



# Detecting Anomalies in Marine Data: A Framework for Time Series Analysis

Nicoletta Del Buono<sup>1,2</sup>, Flavia Esposito<sup>1,2(✉)</sup>, Grazia Gargano<sup>2,3</sup>,  
Laura Selicato<sup>1,2</sup>, Nicolò Taggio<sup>4</sup>, Giulio Ceriola<sup>4</sup>, and Daniela Iasillo<sup>4</sup>

<sup>1</sup> Department of Mathematics, University of Bari Aldo Moro,  
via E. Orabona 4, Bari, Italy

{nicoletta.delbuono,flavia.esposito,laura.selicato}@uniba.it

<sup>2</sup> Members of INDAM-GNCS Research Group, Rome, Italy

<sup>3</sup> Istituto Oncologico Giovanni Paolo II, Bari, Italy

grazia.gargano@oncologico.bari.it

<sup>4</sup> Planetek Italia, Bari, Italy

{taggio,ceriola,iasillo}@planetek.it

**Abstract.** An ensemble framework for the analysis of time series from marine backgrounds is proposed to finally identify and classify anomalies in data time series collected from European Union's Earth Observation Programme Copernicus and Marine-EO project. The framework aims to estimate a prediction model for anomalies detection when new records are explored and then rank the magnitude of the anomalies eventually detected in some biogeochemical parameters of marine and ocean waters, such as chlorophyll-a concentrations, surface temperature profiles and dissolved oxygen.

**Keywords:** Anomaly · Outlier · Anomalies detection · Statistical models · Data pre-processing

## 1 Introduction

Marine environment observations are crucial to acquiring information about water bodies' environmental quality. Knowledge extracted from these data validly supports planning territorial interventions for their restoration. For this reason, it is necessary to obtain periodically and systematically reference information on the physical and biogeochemical state, the variability and dynamics of the ocean, and marine ecosystems. A helpful source of information on these processes is the European Union's Earth Observation Programme Copernicus [4] which monitors our planet by collecting data coming from dedicated satellites, called *Sentinels*, from dozens of other satellites, the so-called *participating missions* and *in situ* sensors. Copernicus marine environment monitoring service is helped by the service Nucleus for European Modelling of the Ocean model (NEMO), which provides a physically consistent description of all relevant ocean variables forecasting their evolution. These services support all marine applications. In particular, using data from satellites and *in situ* sensors, Copernicus

Marine Services [5] provide daily water quality analyses and forecasts which offer an unprecedented capability to observe, understand and anticipate events in the marine environment. It contributes to the protection and sustainable management of marine biological resources, particularly for aquaculture, sustainable fisheries management, or the decision-making process of regional fisheries organizations. It thus enables a wide range of marine and coastal environmental applications.

These data are often time series that can be affected by anomalies. Anomalies or outliers –in the statistical sense– are (rare) points that differ significantly from the expected behavior of other data points. Identifying and detecting anomalies in marine physical and biochemical data can be a challenging task that successfully solving would bring added value to marine studies. Advantages can be obtained to check water quality, control pollution behavior, and climate changes, or assess coastal erosion. Moreover, when sea surface temperatures are under consideration, anomaly detection can help forecast problems in marine ecosystems or the occurrence of tropical cyclones.

The Marine-EO project is part of this context; it is devoted to collecting information about ocean parameters variability in time and space and proposes a set of support services to integrate better the Earth Observation (EO) and Copernicus-enabled resources. In particular, it aims to establish EO-based services, covering the sea-basins of the Mediterranean, North Atlantic, and the Arctic, by adapting Copernicus data and information to provide tempestive and reliable information on the Marine Environment in the thematic areas of Marine monitoring (environmental assessment and fish farming) and Maritime Security (support to navigation in Arctic Sea).

This paper proposes a workflow for the analysis of time series from marine backgrounds. In particular, starting from samples constituting data time series, a predictive model is estimated to adhere to these observations. Moreover, a proper comparison between expected and actual observations helps recognize whenever parameter anomalies are present.

The paper is organized as follows. The proposed new framework for marine anomalies detection is detailed in Sects. 2 and 4, together with some descriptions of data and marine areas which have been considered and inspected (Sect. 3). Some experimental results on physical and biogeochemical parameters of the waters of some interest areas are also reported in Sect. 4. A final Sect. 5 of conclusion and discussions closes the paper.

## 2 The Proposed Framework for Anomalies Detection

The framework focuses on hydrological time series and aims to estimate a prediction model for anomaly detection when new records are explored. The novelty of the proposal stays in its ability to rank the magnitude of the anomalies in the particular context of the water quality time series provided by the Marine-EO project. The complete workflow has been developed in R environment [11]; it integrates some already loaded packages with new designed functions for anomaly

classification. All related codes have been stored on Github<sup>1</sup>. Figure 1 sketches the main steps of the proposed framework.

