

## **Modeling and Prediction of Water Quality: Evidence of Benarpota Station of Sundarbans Region**

***Dulal Chandra Nandi,***

Department of Statistics, Comilla University, Bangladesh

***Ajit Kumar Majumdar,***

Department of Statistics, Jahangirnagar University, Bangladesh

***Md. Farhad Hossain,***

Department of Statistics, Comilla University, Bangladesh

***Shah Eklmur Reza,***

Department of Statistics, Comilla University, Bangladesh

Doi:10.19044/esj.2020.v16n18p242 [URL:http://dx.doi.org/10.19044/esj.2020.v16n18p242](http://dx.doi.org/10.19044/esj.2020.v16n18p242)

---

### **Abstract**

The Sundarbans contain the world's largest coastal mangrove forest, with an area of about 10,000 km<sup>2</sup> (3,900 sq mi), of which about 6,000 km<sup>2</sup> (2,300 sq mi) are located in Bangladesh and about 4,000 km<sup>2</sup> (1,500 sq mi) in India. Salinity will be the major issue in the future which could destroy the natural ecological balance. The major salinity components are electric conductivity (EC) at low tide and high tide with the corresponding chloride in low tide and high tide. This study develops a SARMA (2, 0, 2) (2, 0, 0) [4] model as it provides the minimum value of model selection criterion (AIC, BIC, SIC, RMSE) compare to all other models. It was found that the best-fitted model for EC in a high tide of study area follows a seasonal autoregressive moving average model of order (2, 0, 2)(2, 0, 0) [4] and chloride in high tide follows SARMA (1, 0, 1) (2, 0, 2) [4]. After fitting the time series model and forecasting the fitted model, the study showed that in both electric conductivity (EC) and salinity (Chloride ppm) increasing in the last three decades. EC crosses the poor water quality level which is more than 4 ppt or 2560 ppm. Salinity also crosses the poor water quality level which is more than 5 ppt or 3200 ppm. The findings of this study would be a potential contribution to make a comprehensive management plan for the long-term conservation and protection of the cultural landscape and mangrove wetlands ecosystem in the Sundarbans region.

---

**Keywords:** Surface water quality, Electric Conductivity (EC), Chloride, river system, SARMA

## 1. Introduction

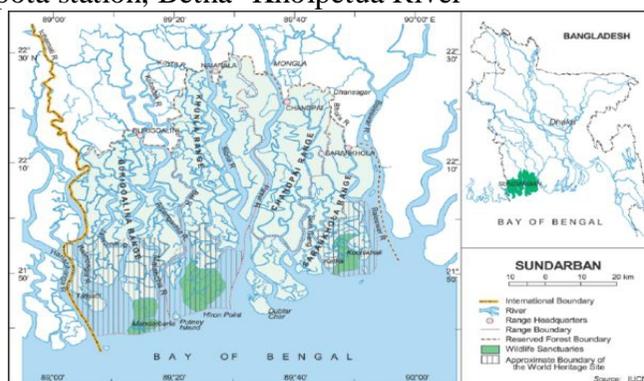
Water is used for various practices, such as drinking, agriculture, and industry. Recently, development of water sports and entertainment has greatly helped attract tourists (Jennings 2007). Among various sources of water supply, due to easy access, rivers have been used more frequently for the development of human societies. According to the Bangladesh Water Development Board (BWDB) about 230 rivers currently flow in Bangladesh (during summer and winter), although the number stated are ambiguous in some sources. This river generally flows freshwater through different small rivers and channels to The Bay of Bengal. The river serves as the main source of water for irrigation, transportation, and fishing, etc. The study of water quality of rivers is a common theme in earth sciences. To evaluate the quality of rivers two approaches are considered, including measuring the water quality components and defining the mechanism of pollution transmission (Kashefipour 2002; Kashefipour & Falconer 2002; Naseri Maleki & Kashefipour 2012; Qishlaqi et al. 2016). Evaluation of water quality is a basic stage for development of agriculture projects in terms of determination of cropping pattern, type of irrigation system, and systems of water purification for industries (Chen et al. 2017). The human activity generated contamination from agriculture, municipal and industrial activities introduces a significant amount of nutrients, toxic metals, and organic materials into the aquatic systems APHA (2006). Discharge of degradable wastewaters in the flowing waters results in dissolved oxygen concentrations due to the metabolism of pollutants and respiration of plants and phytoplankton (Drolc and Konkan, 1996). Good health meets the threshold levels of key parameters temperature, rainfall, pH, carbonaceous biochemical oxygen demand, nitrogen (N), phosphorus (P) and heavy metals (Gopal, 2005). Quantifying material flux in coastal wetlands has been a major focus of research during the past three decades, particularly due to the hypothesized importance of wetlands outwelling as an energy subsidy to coastal oceans (Childers, Day & Mckellar, 2000). Material exchange in wetlands is controlled by the climatic, geomorphologic, hydrologic, biochemical and other physical- forcing regimes of the region (Odum, 1984) often referred to as the energy signature of the ecosystem. By reviewing the time history of water quality components, investigators have attempted to estimate future values. Nowadays, by advancing soft computing techniques in most areas of water and environmental engineering, researchers have attempted to accurately analyse time series of water quality components and their internal relation (May et al. 2008; Palani et al. 2008; Haghiabi 2016a, 2016b; Jaddi & Abdullah 2017). The salinity components such as electric conductivity (EC) and chloride at low tide and high tide from Bay of Bengal create major problem for the entire ecosystems and human beings. So, it should be study by the appropriate model

