

EGU2020-9567

<https://doi.org/10.5194/egusphere-egu2020-9567>

EGU General Assembly 2020

© Author(s) 2021. This work is distributed under the Creative Commons Attribution 4.0 License.



## A machine learning model to link ecological response and anthropogenic stressors: a tool for water management in the Tagus River Basin (Spain)

**Carlotta Valerio**<sup>1,2</sup>, Alberto Garrido<sup>2,3</sup>, Gonzalo Martinez-Muñoz<sup>4</sup>, and Lucia De Stefano<sup>1,2</sup>

<sup>1</sup>Universidad Complutense de Madrid, Department of Geodynamics, Stratigraphy and Paleontology, Madrid, Spain  
(carval02@ucm.es)

<sup>2</sup>Water Observatory, Botín Foundation, Spain

<sup>3</sup>CEIGRAM, ETSIAAB, Universidad Politécnica de Madrid, Madrid, Spain

<sup>4</sup>Escuela Politécnica Superior, Universidad Autónoma de Madrid, Madrid, Spain

Freshwater ecosystems are threatened by multiple anthropic pressures. Understanding the effect of pressures on the ecological status is essential for the design of effective policy measures but can be challenging from a methodological point of view. In this study we propose to capture these complex relations by means of a machine learning model that predicts the ecological response of surface water bodies to several anthropic stressors. The model was applied to the Spanish stretch of the Tagus River Basin. The performance of two machine learning algorithms -Random Forest (RF) and Boosted Regression Trees (BRT) - was compared. The response variables in the model were the biotic quality indices of macroinvertebrates (Iberian Biomonitoring Working Party) and diatoms (Indice de Polluosensibilité Spécifique). The stressors used as explanatory variables belong to the following categories: physicochemical water quality, land use, alteration of the hydrological regime and hydromorphological degradation. Variables describing the natural environmental variability were also included. According to the coefficient of determination, the root mean square error and the mean absolute error, the RF algorithm has the best explanatory power for both biotic indices. The categories of land cover in the upstream catchment area, the nutrient concentrations and the elevation of the water body are ranked as the main features at play in determining the quality of biological communities. Among the hydromorphological elements, the alteration of the riparian forest (expressed by the Riparian Forest Quality Index) is the most relevant feature, while the hydrological alteration does not seem to influence significantly the value of the biotic indices. Our model was used to identify potential policy measures aimed at improving the biological quality of surface water bodies in the most critical areas of the basin. Specifically, the biotic quality indices were modelled imposing the maximum concentration of nutrients that the Spanish legislation prescribes to ensure a good ecological status. According to our model, the nutrient thresholds set by the Spanish legislation are insufficient to ensure values of biological indicators consistent with the good ecological status in the entire basin. We tested several scenarios of more restrictive nutrient concentrations and values of hydromorphological quality to explore the conditions required to achieve the good

ecological status. The predicted percentage of water bodies in good status increases when a high Riparian Forest Quality Index is set, confirming the importance of combining physico-chemical and hydromorphological improvements in order to ameliorate the status of freshwater ecosystems.