

PREDICTION OF HYPOXIA IN A COASTAL LAGOON USING HYDRODYNAMIC MODELLING AND MACHINE LEARNING

Irene Simonetti, University of Florence, irene.simonetti@unifi.it
Lorenzo Cappiotti, University of Florence, lorenzo.cappiotti@unifi.it

INTRODUCTION

Dissolved Oxygen (DO) is a key indicator of water quality in rivers, lakes, and coastal basins. Hypoxia is a condition in which DO level is too low for the proper functioning of the ecosystems, triggering the mortality of fish, losses in biodiversity, and impacting the surviving organisms through sublethal stresses (Vaquer-Sunyer et al, 2008). The upper threshold for hypoxic conditions ranges from 2 to 5 mgDO/l (Gobler & Baumann, 2016), while a conventional threshold of 2 mgDO/l is usually adopted to characterize fisheries collapse (Diaz & Rosenberg, 2008).

Factors governing hypoxia are both of anthropic and natural origin (Caballero-Alfonso et al., 2014). One of the primary anthropogenic drivers is nutrient input, which is in turn related to coastal eutrophication. The main natural drivers of hypoxia are climate-related, e.g., high temperature (decreasing the oxygen solubility), entity of rainfall and runoff (altering the freshwater input), and wind intensity (affecting the oxygen exchange with the atmosphere). Hydrodynamic factors are also relevant (e.g., reduced horizontal advection and increased water residence time). In shallow water, partially enclosed coastal areas, hypoxia is often seasonal, being associated with warmer water temperature, enhanced microbial processes, and higher oxygen demand (Van Helmond et al., 2017).

Coastal lagoons are particularly susceptible to hypoxia phenomena, being characterized by a reduced rate of water exchange with the sea, low thermal inertia due to the limited water depth, and a strong nutrient supply.

Coupled hydrodynamic and biogeochemical physically based numerical models for simulating the most relevant ecological processes involved in the DO dynamics for lagoon coastal systems have been applied in recent years (e.g. Huang et al., 2019; Seiler et al., 2020). Such approaches are generally computationally intensive, require an extensive set of input data, which are not easily accessible, and need time-consuming site-specific calibration procedures.

Machine learning (ML) based approaches to the prediction of water quality-related quantities in lakes or coastal basins gained significant attention in the latest years (e.g. Politikos et al., 2021; Jimeno-Sáez et al., 2020), suggesting a great potential of these techniques for predicting DO. ML-based approaches, however, need to rely on observed values of the input parameters, hindering the possibility of representing the spatial heterogeneity of DO distributions in a coastal basin, unless a wide number of measuring stations are available.

In this work, the possibility of using a Random Forest (RF) algorithm (Breiman, 2001) to predict DO levels in a coastal lagoon is explored. The algorithm has as input data obtained from a hydrodynamic model of the lagoon, aiming to enable a high-resolution description of the system and the identification of hypoxic areas.

STUDY AREA

The study is applied to the Orbetello lagoon, a coastal basin in central Italy (between 42°25' and 42°29' lat. North and 11°10' and 11°17' long. East). The lagoon has an average depth of 1.13 m and an extension of 26 km². At present, water exchange with the Tyrrhenian Sea takes place through three narrow inlets, having a width of 15-20 m and a length of 1 to 2 km. The lagoon constitutes a highly eutrophic environment, subject to frequent episodes of extended hypoxia, which led in some cases to massive deaths of fish.

METHODOLOGY

The methodological approach is schematically reported in Fig. 1.

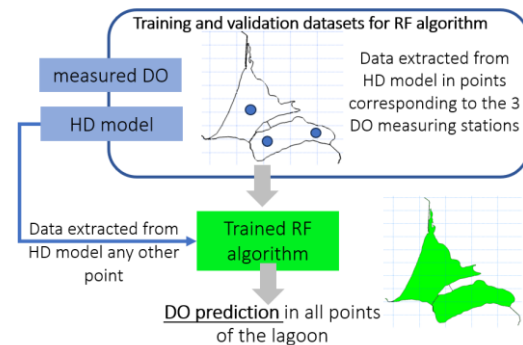


Figure 1 - Sketch of the methodological approach used in this work.

A 2D hydrodynamic model (HD model) is used to reproduce the basic hydrodynamic features of the lagoon, including temperature and salinity. Hourly data from the HD model are extracted in points corresponding to the location of the three DO-measurement stations in the lagoon (Fig. 1), which constitute the training and validation datasets for the RF algorithm. Once the algorithm is trained, it can be used to predict DO levels in any other point of the lagoon, using input data from the HD model in the respective point.

In the HD model, shallow water equations are solved using the Hydrodynamic Flexible Mesh module of MIKE21 software package. As an extension of the model originally presented by Cappiotti et al., 2004, the present model also includes the simulation of the heat exchange with the atmosphere, solving transport-diffusion equations for temperature and salinity. The domain of the numerical model is extended up to the confluence of the lagoon inlets with the open sea. An unstructured grid is used, with a combination of triangular and quadrangular elements with characteristic dimensions varying between 2 m (in the lagoon inlets) and 80 m (in the inner areas). Time series of hourly water levels, temperature, and salinity are used as boundary conditions for the HD

model at each lagoon inlet. These data are extracted from the Mediterranean Sea Physics Analysis and Forecast database of Copernicus Monitoring Environment Marine Service (CMEMS), a coupled hydrodynamic-wave model implemented over the whole Mediterranean Basin with a grid resolution of 1/24°. The atmospheric data (wind, air temperature, relative humidity, solar radiation, precipitation, and clearness) needed to model the heat exchange with the atmosphere are considered uniform over the domain. Such data are based on hourly observations recorded by a meteorological station located at the center of the lagoon.