for the salinity data (EC and chloride) and to predict the future state of river water. Therefore, the objective of the study is to develop statistical models to forecast water quality of the Sundarbans reserve forest considering different aspects of time series model by the best-fitted model among other models.

## 2. Research Method

### 2.1 Location and Study Area of Betna-Kholpetua River

The Sundarbans consists of three coastal districts Bagerhat, Khulna, and Satkhira. The study area is in the western part of Sundarbans in Satkhira district at Benarpota station, Betna- Kholpetua River



**Figure 1:** The Sundarbans and its river system

A complex network of streams and rivers varying considerably in width and depth intersects the entire area. Some of the big rivers are several kilometers in width (Curtis, 1933). The Sundarbans receives a large volume of fresh water from inland rivers flowing from the north and of saline water from the tidal incursions from the sea.

### 2.2 Research Data

Surface water salinity data of the study area from 1980 to 2013 were collected from Bangladesh Water Development Board (BWDB). Salinity analysis is being performed by the concerned Sub – Division office using Cyber Scan con-11 conductivity/ TDs meter. Electric conductivity (EC) and salinity (in ppm) of collected water samples are being observed here and these have been presented in the booklet. The surface water salinity data were collected for the three seasons like winter (November to January), summer (February to April) and pre-monsoon (May to June). The data collected for electrical conductivity (EC) in low tide and high tide depends on salinity (chloride) on low tide and high tide. To understand the present ecological situation, other ongoing research activities and human influences that are harmful to the mangrove wetland ecosystems and relevant for data collection, a reconnaissance survey was conducted. For the analysis of our study, we use

different statistical tools such as IBM SPSS 22.0 for conducting the performance metrics and R 3.1.2. specially library forecast and tseries is used for conducting the time series analysis.

### 2.3 Seasonal ARIMA Processes

If the data shows a strong seasonal pattern, this indicates a high correlation between values observed during the same season in consecutive years. For example, for monthly data, the random variables attached to the January of consecutive years may have a high correlation whereas the June figures may not be correlated with the January figures. In such a case, modeling the data generation process by an ARIMA process of the type considered above may require an unnecessarily large number of parameters to cope with this problem. So far, we have restricted our attention to non-seasonal data and non-seasonal ARIMA models. However, ARIMA models are also capable of modelling a wide range of seasonal data.

A seasonal ARIMA model is formed by including additional seasonal terms in the ARIMA models we have seen so far. It is written as follows:

$$\begin{array}{ccc}
 \text{ARIMA} & \underbrace{(p, d, q)} & \underbrace{(P, D, Q)_m} \\
 & \uparrow & \uparrow \\
 & \text{Non-seasonal part} & \text{Seasonal part of} \\
 & \text{of the model} & \text{of the model}
 \end{array} \tag{1}$$

where m= number of observations per year. We use uppercase notation for the seasonal parts of the model, and lowercase notation for the non-seasonal parts of the model.

