



Long-term forecasting in a coastal ecosystem: case study of a Southern restored Mediterranean lagoon: The North Lagoon of Tunis

Nadia Ben Hadid^{1,2} · Catherine Goyet² · Naceur Ben Maiz³ · Abdesslem Shili¹

Received: 15 March 2021 / Revised: 15 March 2022 / Accepted: 15 March 2022 / Published online: 30 March 2022
© The Author(s) 2022

Abstract

Eutrophication episodes are common in freshwater and coastal environments, causing significant damage to drinking water and aquaculture. Predictive models are efficient approaches for anticipating eutrophication or algal blooms because ecologists and environmentalists can estimate water pollution levels and take appropriate precautionary steps ahead of time. In aquatic ecosystems, chlorophyll-*a* (Chl-*a*) can be employed as a water quality indicator, revealing information on man-made physical, chemical, and biological changes variations or seasonal interventions. In the present study, a Seasonal AutoRegressive Integrated Moving Average (SARIMA) model was developed to forecast monthly Chl-*a* concentrations in the North Lagoon of Tunis, a Ramsar site, and one of the most important lagoons in Tunisia, using approximately three decades of historical data, starting from January 1989 to April 2018. SARIMA (2,0,2)(2,0,2)₁₂ was found to be the best-fitting model for Chl-*a* forecasting in the North Lagoon of Tunis. The resulting SARIMA model was validated with actual monthly Chl-*a* concentrations from our last observations. Furthermore, with only one input variable, the SARIMA model showed greater applicability as a eutrophication early warning system using actual past Chl-*a* data. Finally, the SARIMA model was utilized to anticipate Chl-*a* levels from May 2018 to December 2025 as an early warning system for ecosystem managers and decision-makers for next generations.

Keywords North Lagoon of Tunis · Eutrophication · Chlorophyll-*a* · SARIMA · Forecasting

Introduction

Eutrophication of coastal ecosystems is a global problem, especially in coastal lagoons (Nixon 1995; Cloern 2001). Increased nutrient inputs amplified by urbanization, agriculture or industry, lead to complex direct and indirect natural ecosystem responses (Schramm 1999; Viaroli et al. 2008).

Anoxic crises, toxic algal blooms, loss of biodiversity, and, more broadly, deterioration of ecosystem functions and services are all consequences of anthropogenic eutrophication (Cloern 2001; Zaldívar et al. 2008a,b).

For decades, Mediterranean coastal lagoons have been exposed to anthropogenic eutrophication and are among the most vulnerable systems to such pressures (Viaroli et al. 2005; Zaldívar et al. 2008a, b; Souchu et al. 2010). They are influenced by intensively farmed and densely populated watersheds, particularly during the summer, when the Mediterranean Sea becomes a major tourist attraction across the world (Vogiatzakis et al. 2006). In addition to urban pressure, these ecosystems are particularly vulnerable to eutrophication due to their transitional condition (de Jonge et al. 2002; Newton et al. 2014), The Palavasian complex, for example, is composed of eight lagoons located along the French Mediterranean coast that have seen extensive eutrophication over the last four decades, owing mostly to nutrient over-enrichment from constant sewage discharges (Leruste et al. 2016). According to García-Ayllón (2017), the Mar Menor lagoon, located in the east of Spain's Region

✉ Nadia Ben Hadid
nadiabenhadid01@gmail.com

¹ Département Génie Halieutique et Environnement, UR03AGRO1, Unité de Recherche Ecosystèmes et Ressources Aquatiques, Université de Carthage, Institut National Agronomique de Tunisie, 43 Avenue Charles Nicolle, 1082 Tunis, Tunisia

² ESPACE-DEV, Univ Montpellier, IRD, Univ Antilles, Univ Guyane, Univ Réunion, Montpellier France Laboratoire IMAGES-ESPACE-DEV, Univ. Perpignan Via Domitia, Perpignan, France

³ Société Al Buhaira de développement et d'investissement, Rue de Rodrigo de Freitas, 1053, Les Berges du Lac, Tunis, Tunisia

of Murcia, has undergone a considerable process of severe anthropization during the preceding five decades. The rapid population increase of new jellyfish species, which reached over 100 million specimens every summer, was one of the most surprising indicators (Robledano et al. 2011). Thau Lagoon is another interesting example of Mediterranean coastal lagoon eutrophication. Thau lagoon is a coastal environment located on the Mediterranean French coast that is recognized for sustaining traditional shellfish farming in France. It has been subjected to eutrophication, which has resulted in large anoxic occurrences linked with enormous shellfish mortalities (Derolez et al. 2020). The Ghar el Melh lagoon, located on Tunisia's north Mediterranean coast, is an excellent model for studying the eutrophication process in coastal Mediterranean lagoons. According to Shili et al. (2002), this lagoon had many dystrophic crises between 1994 and 1996. Furthermore, Turki et al. documented the proliferation of toxic algae species in the lagoon in 2007, including *Kryptoperidinium foliaceum*, *Prorocentrum micans*, and *Anabaena sp.*

Despite the availability of a large number of studies on eutrophication processes in natural aquatic habitats (García-Pintado et al. 2007; Trabelsi-Bahri et al. 2013; Derolez et al. 2020), very little progress has been made in anticipating and preventing eutrophication and its consequences in these ecosystems.

Targeted monitoring of major water pollution indicators can detect episodes of eutrophication in coastal areas. For example, chlorophyll-*a* (Chl-*a*) is the most significant pigment in aerobic photosynthetic species, and its measurement is used to evaluate the quantity of phytoplankton biomass in the water, the potential of algal bloom events, and, as a result, the degree of environmental eutrophication (Yu et al. 2019).

Traditionally, Chl-*a* in-situ sampling and measurement strategies have necessitated regular monitoring and laboratory studies (Oh et al. 2007). Due to several constraints, such as (i) field monitoring costs, (ii) staff availability and resources, (iii) field safety issues, and (iv) long time intervals between data collection, reporting, and public notification, these programs have limited capacity for environmental managers to adequately monitor and respond to eutrophication episodes (Oh et al. 2007). As a result, in order to decrease the cost and time required for in-situ monitoring and laboratory studies, an early-warning proactive approach to Chl-*a* is required to avoid or mitigate the incidence of eutrophication and, ultimately, to reduce its risk on water bodies (Oh et al. 2007).

Forecasting potential algal bloom occurrences is a critical emergency management tool for monitoring water quality and protecting the aquatic system (Villanoy et al. 2005; Dippner et al. 2011; Recknagel et al. 2013, Recknagel et al. 2014). This has piqued ecologists' curiosity, resulting in the

creation of eutrophication or algal bloom predictive modeling techniques (Chen et al. 2015).

In general, these models are classified into two types: physical-based approaches and data-driven approaches. Physical-based simulations reveal the fundamental mechanisms for algae growth and outbreak (Hood et al. 2006; Zhang et al. 2013). Many physical, chemical, and biological processes, however, remain unknown due to the complexity of marine ecosystems and the wide variety of factors that require calibration. Such models are mostly used for scenario analysis rather than prediction (Recknagel et al. 2014).

Data-driven models based on autoregression, multivariate regression, piecewise regression, or Artificial Neural Networks (ANN) have been rapidly developed and applied to eutrophication and algal bloom forecasting as an alternative. When an explicit function to characterize the modeled system is not provided, ANNs are effective in predicting data (Chen et al. 2015). In 2007, Oh et al. have applied two ANN models to identify temporal phytoplankton community patterns in the Daechung reservoir (Korea). Wang et al. built an ANN model in 2010 to forecast cyanobacterial blooms based on meteorological variables for use in early warning systems in Dianchi Lake (China). However, because the construction of an ANN model is still based on trial and error, it lacks generic direction (Chen et al. 2015). Furthermore, ANN models are prone to falling into a local optimum (Hecht-Nielsen 1989). Regression models are set to identify governing factors and establish their approximate relationship to dependent variables (Cui et al. 2007; Onderka 2007; Davis et al. 2009; Wilhelm et al. 2011). In 1982, Ouchi had applied principle component analysis to develop a multivariate prediction model for algal biomass in the Northern Hiroshima Bay (Japan). In 2007, Lui et al. developed a vector-based autoregression model for algal bloom forecasting in Hong Kong waters. These models often have a good prediction performance for stationary systems. Eutrophication, on the other hand, is event-driven and characterized by non-stationary characteristics (Hasting 2001; Onderka 2007; Paerl and Huisman 2008; Recknagel et al. 2013). More significantly, aquatic ecological data are typically sparse and incomplete, with either hydro-environmental or biological factors missing (Donner 1982). Despite the availability of tools for dealing with missing and noisy data, such data makes developing multivariate regression models challenging (Donner 1982).