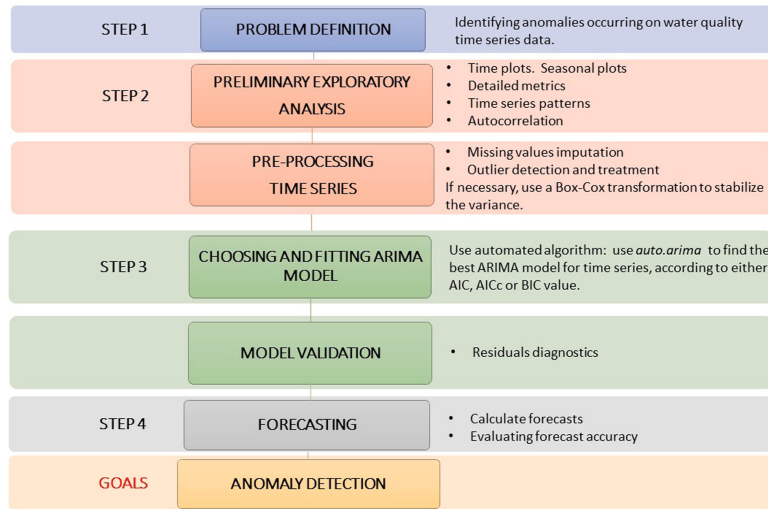


Fig. 1. Workflow of the proposed approach

A preliminary Exploratory and Preprocessing phase is performed to provide a qualitative study of the time series and data cleaning operations managing missing values and outliers. This phase is followed by estimating the data best fit model; after a check of the working hypotheses, a forecasting procedure is applied according to different levels of prediction. The last workflow phase is devoted to detecting and then classifying the anomalies. This step distinguishes different levels of anomalies, from moderate to severe, which vary according to the parameter studied. A detailed description of the anomalies detection procedure is reported in the following sections.

It should be observed that since hydrological data are taken into study, due to their characteristics, the adoption of distribution-free methods is preferred for as long as possible.

### 3 Data Description

Data from two marine protected areas, the island of Zakynthos in the Ionian Sea, the Condor mouth, and an area of high interest for aquaculture, the Ribeira Quente in the Azores of the Atlantic Ocean, are considered. The Mediterranean Sea, with its peculiar characteristics and the presence of physical processes common to other oceans, can be considered a miniature ocean system. Many techniques that are fundamental to the general circulation of the global ocean occur—identically or similarly—within the Mediterranean basin.

<sup>1</sup> <https://github.com/flaespo/AnomaliesMarineData>.

Among the physical and biogeochemical parameters of marine and ocean waters, surface temperature profiles and dissolved oxygen and chlorophyll-a concentrations are analyzed.

Sea surface temperature anomalies are an important predictor of atmospheric and oceanic circulation patterns. Dissolved oxygen can be used to characterize the fitness for the life of aquatic flora and fauna and the pollution level in a water system. In contrast, the concentration of chlorophyll-a in water highlights the level of eutrophication in coastal waters. It is critically important for assessing trophic characteristics of water bodies and is also an excellent indicator for evaluating primary production and the state of ecosystems.

Each of the examined parameters has been studied through its data time series following Marine-EO specifications, covering the 1999–2019 time window. In particular, each data correspond to a time series raster image with a specific space resolution,  $1\text{ Km} \times 1\text{ Km}$  for Chlorophyll-a and sea surface temperature,  $4\text{ Km} \times 4\text{ Km}$  for dissolved oxygen. To harmonize data for a selected area, a monthly sampling (the geometric mean for Chlorophyll-a and the arithmetic mean for the other) has been performed. Finally, the arithmetic mean has been computed for all pixels within the same area. In the following, a complete description of the proposed framework is illustrated when the level of Chlorophyll-a in the Ionian sea on Zakynthos Greek island is considered. The results on other parameters and regions of interest are summarized for space reasons in Appendix A.

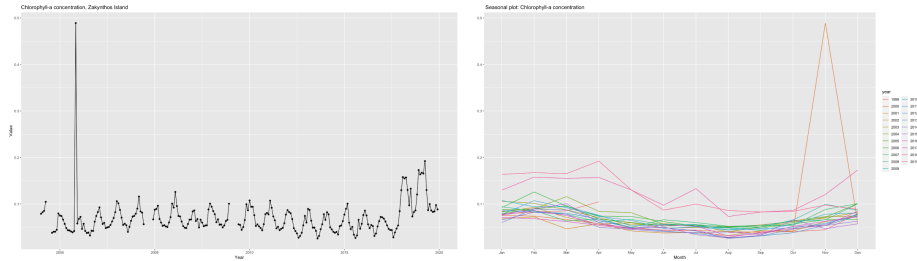
## 4 Methods and Results

This section describes in detail the main phases of the proposed framework for anomalies detection. The methods and the techniques included in it are described referring to the results obtained when data on the level of Chlorophyll-a in Zakynthos island sea are analyzed.

### 4.1 Preliminary Exploratory Analysis Phase

This initial phase in the framework aims to qualitative study and visually understand the general time-series behavior and its possible quirks (i.e., missing data, abrupt changes in time, or variables relationships). This phase provides plotting of the time course of the studied parameters measured monthly. Seasonal plots are also adopted to detect the presence of seasonal fluctuations: yearly reference period is plotted and scanned into monthly sub-periods as well. Figure 2 shows the behavior of Chlorophyll-a monthly acquired in the Ionian sea. A visual inspection of the qualitative graphs reported by the preliminary data analysis phase immediately reveals some interesting features: data present missing values, some observations that significantly differ from the others, and an upward shift in this parameter value has been observed since January 2018.

Time series internal dynamics can also be investigated using the tools present in this phase of the framework. Generally speaking, time series are characterized by systematic patterns, which need to be identified and separated from possible



**Fig. 2.** Time plot and seasonal plots: Chlorophyll-a concentration in Zakynthos island, Greece.

accidental oscillations (erratic components) by equalization or smoothing methods. These two patterns are trend and seasonality. The trend is the underlying monotonic movement, which highlights a structural evolution of the phenomenon due to systematic causes over a long period. On the other hand, seasonality is constituted by variations having similar intensity each year but different magnitude during the same year. Then a residual component includes factors that influence the phenomenon randomly.

