/~piotr/kpss.pdf

Loo, M. van der. (2022, June 16). *simputation Package*. Getting started with simputation. Retrieved August 15, 2022, from https://cran.r-project.org/web/packages/simputation/vignettes/intro.html

Matamoros, A. A., Torres, C. C., Dala, A., Hyndman, R., & O'Hara-Wild, M. (2021, June 17). *Bayesian time series modeling with Stan [R package bayesforecast version 1.0.1]*. The Comprehensive R Archive Network. Retrieved August 16, 2022, from https://cran.microsoft.com/snapshot/2022-05-26/web/packages/bayesforecast/index.html

Murray Darling Basin Authority. (2022a). *About the data: River Murray Data*. Live River Data. Retrieved August 15, 2022, from https://riverdata.mdba.gov.au/about-data

Murray Darling Basin Authority. (2022b). *Salinity*. Australian Government - Murray-Darling Basin Authority. Retrieved August 15, 2022, from https://www.mdba.gov.au/issues-murray-darling-basin/salinity

Murray Darling Basin Authority. (2022c). *Acid sulfate soils*. Australian Government - Murray-Darling Basin Authority. Retrieved August 15, 2022, from https://www.mdba.gov.au/issues-murray-darling-basin/acid-sulfate-soils

Murray-Darling Basin Authority. (2022d). *Drought*. Australian Government - Murray-Darling Basin Authority. Retrieved August 15, 2022, from https://www.mdba.gov.au/issues-murray-darling-basin/drought

Murray-Darling Basin Authority. (2022e). *The murray–darling basin and why it's important*. Australian Government - Murray-Darling Basin Authority. Retrieved August 16, 2022, from https://www.mdba.gov.au/importance-murray-darling-basin

Murray-Darling Basin Authority. (2022f). *Environmental importance*. Australian Government - Murray-Darling Basin Authority. Retrieved August 16, 2022, from https://www.mdba.gov.au/importance-murray-darling-basin/environment

PennState. (2022). *2.3 notational conventions: Stat 510*. PennState: Statistics Online Courses. Retrieved August 16, 2022, from https://online.stat.psu.edu/stat510/lesson/2/2.3

Qiu, D. (2015). Package aTSA. CRAN. Retrieved August 16, 2022, from https://cran.r-project.org/web/packages/aTSA/aTSA.pdf

Sitohang, Y. O., & Darmawan, G. (2017, August). The accuracy comparison between ARFIMA and singular spectrum analysis for forecasting the sales volume of motorcycle in Indonesia. In *AIP Conference Proceedings* (Vol. 1868, No. 1, p. 040011). AIP Publishing LLC.

Statsmodels. (2022). Stationarity and detrending (ADF/KPSS). Retrieved August 16, 2022, from https://www.statsmodels.org/dev/examples/notebooks/generated/stationarity_detrending_adf_kpss.html

Swanson, D. A. (2015). On the relationship among values of the same summary measure of error when used across multiple characteristics at the same point in time: an examination of MALPE and MAPE. *Review of Economics and Finance*, *5*(1).

United States Geological Survey. (2022). *National Water Information System: Web Interface*. USGS current water data for the nation - Daily Streamflow Conditions. Retrieved August 16, 2022, from https://waterdata.usgs.gov/nwis/rt

Wouters, S. (2021). Package 'DecomposeR.' Retrieved August 15, 2022, from https://cran.r-project.org/web/packages/DecomposeR/DecomposeR.pdf

Zivot, E. (2014). *Unit root tests*. University of Washington. Retrieved August 16, 2022, from https://faculty.washington.edu/ezivot/econ584/notes/unitroot.pdf

APPENDIX

The plot of Murray Bridge Location:

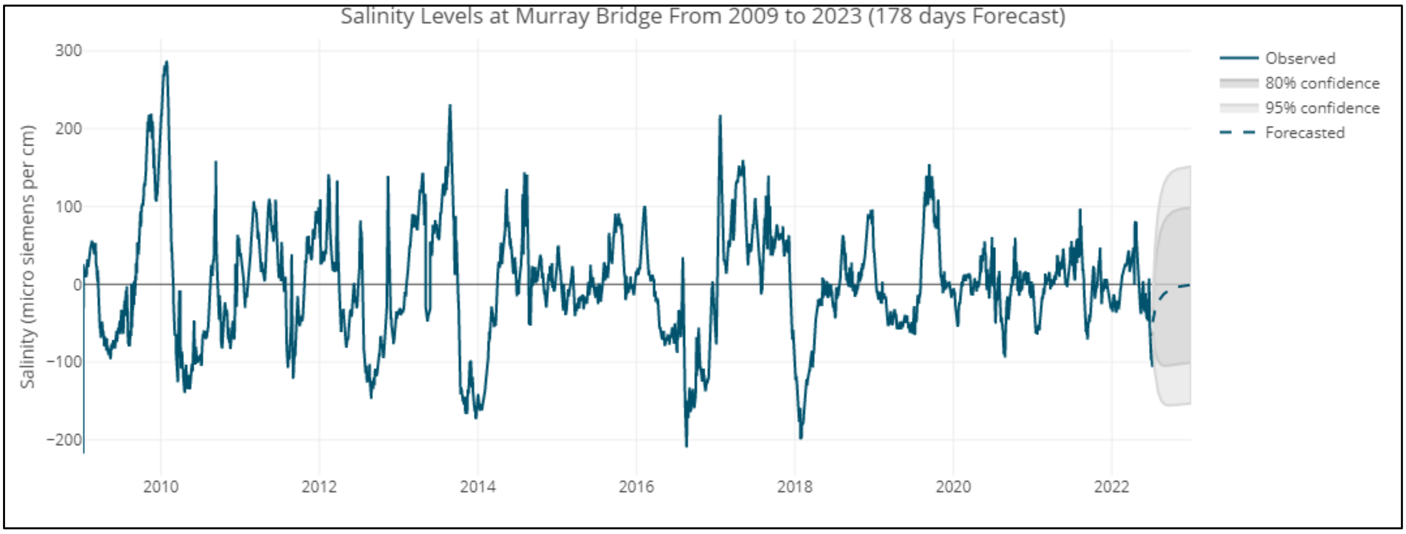

Figure 22: Salinity Levels 178 days Forecast given by the best ARIMA(3,0,2) Model

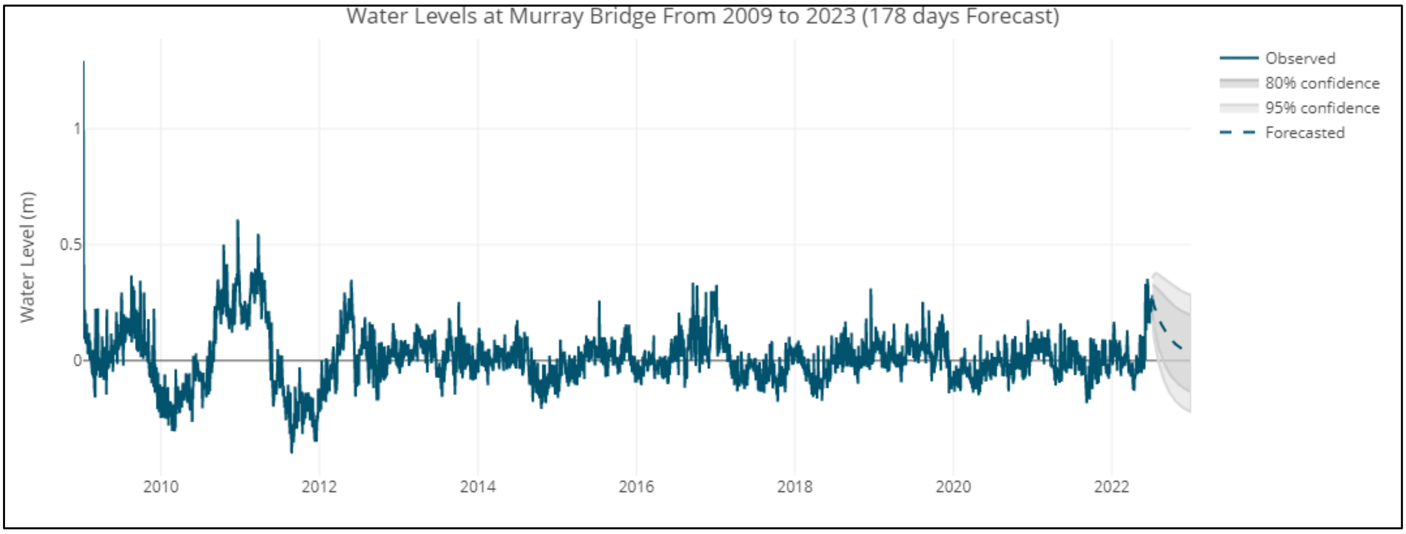

Figure 23: Water Levels 178 days Forecast given by the best ARIMA(1,0,2) Model

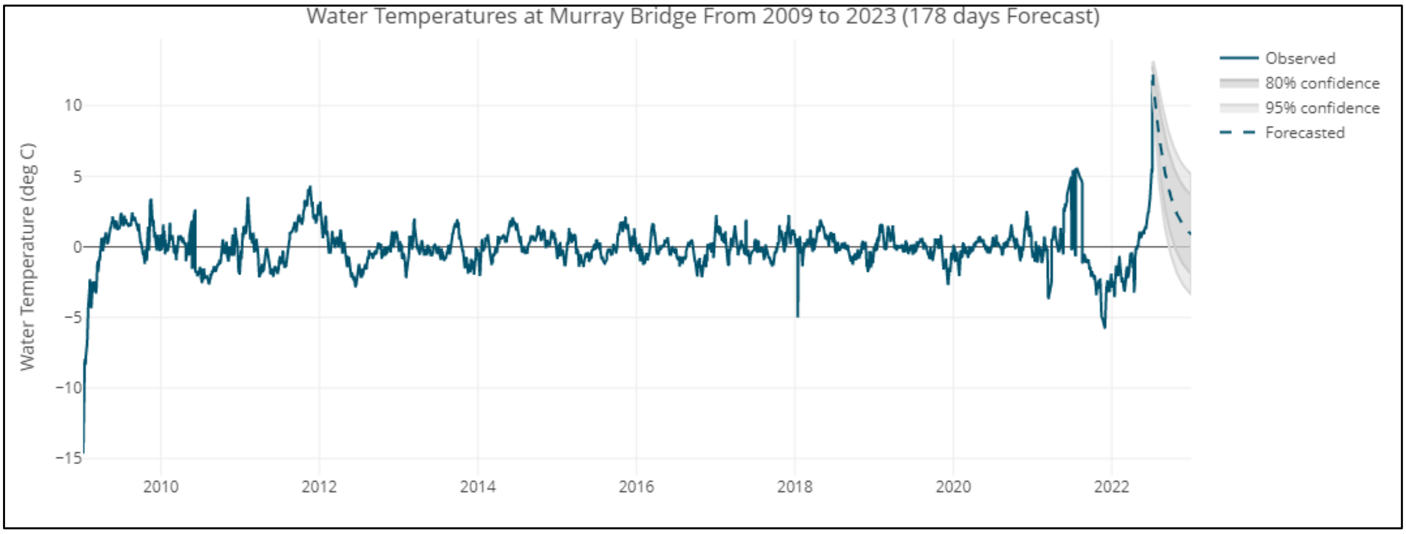

Figure 24: Water Temperatures 178 days Forecast given by the best ARIMA(3,0,2) Model

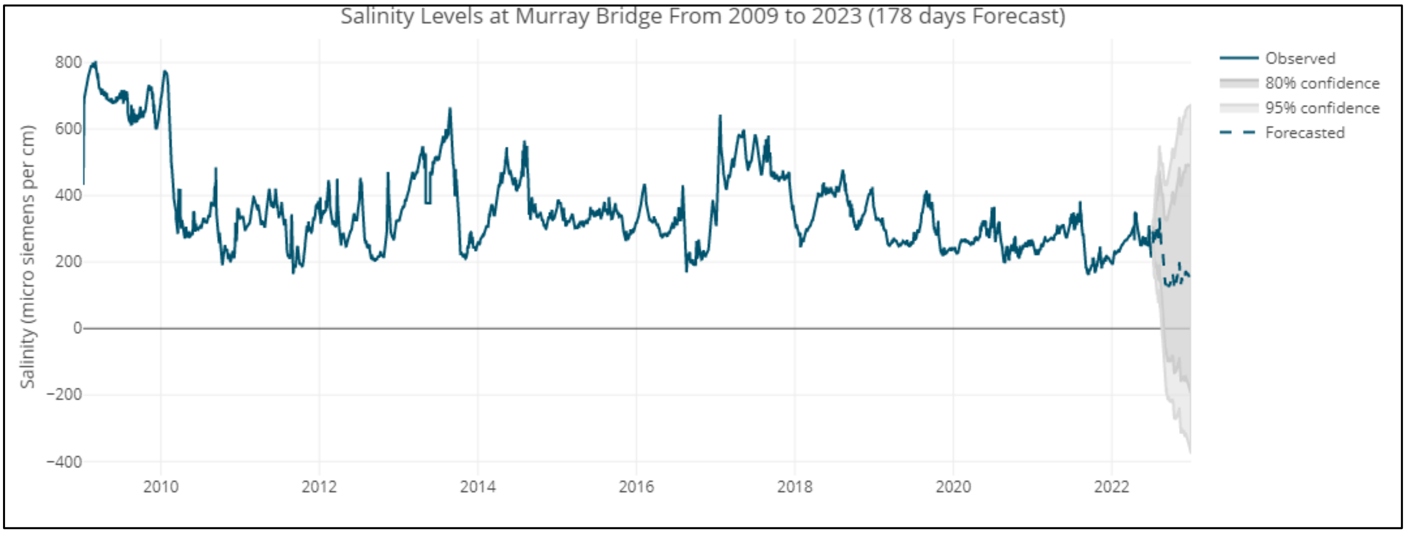

Figure 25: Salinity Levels 178 days Forecast given by the best ARIMA(2,1,2)(0,1,0)[365] (SARIMA) Model