The Box and Jenkins (1976) method has been used widely in a variety of disciplines. The method is a dynamic, computer-based iterative process that generates an autoregressive, integrated moving average model (ARIMA), optimized for seasonal and trend factors (Gaynor and Kirkpatrick 1994).

The Box and Jenkins method is known for: 1) its capacity to deal with complex processes; 2) its adaptability in processing dependent time series data; 3) its advanced

mathematical and statistical techniques and 4) the ease with which it is implemented. Box and Jenkins methodologies often provide the most reliable forecasting models for any data collection (Gaynor and Kirkpatrick 1994).

Armstrong's (1985) comparative test on the ranking of extrapolation methods (from the highest rank as 1 to the lowest rank as 5) for both short and long range rated Box and Jenkins methods as 1.5 for short-range forecast accuracy and 2 for long-range forecast accuracy (Lu and Abou Rizk 2008).

Box and Jenkins approaches often start with the most recent observations and then examine recent forecasting errors to make appropriate modifications for future periods (Lu and Abou Rizk 2008). In doing so, they allow timely adjustments of error levels and provide a more flexible imitation of a particular complex trend or seasonality (Lu and Abou Rizk 2008). In addition, Box and Jenkins models are capable of dealing with those dependent time-series data which are not considered suitable for other methods. For example, a regression model has a standard assumption that the error term should be statistically independent. In reality, many time series related data are dependent or are correlated with each other (Lu and Abou Rizk 2008).

The method begins with the assumption that the process that created the time series may be estimated using an ARMA model if it is stationary, or an ARIMA model if non-stationary (Lu et al. 2008).

Time series analysis is a powerful tool for assessing water quality indicators and predicting future changes. Several studies have used ARIMA model. For instance, for predicting water levels in Lake Malawi (Makwinja et al. 2017), or water salinity in Apalachicola bay in Florida (Sun and Koch 2001) and for forecasting sulphur dioxide in Tehran (Hasanzadeh et al. 2009). Chen et al. (2015) have designed and proved the efficiency of an ARIMA model for predicting daily Chl-*a* concentrations in Taihu Lake (China) in comparison to a multivariate linear regression model (MVLRL). Seasonal changes are common in natural systems generating variations in biomass and ecosystems structures (Prista et al. 2011). The seasonal, autoregressive, integrated moving average (SARIMA) model is composed of ARIMA model, including a seasonal component of the time series data. SARIMA is very frequently used for monthly time series that exhibit a seasonal pattern (Prista et al. 2011). Chl-*a* is a parameter known to be related to temperature (Tizro et al. 2014), which have seasonal characteristics. For that reason, a SARIMA model is implemented to handle the characteristics of seasonal variations, which improves the prediction accuracy (Tizro et al. 2014). In this context, the North Lagoon of Tunis is particularly an interesting case for a Mediterranean coastal lagoon eutrophication study. It is a South Mediterranean shallow coastal lagoon located in the north of Tunisia. This environment has a long history of pollution and was formerly one of the world's most contaminated lagoons

(Harbridge et al. 1976; Affi et al. 2008; Armi et al. 2008). To limit the anthropogenic input, the lagoon has been the subject of a restoration project that has been effective in greatly decreasing environmental deterioration. This operation was implemented in 1985. Before this project was conducted, the above-mentioned lagoon was the main outlet for solid waste and domestic/industrial wastewaters stemming from the city of Tunis (the capital of Tunisia). Dystrophic events, anoxia, fish mortality, and red waters were reported in the environment of the North Lagoon of Tunis (Shili et al., 2002). This coastal area remains relatively sensitive and its environmental monitoring is necessary to ensure the sustainability of its proper ecological functioning. In the light of this, the North Lagoon of Tunis was considered as the study case. Time series predictive modeling of key environmental parameters are good decision-making tools for sustainable management of this vulnerable coastal area in the center of an urban agglomeration.

The objective of this present study was indeed to investigate the applicability of a SARIMA model in algal bloom forecasting using Chl-*a* concentrations as a eutrophication indicator and a period of approximately three decades (January 1989–April 2018) of retrospective Chl-*a* data in the North Lagoon of Tunis.

Materials and methods

Study area

The North Lagoon of Tunis is located in the Southern Mediterranean Sea to the east of the city of Tunis (36°45'-36°52' N and 10°10'-10°20' E). Water is exchanged with the Mediterranean Sea via the Kheireddine canal, which is 800 m long, 40 m wide, and has a mean depth of roughly 2.5 m (Ben Charrada 1992). To reduce the eutrophication, the hydrological system was changed by installing a supplemental separation dam (Fig. 1), which led in a significant gain in biodiversity (Ben Charrada 1992; Shili et al. 2002; Rezgui et al. 2008). The North Lagoon of Tunis was designated a Ramsar Wetland of International Importance in January 2013 (Mdaini et al. 2019).

However, The North lagoon of Tunis is currently a completely artificial environment as a result of the human intervention, and the ecological follow-up is a necessity to ensure the proper ecological functioning of this ecosystem located in the center of the city.

Data collection

In this study, the monthly Chl-*a* data from January 1989 to April 2018 in the North Lagoon of Tunis were obtained from sampling campaigns carried out by Al-Buhaira Invest