The seasonal part of the model consists of terms that are similar to the non-seasonal components of the model, but involve backshifts of the seasonal period. For example, an ARIMA(1,1,1)(1,1,1)<sub>4</sub> model (without a constant) is for quarterly data (m=4), and can be written as

$$(1 - \phi_1 B)(1 - \phi_1 B^4)(1 - B)(1 - B^4)y_t = (1 + \theta_1 B)(1 + \theta_1 B^4)\epsilon_t \tag{2}$$

The additional seasonal terms are simply multiplied by the non-seasonal terms.(Hyndman, R. J., & Athanasopoulos, G. (2013)).

### 2.4 Seasonality Adjustment Procedure

Seasonal adjustment is a statistical method for removing the seasonal component of a time series that exhibits a seasonal pattern. It is usually done when wanting to analyze the trend of a time series independently of the seasonal components.

The investigation of much economic time series becomes problematic due to seasonal fluctuations. Time series is made up of four components:

$S_t$ : The seasonal component

$T_t$ : The trend component

$C_t$ : The cyclical component

$E_t$ : The error or irregular component.

The relation between decomposition of time series components

Additive decomposition:  $Y_t = S_t + T_t + C_t + E_t$ , where  $Y_t$  is the data at time  $t$ .

Multiplicative decomposition:  $Y_t = S_t * T_t * C_t * E_t$ ,

Logs turn a multiplicative relationship into an additive relationship:

$$Y_t = S_t * T_t * C_t * E_t,$$

$$\log Y_t = \log S_t + \log T_t + \log C_t + \log E_t, \quad (3)$$

An additive model appropriate if the magnitude of seasonal fluctuations does not vary with level.

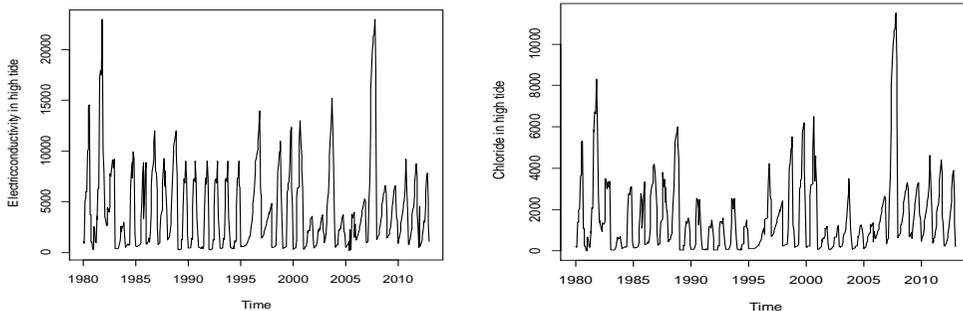
If seasonal fluctuations are proportional to the level of series, then a multiplicative model is appropriate. Multiplicative decomposition is more prevalent in economic series.

If the seasonal component is removed from the original data, the resulting values are called the “seasonally adjusted” data. For an additive model, the seasonally adjusted data are given by  $Y_t - S_t$  and for multiplicative data; the

seasonally adjusted values are obtained using  $\frac{Y_t}{S_t}$ . It can be useful to use seasonal plots and seasonal sub-series plots of the seasonal component. These help us to visualize the variation in the seasonal component over time.

### 3. Results and Discussion

#### 3.1 Time Series Plot of Electric Conductivity (EC) and Chloride in the High Tide of the Study Are



**Figure 2.** Time series plot of EC in high tide

**Figure 3.** Time series plot of Chloride in high tide

It is evident from the above plot electric conductivity and chloride in high tide is stationary with some sorts of seasonal variation over time. Two irregular variations are observed in 1983 and 2008 because there was two

cyclones and tornado type occurrence that occurred in coastal region of Bangladesh. Due to the seasonal effect, seasonal adjustment is necessary for further analysis that is considered in the following sections. Also, ADF and KPSS test suggest electric conductivity and chloride in the high tide of the study area is stationary with different p values.