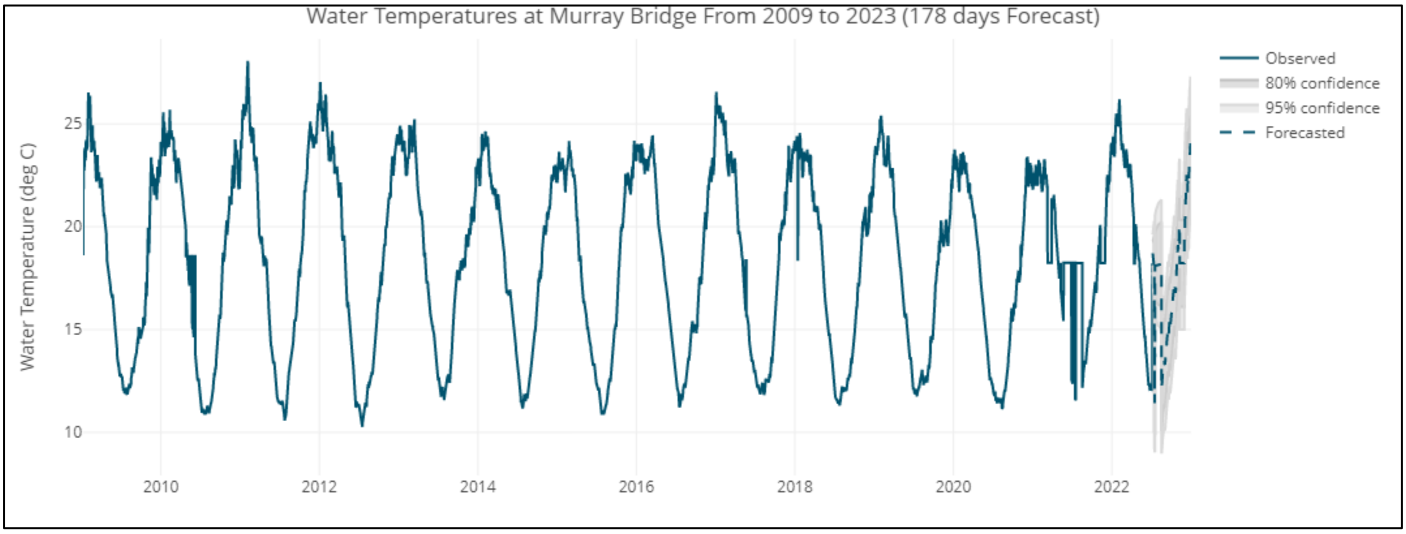

Figure 26: Water Temperatures 178 days Forecast given by the best ARIMA(2,0,2)(0,1,0)[365] (SARIMA) Model

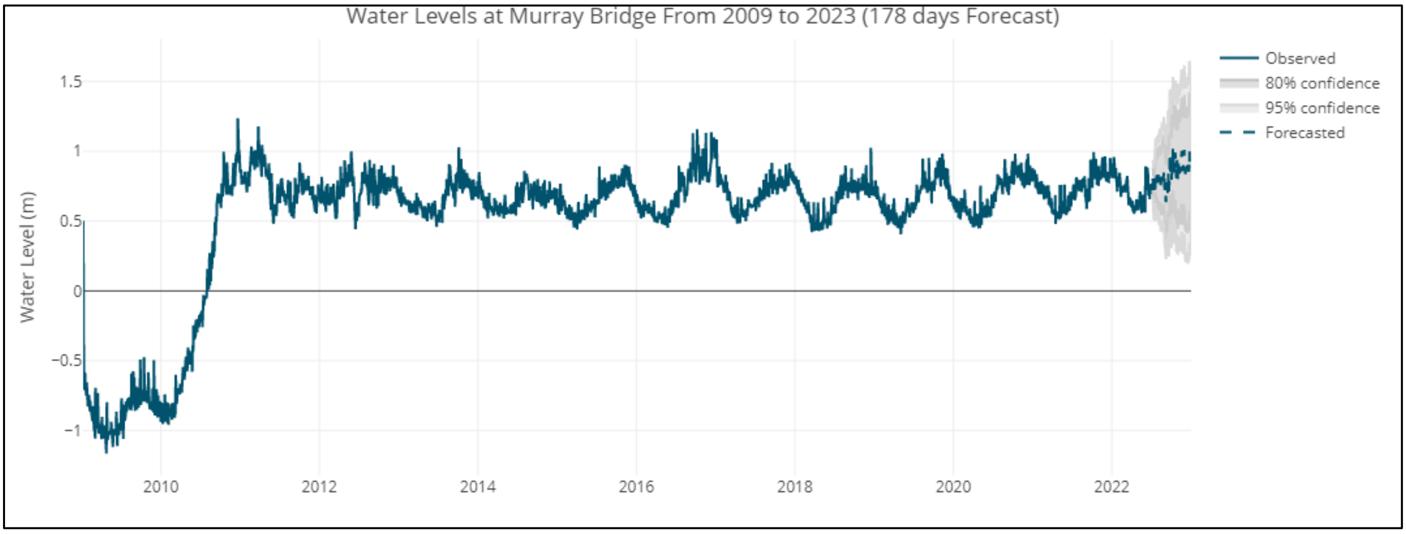

Figure 27: Water Levels 178 days Forecast given by the best ARIMA(1,1,3)(0,1,0)[365] (SARIMA) Model

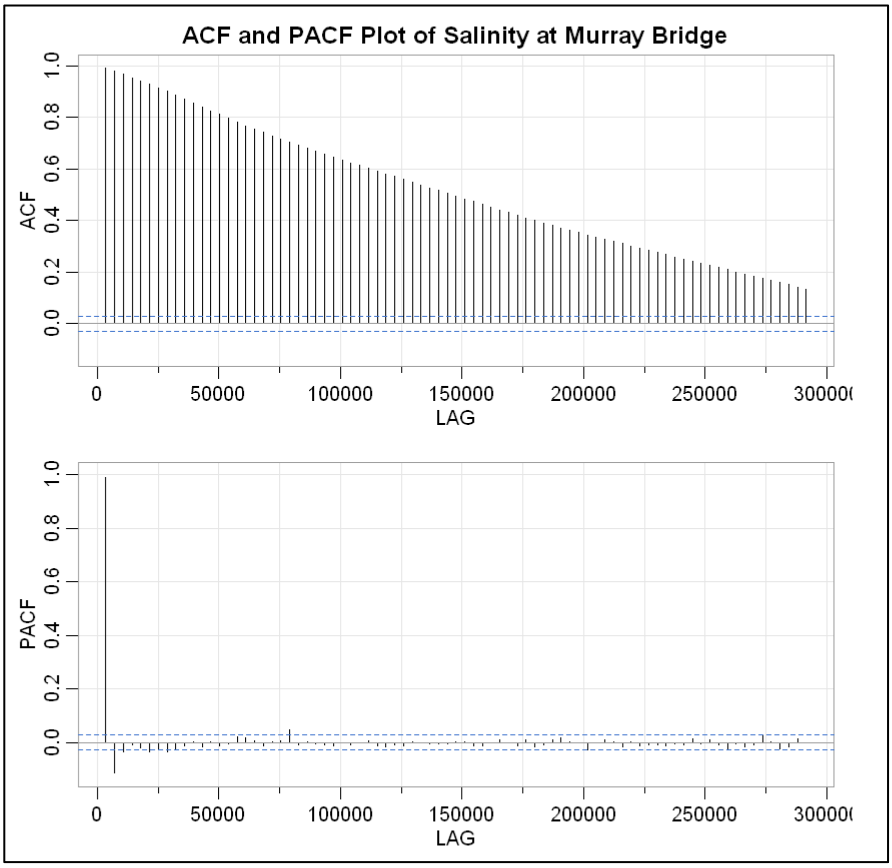

Figure 28: ACF and PACF of Salinity at Murray Bridge

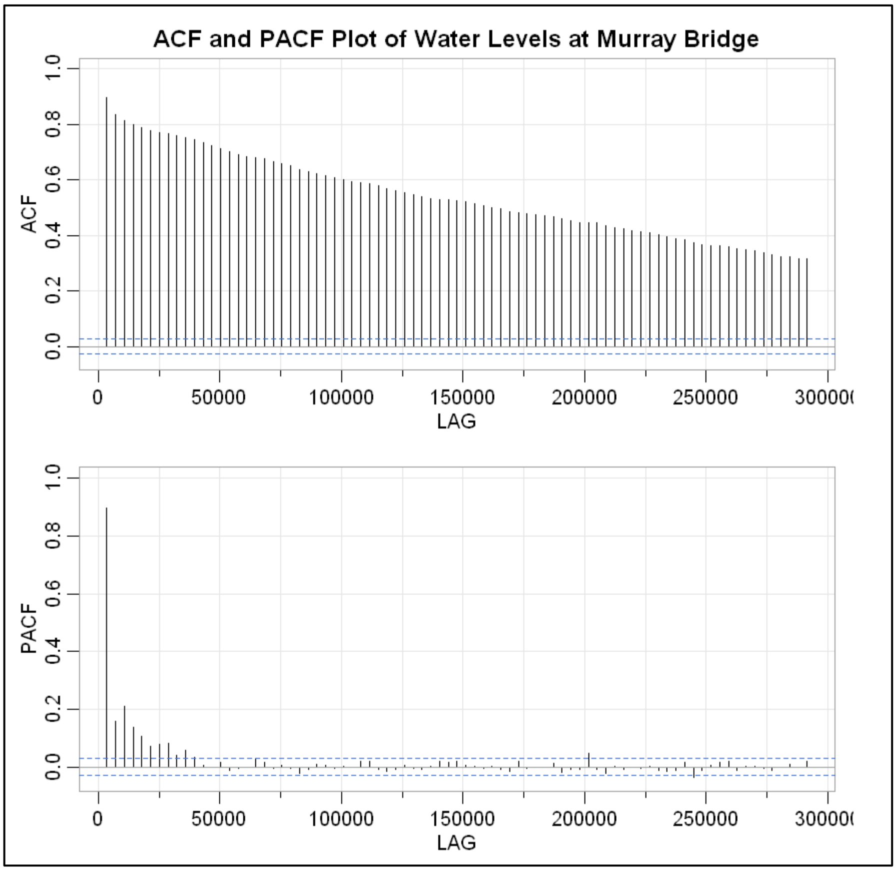

Figure 29: ACF and PACF of Water Levels at Murray Bridge

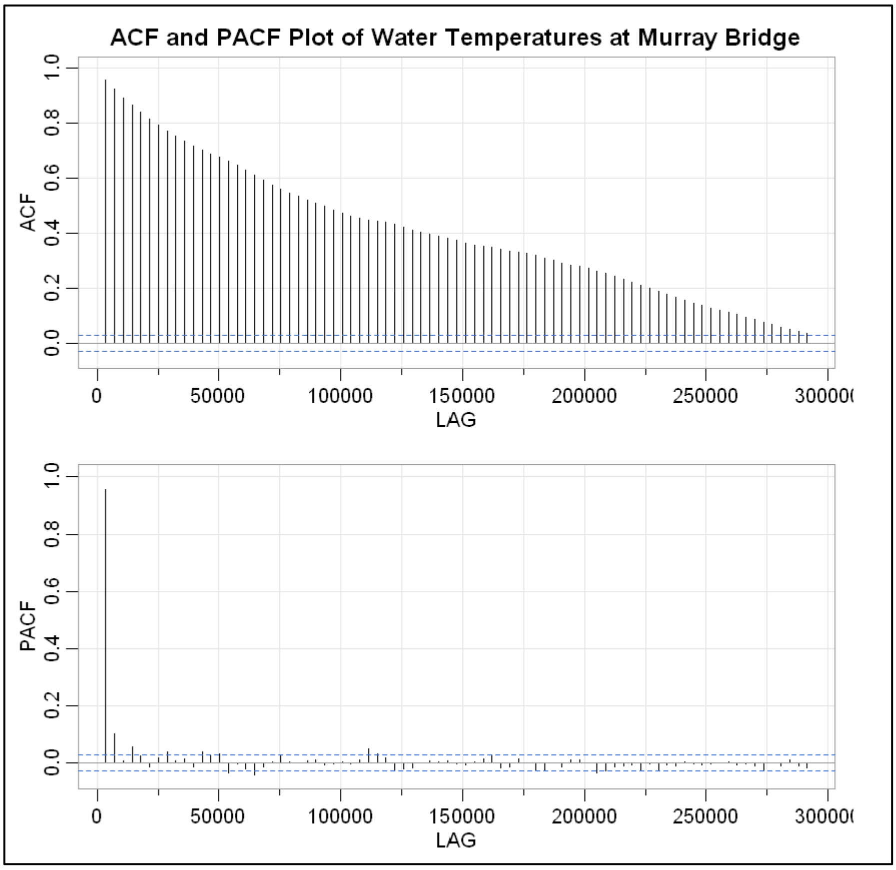

Figure 30: ACF and PACF of Water Temperatures at Murray Bridge

The table data of Murray Bridge Location:

