Ecological Indicators 95 (2018) 765–774

Contents lists available at ScienceDirect

Ecological Indicators

journal homepage: www.elsevier.com/locate/ecolind

Original Articles

Machine learning predictions of trophic status indicators and plankton dynamic in coastal lagoons

Béchir Béjaoui[a], Ennio Ottaviani[b], Enrico Barelli[b], Boutheina Ziadi[c,f], Amel Dhib[c,f], Michel Lavoie[d], Coidessa Gianluca[e], Souad Turki[c], Cosimo Solidoro[e], Lotfi Aleya[f,*]

[a] Laboratoire Milieu Marin, Institut National des Sciences et Technologie de la Mer, Salammbô, Tunisia
[b] OnAIR srl, 26, Via Carlo Barabino, 16129 Genoa, Italy
[c] Institut National des Sciences et Technologies de la Mer, Centre la Goulette, Tunisia
[d] Université Laval – Unité mixte international Takuvik ULaval-CNRS and Québec-Océan, Département de Biologie, Pavillon Alexandre-Vachon, 1045, Avenue de la Médecine, Québec (Qc), Canada
[e] Istituto Nazionale di Oceanografia e di Geofisica Sperimentale—OGS, Borgo Grotta Gigante 42/c, Sgonico, Italy
[f] Laboratoire de Chrono-Environnement, UMR CNRS 6249, Université de Bourgogne-Franche-Comté, France

ARTICLE INFO

Keywords:
Coastal lagoon
Ghar El Melh Lagoon
Nutrients
Eutrophication
Random forest models
Phytoplankton descriptors
TRIX

ABSTRACT

Multivariate trophic indices provide an efficient way to assess and classify the eutrophication level and ecological status of a given water body, but their computation requires the availability of experimental information on many parameters, including biological data, that might not always be available. Here we show that machine learning techniques – once trained against a full data set – can be used to infer plankton biomass information from chemical and physical parameter only, so that trophic index can then be computed without using additional biological data. More specifically, we reconstruct plankton information from chemical and physical data, and this information together with chemical data was used to compute the TRIX, which was eventually used to assess the eutrophication status of the water body. The RF was also used to evaluate the prevailing mechanism (bottom-up versus top-down) controlling plankton dynamic. The case study was a Mediterranean lagoon, the Ghar El Melh Lagoon, which has been used as a natural laboratory to test the effectiveness of the proposed approach. Based on the resulting TRIX values (4.2 in April – 5.7 in December) the Ghar El Melh Lagoon can be classified an eutrophic ecosystem. This modeling process suggests that phytoplankton growth in Ghar El Melh Lagoon is mainly bottom-up control by nutrients availability, whereas the top-down control exerted by the zooplankton is relatively weak. Results highlight that in bottom up controlled lagoon machine learning technique can efficiently be used to compute ecological indicators even with low availability of biological data.

1. Introduction

Coastal lagoons are transitional areas between land and sea, covering 13% of the coastlines of all continents (Knoppers, 1994; Pérez-Ruzafa et al., 2011; Aleya et al., 2018). The lagoon services are well-documented (Esteves et al., 2008; Tavares and Siciliano, 2014; Newton et al., 2018) and classified according to the benefits provided to society (Lundquist et al., 2017). They provide a nursery areas for endangered species (Franco et al., 2006; Giles and Pilditch, 2006; Sousa et al., 2013; Potts et al., 2014), harbor habitats that host rich and uniquely biodiversity communities (Esteves et al., 2008; Whitfield et al., 2008; Tavares et al., 2015), are important fishery and aquaculture sites, and support tourism as well as spiritual and cultural activities (Chan et al., 2012). At the same time, coastal lagoons face a range of threats due to their proximity to human population centers and are among the most impacted ecosystems on Earth (Bulleri and Chapman, 2010; McKinney et al., 2010). Pollution and nutrient loads from industry and agriculture (Canu et al., 2010; Solidoro et al., 2010; Martins et al., 2015, Pastres et al., 2004), and overfishing are examples of anthropogenic pressures, which interacts with ecosystem alteration due to climate change (Cossarini et al., 2008; Anthony et al., 2009; Clausen and Clausen, 2014). The interplay among the multiple factors affecting water quality and eutrophication in coastal lagoons is complex and their cumulative effects are not always easy to predict with current modeling approaches (Vollenweider et al., 1980; Menesguen, 1992, Solidoro et al., 2005).