**Table 1** Stationary test (ADF and KPSS) for first difference of Electric Conductivity in High Tide of Study Area

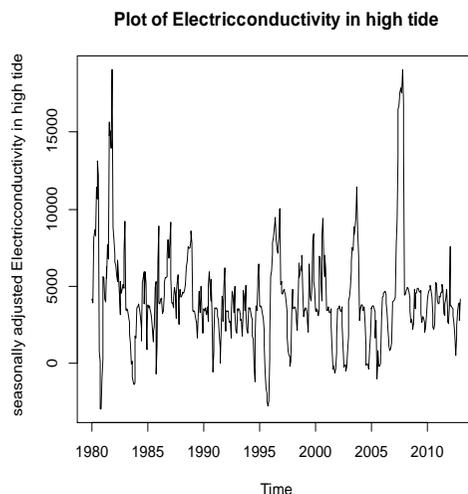
Name of the Test	Value of Test Statistic	P-Value	Alternative Hypothesis	Decision
ADF test	Dickey-Fuller = -7.1701	p-value = 0.01*	stationary	Reject Null hypothesis at 5% level of significance Accept Null at 5% level of significance
KPSS test	KPSS (Trend stationary) = 0.0968	p-value = 0.01*	stationary	

The null hypothesis is that the series is non-stationary, or contains a unit root. Decision Rule: Reject the null hypothesis if the p-value <  $\alpha = 0.05$ ; otherwise do not reject the null hypothesis. \* indicates the rejection of the null hypothesis.

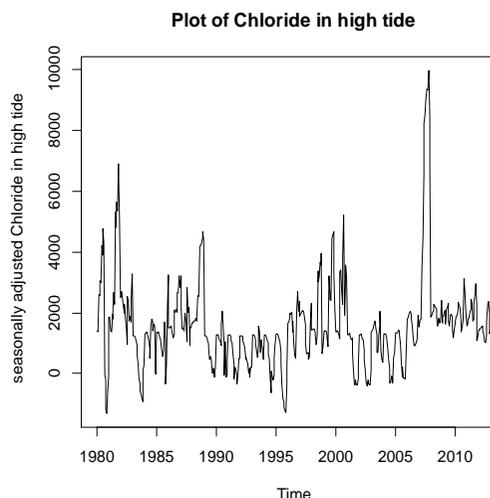
The table 1 suggests that ADF test at 5% level of significance electric conductivity in high tide is stationary. On the other hand KPSS test suggest that the electric conductivity in high tide is trend stationary at 5% level of significance.

### 3.2 Seasonally Adjusted Plot Electric Conductivity and Chloride in High Tide of the Study Area

It is observed from figure 4 and figures 5 the seasonal effect as consequences seasonal adjustment to the series is necessary.



**Figure 4:** Seasonally Adjusted Plot of EC in high tide



**Figure 5:** Adjusted Plot of Chloride in high tide

It is evident from, figures 4 and 5, between 1980 to 1985, the values EC and chloride are high and from 1985 to 1995 the variation of EC and chloride shows moderate variation. In 2009, EC and chloride have irregular variations.

### 3.3 ACF and PACF of Electric Conductivity (EC) and Chloride in a High Tide of the Study Area

To identify the appropriate model of different series preliminary graphs such as the autocorrelation function (ACF) and the partial autocorrelation function (PACF) of different orders are helpful. In the following sections, we look at the lots of the ACF and PACF.

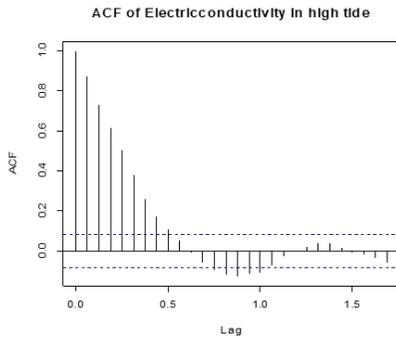


Figure 6: Plot of ACF of EC in high tide

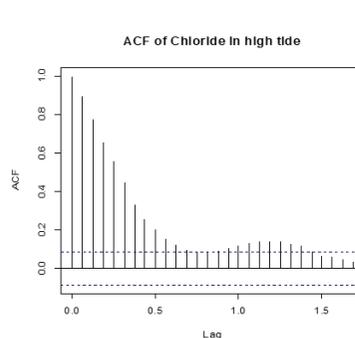


Figure 7: Plot of ACF of Chloride in high tide

Figures 6 and 7 show an exponential decay pattern of electric conductivity (EC) and chloride in a high tide of Benarpota station. But as time increases some spikes fall outside of the confidence region at 5% confidence region which indicates seasonality present in the data. Thus a SARIMA type model may be appropriate.