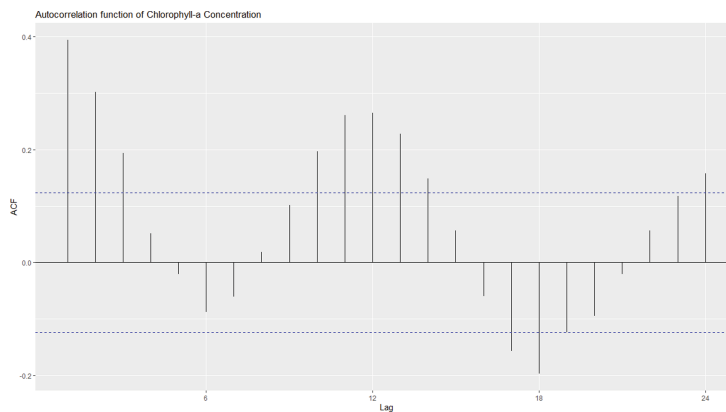
Different time series decomposition methods exist in the literature. The proposed framework includes Seasonal and Trend decomposition with the Local Weighted Regression (STL with LOESS) approach [3]. STL acts by averaging, with some weights, the time series for each month for the seasonal component and then smoothing the remaining part with LOESS for the trend component. This non-parametric methodology provides better flexibility in describing seasonal trends compared to other analytical approaches. It should be observed that for computations, no temporal continuity is required, allowing the presence of missing values, which almost will affect the robustness of the local month average. By the way, the presence of the smoothing procedure reduces the effect of the small number of observations used to calculate the monthly averages. It follows that missing values, even large numbers, do not affect a sufficiently reliable definition of the seasonal components.

The behavior of sea surface temperature is typically influenced by climate change [6] but, in this context, due to the “short time period” (relative to a climate scale) and that the areas are in a coastal zone (apart from the Condor Mouth) it is possible to assume independence from climate change.

Besides the general trend of the time series, it is also interesting to analyze whether the behavior of past data influences the current one. Each time series value can affect its previous one, reflected in the autocorrelation. This phenomenon defines the degree of linear dependence between data in a time series; it has been measured with the autocorrelation coefficient:

$$r_k = \frac{\sum_{t=k+1}^n (y_t - \bar{y})(y_{t-k} - \bar{y})}{\sum_{t=1}^n (y_t - \bar{y})^2}, \tag{1}$$

where  $y_t \in \mathbb{R}^n$  is the time series,  $\bar{y}$  its mean and  $k$  the temporal lag. It is important to observe that for time lags greater than half the series length, statistical significance of  $r_k$  is lost, since it is obtained using a number of data much smaller than the analysed series. Therefore, to provide more reliable values of the autocorrelation coefficient it is advisable to compute it at most up to  $n/4$  in order to have sufficiently long series. Correlogram plots are also present in the framework to graphically inspect autocorrelation. This plot provides a clear representation of the time dependency structure, allowing to visually capture some general physical properties of the analysed hydrological data (such as seasonality, temporal linkage in the short and long term, non-stationarity). The correlogram plot of concentration of Chlorophyll-a in Zakynthos water is reported in Fig. 3.



**Fig. 3.** Correlogram plot of concentration of Chlorophyll-a in Zakynthos water.

Even if the autocorrelation coefficient presents a some kind of variation, it is positive and with maximum values in correspondence to  $k$  values such as to configure an annual periodicity (e.g. for  $k$  equal to 12 or its multiples). While they are lower or negative for other values of  $k$  (e.g. for  $k$  equal to 6 and 18). This reflects strong correlation with values of the same periods of previous years. This behaviour characterizes a phenomenon which varies during each year and in a similar way from one year to the other. The same behaviour highlighted in Fig. 3 was found in most of the time series examined, confirming the prevalence of a seasonal component in analyzed data.

## 4.2 Pre-processing Time Series Phase

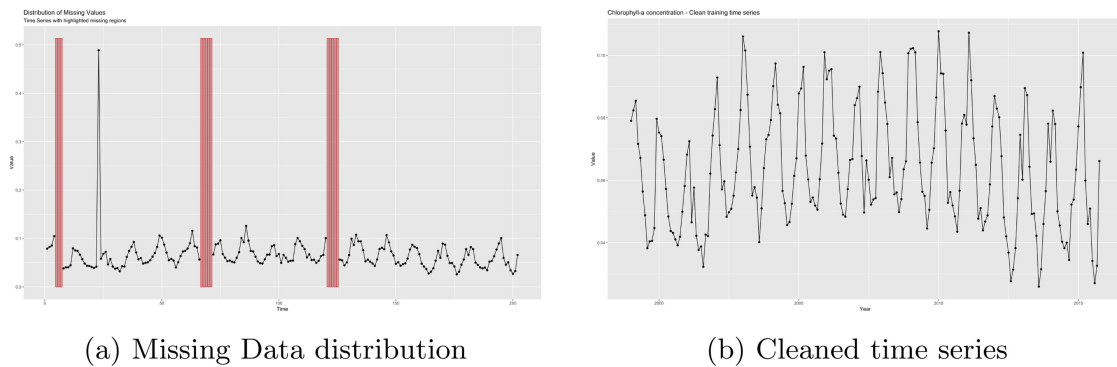
Entering the preprocessing phase requires to split data series with an hold out of 80-20. The first part will be preprocessed and used for fitting the model, whereas the latter for testing and evaluating the accuracy of the constructed model.

Measurement instrument malfunction or interruption of the survey service for maintenance is a very frequent problem in hydrological data time series acquisition; these occurrences cause the presence of missing data and outliers embedded

in the time series which can lead to meaningless, highly biased, and/or even incorrect analyses. Building a prediction model by neglecting the presence of missing data and values that deviate significantly from the general trend in parameter values could cause distortions in the prediction model.