In many studies and applications the trophic status of coastal areas is quantitatively assessed by using multimetric indices, such as trophic index (TRIX), turbidity index (TRBIX) and a general water quality index

---

[*] Corresponding author.
E-mail address: lotfi.aleya@univ-fcomte.fr (L. Aleya).

https://doi.org/10.1016/j.ecolind.2018.08.041
Received 28 February 2018; Received in revised form 12 August 2018; Accepted 16 August 2018

(GWQI) (Vollenweider et al., 1998), or other combination of environmental parameters which have proved to be useful in classifying the trophic levels of marine and coastal ecosystems (Sharma, 1996; Primpas et al., 2010; Khedhri et al., 2016, 2017; Hachani et al., 2018). Multimetric trophic indices integrate the physical (e.g. temperature, salinity, water transparency), chemical (mainly nutrient –N, P, and Si– and dissolved oxygen concentration), and biological (e.g. plankton density, chlorophyll concentration) in new parameters meant to synthetically provide indication on the level of nourishment or food-web abundance within an ecosystem.

Although such multimetric indices have limits (e.g., outlier sensitivity, implicit assumptions on data distribution) (Pavlidou et al., 2015) -they are usually site-specific at geographic level-, their use is efficient in assessing the eutrophication level of a given water body (Béjaoui et al., 2016a). However, their computation requires the availability of experimental information on many different parameters, which might not be available. In particular, biological data are more expensive and usually less frequently measured than chemical or physical ones. Therefore, the possibility to use models to derive trophic indexes by using only a reduced subset of environmental parameters, and possibly without using biological data, is an important – but challenging- issue.

Machine Learning (ML) and data mining techniques are numerical methods can do not rely on a priori identification of interdependency among variables, nor of any conceptual model, and do not require any assumptions on data structure. These approaches have recently gained popularity in marine ecology (Andonegi et al., 2011; Bertoni et al., 2016) and have recently enhanced our ability to predict the effects of anthropogenic activities and climate change on marine ecosystems, including lagoons (Bandelj et al. 2009; Alves et al., 2013). In comparison to other techniques Random Forests (RF) (Breiman, 2001) present many advantages. RF do not assume any probability distribution for variables and they can also manage a large number of predictors, choosing among them the most useful within the given scope (Mulia et al. 2013; Park et al., 2015). In addition, RF is well suited for biological statistical models because they can be trained even on small datasets (Carvalho et al., 2011; Béjaoui et al., 2016a). RF predictions are also highly reliable as they come from an ensemble average of many simple models and thus avoid the over-fitting problem typical of many non-linear regression techniques (Phillips et al., 2008; Huang et al., 2015). Finally, since each tree is built on a random subset of the original data, no separate independent dataset or cross-validation procedure is required for assessment of the model's predictive performance (Were et al., 2015). This combined method for performance evaluation is called *Out-Of-Bag* (OOB) analysis (Breiman, 2001).

In this study we started form the idea that in water bodies in which the plankton dynamic is controlled more by phytoplankton productivity than from grazing pressures (i.e. plankton is bottom-up rather than top-down controlled) plankton biomass can be inferred by chemical and physical parameters only, and therefore multimetric trophic index can be computed without using biological data. More specifically, the objective is to build a comprehensive predictive model of eutrophication status of a coastal lagoon by coupling a Random Forest (RF) model and the multitrophic index (TRIX). The RF was used to reconstruct plankton information from chemical and physical data, to evaluate the prevailing mechanism (bottom-up versus top-down) controlling plankton dynamic, and to compute – together with chemical data- the TRIX, which was eventually used to assess the eutrophication status of the water body. The case study was a Mediterranean lagoon, the Ghar El Melh Lagoon, which has been used as a natural laboratory to test the effectiveness of the proposed approach.

