CLASSIFICATION OF SATELLITE DERIVED CHLOROPHYLL A SPACE-TIME SERIES BY MEANS OF QUANTILE REGRESSION: AN APPLICATION TO THE ADRIATIC SEA

P. Girardi[1], R. Pastres[1], C. Gaetan[1], A. Mangin[2], and M. A. Taji[3]

[1]*Ca' Foscari University Venice, Italy*
[2]*ACRI-ST Sophia Antipolis, France*
[3]*ACRI-EC Casablanca, Morocco*

ABSTRACT

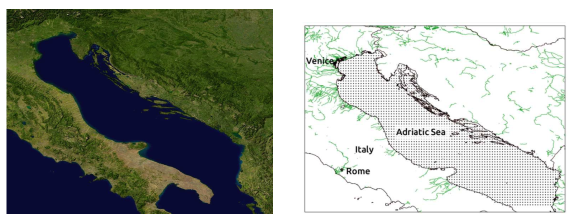
In this paper, we present the results of a classification of Adriatic waters, based on spatial time series of remotely sensed Chlorophyll type-a. The study was carried out using a clustering procedure combining quantile smoothing and an agglomerative clustering algorithms. The smoothing function includes a seasonal term, thus allowing one to classify areas according to "similar" seasonal evolution, as well as according to "similar" trends. This methodology, which is here applied for the first time to Ocean Colour data, is more robust with respect to other classical methods, as it does not require any assumption on the probability distribution of the data. This approach was applied to the classification of an eleven year long time series, from January 2002 to December 2012, of monthly values of Chlorophyll type-a concentrations covering the whole Adriatic Sea. The data set was made available by ACRI (http://hermes.acri.fr) in the framework of the GlobColour Project (http://www.globcolour.info). Data were obtained by calibrating Ocean Colour data provided by different satellite missions, such as MERIS, SeaWiFS and MODIS. The results clearly show the presence of North-South and West-East gradient in the level of Chlorophyll, which is consistent with literature findings. This analysis could provide a sound basis for the identification of "water bodies" and of Chlorophyll type-a thresholds which define their Good Ecological Status, in terms of trophic level, as required by the implementation of the Marine Strategy Framework Directive. The forthcoming availability of Sentinel-3 OLCI data, in continuity of the previous missions, and with perspective of more than a 15-year monitoring system, offers a real opportunity of expansion of our study as a strong support to the implementation of both the EU Marine Strategy Framework Directive and the UNEP-MAP Ecosystem Approach in the Mediterranean.

Key words: Functional data analysis, Quantile regression, Clustering, Satellite data.

Proc. 'Sentinel-3 for Science Workshop', Venice, Italy 2–5 June 2015 (ESA SP-734, December 2015)

1. INTRODUCTION

With the scope to achieving a "good ecological status" for all bodies of surface water, in the year 2008 the Marine Strategy Framework Directive (MSFD) prescribes monitoring from the coast to the Exclusive Economic Zone (EEZ), which can reach up to 200 nautical miles from the coast [Olenin et al. (2010)]. The evaluation of biological, physico-chemical and hydromorphological quality elements are key activities to evaluate of the water status. Many Pressures may influence the trophic



(a) (b)

Figure 1. Adriatic Sea(a) and satellite grid-point (b)

status of marine ecosystems, including an excessive loads of nutrients from human activity (sewage effluents, aquaculture farms and industrial activities) or natural causes (river fluxes, lagoon, etc. . .). Nutrient enrichment, in some instancese, give rise to eutrophication, which causes many adverse effects for the marine ecosystem. Chlorophyll type-a (Chl-a) concentration is an indicator of phytoplankton biomass and a well established indicator of eutrophication.
Among the European waters, the Adriatic Sea presents some peculiarities: a) it is an almost land-locked basin separated from the central Mediterranean by the Strait of Otranto; b), it encompasses a significant diversity of properties, from eutrophic or oligotrophic; c) water quality in the Italian coastal zone is affected by discharges of several rivers and, in particular, by that of the Po river. Due these conditions, the Adriatic Sea hosts a variety of habitats and presents North-South as well as East-West gradients of nutrient concentrations and physico-chemical properties, as well as sign of degradation of its environmental status [Giani et al. (2012); Lotze et al. (2006);

Diaz & Rosenberg (2008)].