The proposed framework is equipped with appropriate tools able to imputing the missing values and replacing the outliers with estimates that are as consistent as possible with the data trend.

Missing data were linearly imputed from STL deseasoned series. Once the series is completed the seasonal component is added. Figure 4a shows the distribution of the missing data for the Chlorophyll-a concentration in Zakynthos waters.



**Fig. 4.** Analysis of missing data on the parameter Chlorophyll-a concentration in Zakynthos sea

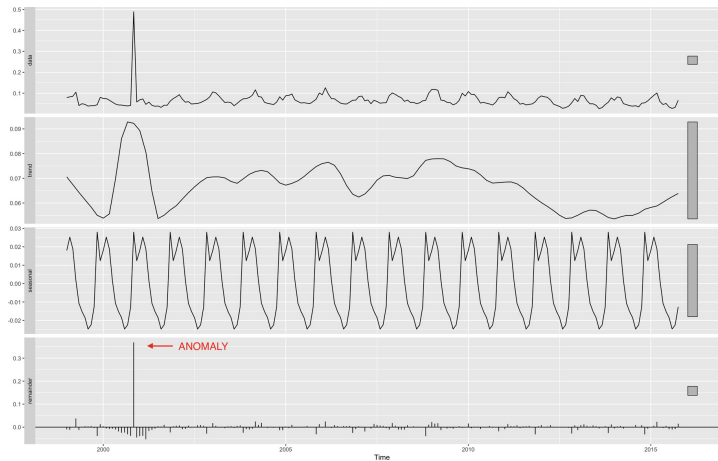
Similarly, anomalies are found from residual of STL decomposition, by looking for values outside the default range  $\pm 2(q_{0.9} - q_{0.1})$ , where  $q_p$  is the  $p$ -th residual quantile, estimated taking into account the behavior of the parameter. Figure 5 shows an example of anomalies for the Chlorophyll-a in Zakynthos water on its STL decomposition in which one anomaly is founded and corrected.

Once the time series data have been completed, the framework also checks if appropriate transformations are additionally required to finally prepare the series and then use it to construct the anomalies detection model. The Box and Cox transformation [1] described by the following equation is eventually adopted to stabilise the variance of a given time series  $\{y_t\}_{t=t_0, \dots, t_{fin}}$

$$w_t = \begin{cases} \log(y_t) & \text{se } \lambda = 0, \\ (y_t^\lambda - 1)/\lambda & \text{se } \lambda \neq 0. \end{cases} \quad (2)$$

Equation 2 depends on the parameter  $\lambda$  that gives good results when the seasonal variation remains constant. In the proposed framework  $\lambda$  is estimated according to the proposal in [7]. It is worthy to note that at the end of the process, output needs to be back-transformed to obtain predictions readable on the original scale.

Time series can be stabilised to be stationary, this is an assumption most prediction models are based on. Different reasons can cause non-stationary data time



**Fig. 5.** Example of anomalies identified for the Chlorophyll-a related to Zakynthos sea.

series: they could be ascribable to climatic change over a long period, they could be determined by drifts of the measuring instrument or even they could be associated with a change in the type of measuring instrument (i.e. the shift of the point of measurement). Verification of the stationarity of a hydrological data set should therefore be the preparatory step for any hydrological assessment. Exploratory and visual data analysis could provide information on non-stationary characteristics, at least for the most evident ones. Time and seasonal plots, such as correlogram, can be useful to identify non-stationary time series. Regarding this latter, the autocorrelation function of a stationary time series, tends to zero rather quickly, on the contrary of the slow decrease for non-stationary data. In this context, in addition to graphical tools, it is therefore particularly important to have statistical methodologies able to verify the stationarity for a long series of data, and in lacking of the basics hypothesis, it is necessary to make corrections. These approaches can be useful to understand and evaluate the state of the ocean ecosystem but also to implement timely action to restore it. One way to make a time series stationary is the so-called differencing, which is about calculating the differences between consecutive observations. By using the lag operator  $B$  on the time series  $y_t$ , it can be defined the  $d$ -order differencing as  $y_t^d = (1 - B)^d y_t$ , and the seasonal differencing as  $y_t'' = (1 - B)(1 - B^m)y_t$  where  $m$  is the number of seasons. The first difference represents changes between two consecutive observations, whereas the latter changes between years. One way to statistically determine whether differentiation is necessary, is to use a unit root test. In our workflow, it is automatically included into the model the well known Kwiatkowski-Phillips-Schmidt-Shin (KPSS) test [9], that checks for null hypothesis regarding stationary data. Several KPSS allow to determine the optimal number needed to make stationary the time series.

Figure 4b shows the output from the Pre-processing time series phases for the Chlorophyll-a concentration in the Ionian sea. Other delays in differencing are unlikely to make interpretative sense and should, therefore, be avoided. Moreover, care should be taken not to apply more lags than necessary, as this

may induce false dynamics or autocorrelations that do not really exist in the evaluated time series.

### 4.3 Fitting and Evaluation Model Generation Phase

Once the preliminary exploratory and preprocessing phases end, it enters the model fitting step. This phase is devoted to find the appropriate prediction model best fitting the hydrological time series. This will be subsequently adopted to produce predictions necessary for the anomalies detection. This latter phase, in fact, is performed on the basis of comparisons between the expected and the actual values to recognise whenever the presence of anomalous parameter occurs.