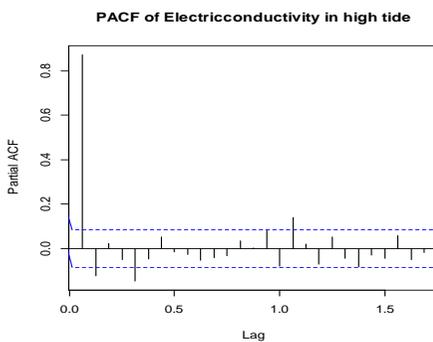


Figure 8: Plot of PACF of EC in high tide

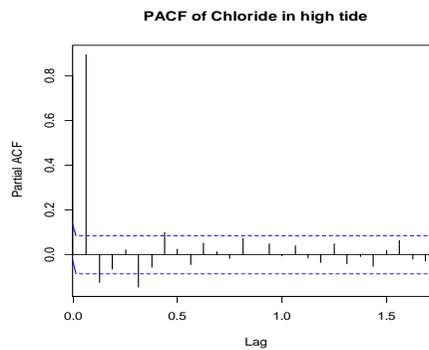


Figure 9: Plot of PACF of Chloride in high tide

It is evident from the partially autocorrelation function (PACF) (figure 8 and figure 9) that, there is a significant spike. So PACF indicates an AR model of order one but some of the spikes of PACF cross the confidence interval which indicates seasonality is responsible for such behavior. Thus the

seasonal Autoregressive Integrated Moving Average (SARIMA) type of model may be appropriate.

### 3.4 Model Accuracy Checking for Electric Conductivity (EC) and Chloride in the High Tide of the Study Area

ACF and PACF suggest the seasonal ARIMA model commonly known as the SARIMA model and evaluate the final model for electric conductivity in high tide different model selections criterion such as mean error (ME), root mean square error (RMSE), mean absolute error (MAE), mean percentile error (MPE), mean absolute percentile error (MAPE), mean absolute scaled error (MASE) are used in the following section.

Model	ME	RMSE	MAE	MPE	MAPE	MASE	ACF
SARIMA(1,1,2),(0,1,1)[4]	6.5355	16.5441	10.1935	-1.910	6.2944	0.3494	-0.002
SARIMA(1,1,2),(1,1,1)[4]	7.6710	15.9194	10.0491	-1.630	5.9756	0.3445	-0.004
SARIMA(2,0,0),(2,0,0)[4]	3.6377	15.4699	9.9702	-8.990	17.969	0.3418	0.006
.....	.....	.....	.....	.....	.....	.....	.....
SARIMA(1,1,2),(1,1,2)[4]	7.8726	15.8263	10.0525	-1.563	5.8588	0.3446	-0.005

**Table 2:** Model accuracy checking of electric conductivity in high tide

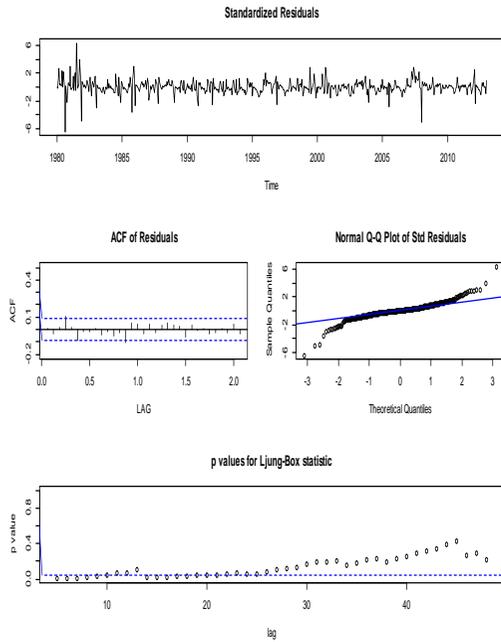
Model	ME	RMSE	MAE	MPE	MAPE	MASE	ACF
SARIMA(2,0,2),(2,0,1)[4]	1.9444	9.2021	8.7521	2.5325	10.123	0.7324	-0.003
SARIMA(1,0,2),(2,0,1)[4]	3.1682	6.5052	4.0829	1.2973	17.660	0.3858	-0.002
SARIMA(1,0,1),(2,0,2)[4]	0.7245	6.4925	4.0566	1.2170	17.745	0.3570	-0.001
....	....	...	...	...	...	...	...
SARIMA(1,0,2),(1,0,1)[4]	1.840	6.5054	4.1438	2.1950	17.648	0.3589	-0.002