The RF algorithm proposed by Breiman, 2001 is used as a classification method, i.e., the output of the RF algorithm is the predicted class of DO, rather than its exact value. DO data measured at the three stations in the lagoon have been divided into 4 classes: (i) very high: DO>8 mg/l; (ii) medium: 5 mg/l<DO<8 mg/l; (iii) low: 3 mg/l<DO<5 mg/l; (iv) very low: DO<3 mg/l. The training dataset consists of hourly values of temperature, salinity, water velocity, water levels, and a parameter derived from the Water Residence Time (WRT) extracted from the HD model in the three points in which DO is measured. Day length and other measured atmospheric data (assumed constant over the whole lagoon area) have been also used as candidate predictors. 18 months of hourly data have been used, For training the classification algorithm, the dataset was split into training (70% of data) and testing (30% of data) subsets. Starting from the complete parameter set in the training dataset, a recursive feature elimination procedure is used to select the most influential ones to train the RF algorithm and to be later used as predictors of the DO class.

PRELIMINARY RESULTS

The preliminary application of the overall methodology shows that water temperature, day length, water salinity, water surface elevation, and WRT are an appropriate set of predictors for the RF algorithm. Training the model using this set of predictors, a quite satisfactory performance of the RF algorithm to classify DO levels is achieved (Fig. 2). The predictive performance is evaluated in terms of confusion matrix, indicating the fraction of the total data that the algorithm can identify correctly (true positives, TP), for each DO class. The overall accuracy is around 90%. The algorithm can predict the occurrence of “very low” DO cases with an

		RF algorithm			
		High	Medium	Low	Very Low
DO Observations	High	94%		6%	
	Medium	6%	91%	3%	
	Low		21%	75%	4%
	Very Low		1%	22.3%	76.7%

Figure 2 - Confusion matrix for true positive rates of classification with the RF algorithm.

accuracy of around 77%. Under the assumption that the RF algorithm trained in the three points for which DO measurements are available can represent DO dynamics also in different lagoon points, the trained RF algorithm can be applied to predict the class of DO over the whole lagoon basin, since the needed predictors are known from the hydrodynamic model.

CONCLUDING REMARKS

The use of hydrodynamic modeling (forced with measured data from local stations and numerical models at the regional scale) combined with an RF algorithm seems to be a promising approach for predicting DO levels in coastal basins, possibly providing an indication of such parameters throughout the entire lagoon, allowing to consider the spatial heterogeneity of the coastal basin. Such an approach could be applied to forecast hypoxia events and to localize the most critical areas of the basin. Further studies are at present going on to refine the set-up of the RF algorithm, possibly improving its predictive performance.

REFERENCES

Gobler, Baumann (2016): Hypoxia and acidification in ocean ecosystems: coupled dynamics and effects on marine life, *Biol. Lett.*, vol. 12.

Vaquero-Sunyer & Duarte (2008): Thresholds of hypoxia for marine biodiversity, *Proc. Natl. Acad. Sci.*, vol. 105.

Diaz & Rosenberg, (2008): Spreading Dead Zones and consequences for Marine Ecosystems, *Science*, vol. 321.

Caballero-Alfonso, Carstensen, Conley (2015): Biogeochemical and environmental drivers of coastal hypoxia, *J. Mar. Syst.*, vol. 141.

Van Helmond, et al. (2017): Seasonal hypoxia was a natural feature of the coastal zone in the Little Belt, Denmark, during the past 8 ka, *Mar. Geol.*, vol. 387.

Seiler, Fernandes, Siegle (2020): Effect of wind and river discharge on water quality indicators of a coastal lagoon, *Regional Studies in Marine Science*, 40, 101513.

Huang, Hu, Li, Wang, Xu, Liang, Liu (2019): Effects of Physical Forcing on Summertime Hypoxia and Oxygen Dynamics in the Pearl River Estuary, *Water*, 11, 2080.

Lenzi et al. (2015): A new approach to macroalgal bloom control in eutrophic, shallow-water, coastal areas», *J. Environ. Mangem.* vol. 150.

Cappiotti & Scroccaro (2004): The hydrodynamics of the Orbetello Lagoon, Measurements and numerical simulations”. *Proceedings of the 29° Convegno di Idraulica e Costruzioni Idrauliche*, Trento, Italy.

Politikos, Petasis, Katselis (2021): Interpretable machine learning to forecast hypoxia in a lagoon, *Ecological Informatics*, vol. 66.

Jimeno-Sáez, Senent-Aparicio, Cecilia, Pérez-Sánchez, (2020): Using Machine-Learning Algorithms for Eutrophication Modeling: Case Study of Mar Menor Lagoon (Spain), *Int. J. Env. Res. Publ. Health*, 17.

Breiman (2001): *Random Forest, Machine Learning*, vol 45.