Since the seasonal nature of the analysed data, we used as fitting mathematical model SARIMA (Seasonal AutoRegressive Moving Average) method. It assumes stationarity of the time series and it is composed by two main parts: the AutoRegressive (AR) and the Moving Average (MA) models. The AR part allows to incorporate the effect of past values while the MA one represents a part of the time series not explained by trend or seasonality, permitting to set the model error as a linear combination of the error values observed in the past. An integrated part is added to the union of these two models, which allows to manage the differentiation of a series. It comes from ARIMA model that, depending on some parameters  $(p, d, q)$ , can be written as:

$$(1 - \phi_1 B - \dots - \phi_p B^p) (1 - B)^d y_t = c + (1 + \theta_1 B + \dots + \theta_q B^q) \varepsilon_t, \tag{3}$$

$\uparrow$   
AR(  $p$  )

$\uparrow$   
 $d$  differences

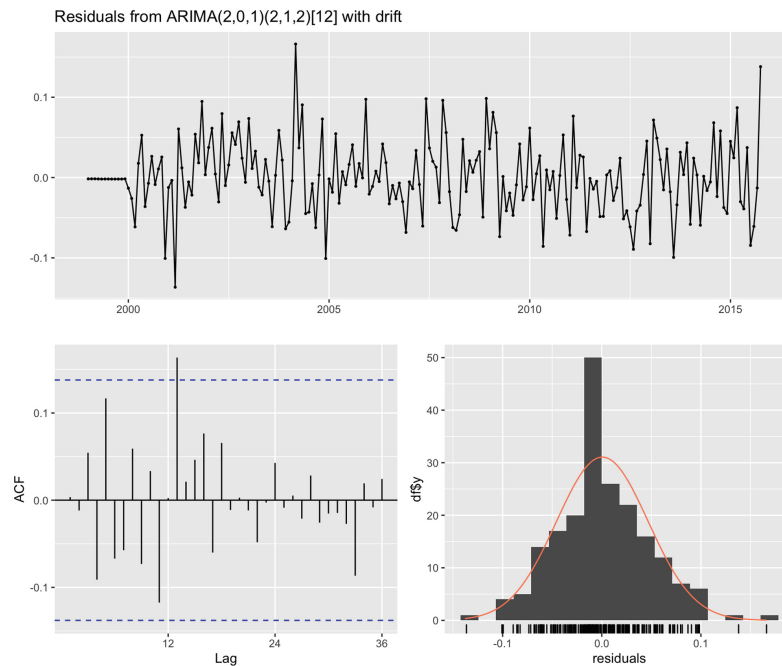
$\uparrow$   
MA(  $q$  )

where  $\mu$  is the mean of  $(1 - B)^d y_t$ ,  $\varepsilon_t$  is the white noise and  $\phi$  and  $\theta$  the model coefficients in  $\mathbb{R}$ . The optimal values of the model hyperparameters  $(p, d, q)$  are estimated using a variation of the algorithm proposed in [8] and combining test for unit roots and the minimization of both AICc (*corrected Akaike's Information Criterion*) and MLE (*Maximum Likelihood Estimation*). These processes find the order of the model  $(p, d, q)$ , estimate the parameters  $c, \phi_1, \dots, \phi_p, \theta_1, \dots, \theta_q$  and then compute the best model fitting given data. SARIMA includes seasonal terms, similar to the not seasonal one, but with a back shift of the seasonal period, that is  $ARIMA((p, d, q), (P, D, Q)_m)$ , being  $P, D, Q$  and  $m$  (the number of time steps for a single seasonal period) four seasonal elements that are not part of ARIMA and must be configured. To check the correctness of trained SARIMA model in capturing right information embedded in data, a residuals analysis is also performed. In fact, a good forecasting method produces uncorrelated and zero mean residuals, while the presence of correlations between the residuals indicates information left in it that should be used in the forecasting calculation. If the residuals have a non-zero mean, then the forecasts are biased. Graphical analysis based on correlogram and the Ljung-Box<sup>2</sup> statistical test [10] were used together with a  $T$ -test for the hypothesis of zero mean residuals.

---

<sup>2</sup> LB test computes the statistics  $T_{L\&B} = n(n + 2) \sum_{k=1}^h (n - k)^{-1} r_k^2 \sim \chi(h)$ , with  $h$  degrees of freedom usually chosen as  $h = \min(2m, n/5)$  [8].

Figure 6 reports the residual analysis performed on the estimated model for the Chlorophyll-a concentration in Zakynthos sea water.



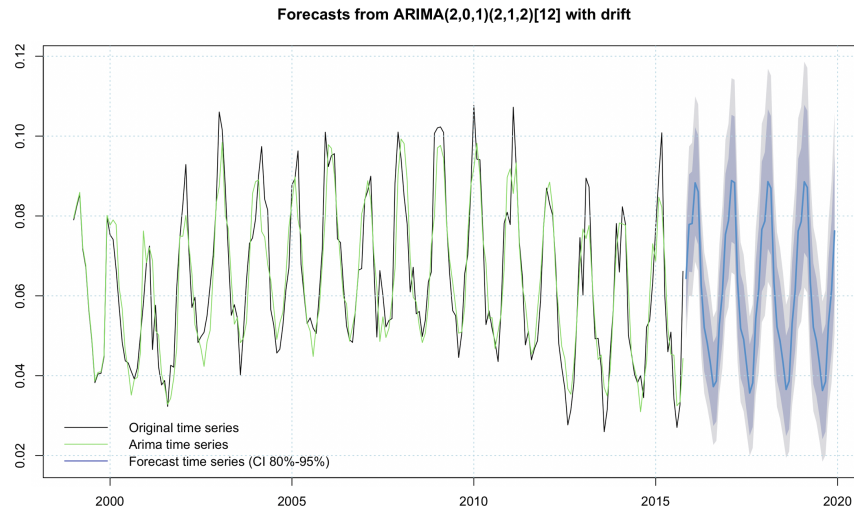
**Fig. 6.** Residual behaviour of the estimated model for the Chlorophyll-a concentration in Zakynthos sea water.