The implementation of the MSFD does not strictly require the definition of reference conditions for each assessment unit, called “water body”. However, reference conditions combined with thresholds which define the acceptable deviation from them are the most common approach to define the boundary between the “good” or “not good” environmental status. Identifying water bodies and reference

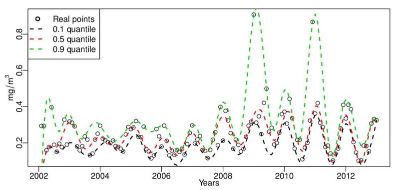


Figure 2. An example of Chl-a time series with three different quantile regression curves.

conditions is, in fact, a critical step in the application of any integrative assessment methodology: in this paper, we suggest an innovative procedure for guiding this process, based on an objective classification process, based on the concept of similarity between the whole time series of data, rather than on the proximity of summary statistical indeces. Functional Data Analysis (FDA) is a branch of statistics where the latent model is given by smooth curves or continuous functions [James & Sugar (2003)]. We use Quantile Smoothing Regression (QSR) relaxing assumptions on distribution [Koenker et al. (1994)] with a flexible approach. Chl-a time series shows seasonal variations and changes in amplitude and this change may elegantly estimates by seasonal modulation model [Eilers et al. (2008)]. The aim of this paper is to present a flexible functional clustering procedure applied to spatial time series. This technique includes the quantile regression with a seasonal modulation component. It was applied to time series from 2002 to 2012 of Chl-a concentrations.

2. FUNCTIONAL CLUSTERING BY QUANTILE SMOOTHING REGRESSION

Time series could be seen as observations of a continuous function collected at certain points Ramsay & Silverman (2005). In some cases, the interest of the researchers focus on quantiles; the use of quantiles may be dictated by practical uses (several specifications in environmental sciences are based on a certain quantile) or the need for a more robust and flexible approach, free of assumption on the probability distribution of the variable of interest. In addition, the quantile regression permits to obtain a curve with the scope to describe a certain quantile (or a series of quantiles). Quantile regression [Koenker et al. (1994)] estimates the considered smoothing curves.
The variable Y is associated to a vector of covariates X and assuming linear dependence on covariates, the τ-th

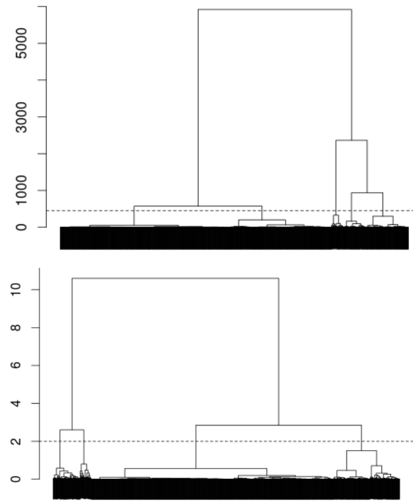


Figure 3. Dendrogram based on the distance matrix of trend (a) and seasonal coefficients (b). The dashed line represented the cut-off threshold

quantile regression is defined by $Q_{Y|X}(\tau|x) = x^T\theta$ and the regression coefficients θ^* are obtained by

$$\theta^* = \arg\min_\theta E[\rho_\tau(Y - X^T\theta)|X = x],$$

where τ is the tilded absolute value function $\rho_\tau(x) = x * (\tau - I(x < 0))$. We replace the parametric part with a smoothing function using the generalized regression quantiles [Guo et al. (2015)]. The coefficient estimation is obtained minimizing the unconditional expect loss function $l_\tau = \arg\min_f E[\rho_\tau(Y - \sum_{i=1}^k \theta_i\beta_i(x))]$, where k is the number of knots and $\beta_i(x)$ are the B-splines basis. The generalized QSR can be explained as:

$$\theta^* = \arg\min_\theta \sum_{j=1}^{n}(\rho_\tau(y_j - \sum_{i=1}^{k}\theta_i\beta_i(x)))$$

Finally, we include in the QSR a truncated Fourier series [Eilers et al. (2008)]. The previous formula became:

$$(\theta^*, \alpha^*, \phi^*) =$$

$$= \arg\min_{\theta,\alpha,\phi} \sum_{j=1}^{n} \rho_\tau(y_j - \sum_{i=1}^{k}\theta_i\beta_i(x_j) - S_i(x_j, \alpha, \phi)),$$

where the seasonal part $S_i(x_j, \alpha, \phi))$ is

$$S_i(x_j, \alpha, \phi) =$$

$$= \alpha(\beta_i(x_j)sin(\frac{2\pi}{12}x_j)) + \phi(\beta_i(x_j)cos(\frac{2\pi g}{12}x_j)).$$

The coefficients θ estimates the trend; α and ϕ describe the changes in amplitude of the sine and cosine waves, respectively. The clustering procedure is based on a two-stage approach [Abraham et al. (2003)]: in the first stage we estimate the coefficients θ, α and ϕ. In the second step, we calculate two different distance matrix based on L2-norm, the first one using the trend coefficients and the second one considering the seasonal component $S_i(x_j, \alpha, \phi))$. A Ward's agglomerative hierarchical method aggregates similar curves and the number of cluster is chosen according to the tree dendrogram.