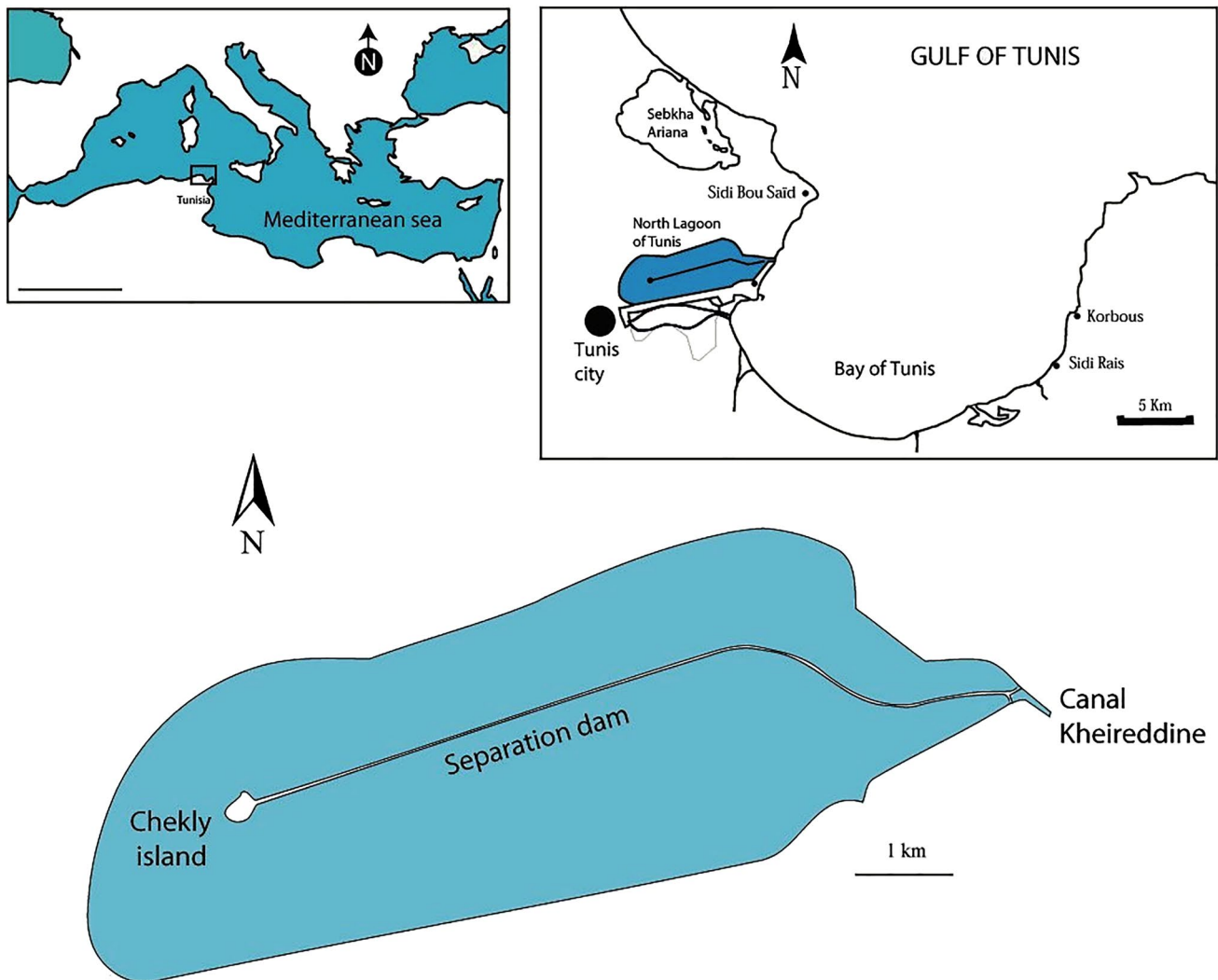


Fig. 1 Location of the study area.

company that is in charge of the lagoon ecological stability, as a part of the monitoring program.

Box and Jenkins methodology

Time series data is a collection of measurements recorded sequentially and equally spaced across a discrete time interval. The fundamental assumption in time series modeling is that past characteristics will continue to occur in the future (Raman et al. 2018). The autoregressive integrated moving average (ARIMA) model developed by Box and Jenkins (1976) is the most widely used approach for time series analysis. The box and Jenkins approach rely on a mathematical model developed to characterize the time series data based on autocorrelation analysis (Chen et al. 2015). Following the creation of the model, it is utilized to anticipate future values using historical and current time series data (Chen

et al. 2015). A major characteristic of time series models is the assumption of some form of statistical equilibrium (Hyndman and Athanasopoulos 2013). An assumption of this kind is that of stationarity (Pai and Lin 2005). Forecasting is based on a linear combination of past observations, which necessitates a stable time series with no visible trend (Pai and Lin 2005). The ARIMA model presumes that the process remains in statistical equilibrium with constant probability features across time (Box et al. 2008). If the mean or variance changes over time, the series may need to be transformed to make it stationary before modeling (Allard 1998).

Time series data are frequently non-stationary, with mean and variance fluctuations across time. Previously, researchers discovered that by differencing the time series, they could remove trend components in the mean (Helfenstein 1991). Typically, one or two orders of differencing are sufficient to prepare data for the technique (Dindarloo 2015). In addition to

the trend components, time series related to the ecology field often show seasonal patterns. Box and Jenkins have developed a method to handle time series that contain seasonality (Box et al. 2008). In this case, the model is known as SARIMA model with S observations per period. It is represented by SARIMA $(p,d,q) (P,D,Q)_S$, which has the following form:

$$\varphi(B)\Phi(B^S)(1 - B^S)^D(1 - B)^dZ_t = \theta(B)\Theta(B^S)_{et} \tag{1}$$

With

$$\varphi(B) = 1 - \varphi_1B - \dots - \varphi_pB^p \tag{2}$$

$$\theta(B) = 1 - \theta_1B - \dots - \theta_qB^q \tag{3}$$

$$\Phi(B^S) = 1 - \Phi_1B^S - \dots - \Phi_pB^{Sp} \tag{4}$$

$$\Theta(B^S) = 1 - \Theta_1B^S - \dots - \Theta_QB^{SQ} \tag{5}$$

Where p is the autoregressive order, d is the number of differencing operations, q is the moving average order and $P, D,$ and Q are the corresponding seasonal orders.

To build an ARIMA or SARIMA model for a time series, Box and Jenkins (1976) proposed an iterative approach (Tiao 2001) involving four steps: stationarity check, identification and estimation, diagnostic and residual check and prediction (Fig. 2).

This technique has gained popularity, allowing for the practical use of time series models for forecasting. The stationarity check step entails appropriate time series differencing. It is carried out to attain stationarity and normality; the time series' stationarity is evaluated using the Augmented Dickey Fuller (ADF) test. This test's null hypothesis is that the data is non-stationary. If the null hypothesis is rejected, that implies stationary data with a p value of equal to or less than 0.05.

The transformed data correlation structure is identified by studying its autocorrelation (ACF) and partial autocorrelation (PACF) functions (Mishra and Desai 2005). The goal is to restrict the field of parsimonious models worthy of further research (Tiao 2001). The ACF and PACF graphs are used to identify the six parameters (p, d, q, P, D, Q) of the SARIMA model and various models' parameters can be selected. The lowest Aikake information criterion (AIC) and Bayesian information criterion (BIC) proposed by Akaike (1973) and Schwartz (1978), respectively, are used to select the best-fit model (Fraley and Raftery 1998) among the candidate ones that were created in the previous step.

AIC estimates the relative amount of information lost by a given model: the less information a model loses, the greater the quality of that model (Aho et al. 2014).

$$AIC = \ln(\sigma_r^2) + 2r/T$$

The BIC is a model selection criteria that is closely connected to the AIC. It is preferable to use the model with the lowest BIC (Aho et al. 2014).

$$BIC = \ln(\sigma_r^2) + rn(T)/T$$

Where, σ_r^2 is the maximum likelihood estimate of the innovation variance, r is the number of parameters in the model and T is the size of the sample series.

After an appropriate model is chosen, several tests are required (model fitness and residual checking) to verify whether the model is adequate for describing the studied process.

To do so, the ACF and quantile-to-quantile (Q-Q) figures of residuals are plotted. In addition, the forecast accuracy of the model is evaluated by splitting the data into two sets. The last observations in our data set are used to compare between simulated values and actual ones. Finally, the chosen SARIMA model is applied to predict the monthly Chl-*a* future values. The long time series, which covers approximately three decades of monthly Chl-*a* historical data, enabled us to perform multiple time steps ahead (Wei 1990) from May 2018 to December 2025.

The modeling of the SARIMA model was performed using Econometrics toolbox in the MATLAB software MATLAB® software (version 7,860,349 (R2020b), The Mathworks, MA, USA).

Results and discussion

The time series of monthly Chl-*a* values are shown in Fig. 3. The autocorrelation properties of the Chl-*a* time series in the North Lagoon of Tunis (Fig. 4) shows that the autocorrelation coefficients declined relatively quickly. It was preferable to do an ADF test to ensure the stationarity of the Chl-*a* concentration data (Table 1).

The ADF test confirmed the stationarity of the Chl-*a* time series with a p value <0.05 . According to that, the time series does not need transformation. It was not necessary to transform the data by differencing. Instead, we used the data on the original scale. Fitting a SARIMA model directly is advantageous for forecasting because forecasts are returned on the original scale.

Based on the PACF (Fig. 4), all SARIMA $(p, d, q, P, D, Q)_S$ models in which the autocorrelation delay p and seasonal autocorrelation delay P was less than or equal to 4 and the moving average q and seasonal moving average Q was less than or equal to 4, were tested. As mentioned above, Chl-*a* time series data contain a seasonal component. To analyze its time series from January 1989 through April 2018, we defined $S = 12$ because we have 12 observations per year.