Table 9: Different ARIMA Models Trained for Salinity

ARIMA Models	AIC	AICc	BIC
ARIMA(1,0,1)	35829.58	35829.58	35849.08
ARIMA(1,0,2)	35819.44	35819.45	35845.45
ARIMA(3,0,1)	35820.71	35820.72	35853.22
ARIMA(3,0,2)	35810.45	35810.47	35849.47

Table 10: Different ARIMA Models Trained for Water Level

ARIMA Models	AIC	AICc	BIC
ARIMA(1,0,1)	-16044.42	-16044.41	-16024.91
ARIMA(1,0,2)	-16405.48	-16405.47	-16379.47
ARIMA(2,0,0)	-16003.85	-16003.84	-15984.34
ARIMA(4,0,0)	-16298.26	-16298.25	-16265.75

Table 11: Different ARIMA Models Trained for Water Temperature

ARIMA Models	AIC	AICc	BIC
ARIMA(1,0,1)	3522.903	3522.908	3542.41
ARIMA(1,0,2)	3524.601	3524.609	3550.61
ARIMA(3,0,1)	3515.23	3515.242	3547.741
ARIMA(3,0,2)	3489.25	3489.27	3528.26

Table 12: Different SARIMA Models Trained for Salinity

SARIMA Models	AICc
ARIMA(2,1,1)(0,1,0)[365]	33958.24
ARIMA(2,1,2)(0,1,0)[365]	33950.11
ARIMA(3,1,1)(0,1,0)[365]	33951.4
ARIMA(3,1,3)(0,1,0)[365]	33953.36

Table 13: Different SARIMA Models Trained for Water Level

SARIMA Models	AICc
ARIMA(1,1,3)(0,1,0)[365]	-10504.28
ARIMA(2,1,2)(0,1,0)[365]	-10504.16
ARIMA(2,1,4)(0,1,0)[365]	-10501.41
ARIMA(3,1,2)(0,1,0)[365]	-10502.08

Table 14: Different SARIMA Models Trained for Water Temperature

SARIMA Models	AICc
ARIMA(1,0,1)(0,1,0)[365]	5606.28
ARIMA(1,0,2)(0,1,0)[365]	5604.333
ARIMA(2,0,1)(0,1,0)[365]	5549.63
ARIMA(2,0,2)(0,1,0)[365]	5549.041

The plot of Colignan Location:

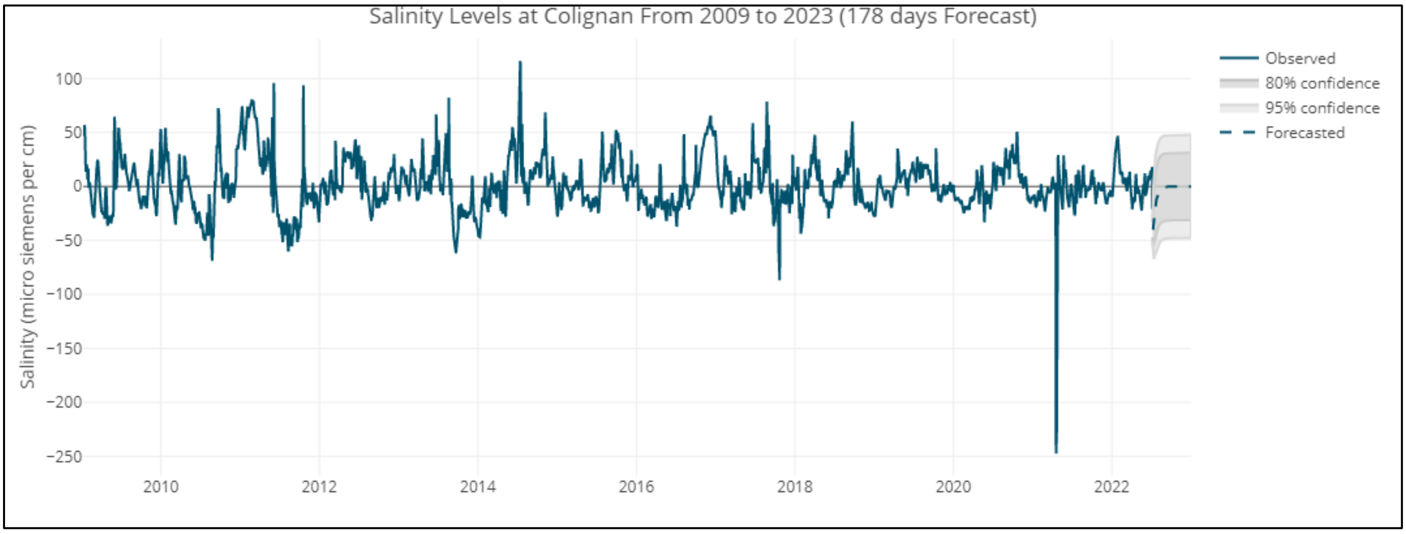

Figure 31: Salinity Levels 178 days Forecast given by the best ARIMA(3,0,1) Model

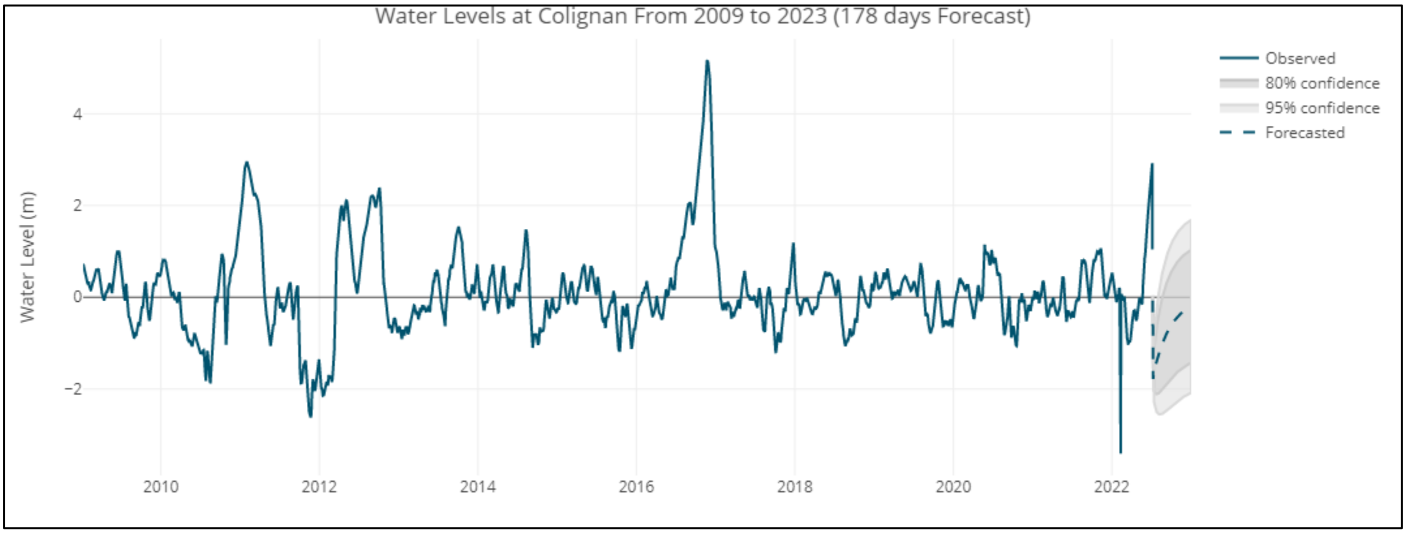

Figure 32: Water Levels 178 days Forecast given by the best ARIMA(1,0,4) Model

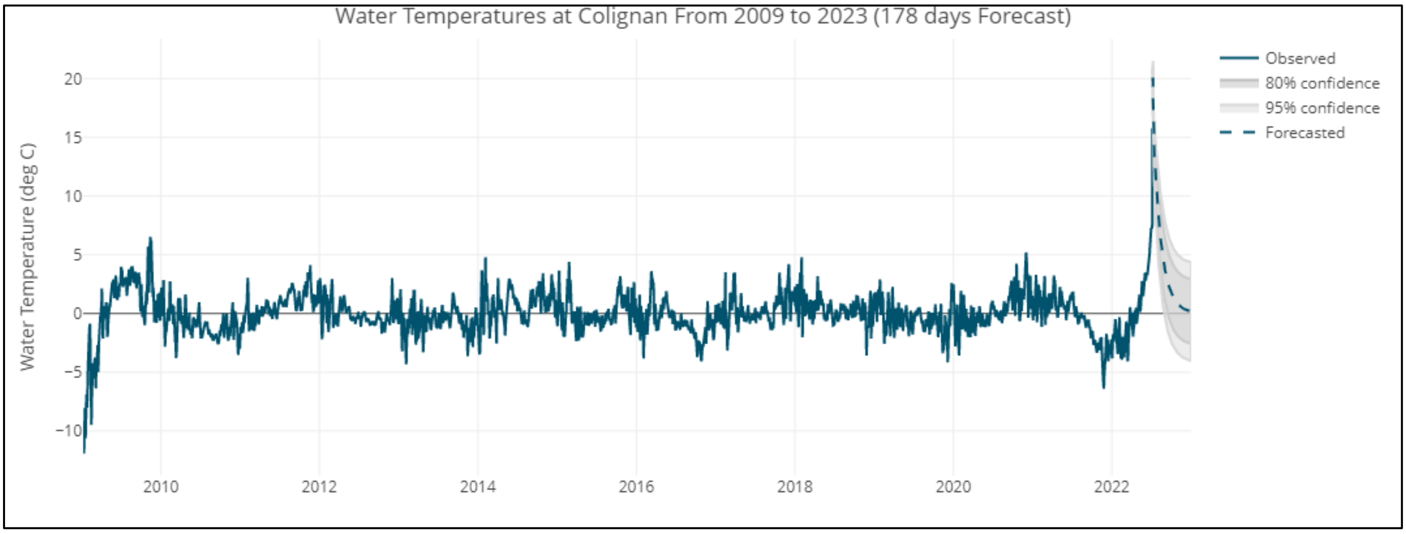

Figure 33: Water Temperatures 178 days Forecast given by the best ARIMA(3,0,1) Model

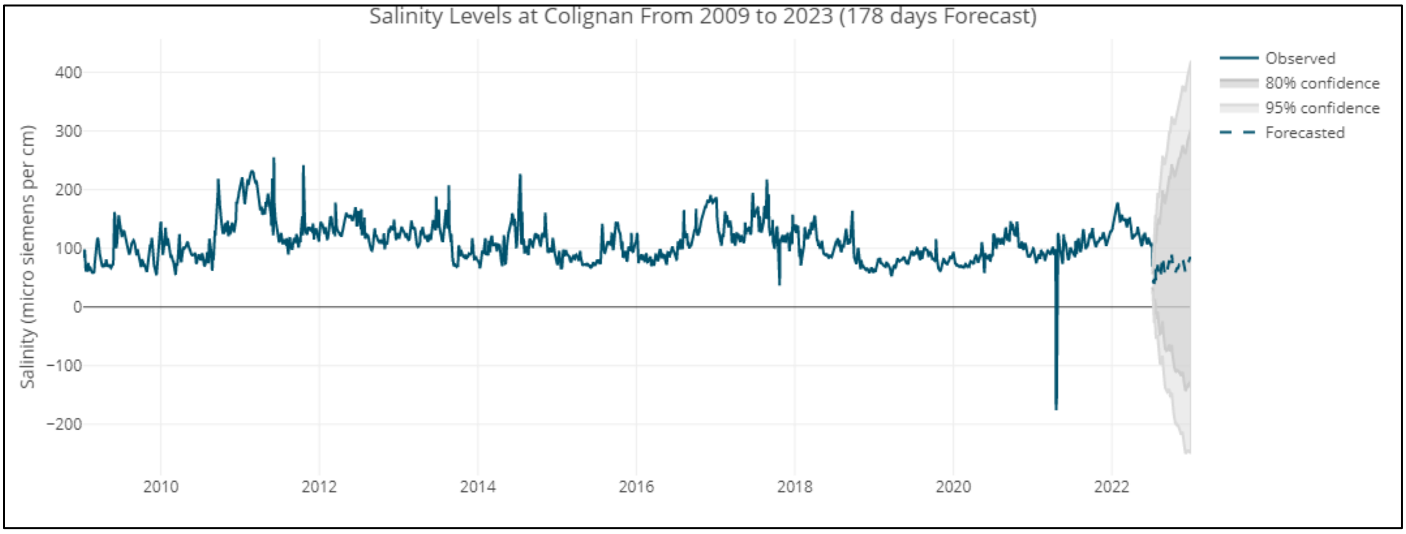

Figure 34: Salinity Levels 178 days Forecast given by the best ARIMA(0,1,1)(0,1,0)[365] (SARIMA) Model