**Table 3:** Model accuracy checking of chloride in high tide

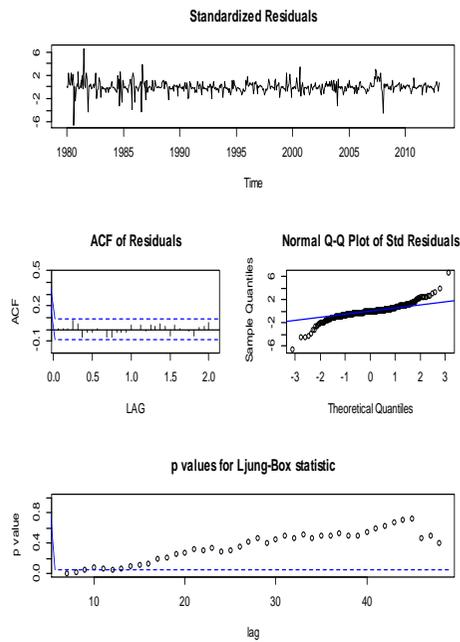
It is evident from the table 2 and 3 the electric conductivity in high tide at Benarpota station of Betna-Kholpetua river of Khulna region with the SARIMA(2, 0, 0) (2, 0, 0) [4] model is found to be the best-fitted model considering different combination orders according to model performance criteria . From the model accuracy properties, it is found SARIMA (2, 0, 0) (2, 0, 0) [4] model, the RMSE, MAE, ME, MPE, MASE and ACF are at a minimum compared to other models. The correlogram of both autocorrelation and partial autocorrelation for chloride in high tide gives the impression that the residuals estimated from SARIMA (1, 0, 1) (2, 0, 2) [4] model are purely random. Hence, there is no need to look for another SARIMA model.

### 3.5 Diagnostic Checking of the Fitted Model

Diagnostic checking of the fitted model to evaluate the appropriateness of the model.



**Figure 10:** Standard Residuals plot of EC in high tide

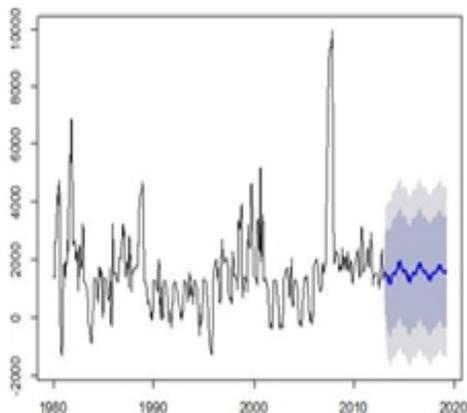


**Figure 11:** Standard Residuals plot of Chloride in high tide

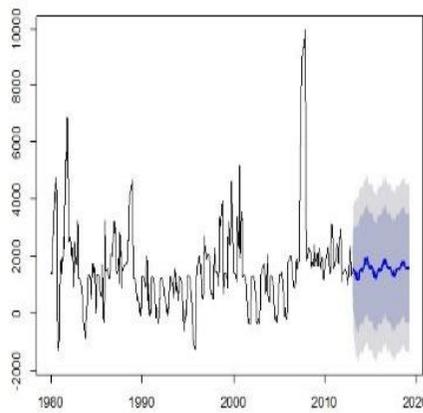
It is evident from the residual plots (figure 10 and figure 11), the noise follows IID (0, 1) which indicates our fitted model SARIMA (2, 0, 0) (2, 0, 0) [4] is the best-fitted model for EC and chloride in the high tide. It is observed that, SARIMA1 (1, 0, 1) (2, 0, 0)[4] model is the best-fitted model considering other model specifications criteria.