As it can be observed the series presents a little but significant peak, still coherent with white noise assumption. Moreover, statistical tests confirm the absence of residual correlation and zero mean ( $p$ -values equals to 0.1288 and 0.8660, for LB and  $t$ -test, respectively).

It should be noted that since real hydrological data are considered, it may happen that the lack of correlation between the residuals is not verified. In this case the model can still be used to make forecasts, but the forecast intervals may not be accurate due to the correlated residuals.

#### 4.4 Forecasting and Model Evaluation

After the model of time series data is estimated and the white noise assumption on residuals checked, the framework proceeds with the forecasting phase using as confidence interval 80% and 95% as suggested in literature [2]. As an example, Fig. 7 reports the estimated model and the forecasting for the Chlorophyll-a concentration on the Zakynthos sea water.



**Fig. 7.** Estimated model and the forecasting for the Chlorophyll-a concentration on the Zakynthos sea water forecasting.

**Table 1.** Performance values of the proposed predictive model for Chlorophyll-a concentration in the Zakhyntos waters.

	RMSE	MAE	MPE	MAPE	MASE
Training set	0.0079	0.0058	-1.1484	9.4642	0.6455
Test set	0.0472	0.0337	18.3203	30.9848	3.7262

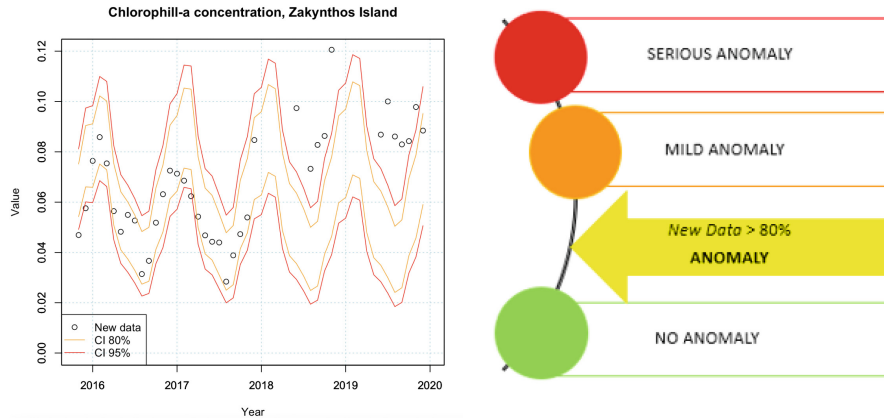
The error<sup>3</sup> model is  $e_{T+h} = y_{T+h} - \hat{y}_{T+h|T}$ , where  $\{y_1, \dots, y_T\}$  is the training set and  $\{y_{T+1}, y_{T+2}, \dots\}$  the test set. For seasonal time series of length  $n$  and seasonal period  $m$ , errors are measured using: i) Standard and Root Mean Squared Error (MSE/RMSE); ii) Mean Absolute Percentage Error (MAPE) and iii) Standard and Scaled Mean Absolute Error (MAE/MASE).

Table 1 reports the performances of predictive model for Chlorophyll-a concentration in the Zakhyntos waters. All the measures confirm the goodness of the estimated model.

#### 4.5 Anomalies Detection

To detect if a sample in the data time series is an anomaly, the framework compares the new value with the forecast results constructed as previously described. All values which fall outside the confidence interval of 80% were labelled as anomalous points. The anomalies detection phase of the proposed framework is also able to classify the levels of anomalies from moderate to severe (as illustrated

<sup>3</sup> This differ from residual, computed on training set, that are actual errors from the fitted and real values  $e_t = y_t - \hat{y}_t$ .



**Fig. 8.** Anomalies detection phase: new data are depicted as black circle points, while regions of serious and mild anomalies are bounded by red and orange solid lines, respectively (Color figure online)

in Fig. 8 where new data are depicted as black circle points, while regions of serious and mild anomalies are bounded by red and orange solid lines, respectively). These levels strictly depend on the nature of the parameter under studies. In the context described in the paper, for Chlorophyll-a concentration, if the value is below the prediction it is never labelled as severe as shown in Fig. 8. Appendix A described, with the help of visual panels, the analysis conducted for other parameters under study. From these it can be seen the levels of severity depends on the expert of domain. When surface temperature is taken into account, the severity degree of the anomaly occurs either the anomaly is above or below the prediction value, whereas for dissolved oxygen concentration if the value is above the prediction the anomaly is never severe.

Identification of anomalies is of great importance in order to evaluate impact of human activities on coastal environment and vice versa, in a framework of integrated coastal management. In Marine-EO the focus was put on Marine Protected Areas (MPA) and on aquaculture farms. For what concerns the former, to estimate the number and type of anomalies in a given time frame is important in order to evaluate the eventual impact of external factors to the health of the MPA and possibly to identify them by crossing with other information (e.g. changes occurring in surrounding areas or implementation of access policies). This will allow then to define actions to mitigate such impact on the coastal environment. Concerning aquaculture farms, it is important for the farmers to be rapidly informed of specific anomalies in order to be able to take measures to avoid production losses. In this case they are interested in those specific anomalies that can impact the fishes: different type of anomalies can have different impact. For example lower oxygen concentration is extremely dangerous, because

it can result in fishes death, while higher oxygen concentration usually has not a relevant impact. A “positive” sea surface temperature anomaly, together with similar chlorophyll anomaly, can lead to dangerous phenomena like algal bloom or foster disease spreading. While low temperature can result in a decrease of growth rate.