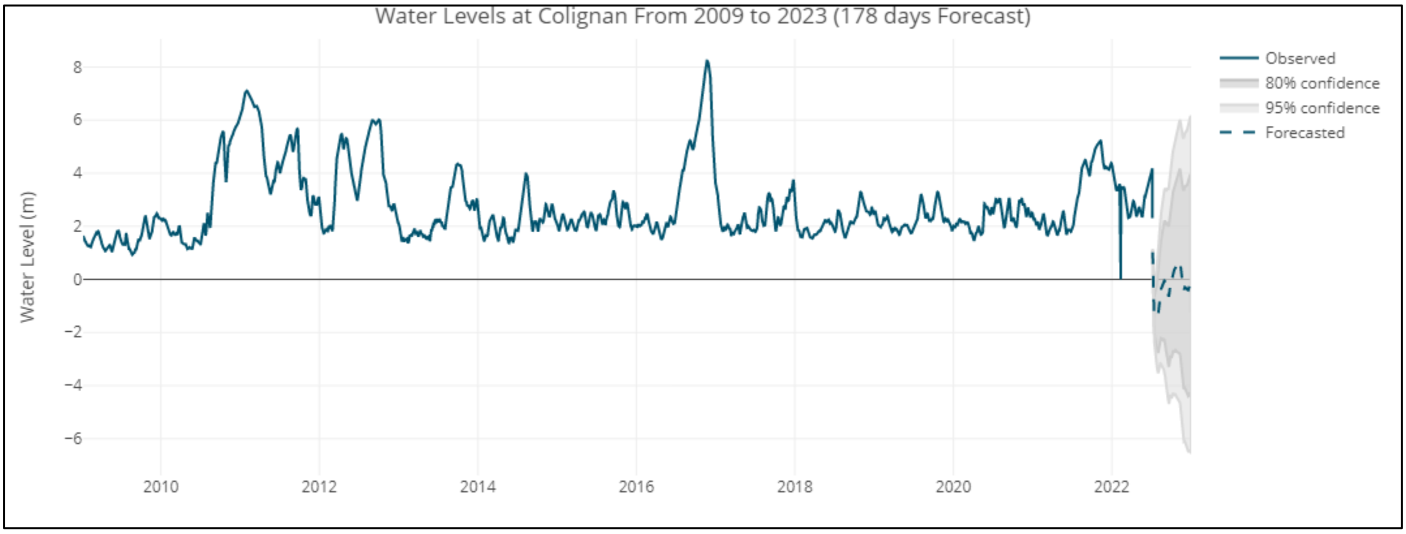

Figure 35: Water Levels 178 days Forecast given by the best ARIMA(3,1,5)(0,1,0)[365] (SARIMA) Model

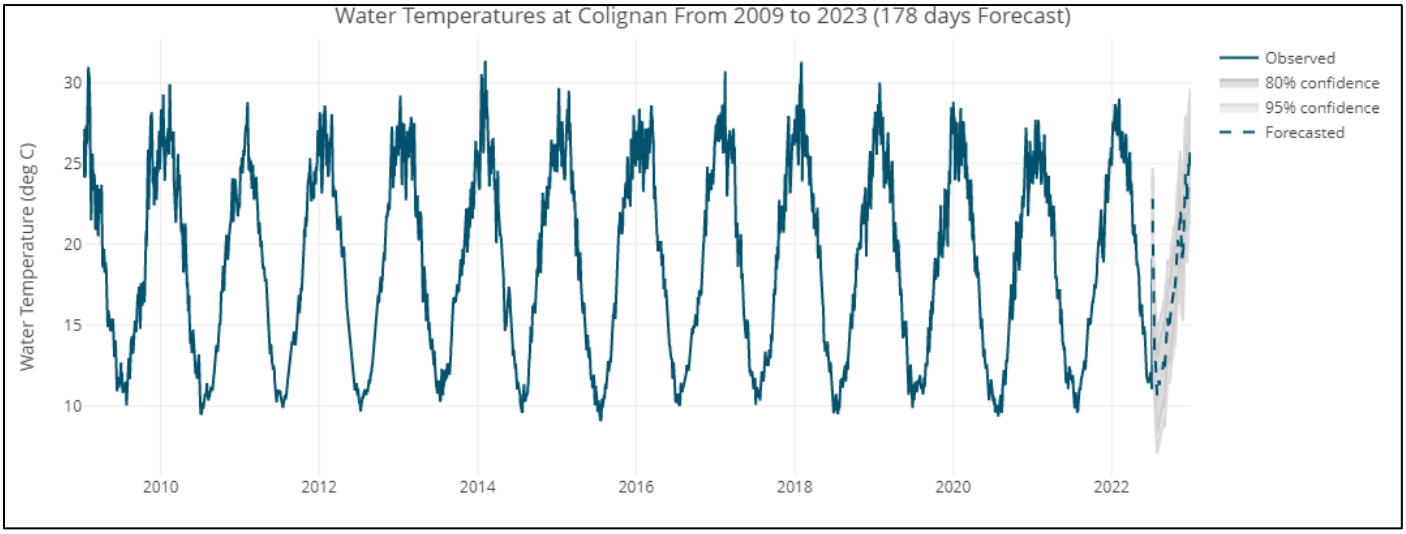

Figure 36: Water Temperatures 178 days Forecast given by the best ARIMA(1,0,5)(0,1,0)[365] (SARIMA) Model

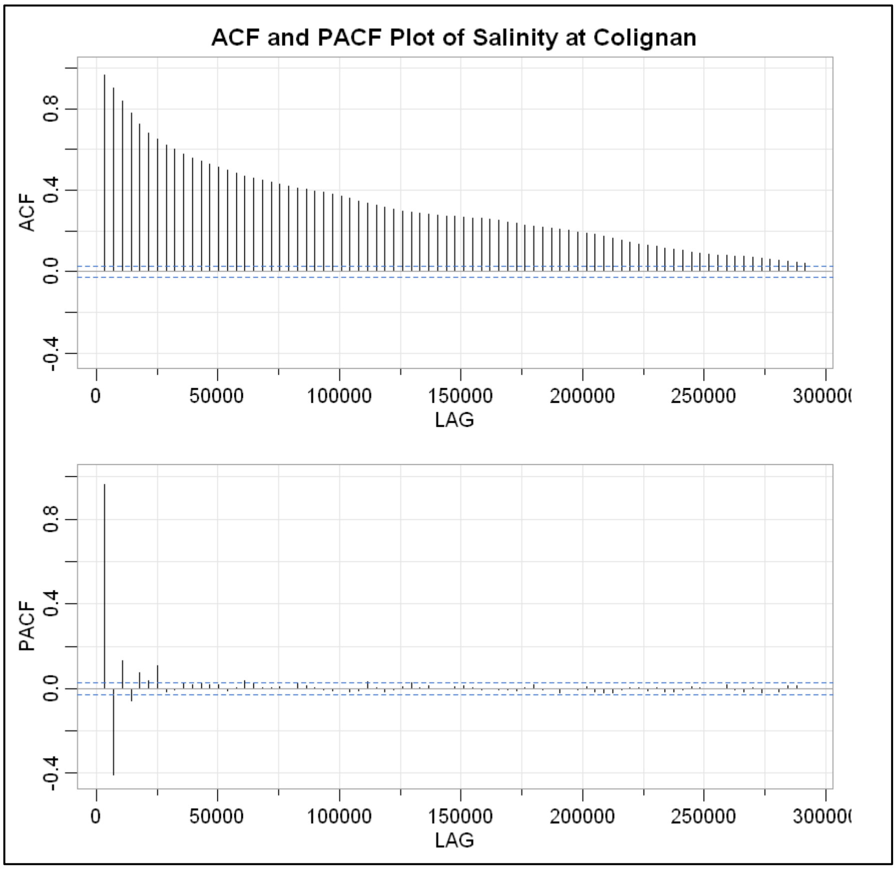

Figure 37: ACF and PACF of Salinity at Colignan

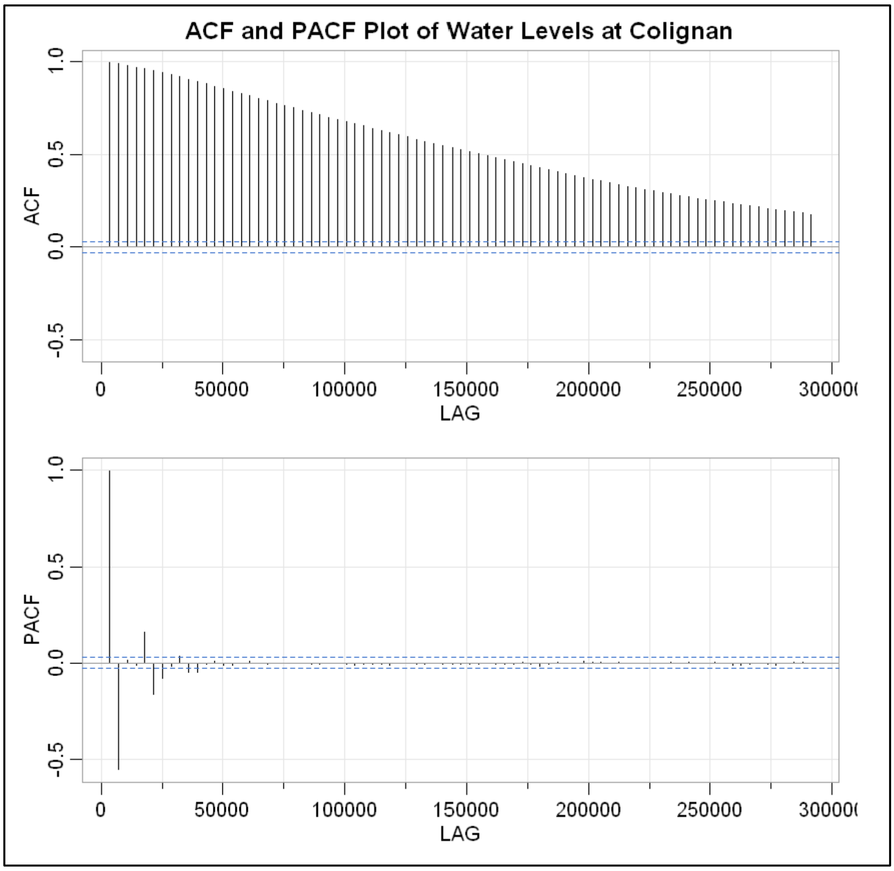

Figure 38: ACF and PACF of Water Levels at Colignan

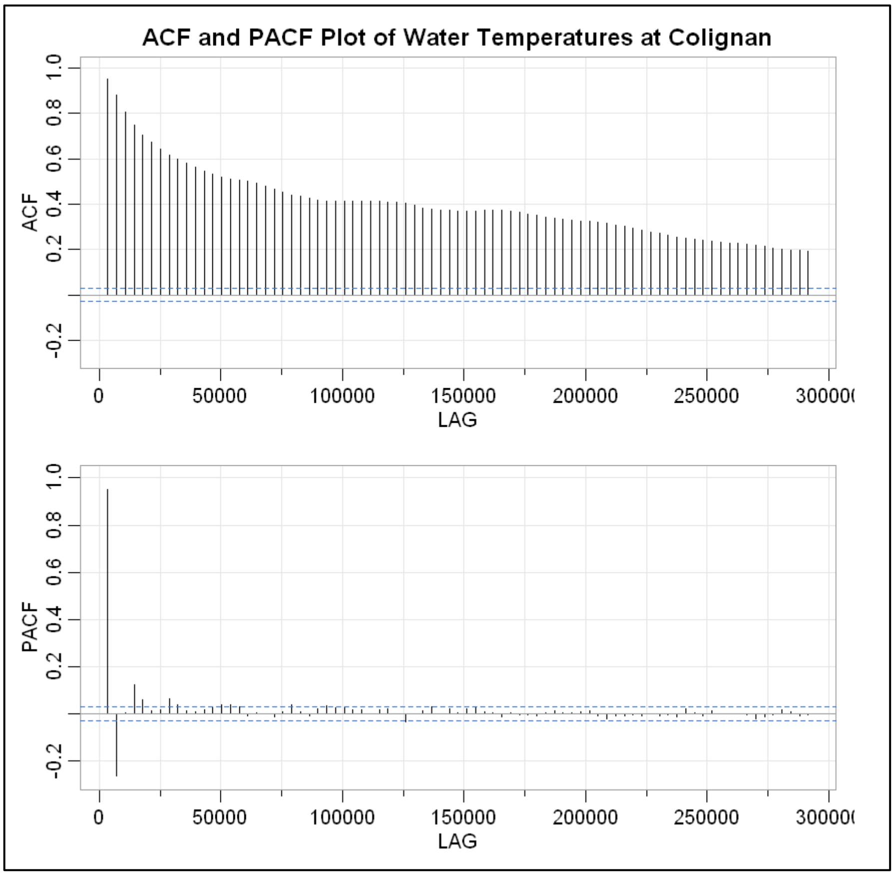

Figure 39: ACF and PACF of Water Temperatures at Colignan

The table data of Colignan Location:

Table 15: Different ARIMA Models Trained for Salinity

ARIMA Models	AIC	AICc	BIC
ARIMA(1,0,1)	31026.25	31026.25	31045.75
ARIMA(1,0,2)	31022.64	31022.65	31048.65
ARIMA(3,0,1)	31009.09	31009.1	31041.6
ARIMA(3,0,2)	31009.84	31009.86	31048.86