3. CLUSTERING CHL-A CONCENTRATIONS ON THE ADRIATIC SEA

The methodology presented in the previous sections was applied in the first step to a comprehensive sets of time series, i.e. the mean monthly values of the Chl-a concentrations in the Adriatic Sea (from January, 2002 to December 2012).

For each month, gridded data with resolution of 192×240 points (4km scale and $46\,080$ values) are available, but we rescaled the resolution to 96×120 points (8km scale and $11\,520$ values); 2168 time series cover the whole Adriatic Sea. The presented results were obtained considering the 0.5 quantile. The dendrograms suggested 5 and 4 clusters for the trend component and the seasonal part, respectively.

Trend classification - 5 clusters
In the Figure 4 we present the results of clustering methodology based on the trends behaviour and the Figure 5 the related estimated curves. The cluster 1 ($\#1$) shows chl-a values are lower because the sites belonging to this cluster are located far from sources of land based sources of Nitrogen and Phosphorous. Cluster 2 encompass several areas disseminated in different zones of the Adriatic Sea: the coastal area in the southern Italy, the Albanian coast and areas in the northern part of the Adriatic Sea between 44th and 45th parallel North and certain areas of the Dalmatian coasts. Although both the trend is analogue to the first cluster, the values are higher ranging from 0.1 to 0.7 mg/m^3. The cluster 3 covers a smaller area compared to the previous ones, including only the coasts of the intermediate Adriatic Sea and a little portion of the north Adriatic Sea far from the coast. This cluster presents an average value of about 1.7 mg/m^3 with a peak value in the year 2010. Clusters 4 and 5 cover to the northern part and most of the Italian coastal zone. Although, the cluster 5 reports around 1.5-fold the values of the cluster 4, they display a similar temporal behavior with a pronounced maximum value in the year 2010.

Seasonal classification – 4 clusters
Considering the classification of the seasonal distance matrix, it suggested the presence of 4 clusters. The Figures 6 and 7 represent the spatial distribution of these 4 clusters and their seasonal evolution. The first cluster (#1) covers the 87% of the Adriatic Sea; the seasonal component is approximately constant over the years with a peak in the early spring and a minimum in the summer. The cluster #2 covers to the low-intermediate Adriatic Italian coasts

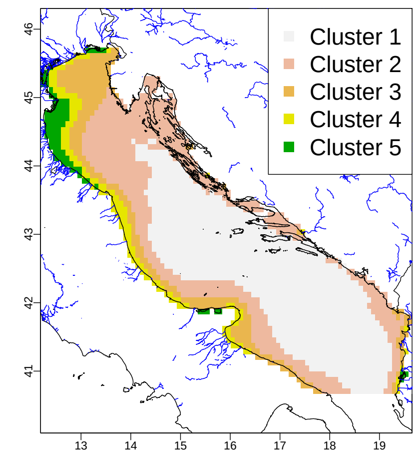


Figure 4. Spatial distribution of the classification estimated on the basis of trend in the QSR based on the 0.5 quantile.

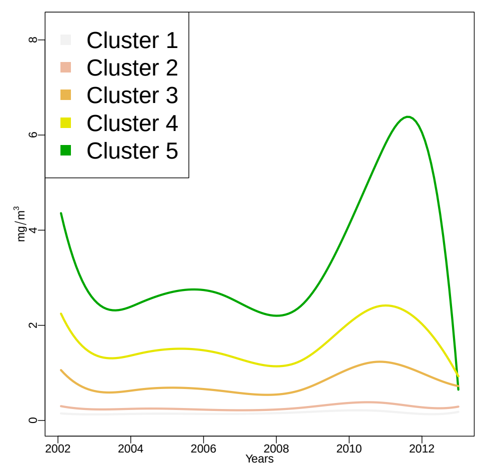


Figure 5. Temporal behaviors estimated by the trend curves in the QSR based on the 0.5 quantile.