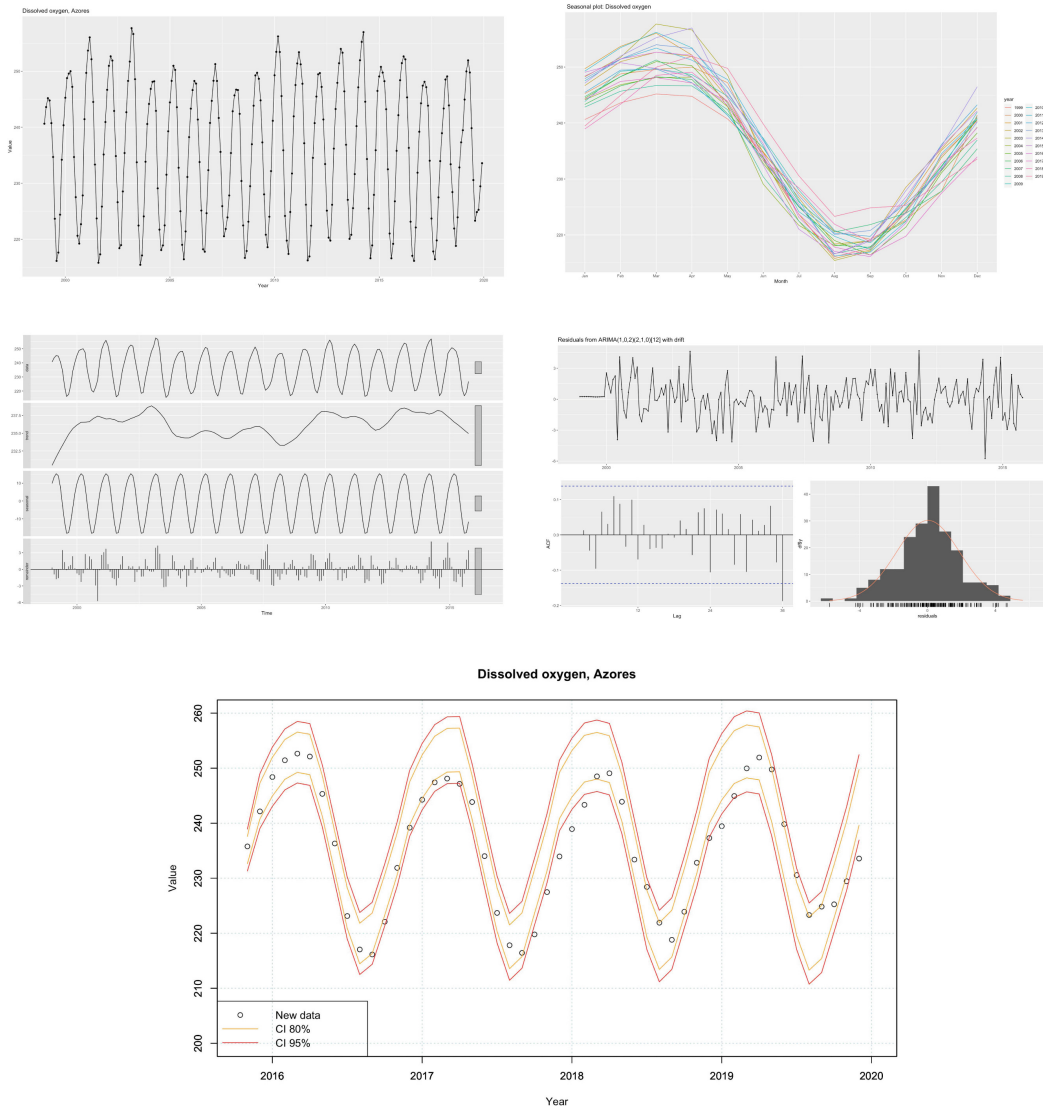
## 5 Discussions

This paper briefly illustrates a novel framework for anomalies detection in data time series related to some physical and biogeochemical parameters of marine and oceanic waters, such as surface temperature profiles, dissolved oxygen and Chlorophyll-a concentrations of the waters of some areas of interest. The peculiarity of the approach is the possibility of classify a new value as an anomaly on the basis of a comparison with appropriate forecasting trained model and also indicating the level of detected anomaly point from severe to mild according to some threshold defined in agreement with the domain experts.