Table 16: Different ARIMA Models Trained for Water Level

ARIMA Models	AIC	AICc	BIC
ARIMA(1,0,3)	-13497.75	-13497.74	-13465.24
ARIMA(1,0,4)	-13505.02	-13505	-13466
ARIMA(2,0,3)	-13501.59	-13501.57	-13462.57
ARIMA(3,0,2)	-13148.45	-13148.44	-13109.44

Table 17: Different ARIMA Models Trained for Water Temperature

ARIMA Models	AIC	AICc	BIC
ARIMA(1,0,3)	5120.625	5120.637	5161.655
ARIMA(1,0,4)	5079.161	5079.178	5118.175
ARIMA(3,0,1)	5026.96	5026.97	5059.47
ARIMA(4,0,0)	5033.769	5033.781	5066.281

Table 18: Different SARIMA Models Trained for Salinity

SARIMA Models	AICc
ARIMA(1,1,0)(0,1,0)[365]	31150.35
ARIMA(0,1,1)(0,1,0)[365]	30978.41
ARIMA(1,1,1)(0,1,0)[365]	30981.26
ARIMA(1,1,2)(0,1,0)[365]	30983.25

Table 19: Different SARIMA Models Trained for Water Level

SARIMA Models	AICc
ARIMA(1,1,0)(0,1,0)[365]	-9323.48
ARIMA(0,1,1)(0,1,0)[365]	-8724.894
ARIMA(3,1,5)(0,1,0)[365]	-9659.549
ARIMA(5,1,4)(0,1,0)[365]	-9656.908

Table 20: Different SARIMA Models Trained for Water Temperature

SARIMA Models	AICc
ARIMA(1,0,4)(0,1,0)[365]	7157.303
ARIMA(1,0,5)(0,1,0)[365]	7123.257
ARIMA(2,0,4)(0,1,0)[365]	7167.581
ARIMA(2,0,5)(0,1,0)[365]	7125.017

The plot of Albury Location:

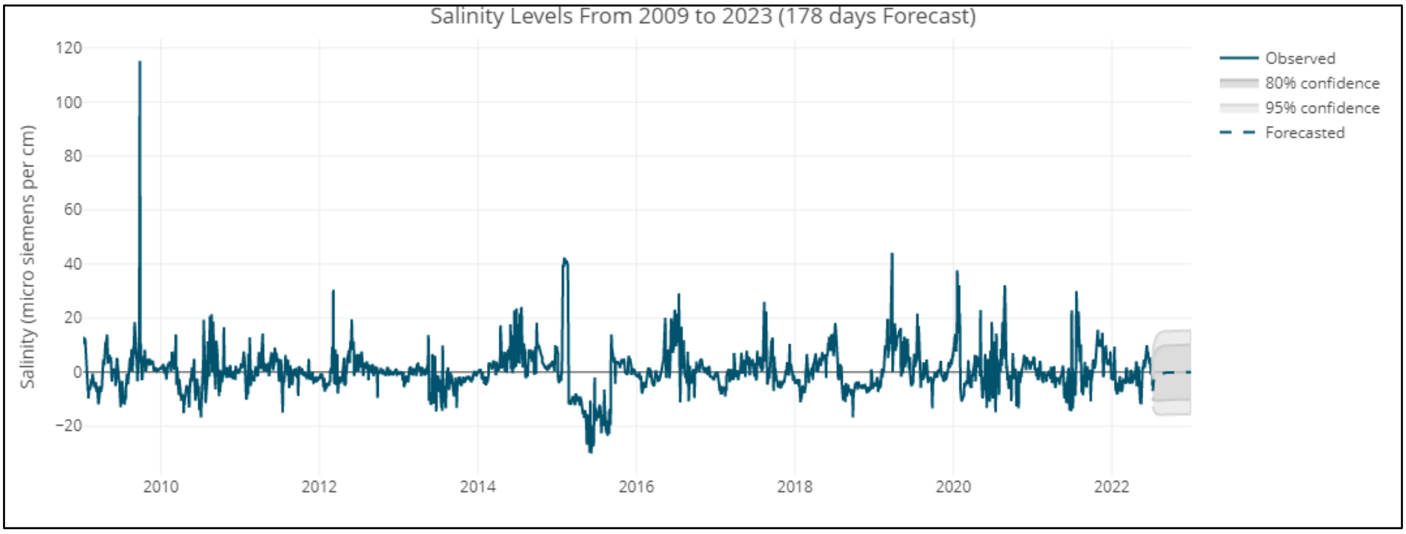

Figure 40: Salinity Levels 178 days Forecast given by the best ARIMA(2,0,3) Model

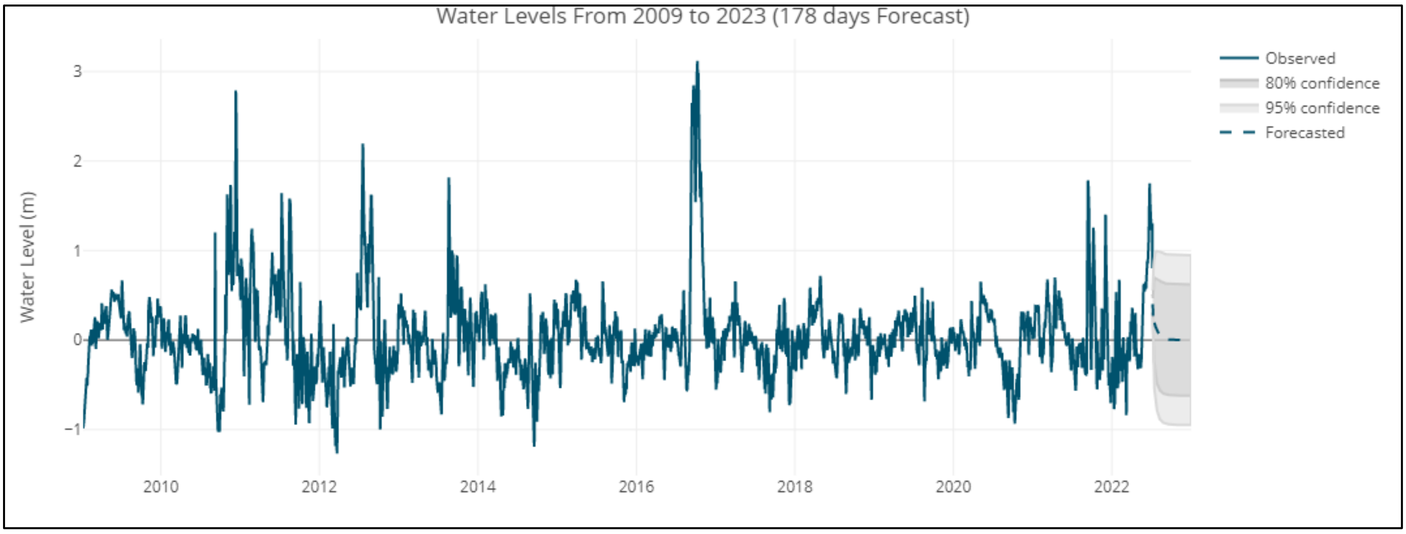

Figure 41: Water Levels 178 days Forecast given by the best ARIMA(5,0,0) Model

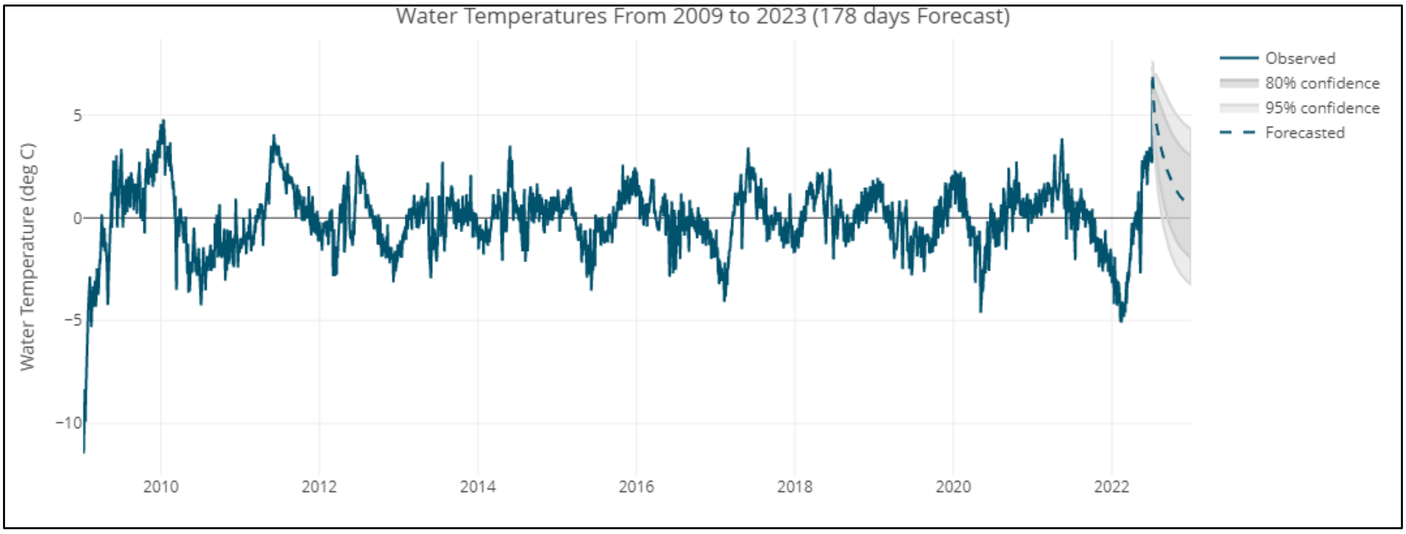

Figure 42: Water Temperatures 178 days Forecast given by the best ARIMA(1,0,4) Model

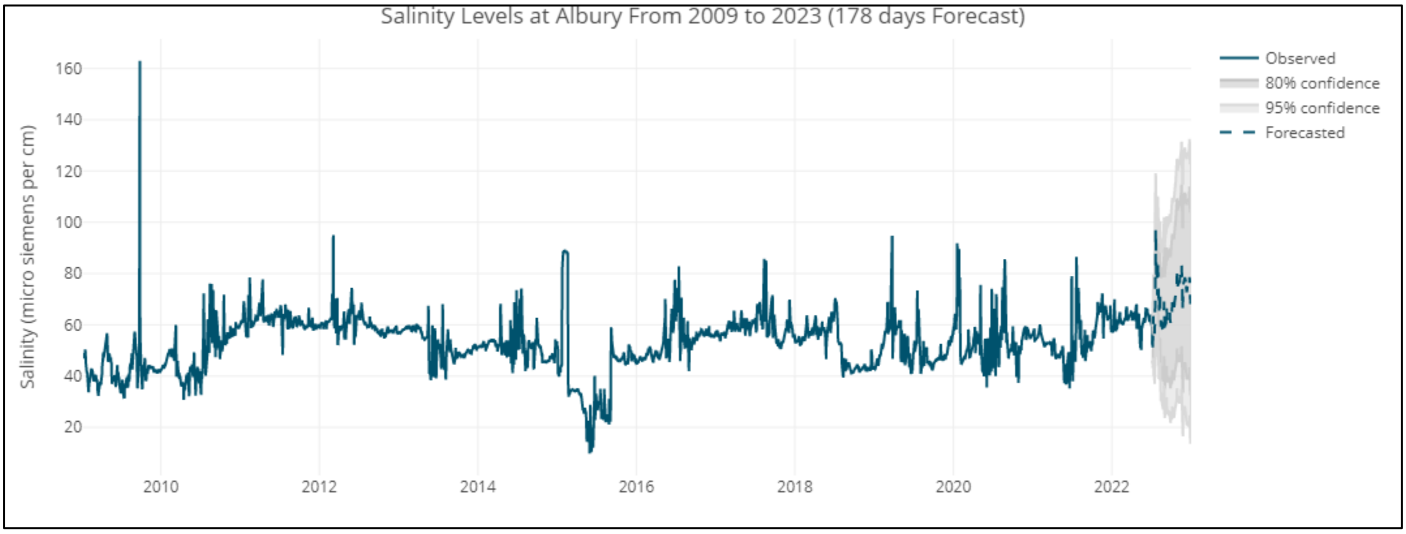

Figure 43: Salinity Levels 178 days Forecast given by the best ARIMA(1,1,2)(0,1,0)[365] (SARIMA) Model