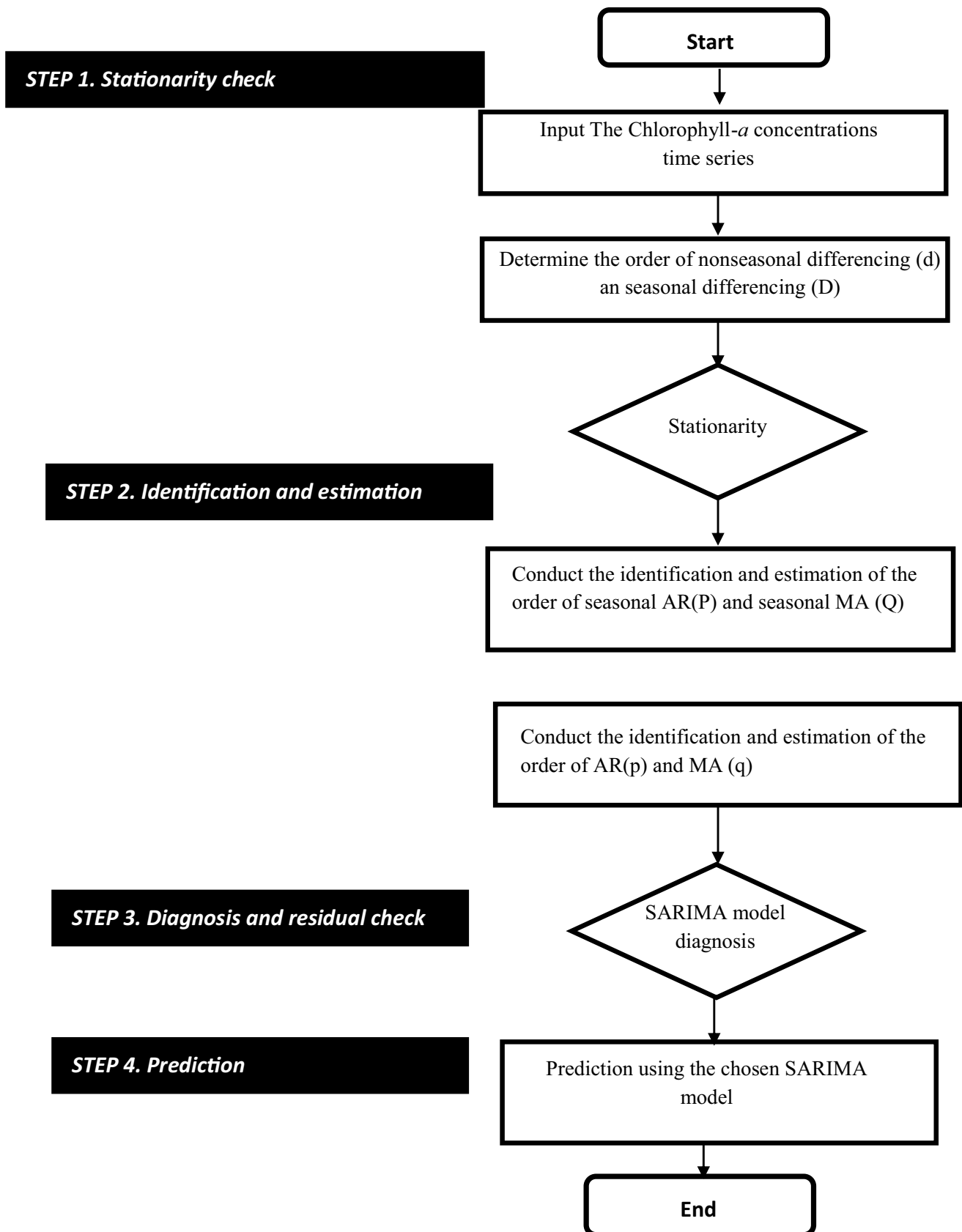


Fig. 2 The prediction process using the SARIMA model

Fig. 3 Temporal distribution of Chl-*a* concentrations in the North lagoon of Tunis

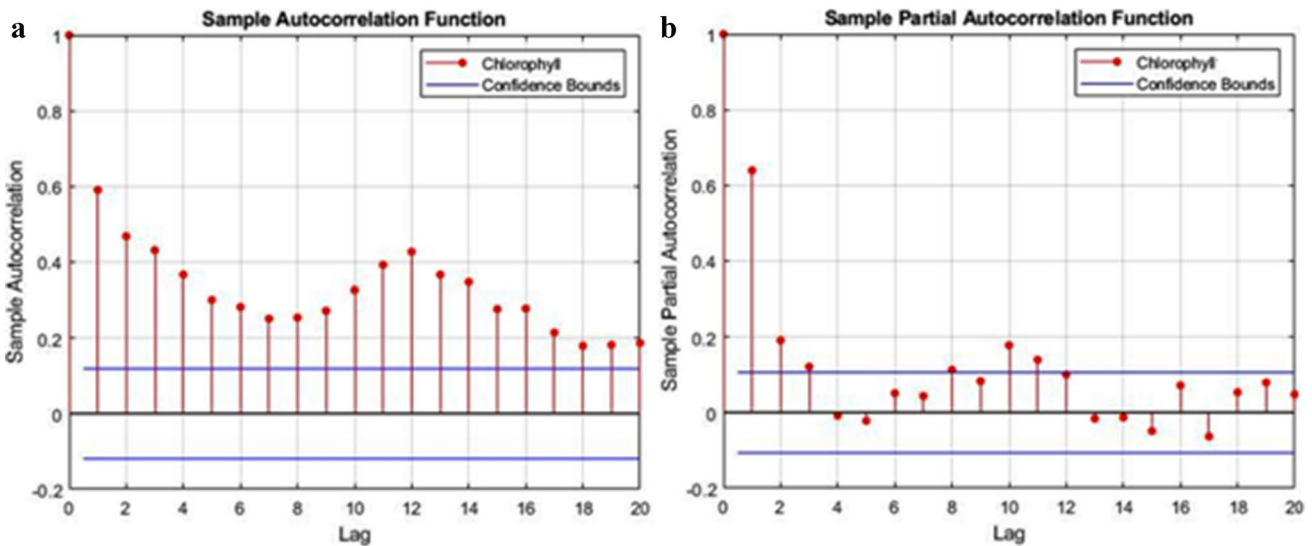
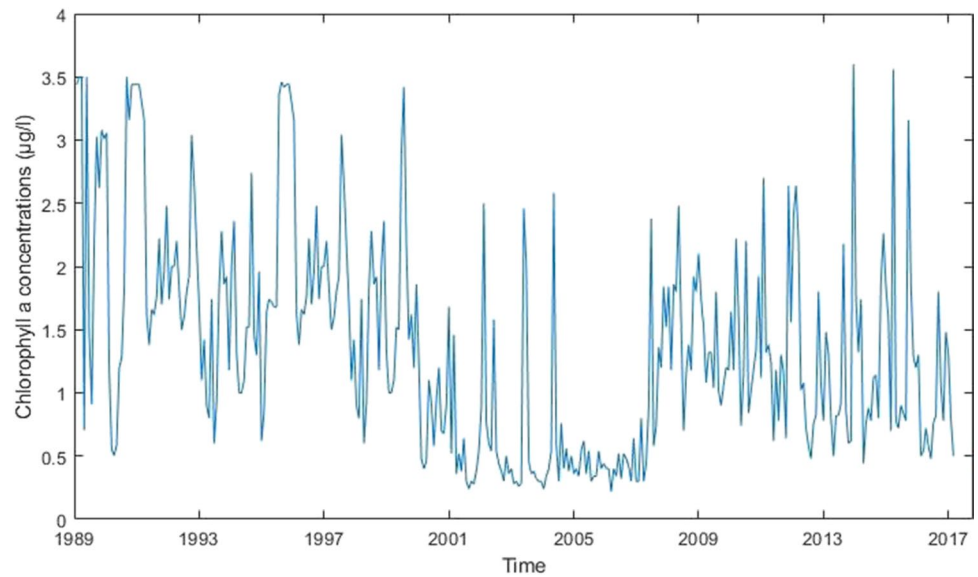


Fig. 4 ACF (a) and PACF (b) of the monthly Chl-*a* variations in the North Lagoon of Tunis

Table 1 ADF test results applied to the Chl-*a* original time series.

| Null Rejected | P value | Statistics Tests | Critical value | Significance level |
|---------------|---------|------------------|----------------|--------------------|
| True | 0.001 | -4.599 | -1.941 | 0.0500 |

It was found the minimal AIC and BIC values when performing a SARIMA (2,0,2)(2,0,2)₁₂. With AIC = 628.91, BIC = 666.78 and R² = 0.52. To avoid making remarkable changes in the original data, it is better to keep the number of parameters to a minimum, so that the values of p, P, q, Q, d, and D selected are less than or equal to 2 (Hintze 2007).

It was observable that the SARIMA (2,0,2)(2,0,2)₁₂ developed in this study outperformed the SARIMA (1,0,0)(2,0,0)₁₂ with AIC = 1593, BIC = 1612 and R² = 0.46 reported by Raman et al. (2018).