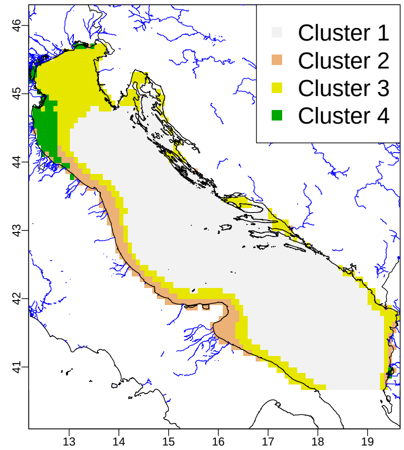


Figure 6. Spatial distribution of the classification estimated on the basis of seasonality in the QSR based on the 0.5 quantile.

and its oscillation was similar to the previous cluster. The cluster #3 included the Northern Adriatic Sea and some areas near to the Albanian coasts. Changes in amplitude were observed at the beginning and at the end of the period. The last cluster (#4) is closed to the Italian coast at south of the Po river and the Karavasta Lagoon in Albania. The sites classified in this cluster reported the highest seasonal oscillation.
The presence of North-South and West-East gradient in the level of Chlorophyll was consistent with literature findings. Furthermore, it also highlighted the presence of a more pronounced seasonality along the Northern Italian coast, probably driven by the nutrients apportioned by river discharges.

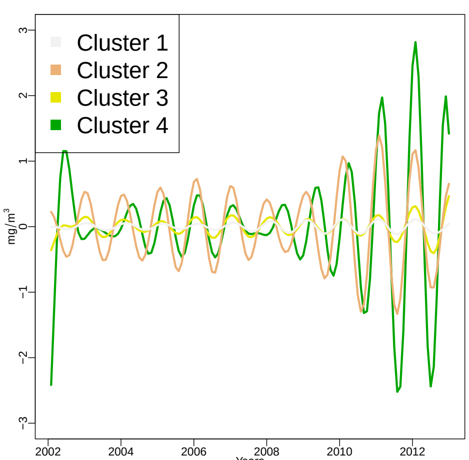


Figure 7. Seasonal curves estimated by QSR based in the 0.5 quantile.

4. DISCUSSIONS

This analysis could provided a sound basis for the identification of "water bodies" and of chl-a thresholds which define their Good Environmental Status, in terms of trophic level, as required by the implementation of the Marine Strategy Framework Directive. At present, such thresholds are selected on the basis of mean values, without taking trend and seasonality into account.
This paper proposes and develops a robust approach to define similar areas in relation to temporal profile of a variable. The presented technique is applied to glob-colour data related to chlorophyll type-a concentration in the Adriatic Sea. This methodology allows to classify time series taking in to account a modulation in the seasonal signal and a certain quantile of the parameter distribution. In many fields and particularly in environmental policy, threshold values based on percentile were used Reich (2012); Schmidt et al. (2012). In addition the spatial dependence can be easily incorporated by existing methods [Giraldo et al. (2012)]. The forthcoming availability of Sentinel-3 OLCI data, in continuity of the previous missions, and with perspective of more than a 15-year monitoring system, offers a real opportunity of expansion of our study as a strong support to the implementation of both the EU Marine Strategy Framework Directive and the UNEP-MAP Ecosystem Approach in the Mediterranean.

REFERENCES

Abraham, C., Cornillon, P. A., Matzner-Løeber, E., & Molinari, N. 2003, Scandinavian Journal of Statistics, 30, 581

Diaz, R. J. & Rosenberg, R. 2008, science, 321, 926

Eilers, P. H., Gampe, J., Marx, B. D., & Rau, R. 2008, Statistics in Medicine, 27, 3430

Giani, M., Djakovac, T., Degobbis, D., et al. 2012, Estuarine, Coastal and Shelf Science, 115, 1

Giraldo, R., Delicado, P., & Mateu, J. 2012, Statistica Neerlandica, 66, 403

Guo, M., Zhou, L., Huang, J. Z., & H´ardle, W. K. 2015, Statistics and Computing, 25, 189

James, G. M. & Sugar, C. A. 2003, Journal of the American Statistical Association, 98, 397

Koenker, R., Ng, P., & Portnoy, S. 1994, Biometrika, 81, 673

Lotze, H. K., Lenihan, H. S., Bourque, B. J., et al. 2006, Science, 312, 1806

Olenin, S., Alemany, F., Cardoso, A., et al. 2010, Task Group, 2

Ramsay, J. O. & Silverman, B. W. 2005, Functional Data Analysis, 2nd edn., Springer Series in Statistics (Springer)

Reich, B. J. 2012, Journal of the Royal Statistical Society Series C, 61, 535

Schmidt, T. S., Clements, W. H., & Cade, B. S. 2012, Freshwater Science, 31, 709