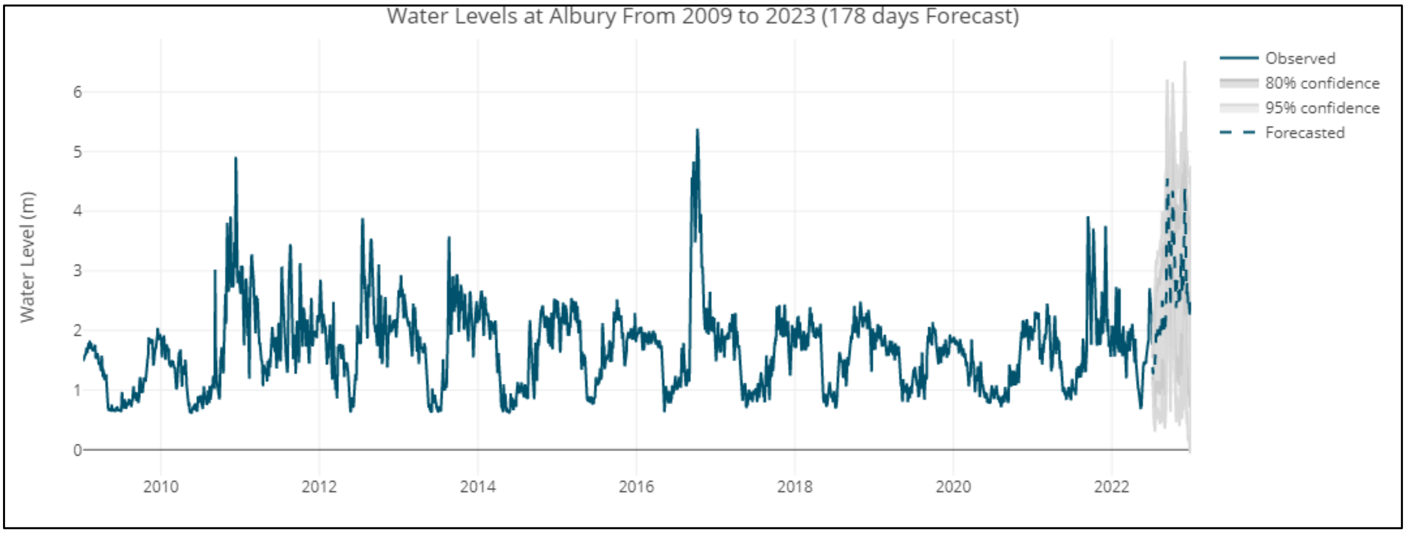

Figure 44: Water Levels 178 days Forecast given by the best ARIMA(2,1,2)(0,1,0)[365] (SARIMA) Model

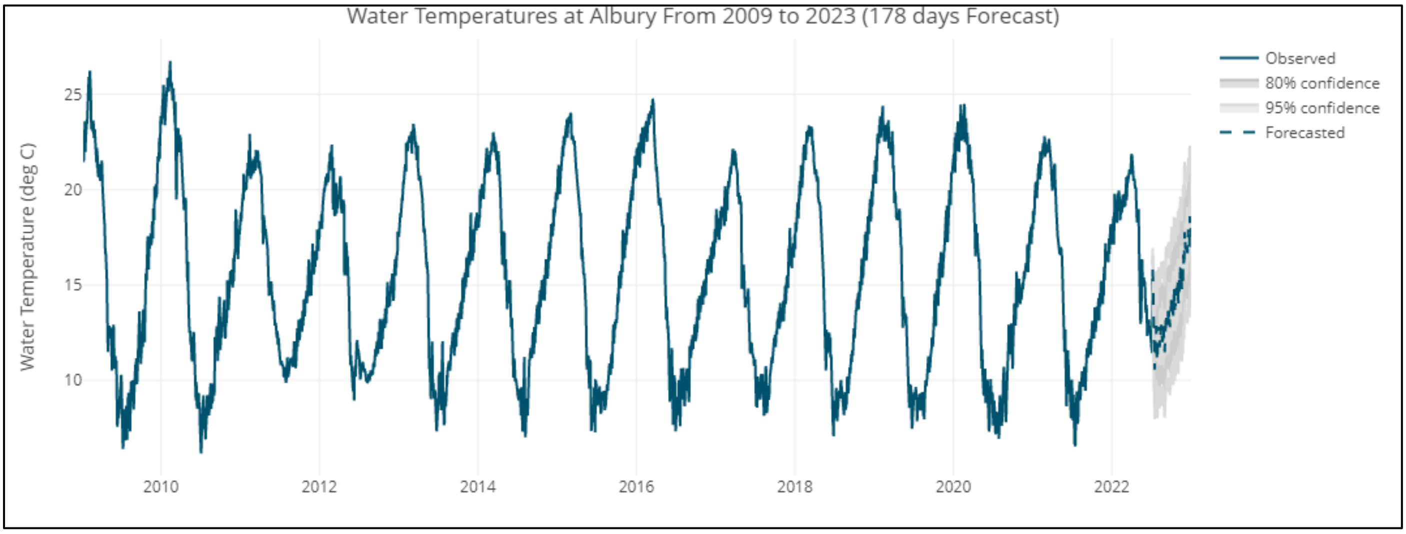

Figure 45: Water Temperatures 178 days Forecast given by the best ARIMA(2,0,2)(0,1,0)[365] (SARIMA) Model

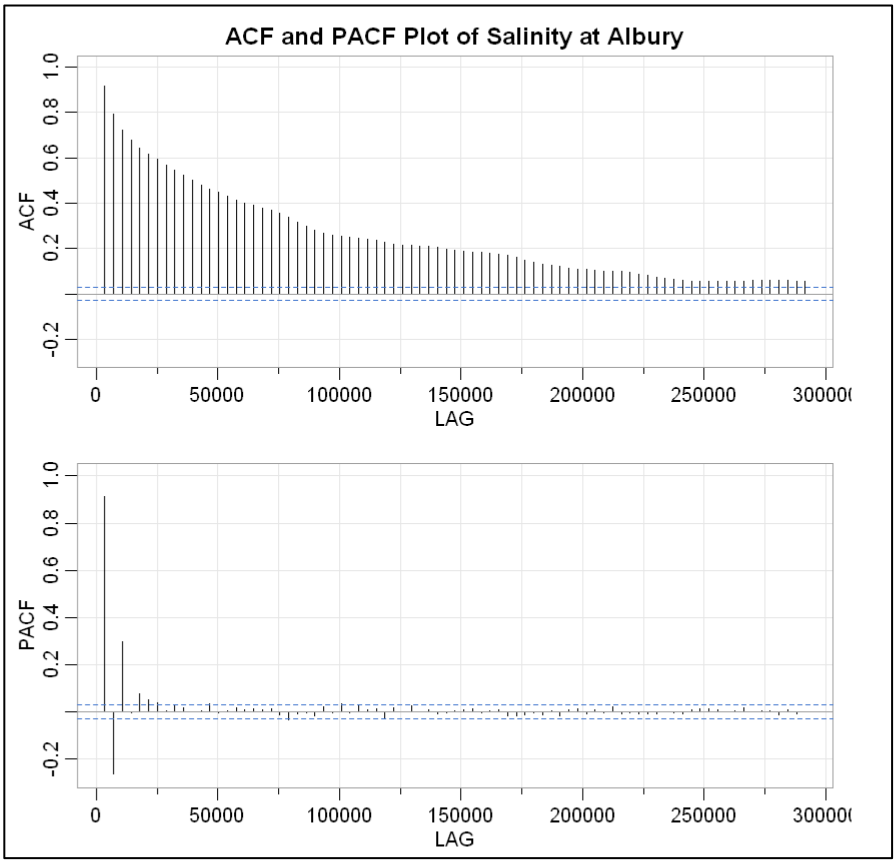

Figure 46: ACF and PACF for Salinity at Albury

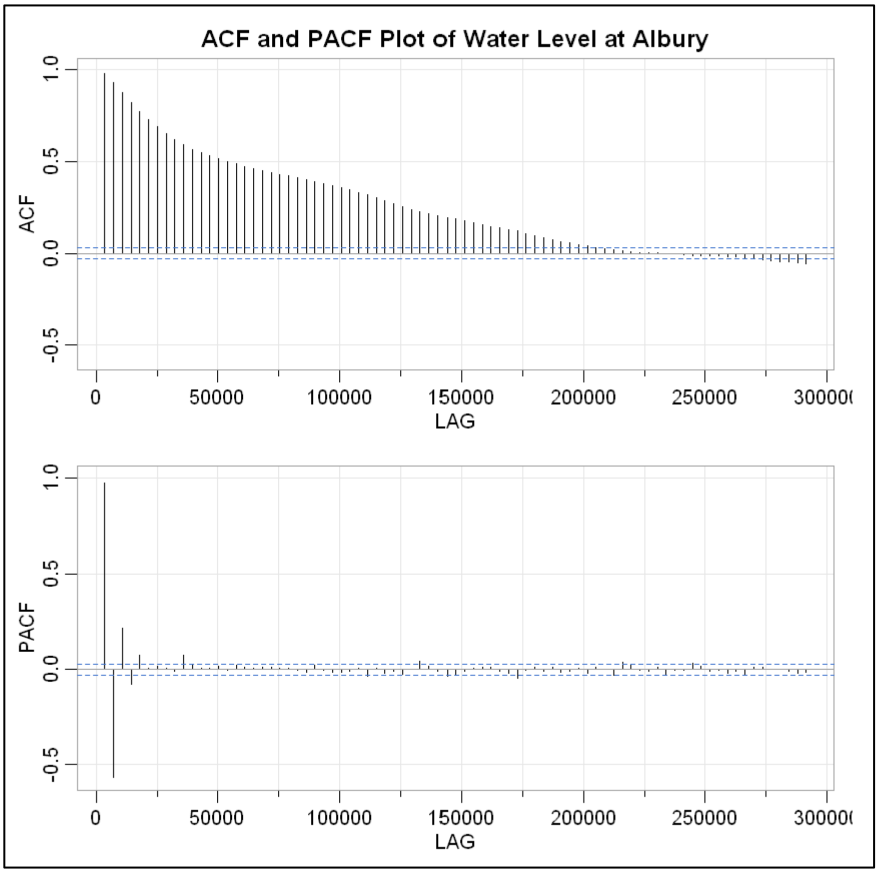

Figure 47: ACF and PACF for Water Level at Albury

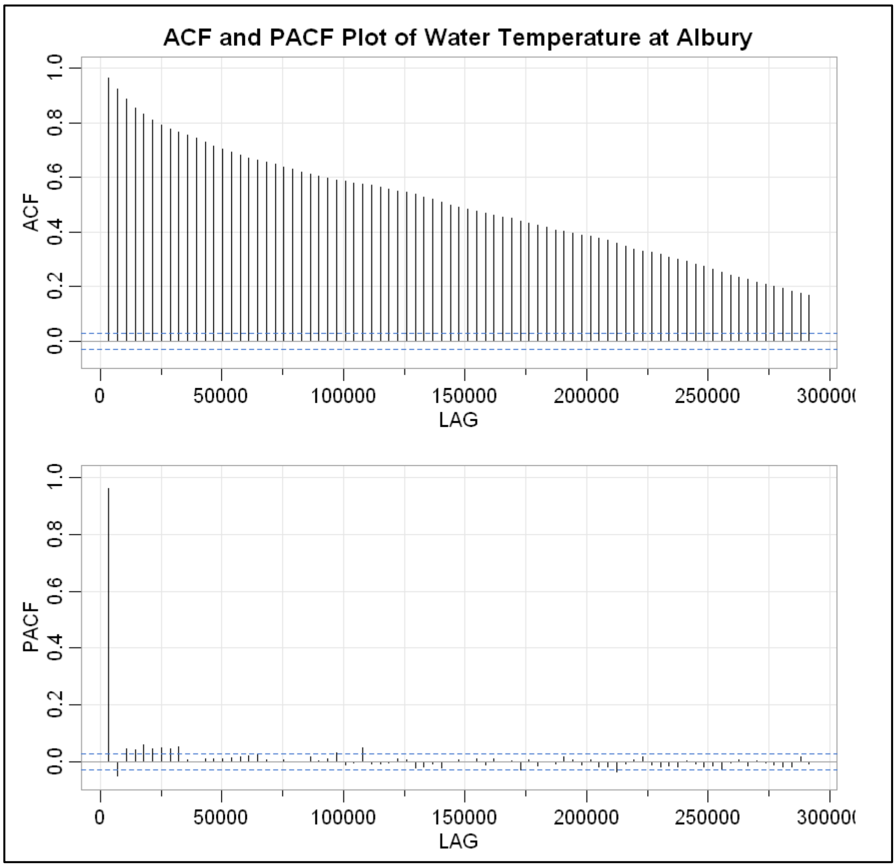

Figure 48: ACF and PACF for Water Temperature at Albury

The table data of Albury Location:

Table 21: Different ARIMA Models Trained for Salinity

ARIMA Models	AIC	AICc	BIC
ARIMA(1,0,1)	24716.84	24716.84	24736.35
ARIMA(1,0,4)	24514.41	24514.42	24553.42
ARIMA(2,0,3)	24512.76	24512.78	24551.78
ARIMA(4,0,0)	24562.63	24562.64	24595.14

Table 22: Different ARIMA Models Trained for Water Levels

ARIMA Models	AIC	AICc	BIC
ARIMA(1,0,1)	-10724.25	-10724.25	-10704.74
ARIMA(1,0,4)	-10920.07	-10920.06	-10881.06
ARIMA(2,0,3)	-10915.79	-10915.78	-10876.78
ARIMA(5,0,0)	-10924.34	-10924.32	-10885.32

Table 23: Different ARIMA Models Trained for Water Temperatures

ARIMA Models	AIC	AICc	BIC
ARIMA(1,0,1)	5103.846	5103.851	5123.353
ARIMA(1,0,4)	5034.9	5034.91	5073.91
ARIMA(2,0,1)	5094.072	5094.081	5120.082
ARIMA(5,0,0)	5051.681	5051.698	5090.695

Table 24: Different SARIMA Models Trained for Salinity

SARIMA Models	AICc
ARIMA(1,1,1)(0,1,0)[365]	25001
ARIMA(1,1,2)(0,1,0)[365]	24692.53
ARIMA(2,1,2)(0,1,0)[365]	24694.31
ARIMA(2,1,3)(0,1,0)[365]	24694.98