In addition, the developed model performed better than the SARIMAX (1,0,0)(2,0,0)₁₂ with AIC = 1114.2,

BIC = 1141.0 and R² = 0.52 reported by Raman et al. (2018). Even though the two models have the same coefficient of determination (R²), the AIC and BIC of our SARIMA (2,0,2)(2,0,2)₁₂ are the lowest.

It is important to mention, that SARIMAX is composed of a SARIMA model with external factors derived using PCA analysis.

Table 2 Parameter estimates and their testing results of the SARIMA (2, 0, 2) (2, 0, 2)₁₂ model

| Parameter | Coefficient | Standard Error | T statistics | p value |
|-----------|-------------|----------------|--------------|------------|
| Constant | 0.0448 | 0.0031 | 2.4357 | 0.01511 |
| AR (1) | 0.0896 | 0.0400 | 1.9890 | 0.03227 |
| AR (2) | 0.0812 | 0.0333 | 2.2383 | 0.02156 |
| SAR (12) | 0.0967 | 0.0435 | 2.2250 | 0.02156 |
| SAR (24) | 0.6763 | 0.0342 | 19.7782 | 4.5892e-87 |
| MA (1) | -0.0686 | 0.3935 | -0.0217 | 0.0382 |
| MA (2) | -0.3391 | 0.1912 | -1.7739 | 0.0491 |
| SMA (12) | -0.0983 | 0.0577 | 1.7045 | 0.0488 |
| SMA (24) | -0.6917 | 0.0578 | -11.6159 | 5.3588e-33 |
| Variance | 0.3302 | 0.0225 | 14.6879 | 7.7121e-41 |

Table 3 Comparison of predicted and observed monthly Chl-*a* variations using the SARIMA (2, 0, 2)(2, 0, 2)₁₂ model for the data from January 2017 to April 2018

| Month | Pre-dicted Values | Observed Values |
|-----------|-------------------|-----------------|
| January | 1.44 | 1.2 |
| February | 1.22 | 1.3 |
| March | 1.15 | 0.5 |
| April | 0.84 | 0.54 |
| May | 0.68 | 0.72 |
| June | 0.74 | 0.58 |
| July | 0.57 | 0.48 |
| August | 0.91 | 0.76 |
| September | 0.74 | 0.82 |
| October | 1.52 | 1.8 |
| November | 0.87 | 1.06 |
| December | 0.62 | 0.78 |
| January | 1.23 | 1.48 |
| February | 0.98 | 1.3 |
| March | 0.76 | 0.80 |
| April | 0.63 | 0.5 |

Estimation of the SARIMA (2,0,2)(2,0,2)₁₂ model parameters and their testing results are presented in Table 2. All estimated coefficients are statistically significant (p value < 0.05).

The ACF of SARIMA (2,0,2)(2,0,2)₁₂ residuals are presented in Fig. 5. ACF residuals suggested autocorrelations near 0. This indicates that the residuals did not deviate significantly from a 0 mean. In other words, it means that the residuals are not correlated.

The residual Q-Q plot (Fig. 6) suggests that the residuals are approximately normally distributed, with a slightly heavier tail. Given that, we can conclude that residuals of the SARIMA (2,0,2)(2,0,2)₁₂ model are uncorrelated and normally distributed.

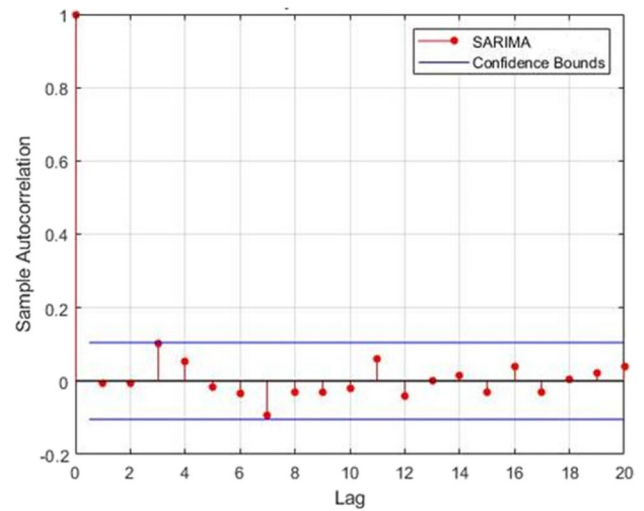


Fig. 5 ACF of residuals

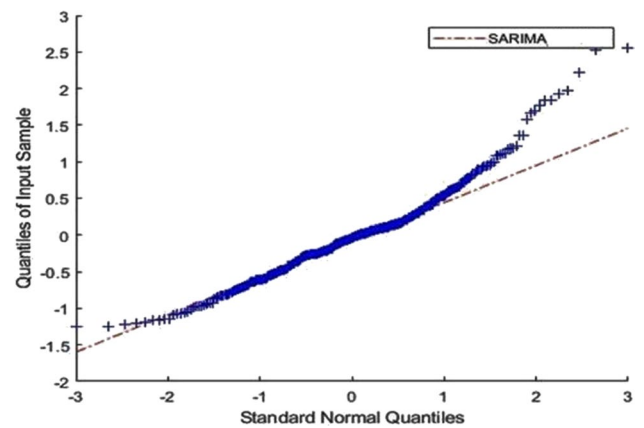


Fig. 6 Residuals (Q-Q) plot

Comparison of observed Chl-*a* data with the fitted ones by the SARIMA (2,0,2)(2,0,2)₁₂ model is presented in Fig. 7. There is a fairly good match between the observed values and the fitted values. To check the forecasting accuracy, Table 3, shows a comparison between the predicted values and the observed ones for the period from January 2017 to April 2018.

In this study, we have applied a SARIMA (p, d, q) (P, D, Q)_S model to analyze Chl-*a* variations of monthly collected data in the North Lagoon of Tunis from 1989 through 2018, with the purpose of assessing and forecasting eutrophication and contributing to prevent any deterioration in the studied ecosystem. We have developed a SARIMA that closely fits Chl-*a* observed data. According to our results, the SARIMA model developed in this study was reliable with high validity, which suggests that our model could be an appropriate statistical tool to predict the future changing trends of Chl-*a* values or other key parameters, thus preventing high

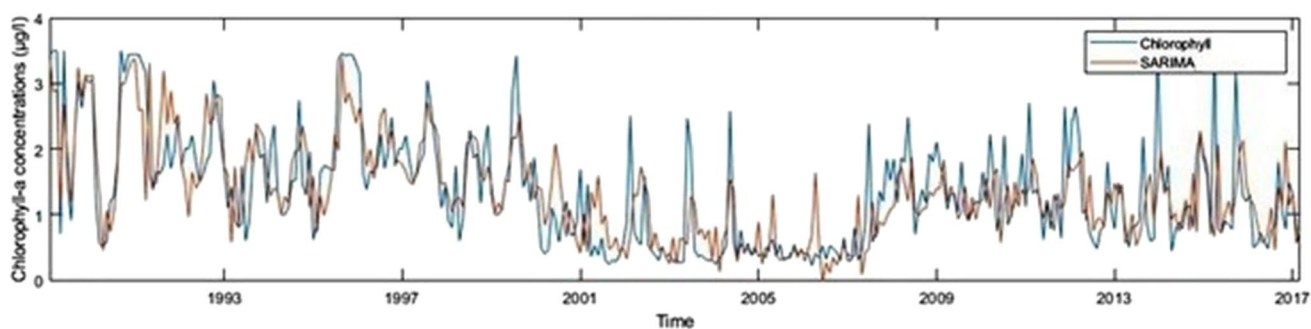


Fig. 7 SARIMA (2,0,2) (2,0,2)₁₂ model fit of Chl-*a* content time series in the North Lagoon of Tunis from January 1989 through April 2018

eutrophication scenarios in the North Lagoon of Tunis or other ecosystems. SARIMA offers various advantages over other prediction methods, such as moving average, exponential smoothing, neural networks, and fuzzy logic, owing to its forecasting power, particularly over long time periods, and its richer information on time-related changes (Linthicum et al. 1999; Box et al. 2008). Once we obtained the satisfactory model, we have used it to forecast future values of the Chl-*a* in the ecosystem. Fig. 8 shows the forecast results of Chl-*a* concentrations in the North Lagoon of Tunis. The model seems to provide realistic predictions for the future.