The Forecasted Plot of Electric Conductivity (EC) in a High Tide of Benarpota Station

We have identified that SARIMA (2, 0, 0), (2, 0, 0) [4] performs better than any time series model. To evaluate the performance of the model, visual inspection is necessary. And we plot forecasted electric conductivity in high tide in figure 10 and chloride in figure 11.



**Figure 12:** SARIMA (2, 0, 0), (2, 0, 0) [4] of EC in high tide



**Figure 13:** SARIMA(1, 0, 1) (2, 0, 2) [4] of Chloride in high tide

Figures 12 and 13 suggest that, the forecasted seasonally adjusted values of electric conductivity in the high tide for Benarpota station. The shaded region represents the 90 percent confidence interval. The forecasted values are found to be 4472.152, 4578.796, 4564.223... 5312.214, 5769.76... 5544.860, 5546.691, 5498.152-mile Siemens per centimeter ( $\mu S/cm$ ) and the forecasted seasonally adjusted of chloride in the high tide for Benarpota station. The shaded region represents the 90 percent confidence interval. The forecasted values are found to be 1472.152, 1578.796, 1564.223... 2312.214, 2769.76... 4544.860, 4546.691, and 4498.152 parts per million (ppm). This study will compare the forecasted value with the actual value for the year 2014 to 2019 in further. If these values are basic identical then we may conclude that the fitted model is best fitted model.

### 3.7 Developing Model of Electric Conductivity and Chloride in High Tide

Comparing the forecasting performance of competitive models is one of the most important aspects of any forecasting exercise. The literature on forecast has proposed a good number of models. This gives a wide range of models to choose from among the models. This section describes the most important forecast evaluation models.

**Table 4.** Estimates of Parameters of SARIMA (2, 0, 0) (2, 0, 0) [4] Model for EC High Tide

Coefficients	ar1	ar2	sar1	sar2
Value of Coefficient	0.9720	-0.1153	-0.1862	-0.1394
Standard Errors	0.0086	0.0137	0.0156	0.0085
Intercept	4066.3694	355.2166		
$\sigma^2$ Estimated	239.320			
Log likelihood		:-46.368		

The above table suggest the coefficients of SARIMA (2, 0, 0) (2, 0, 0) [4] with AIC=9285.75, AICc=9285.95 and BIC= 9311.41. These coefficients of the model can be applied for forecasting the future values. Also, too small value of log likelihood suggests that the outcome is more reliable and the model is well fit for the under fitting cases.

**Table 5.** Estimates of Parameters of Seasonal ARIMA (1, 0, 1) (2, 0, 2)[4] Model for Chloride in High Tide

Coefficients	ar1	ma1	sar1	sar2	sma1	sma2
Value of Coefficient	0.8822	0.1278	-0.3791	-0.2505	0.3944	0.2205
Standard Errors	0.0224	0.0450	0.1121	0.0785	0.1197	0.0857
Intercept	1539.4261	355.2166				
$\sigma^2$ Estimated	239.320	246.5218				
Log likelihood			: -41.429			

The above table suggest the coefficients of SARIMA (1, 0, 1) (2, 0, 2) [4] model with AIC=8303.53, AICc=8303.8 and BIC=8337.69. These coefficients of the model can be applied for forecasting the future values. Besides, too small value of log likelihood suggests that the outcome is more reliable and the model is well fit for the under fitting cases

## Conclusion

This study observes electric conductivity (EC) in high tide over time and the pattern electric conductivity is highly affected by seasonality. Augmented Dickey-Fuller test and KPSS test which shows that the series are stationary. Autocorrelation function (ACF) for EC in high tide decay exponentially and few spikes falls outside confidence interval and partially autocorrelation function (PACF) has a significant spike which suggests seasonality is present in the data. So we consider the seasonal ARIMA model commonly known as the SARIMA model. This study considers all possible combinations of all orders of AR and MA. After conducting the iterative procedure we find SARMA (2, 0, 2) (2, 0, 0) [4] provides the minimum value of model selection criterion (AIC, BIC, SIC, RMSE) compare to all other models. It is clear that the best-fitted model for EC in a high tide of study area follows a seasonal autoregressive moving average model of order (2, 0, 2)(2, 0, 0) [4] and chloride in high tide follows SARMA (1, 0, 1) (2, 0, 2) [4]. Forecasted values of all the surface water variables show an alarming situation. Electric conductivity, as well as chloride, is also increasing over time. This means that according to this set of data, the longer, EC and salinity have a positive impact on surface water. The findings of this study would be a potential contribution to make a comprehensive management plan for the long-term conservation and protection of the cultural landscape and mangrove wetlands ecosystem in the Sundarbans region. This study has been done for the years 1980 to 2013 and further investigate the performance of the

SARIMA model should be tested for others salinity data after 2013 time period.