Table 25: Different SARIMA Models Trained for Water Levels

SARIMA Models	AICc
ARIMA(1,1,1)(0,1,0)[365]	-5112.618
ARIMA(1,1,2)(0,1,0)[365]	-5110.829
ARIMA(2,1,1)(0,1,0)[365]	-5110.046
ARIMA(2,1,2)(0,1,0)[365]	-5236.531

Table 26: Different SARIMA Models Trained for Water Temperatures

SARIMA Models	AICc
ARIMA(1,0,1)(0,1,0)[365]	7635.622
ARIMA(1,0,2)(0,1,0)[365]	7618.019
ARIMA(2,0,1)(0,1,0)[365]	7627.981
ARIMA(2,0,2)(0,1,0)[365]	7544.7

The plot of Biggara Location:

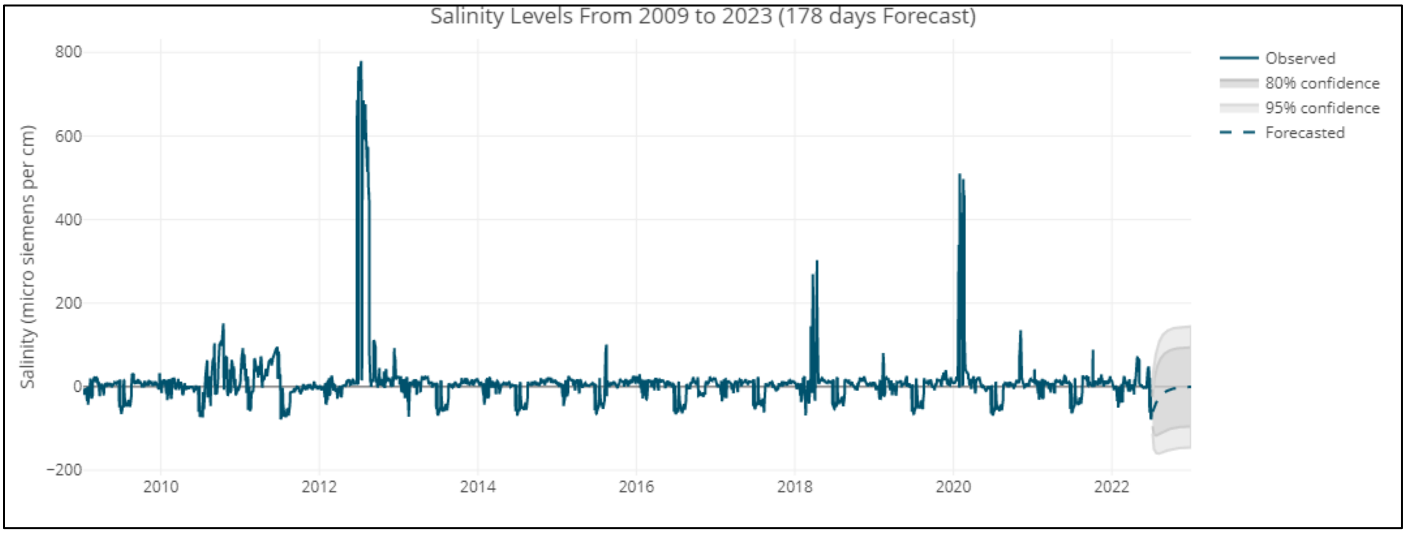

Figure 49: Salinity Levels 178 days Forecast given by the best ARIMA(4,0,1) Model

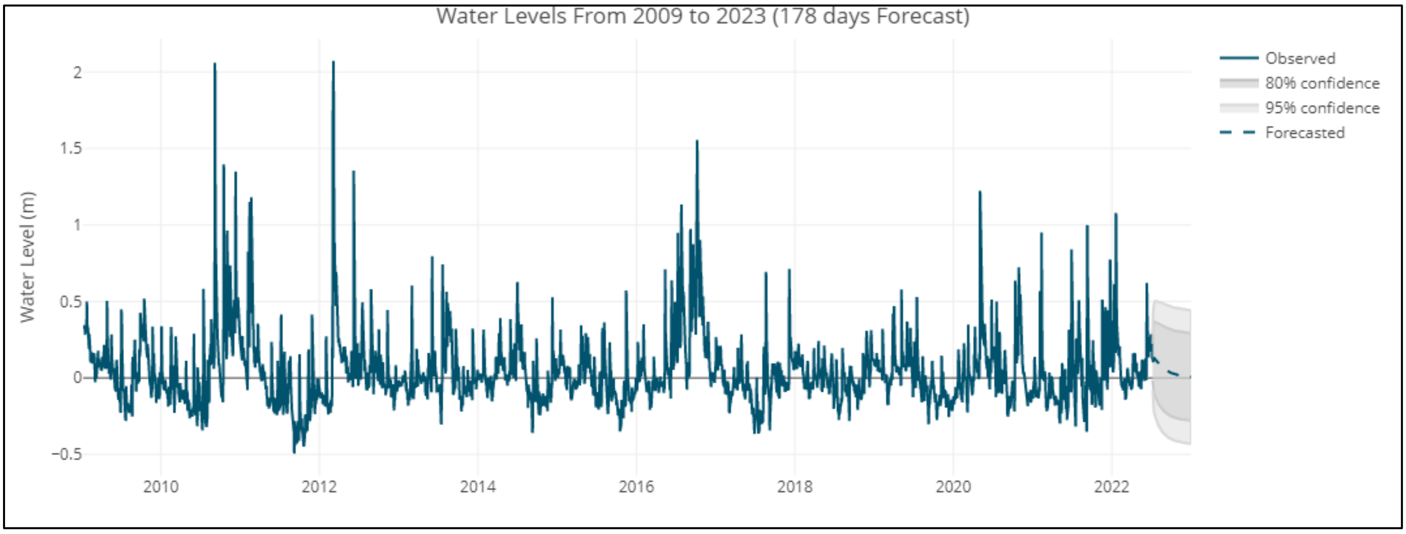

Figure 50: Water Levels 178 days Forecast given by the best ARIMA(3,0,2) Model

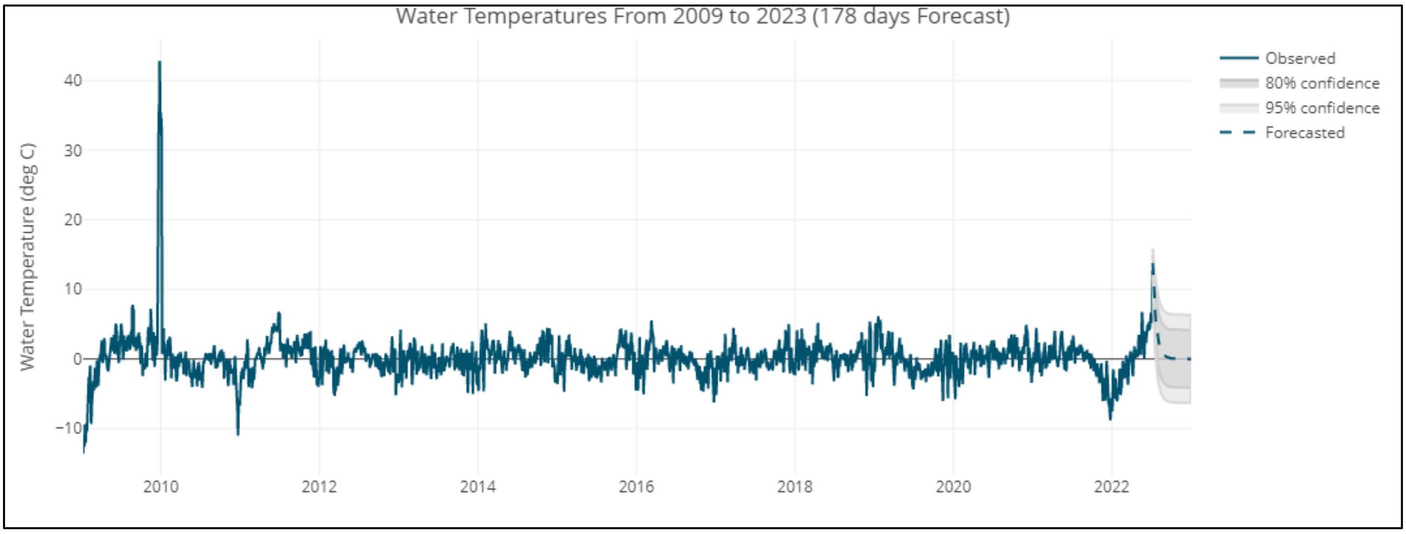

Figure 51: Water Temperatures 178 days Forecast given by the best ARIMA(1,0,4) Model

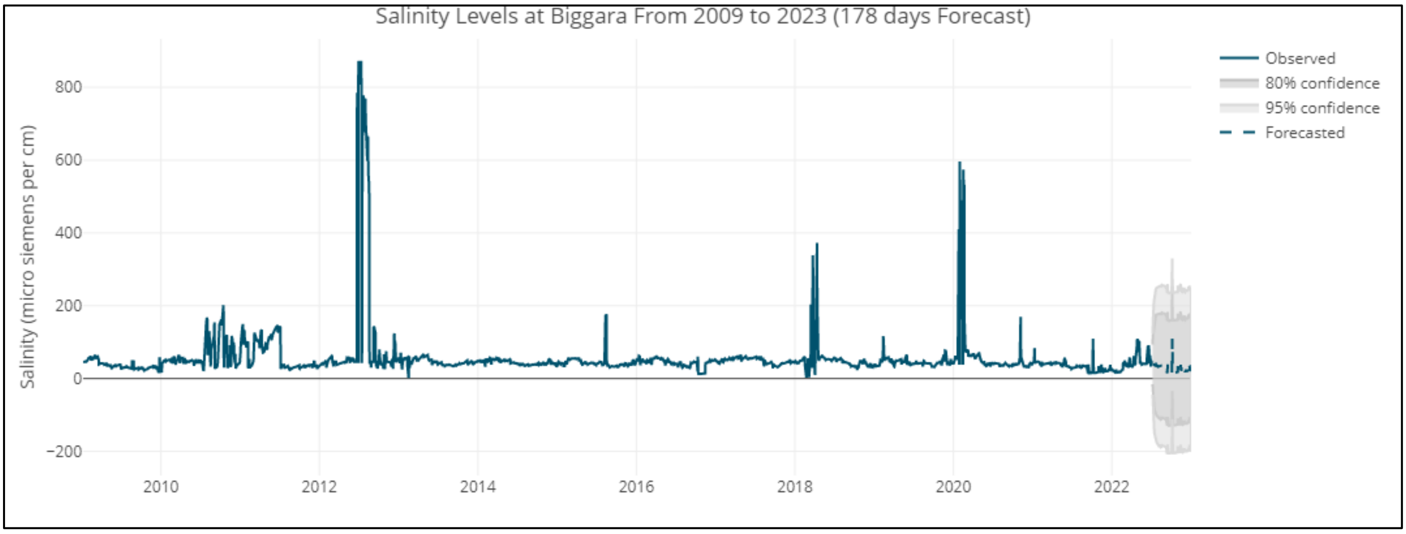

Figure 52: Salinity Levels 178 days Forecast given by the best ARIMA(1,0,2)(0,1,0)[365] (SARIMA) Model

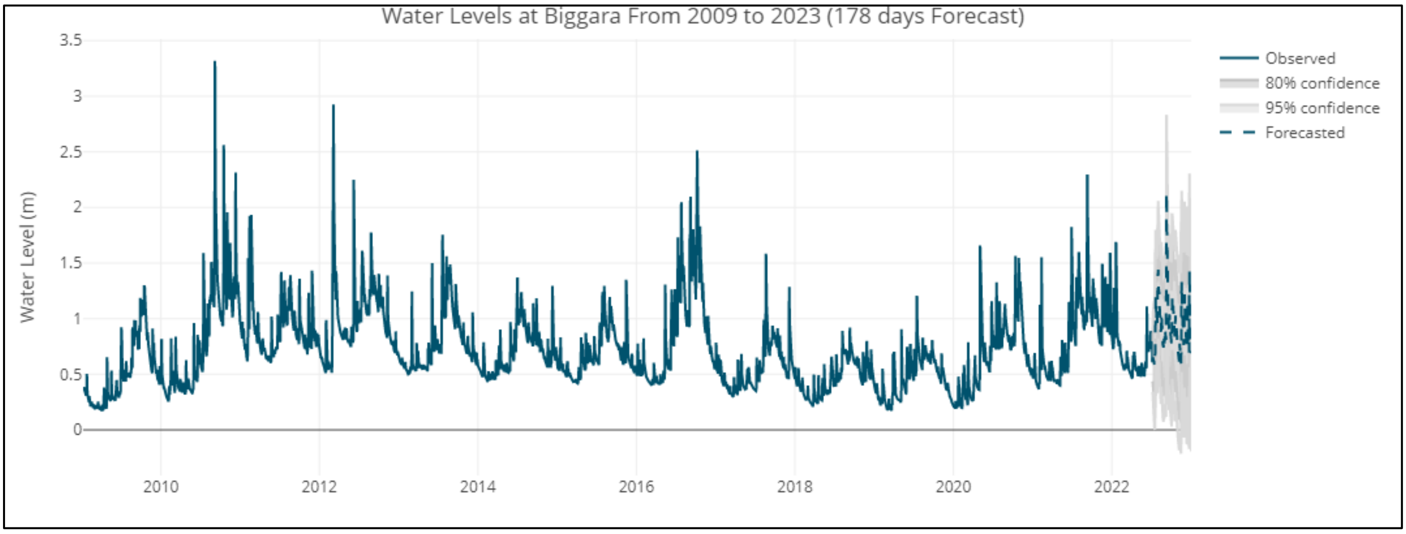

Figure 53: Water Levels 178 days Forecast given by the best ARIMA(2,1,2)(0,1,0)[365] Model