To the best of our knowledge, this is the first study that has applied the SARIMA model to fit and forecast monthly Chl-*a* concentrations in the North Lagoon of Tunis.

Forecasting is essential in the creation and implementation of policies, particularly in the management of fisheries and aquatic resources. Accordingly, multiple time series model have been developed for forecasting purposes, including ARIMA model, Neural Networks, multiple linear

regression (MLR), nonlinear regression (NLR), smoothing models, generalised auto regressive conditional heteroscedasticity (GARCH), Gaussian autoregressive models (Stergiou 1991; Stergiou et al. 1997; Romilly 2005). According to the studies, ARIMA validation errors are substantially smaller and it is a superior forecasting model (Raman et al. 2018).

Similarly, ARIMA model has been used in various studies in lagoons, such as on modeling of the great lakes freeze for ferrous scrap (Albertson and Ayleen 1996), and on simulating spatio-temporal behaviour of atmospheric temperature data of the great lakes US (Agrawal 2011). ARIMA's reliable forecasts with sufficient precision demonstrated an essential role for managers of aquatic natural environments (Raman et al. 2018).

It is known that the SARIMA modeling process needs a large amount of data. An early study suggested that a minimum of 50 observations are needed to build a reasonable SARIMA model (Wei 1990), understanding the majority

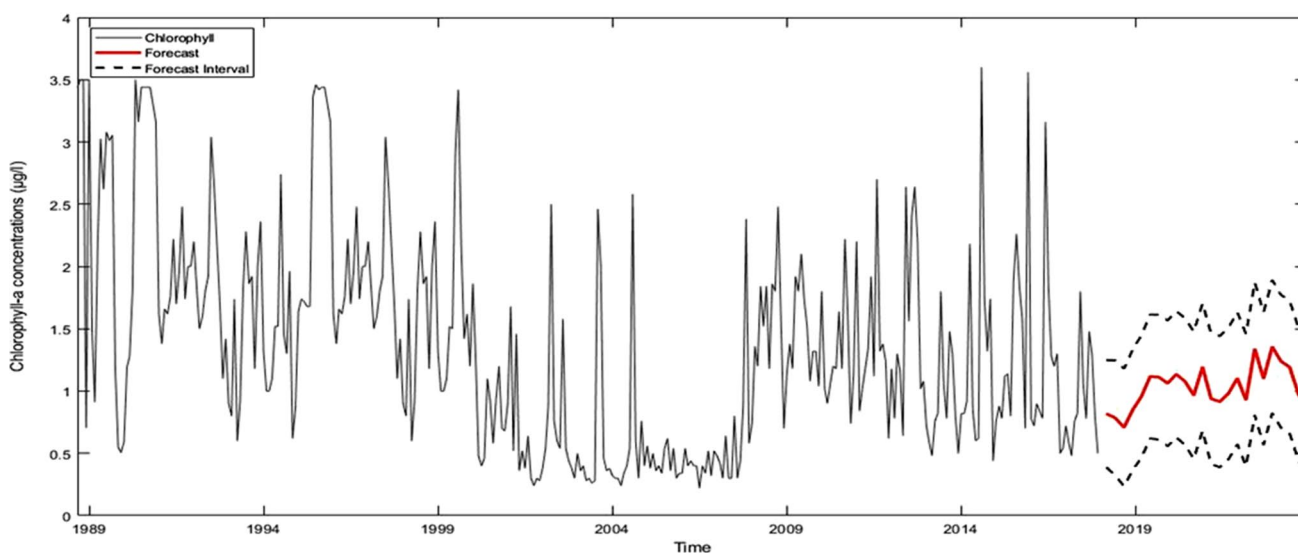


Fig. 8 Time series plot of Chl-*a* concentrations in the North Lagoon of Tunis with forecasts at 75% confidence interval

of its variation and accurately predicting the data's seasonality and correlation structure (Stergiou 1991; Pajuelo and Lorenzo 1995). To obtain a stable and precise SARIMA model, we had to collect over 350 observations of Chl-*a* data during the past three decades without interruption to be mentioned. We can state that the results of our study are robust enough.

It is important to mention that, using SARIMA model is money and time-saving for researchers, because it needs only one input variable and provides predictions with acceptable accuracy, which facilitates its applicability for algal blooms early warning use, especially, in vulnerable ecosystems.

Conclusion

Thanks to its biological stability, the North Lagoon of Tunis has a significant social and ecological value. Tourism (water sports), fishing, and marine bird conservation are only a few of the benefits provided by this ecosystem. As a result, in order to apply long-term management and conservation strategies, we must maintain a close eye on the eutrophication process and potential algal blooms episodes.

Forecasting Chl-*a* levels as eutrophication indicator is widely used in coastal ecosystems, as it gives the ability to the managers to identify future problems before they arise, make informed management decisions, develop data-driven strategies and evidence-based conservation programs.

In our study, we have used the SARIMA model to forecast future values of Chl-*a* in the North Lagoon of Tunis as a eutrophication indicator. This model was applied for a time window of approximately three decades. Different SARIMA (p, d, q) (P, D, Q)_S models were implemented. The chosen one, a SARIMA (2,0,2) (2,0,2)₁₂ with the lowest AIC and BIC, was used for forecasting. The goodness of fit was analyzed by comparing with the actual data from the last observations and checking the residuals. The residual diagnostic indicated that they are uncorrelated and normally distributed.

The forecasting results are quite satisfactory since the forecasting period seems to reproduce relatively well the normal Chl-*a* monthly content in the lagoon. Despite the fact that the restored North Lagoon of Tunis was classified as a Wetland of International Importance (Ramsar site), given its history and importance, this ecosystem is still fragile. The SARIMA model applied to historical data of Chl-*a* or any other key parameter could be an important tool to providing evidence that guides prevention and control interventions for the ecosystem.

Acknowledgments We are grateful to the Al-Buhaira Invest Company for providing us with monitoring data.

Code availability The Econometrics toolbox in the MATLAB software MATLAB® software (version 7,860,349 (R2020b), The Mathworks, MA, USA).

Author contribution NBH wrote the paper, collected data, conceived, designed and performed the analysis. CG helped in the correction of the paper and was a major contributor in writing the manuscript. NBM provided the data as a part of the monitoring program of the ecosystem. AS helped in the supervision of this work and was a major contributor in writing the manuscript. All authors read and approved the final manuscript.

Data availability The data provided by Al-Buhaira Invest company to the first author are acknowledged.

Declarations

Conflicts of interest/ Competing interests No conflict of interest associated with this publication

Ethics approval Not applicable

Consent to participate Not applicable

Consent for publication Not applicable

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Afli A, Ayari R, Brahim M (2008) Trophic organization of the macrozoobenthic assemblages within coastal areas subjected to anthropogenic activities. *J Mar Biol Assoc UK* 88:663–674. <https://doi.org/10.1017/S0025315408001318>
- Agrawal A (2011) A new approach to spatio temporal kriging and its application. Thesis master of science. Graduate School of the Ohio State University, p 112
- Aho K, Derryberry D, Peterson T (2014) Model selection for ecologists: the worldviews of AIC and BIC. *Ecology* 95(3):631–636. <https://doi.org/10.1890/13-1452.1>
- Akaike H (1973) Maximum likelihood identification of Gaussian autoregressive moving average models. *Biometrika* 60(2):255–265
- Albertson K, Aylen J (1996) Modelling the great lakes freeze: forecasting and seasonality in the market for ferrous scrap. *Int J Forecast* 12(3):345–359. [10.1016/0169-2070\(96\)00669-3](https://doi.org/10.1016/0169-2070(96)00669-3)
- Allard R (1998) Use of time-series analysis in infectious disease surveillance. *Bull World Health Organ* 76(4):327–333
- Armi Z, Turki S, Trabelsi E, Ben Maiz N (2008) Nutrient loading and occurrence of potentially harmful phytoplankton species in the