### References:

1. APHA (2006): Standard Methods for the Examination of Water and Wastewater. Asian Wetland Bureau (AWB, 1991). Bangladesh Forestry III project, Environment Component, Project Preparation Report.
2. Chen, X., Chen, Y., Shimizu, T., Niu, J., Nakagami, K. i., Qian, X., Jia, B., Nakajima, J., Han, J. & Li, J. 2017 Water resources management in the urban agglomeration of the Lake Biwa region, Japan: an ecosystem services-based sustainability assessment. *Sci. Total Environ.* 586 (Suppl. C), 174–187.
3. Childers, D. L. Day, L. W. Mckellar, H. N. (2000): Twenty more years of marsh and estuarine flux and Controversies in tidal and marsh ecology. Pp. 385 – 414
4. Curtis, S. J. (1933): Working plan for the forest of the Sundarbans division for the period from April 1, 1931, to March 31, 1951. Vol.1, Bengal Government Press, Calcutta, India
5. Drolc, A. and Konkan, J. Z. Z. (1996): Water quality modeling of the river Sava, Slovenia. *Water Res.* 30(5): 2587 – 2592.
6. Gopal, B., (2005). Does inland aquatic biodiversity have a future in Asian developing countries *Hydrobiologia*: 69–75.
7. Gujrati N. Damodarand Sangeetha (2007), *Basic Econometrics*, 4<sup>th</sup> Edition; McGraw-Hill.
8. Haghiabi, A. H. 2016a Modeling river mixing mechanism using data driven model. *Water Resour. Manage.* 31(3), 811–824.
9. Haghiabi, A. H. 2016b Prediction of longitudinal dispersion coefficient using multivariate adaptive regression splines. *J. Earth Syst. Sci.* 125 (5), 985–995
10. Hyndman, R. J., & Athanasopoulos, G. (2013). *Forecasting: principles and practice*. Retrieved from <http://otexts.org/fpp/> (this is an online book).
11. Jaddi, N. S. & Abdullah, S. 2017 A cooperative-competitive master-slave global-best harmony search for ANN optimization and water-quality prediction. *Appl. Soft Comput.* 51, 209–224.
12. Jennings, G. 2007 *Water-based Tourism, Sport, Leisure, and Recreation Experiences*. Elsevier, Oxford.
13. Kashefipour, S. M. 2002 *Modelling Flow, Water Quality and Sediment Transport Processes in Reverine Basins*. PhD thesis, Cardiff University, Cardiff

14. Kashefipour, S. M. & Falconer, R. A. 2002 Longitudinal dispersion coefficients in natural channels. *Water Res.* 36 (6), 1596–1608.
15. May, R. J., Dandy, G. C., Maier, H. R. & Nixon, J. B. 2008 Application of partial mutual information variable selection to ANN forecasting of water quality in water distribution systems. *Environ. Model. Softw.* 23 (10–11), 1289–1299
16. Naseri Maleki, M. & Kashefipour, S. M. 2012 Application of numerical modeling for solution of flow equations and estimation of water quality pollutants in rivers (Case study: Karkheh River). *Civil Environ. Eng.* 42.3 (68), 51–60.
17. Odum, (1984): The detritus-based food web of an estuarine mangrove community, In *Estuarine Research*; (ed.) Cronin L. E, Academic Press, New York, Pp. 265 – 286.
18. Palani, S., Liong, S.-Y. & Tkalich, P. 2008 An ANN application for water quality forecasting. *Mar. Pollut. Bull.* 56 (9), 1586–1597.
19. Qishlaqi, A., Kordian, S. & Parsaie, A. 2016m Hydrochemical evaluation of river water quality – a case study. *Appl. Water Sci.* 7(5), 2337–2342.