2. Materials and methods

2.1. Study area

Ghar El Melh Lagoon, formerly known as Porto Farina Lake (Cézilly, 1912), is a Mediterranean lagoon located on the northeastern coast of Tunisia and to the northwest of the Gulf of Tunis (Fig. 1). It has an area of about 31 km^2 and a mean depth of about 1 m. The lagoon is composed of three parts: 1) the central sector, elliptically-shaped whose surface area is about 26 km^2; 2) the Sidi Ali El Mekki sector, located in the north, triangular in shape and with a surface area of 2.7 km^2, isolated from the central sector to the west by embankments situated on an old coastal strip and joined to the central lagoon by a narrow channel; and 3) the Sebkha El Ouafi sector, located further to the south with an area of about 2.5 km^2 which is permanently connected to the central lagoon (Dhib et al., 2013).

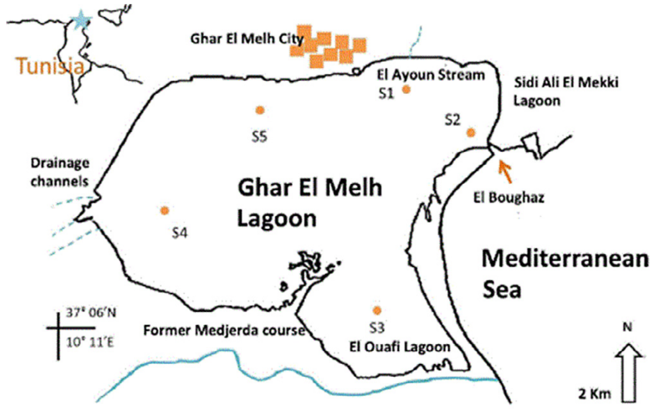

Fig. 1. Geographic localization of Ghar El Melh and location of sampling stations.

The lagoon is connected with the sea through an inlet named El Boughaz (width: 85 m, mean depth: 2.5 m) that passes through the coastal sand bars (Chakroun, 2004). The construction of a fishing port in 1974 created an imbalance in the sediment dynamics and accelerated sedimentation in the inlet (Oueslati et al., 2015). As a result, the lagoon was also subjected to continuous alluvial deposits formed of fine silt–-clay material which, in turn, necessitated dredging operations in the inlet near the fishing port to facilitate water exchange with the sea and decrease the sedimentation rate at the harbor entrance (Oueslati et al., 2015).

Previous studies have shown that the lagoon sediments are contaminated (Oueslati et al., 2010a,b) and the concentrations of some metals exceed levels set out in environmental guidelines. This high contamination has been attributed to both fishing and industry and to the Medjerda River, which formerly flowed into the lagoon. To reduce the Mejerda's sediment inputs the river channel was diverted 10 km to the south at Qualaat El Andalous.

Ghar El Melh Lagoon has provided a variety of important services for human communities since Roman/Phoenician times. It has long been known as an important fishing site where the main species caught are muges, sole, wolf, sea bream, bigerans (*small Mugilidae*), eels and mules (DGPA, 2017).

Ghar El Melh Lagoon is a site of high biodiversity (Ayache et al., 2009), but also an example of an impacted lagoon due to anthropogenic activities such as domestic and industrial wastewater discharge (Dhib et al., 2013). The construction of a fishing port in 1997 has affected sedimentation rate within the lagoon and the inlet, causing a reduction in water circulation and an increase in water residence time (Cataudella et al., 2015; Oueslati et al., 2015). All these factors have led to the degradation of water and sediment quality as well as recurrent eutrophication of the lagoon (Dhib et al., 2013, 2016a,b; Ziadi et al., 2014) and consequently a dramatic decline in fishery productivity (Ayache et al., 2009; DGPA, 2017).

The lagoon experienced several dystrophic crises during the period 1994–1996 (DGPA, 2017). The greatest environmental crisis, marked by the closure of the sea-lagoon channel, was recorded in May 1995 and