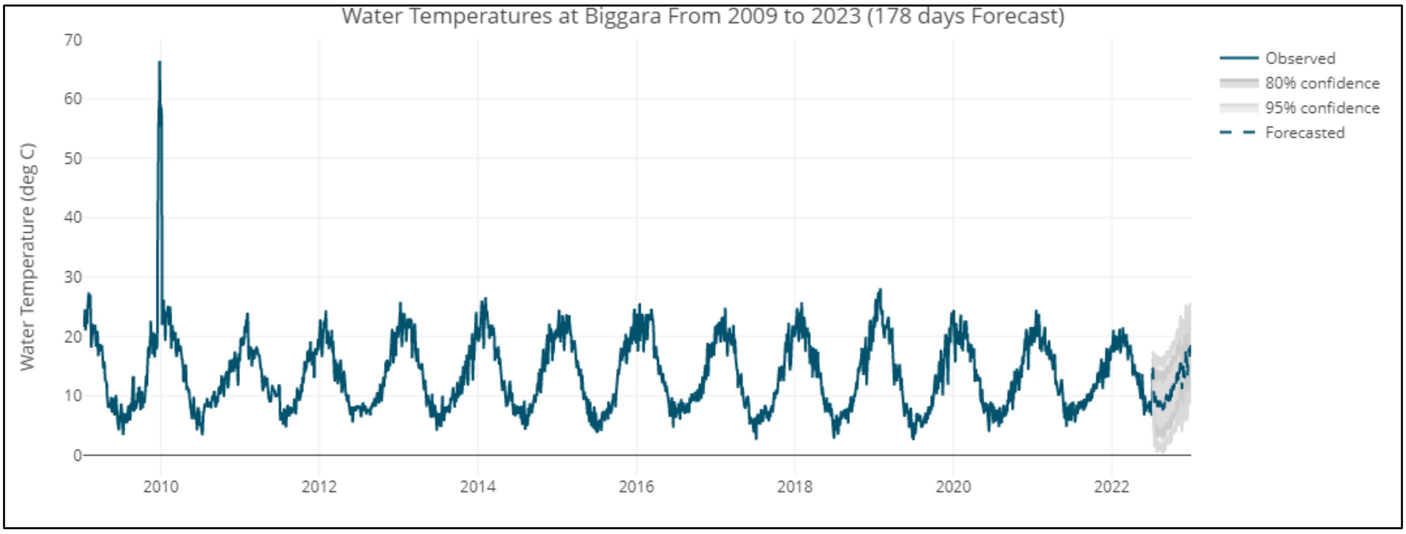

Figure 54: Water Temperatures 178 days Forecast given by the best ARIMA(1,0,2)(0,1,0)[365] Model

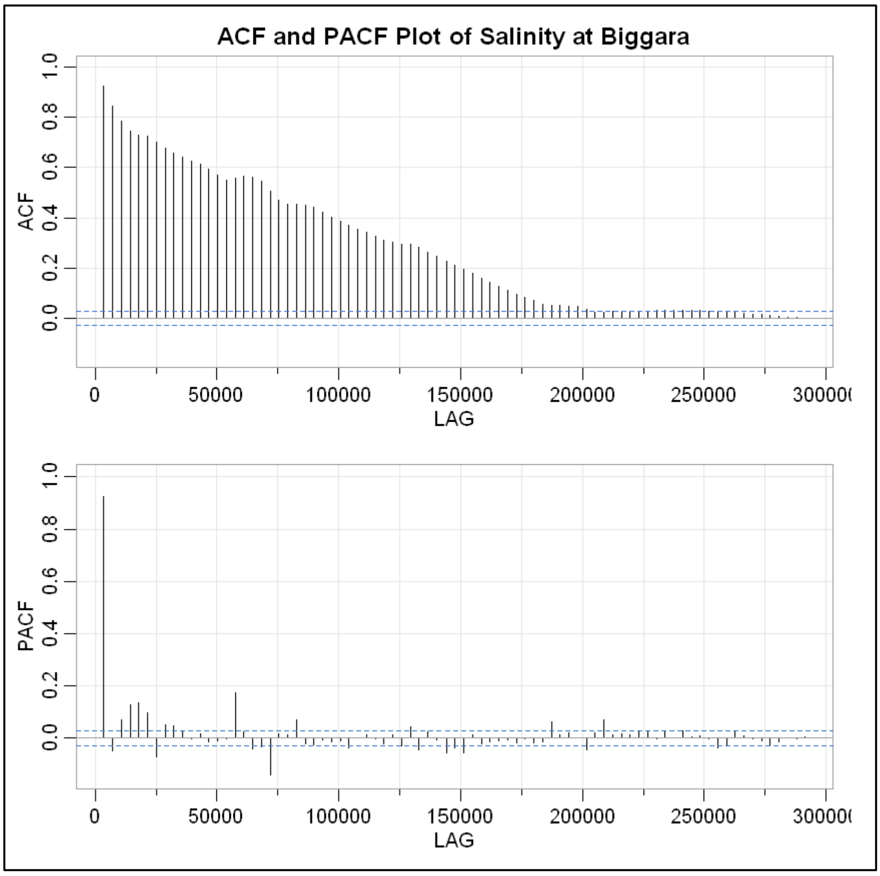

Figure 55: ACF and PACF for Salinity at Biggara

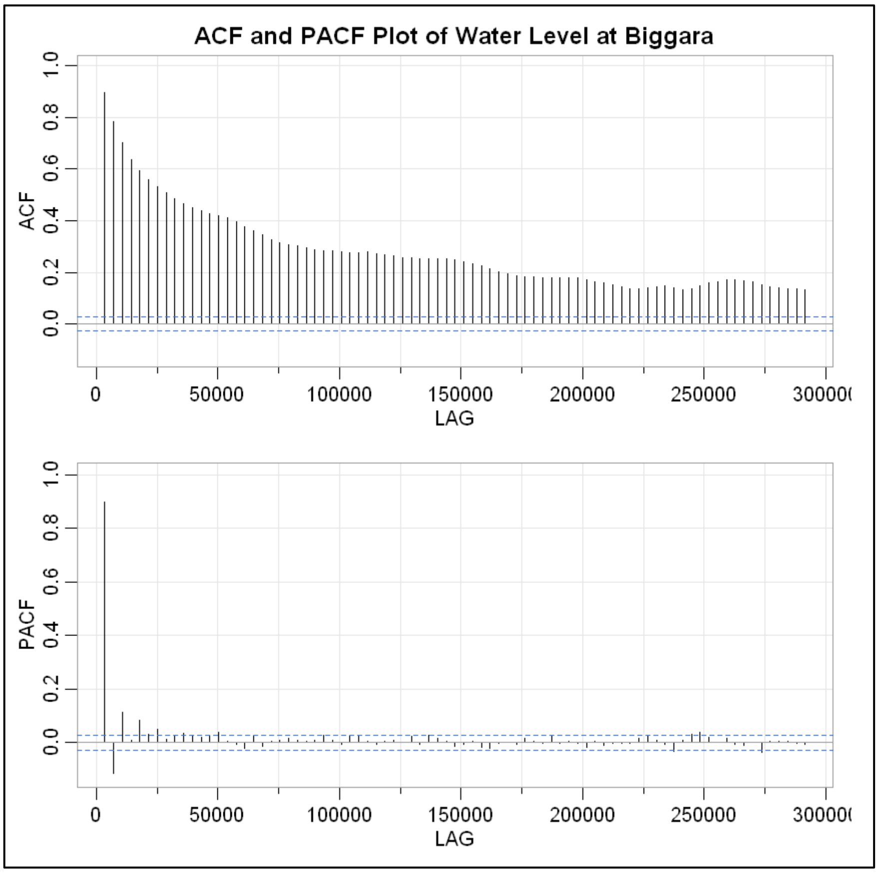

Figure 56: ACF and PACF for Water Level Biggara

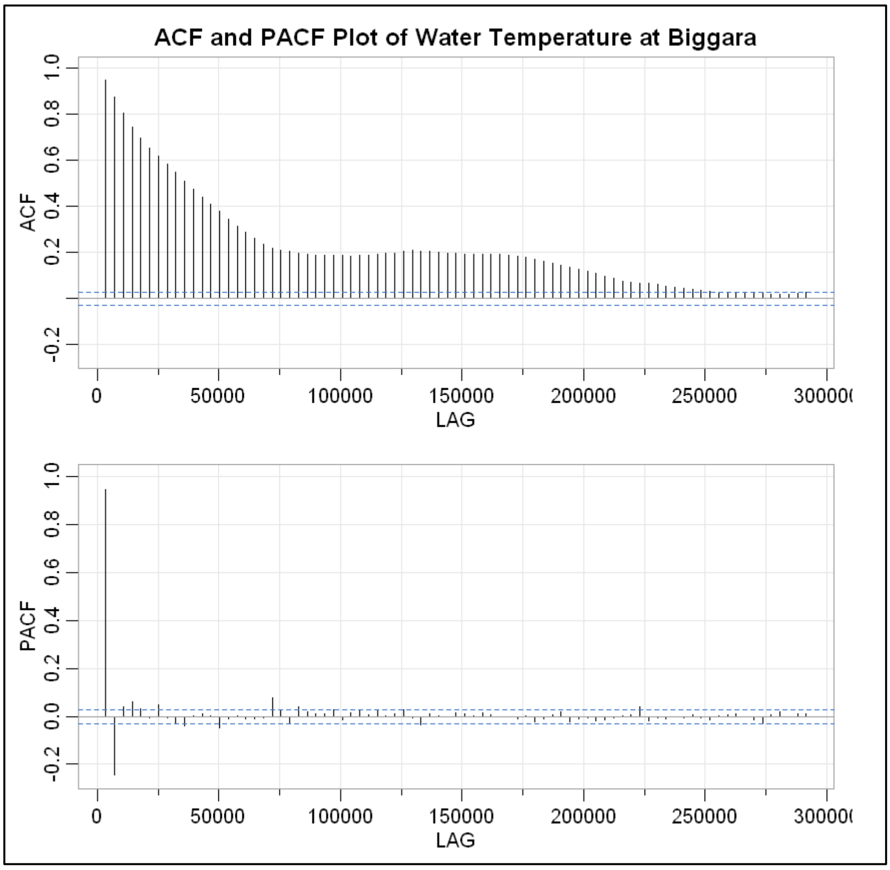

Figure 57: ACF and PACF for Water Temperature Biggara

The table data of Biggara Location:

Table 27: Different ARIMA Models Trained for Salinity

ARIMA Models	AIC	AICc	BIC
ARIMA(1,0,1)	46790.26	46790.26	46809.76
ARIMA(1,0,4)	46596.72	46596.74	46635.73
ARIMA(2,0,1)	46788.94	46788.95	46814.95
ARIMA(4,0,1)	46587.95	46587.97	46626.97

Table 28: Different ARIMA Models Trained for Water Levels

ARIMA Models	AIC	AICc	BIC
ARIMA(1,0,1)	-9142.337	-9142.332	-9122.83
ARIMA(1,0,4)	-9235.177	-9235.16	-9196.164
ARIMA(3,0,2)	-9258.58	-9258.57	-9219.57
ARIMA(4,0,1)	-9189.403	-9189.386	-9150.389

Table 29: Different ARIMA Models Trained for Water Temperatures

ARIMA Models	AIC	AICc	BIC
ARIMA(1,0,1)	13179.3	13179.31	13198.81
ARIMA(1,0,4)	13137.73	13137.75	13176.75
ARIMA(3,0,2)	13142.11	13142.12	13181.12
ARIMA(4,0,1)	13142.06	13142.08	13181.07

Table 30: Different SARIMA Models Trained for Salinity

SARIMA Models	AICc
ARIMA(1,0,1)(0,1,0)[365]	43730.12
ARIMA(1,0,2)(0,1,0)[365]	43711.71
ARIMA(2,0,1)(0,1,0)[365]	43730.07
ARIMA(2,0,2)(0,1,0)[365]	43724.3

Table 31: Different SARIMA Models Trained for Water Levels

SARIMA Models	AICc
ARIMA(1,1,1)(0,1,0)[365]	-3197.952
ARIMA(1,1,2)(0,1,0)[365]	-3467.669
ARIMA(2,1,1)(0,1,0)[365]	-3452.789
ARIMA(2,1,2)(0,1,0)[365]	-3494.824

Table 32: Different SARIMA Models Trained for Water Temperature

SARIMA Models	AICc
ARIMA(1,0,1)(0,1,0)[365]	13688.92
ARIMA(1,0,2)(0,1,0)[365]	13665.81
ARIMA(2,0,1)(0,1,0)[365]	13673.46
ARIMA(2,0,2)(0,1,0)[365]	13667.57