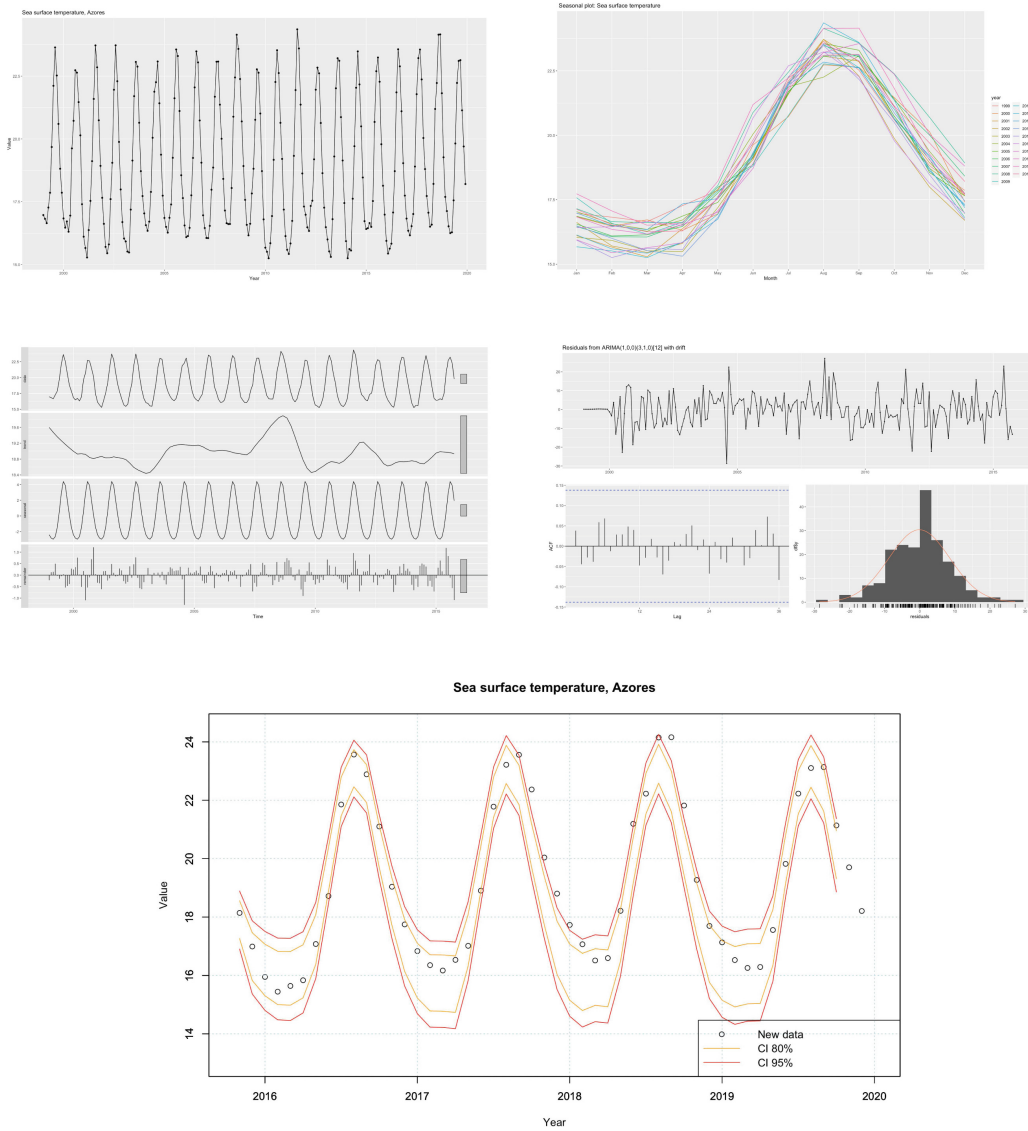
**Acknowledgments.** This work was supported in part by the GNCS-INDAM (Gruppo Nazionale per il Calcolo Scientifico of Istituto Nazionale di Alta Matematica) Francesco Severi, P. le Aldo Moro, Roma, Italy. The author F.E. was funded by REFIN Project, grant number 363BB1F4, Reference project idea UNIBA027 “Un modello numerico-matematico basato su metodologie di algebra lineare e multilineare per l’analisi di dati genomici”.

## A Appendix

The proposed framework was also applied to time series recording dissolved oxygen and sea surface temperature values in some marine zones of the archipelago of the Azores. Figures 9 and 10 report the results for the anomalies detection of the dissolved oxygen in the Condor Mouth (Azores) and of the Sea Surface Temperature in the Ribeira Quente (Azores), respectively.



**Fig. 9.** Panel for the analysis of dissolved oxygen time series in the Condor Mouth, Azores. From left to right and from top to bottom, the five plots illustrate: time and seasonal plots of dissolved oxygen; the time series decomposition; the residual behaviour of the estimated model; the anomalies detection on new data (black circle points) through the estimate model.



**Fig. 10.** Panel for the analysis of Sea Surface Temperature time series archipelago of the Azores. From left to right and from top to bottom, the five plots illustrate: time and seasonal plots of sea surface temperature; the decomposition of the time series; the residual behaviour of the estimated model; the anomalies detection on new data (black circle points) through the estimate model.

## References

1. Box, G.E.P., Cox, D.R.: An analysis of transformations. *J. R. Stat. Soc. Series. B Stat. Methodol.* **26**(2), 211–252 (1964)
2. Brockwell, P.J., Davis, R.A.: *Introduction to Time Series and Forecasting*. STS, Springer, Cham (2016). <https://doi.org/10.1007/978-3-319-29854-2>
3. Cleveland, R.B., Cleveland, W.S., McRae, J.E., Terpenning, I.: STL: a seasonal-trend decomposition procedure based on loess. *J. Offic. Stat.* **6**(1), 3–73 (1990)
4. European Commission: Copernicus Europe’s eyes on earth. <https://www.copernicus.eu/en>. Accessed 20 Apr 2022

5. European Commission: Copernicus marine service. <https://marine.copernicus.eu/en>. Accessed 20 Apr 2022
6. U. EPA: Climate change indicators: Sea surface temperature. <https://www.epa.gov/climate-indicators/climate-change-indicators-sea-surface-temperature>. Accessed 20 Apr 2022
7. Guerrero, V.M.: Time-series analysis supported by power transformations. *J. Forecast.* **12**(1), 37–48 (1993)
8. Hyndman, R., Khandakar, Y.: Automatic time series forecasting: the forecast package for R. *J. Stat. Softw.* **26** (2008)
9. Kwiatkowski, D., Phillips, P.C., Schmidt, P., Yongcheol, S.: Testing the null hypothesis of stationarity against the alternative of a unit root: how sure are we that economic time series have a unit root? *J. Economet.* **54**(1), 159–178 (1992)
10. Ljung, G.M., Box, G.E.P.: On a measure of lack of fit in time series models. *Biometrika* **65**(2), 297–303 (1978)
11. R Core Team: R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria (2017)