- North Lake of Tunis (Tunisia). *Cah Biol Mar* 49:311–321. <https://doi.org/10.21411/CBM.A.12B953FC>
- Armstrong JS (1985) Long range forecasting: from crystal ball to computer. John Wiley, New York
- Ben Charrada R (1992) Le lac de Tunis après les aménagements. Paramètres physicochimiques de l'eau et relation avec la croissance des macroalgues. *Mar Life* 1:29–44
- Box GEP, Jenkins GM (1976) Time Series Analysis: Forecasting and Control. Holden-Day, San Francisco, CA
- Box GEP, Jenkins GM, Reinsel GC, Ljung GM (2008) Time Series Analysis: Forecasting and Control, 4th edn. John Wiley & Sons, Hoboken, NJ, USA
- Chen Q, Guan T, Yun L, Li R, Recknagel F (2015) Online forecasting chlorophyll a concentrations by an auto-regressive integrated moving average model: Feasibilities and potentials. *Harmful Algae* 43:58–65 <https://doi.org/10.1016/j.hal.2015.01.002>
- Cloern JE (2001) Our evolving conceptual model of the coastal eutrophication problem. *Mar Ecol Prog Ser* 210:223–253. <https://doi.org/10.3354/meps210223>
- Cui L, You L, Huang Z (2007) TN/TP ratio of lake water and its implication in algae bloom of Beijing's urban lakes. *Environ Sci Technol* 30(10):47–49
- Davis TW, Berry DL, Boyer GL, Gobler CJ (2009) The effects of temperature and nutrients on the growth and dynamics of toxic and non-toxic strains of Micro-cystis during cyanobacteria blooms. *Harmful Algae* 8:715–725. <https://doi.org/10.1016/j.hal.2009.02.004>
- de Jonge VN, Elliott M, Orive E (2002) Causes, historical development, effects and future challenges of a common environmental problem: eutrophication. *Hydrobiologia* 475(476):1–19. <https://doi.org/10.1023/A:1020366418295>
- Derolez V, Soudant D, Malet N, Chiantella C, Richard M, Abadie E, Aliaume C, Bec B (2020) Two decades of oligotrophication: Evidence for a phytoplankton community shift in the coastal lagoon of Thau (Mediterranean Sea, France). *Estuar Coast Shelf Sci* 241:106810. <https://doi.org/10.1016/j.ecss.2020.106810>
- Dindarloo S (2015) Reliability forecasting of a load-haul-dump machine: a comparative study of ARIMA and neural networks. *Qual Reliab Eng Int* 32(4):1545–1552. <https://doi.org/10.1002/qre.1844>
- Dippner JW, Nguyen-Ngoc L, Doan-Nhu H, Subramaniam A (2011) A model for the prediction of harmful algae bloom in the Vietnamese upwelling area. *Harmful Algae* 10:606–611. <https://doi.org/10.1016/j.hal.2011.04.012>
- Donner A (1982) The relative effectiveness of procedures commonly used in multiple regression analysis for dealing with missing values. *Am Stat* 36:378–381
- Fraley C, Raftery AE (1998) How many clusters? Which clustering method? Answers via model-based cluster analysis. *Comput J* 41:578–588
- García-Ayllón S (2017) Diagnosis of complex coastal ecological systems: Environmental GIS analysis of a highly stressed Mediterranean lagoon through spatiotemporal indicators. *Ecol Indic* 83:451–462 <https://doi.org/10.1016/j.ecolind.2017.08.015>
- García-Pintado J, Martínez-Mena M, Barberá GG, Albaladejo J, Castillo VM (2007) Anthropogenic nutrient sources and loads from a Mediterranean catchment into a coastal lagoon: Mar Menor, Spain. *Sci Total Environ* 373:220–239. <https://doi.org/10.1016/j.scitotenv.2006.10.046>
- Gaynor PE, Kirkpatrick RC (1994) Introduction to Time Series Modeling and Forecasting in Business and Economics. McGraw-Hill College
- Harbridge W, Pilkey OH, Whaling P, Swetland P (1976) Sedimentation in the Lake of Tunis: a lagoon strongly influenced by man. *Environ Geol* 1:215–225. <https://doi.org/10.1007/bf02407508>
- Hasting A (2001) Transient dynamics and persistence of ecological systems. *Ecol Lett* 4:215–220. <https://doi.org/10.1046/j.1461-0248.2001.00220.x>
- Hassanzadeh S, Hosseinibalam F, Alizadeh R (2009) Statistical models and time series forecasting of sulfur dioxide: a case study Tehran. *Environ Monit Assess* 155(1):149–155. <https://doi.org/10.1007/s10661-008-0424-1>
- Hecht-Nielsen R (1989) Theory of the backpropagation neural network. *Int Jt Conf Neural Netw* 1:593–605. <https://doi.org/10.1109/IJCNN.1989.118638>
- Helfenstein U (1991) The use of transfer function models, intervention analysis and related time series methods in epidemiology. *Int J Epidemiol* 20(3):808–815. <https://doi.org/10.1093/ije/20.3.808>
- Hintze J (2007) NCSS 2007. NCSS, LLC. Kaysville, Utah, USA. www.ncss.com. Accessed February/Mar 2021
- Hood RR, Zhang X, Glibert PM, Roman MR, Stocker DK (2006) Modeling the influence of nutrients, turbulence and grazing on Pfiesteria population dynamics. *Harmful Algae* 5:459–479. <https://doi.org/10.1016/j.hal.2006.04.014>
- Hyndman RJ, Athanasopoulos G (2013) Forecasting: Principles and Practice. O Texts Publishers: Monash University, Australia. <https://www.otexts.org/book/fpp>. Accessed February/Mar 2021
- Leruste A, Malet N, Munaron D, Derolez V, Hately E, Collos Y, De Wit R, Bec B (2016) First steps of ecological restoration in Mediterranean lagoons: Shifts in phytoplankton communities. *Estuar Coast Shelf Sci* 180:190–203. <https://doi.org/10.1016/j.ecss.2016.06.029>
- Linthicum KJ, Anyamba A, Tucker CJ, Kelley PW, Myers MF, Peters CJ (1999) Climate and satellite indicators to forecast Rift Valley fever epidemics in Kenya. *Science* 285(5426):397–400. <https://doi.org/10.1126/science.285.5426.397>
- Lu Y, Abou Rizk SM (2008) Automated Box–Jenkins forecasting modelling. *Autom Constr* 18:547–558
- Lui GCS, Li WK, Leung KMY, Lee JHW, Jayawardena AW (2007) Modelling algal blooms using vector autoregressive model with exogenous variables and long memory filter. *Ecol Model* 200:130–138. <https://doi.org/10.1016/j.ecolmodel.2006.06.017>
- Makwinja R, Phiri T, Kosamu IB, Kaonga CC (2017) Application of stochastic models in predicting Lake Malawi water levels. *Int J Water Resour Environ Eng* 9(9):191–200. <https://doi.org/10.5897/IJWREE2017.0740>
- Mdaini Z, El Gafsi M, Tremblay J, Pharand P, Gagné JP (2019) Spatio-temporal variability of biomarker responses and lipid composition of *Marphysasanguinea*, Montagu (1813) in the anthropic impacted lagoon of Tunis. *Mar Pollut Bull* 144:275–286. <https://doi.org/10.1016/j.marpolbul.2019.04.065>
- Mishra AK, Desai VR (2005) Drought forecasting using stochastic models. *Stoch Environ Res Risk A* 19:326–339
- Newton A, Icely J, Cristina S et al (2014) An overview of ecological status, vulnerability and future perspectives of European large shallow, semi-enclosed coastal systems, lagoons and transitional waters. *Estuar Coast Shelf Sci* 140:95–122. <https://doi.org/10.1016/j.ecss.2013.05.023>
- Nixon SW (1995) Coastal marine eutrophication: A definition, social causes, and future concerns. *Ophelia* 41(1):199–219. <https://doi.org/10.1080/00785236.1995.10422044>
- Oh HM, Ahn CY, Lee JW, Chon TS, Choi KH, Park YS (2007) Community patterning and identification of predominant factors in algal bloom in Daechung Reservoir (Korea) using artificial neural networks. *Ecol Model* 203:109–118. <https://doi.org/10.1016/j.ecolmodel.2006.04.030>
- Onderka M (2007) Correlations between several environmental factors affecting the bloom events of cyanobacteria in Liptovska Mara reservoir (Slovakia)-a simple regression model. *Ecol Model* 209:412–416. <https://doi.org/10.1016/j.ecolmodel.2007.07.028>

- Omer FD (2010) A hybrid neural network and ARIMA model for water quality time series prediction. *Eng Appl Artif Intell* 23:586–594
- Ouchi A (1982) Forecast of a red tide occurrence in the northern part of Hiroshima Bay-II. Prediction of red tide occurrence by means of multiple linear regression model. *Bull Jpn Soc Sci Fish* 48:1245–1250
- Paerl HW, Huisman J (2008) Blooms like it hot. *Science* 320(5872):57–58. <https://doi.org/10.1126/science.1155398>
- Pai PF, Lin CS (2005) A hybrid ARIMA and support vector machines model in stock price forecasting. *Omega* 33:497–505. <https://doi.org/10.1016/j.omega.2004.07.024>
- Pajuelo JG, Lorenzo JM (1995) Analysis and forecasting of the demersal fishery of the Canary Islands using an ARIMA model. *Sci Mar* 59:155–164
- Prista N, Diawara N, Costa MJ, Jones C (2011) Use of SARIMA models to assess data-poor fisheries: a case study with a sciaenid fishery of Portugal. *Fish Bull* 109(2):170–185
- Raman RK, Mohanty SK, Bhatta KS, Karna SK, Sahoo AK, Mohanty BP, Das BK (2018) Time series forecasting model for fisheries in Chilika lagoon (a Ramsar site, 1981), Odisha, India: a case study. *Wetl Ecol Manag* 26:677–687. <https://doi.org/10.1007/s11273-018-9600-4>
- Recknagel F, Ostrovsky I, Cao HQ, Zohary T, Zhang XQ (2013) Ecological relationships, thresholds and time-lags determining phytoplankton community dynamics of Lake Kinneret, Israel elucidated by evolutionary computation and wavelets. *Ecol Model* 255:70–86. <https://doi.org/10.1016/j.ecolmodel.2013.02.006>
- Recknagel F, Orr P, Cao H (2014) Inductive reasoning and forecasting of population dynamics of *Cylindrospermopsis raciborskii* in three sub-tropical reservoirs by evolutionary computation. *Harmful Algae* 31:26–34. <https://doi.org/10.1016/j.hal.2013.09.004>
- Rezgui A, Maiz N, Moussa M (2008) Fonctionnement hydrodynamique et écologique du Lac Nord de Tunis par modélisation numérique. *Rev Sci l'Eau* 21:349–361. <https://doi.org/10.7202/018781ar>
- Robledano F, Esteve MA, Martínez-Fernández J, Farinós P (2011) Determinants of wintering waterbird changes in a Mediterranean coastal lagoon affected by eutrophication. *Ecol Indic* 11:395–406. <https://doi.org/10.1016/j.ecolind.2010.06.010>
- Romilly P (2005) Time series modelling of global mean temperature for managerial decision-making. *J Environ Manag* 76:61–70
- Schramm W (1999) Factors influencing seaweed responses to eutrophication: some results from EU-project EUMAC. *J Appl Phycol* 11:69–78
- Schwartz G (1978) Estimating the dimension of a model. *Ann Stat* 6:461–464
- Shili A, Trabelsi EB, Ben Maiz N (2002) Seasonal dynamics of macroalgae in the South Lake of Tunis. *J Coast Conserv* 8:127–134
- Souchu P, Bec B, Smith VH, Laugier T, Fiandrino A, Benau L, Orsoni V, Collos Y, Vaquer A (2010) Patterns in nutrient limitation and chlorophyll-*a* along an anthropogenic eutrophication gradient in French Mediterranean coastal lagoons. *Can J Fish Aquat Sci* 67:743–753. <https://doi.org/10.1139/F10-018>
- Stergiou KI (1991) Describing and forecasting the sardine-anchovy complex in the eastern Mediterranean using vector autoregressions. *Fish Res* 11:127–141
- Stergiou KI, Chritou ED, Petrakis G (1997) Modelling and forecasting monthly fisheries catches: comparison of regression, univariate and multivariate time series methods. *Fish Res* 29:55–95
- Sun H, Koch M (2001) Case study: analysis and forecasting of salinity in Apalachicola Bay, Florida, using Box-Jenkins ARIMA models. *J Hydraul Eng* 127(9):718–727. [https://doi.org/10.1061/\(ASCE\)0733-9429\(2001\)127:9\(718\)](https://doi.org/10.1061/(ASCE)0733-9429(2001)127:9(718))
- Tiao GC (2001) Time Series: ARIMA Methods. *Comput Sci*:363–407
- Tizro AT, Ghashghaie M, Georgiou P, Voudouris K (2014) Time series analysis of water quality parameters. *J Appl Res Water Wastewater* 1(1):40–50
- Trabelsi-Bahri EL, Armi Z, Trabelsi-Annabi N, Shili A, Ben Maiz N (2013) Water quality variables as indicators in the restoration impact assessment of the north lagoon of Tunis, South Mediterranean. *J Sea Res* 79:12–19. <https://doi.org/10.1016/j.seares.2013.01.003>
- Turki S, Balti N, Ben Salah C (2007) First detection of *K. foliaceum* (Stein1883) in Tunisian waters. *Harmful Algae News* 35:9–10
- Viaroli P, Bartoli M, Azzoni R, Giordani G, Mucchino M, Naldi M, Nizzoli D, Taje L (2005) Nutrient and iron limitation to Ulva blooms in a eutrophic coastal lagoon (Sacca di Goro, Italy). *Hydrobiologia* 550:57–71. <https://doi.org/10.1007/s10750-005-4363-3>
- Viaroli P, Bartoli M, Giordani G, Naldi M (2008) Community shifts, alternative stable states, biogeochemical controls and feedbacks in eutrophic coastal lagoons: a brief overview. *Aquat Conserv Mar Freshwat Ecosyst* 18:105–117. <https://doi.org/10.1002/aqc>
- Villanoy CL, Azanza RV, Altemerano A, Casil AL (2005) Attempts to model the bloom of Pyrodinium, a tropical toxic dinoflagellate. *Harmful Algae* 5:156–183. <https://doi.org/10.1016/j.hal.2005.07.001>
- Vogiatzakis IN, Mannion AM, Griffiths GH (2006) Mediterranean ecosystems: problems and tools for conservation. *Prog Phys Geogr* 30(2):175–200. <https://doi.org/10.1191/0309133306pp472ra>
- Wang Z, Huang K, Zhou PJ, Guo HC (2010) A hybrid neural network model for cyanobacteria bloom in Dianchi Lake. *Procedia Environ Sci* 2:67–75
- Wei WWS (1990) Time series analysis: univariate and multivariate methods, 2th edn. Addison-Wesley Publishing Company, New York
- Wilhelm SW, Farnsley SE, LeCleir GR (2011) The relationships between nutrients, cyanobacterial toxins and the microbial community in Taihu (Lake Tai), China. *Harmful Algae* 10:207–215. <https://doi.org/10.1016/j.hal.2010.10.001>
- Yu L, Liang S, Chen R, Lei KK (2019) Predicting monthly biofuel production using a hybrid ensemble forecasting methodology. *Int J Forecast*. <https://doi.org/10.1016/j.ijforecast.2019.08.014>
- Zaldívar JM, Cardoso AC, Viaroli P, de Wit R, Ibanez C, Reizopoulou S, Razinkovas A, Basset A, Holmer M, Murray N (2008a) Eutrophication in transitional waters: an overview. *Transiti. Waters. Monogr* 1:1–78. <https://doi.org/10.1285/i18252273v2n1p1>
- Zaldívar JM, Strozzi F, Dueri S, Marinov D, Zbilut JP (2008b) Characterization of regime shifts in environmental time series with recurrence quantification analysis. *Ecol Model* 210:58–70. <https://doi.org/10.1016/j.ecolmodel.2007.07.012>
- Zhang H, Hu W, Gu K, Li Q, Zheng D, Zhai S (2013) An improved ecological model and software for short-term algal bloom forecasting. *Environ Model Softw* 48:152–162. <https://doi.org/10.1016/j.envsoft.2013.07.001>

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.