biology and life sciences forum

MDPI

Abstract

Machine Learning Approach towards the Early Warning of Cyanobacterial Blooms in Drinking Water Reservoirs †

Claudia Fournier [1,*,‡], Antonio Quesada [1], Ana Justel [2], Agustín Pedro Monteoliva [3], Jordi Cirera [4], Carolina Solà [5], Antoni Munné [5], Juan Carlos Garcia [6], José Javier Rodríguez [6] and Samuel Cirés [1]

1 Department of Biology, Universidad Autónoma de Madrid, 28049 Madrid, Spain; antonio.quesada@uam.es (A.Q.); samuel.cires@uam.es (S.C.)
2 Department of Mathematics, Universidad Autónoma de Madrid, 28049 Madrid, Spain; ana.justel@uam.es
3 ECOHYDROS S.L., 39600 Cantabria, Spain; apmonteoliva@ecohydros.com
4 AECOM URS España S.L.U., 28014 Madrid, Spain; jordi.cirera@aecom.com
5 Catalan Water Agency, 08036 Barcelona, Spain; csolao@gencat.cat (C.S.); anmunne@gencat.cat (A.M.)
6 Ter-Llobregat Water Supply Public-Law Entity, 08970 Barcelona, Spain; jcgarcia@atl.cat (J.C.G.); jjrodriguez@atl.cat (J.J.R.)
* Correspondence: claudia.fournier@uam.es
† Presented at the 7th Iberian Congress on Cyanotoxins/3rd Iberoamerican Congress on Cyanotoxins, Ponta Delgada, Portugal, 18–20 July 2022.
‡ Presenting author (oral communication).

Abstract: Cyanobacterial harmful algal blooms (CyanoHABs) are expanding globally, representing a major risk for lakes and reservoirs due to their toxicity and economic impacts. Therefore, anticipating their occurrence and understanding the main factors related to CyanoHABs are critical to improve decision-making processes and water resource management. In this context, we present two modelling options for the analysis and prediction of cyanoHABs in two drinking water reservoirs from Spain. This case represents a unique opportunity to combine efforts from different academic disciplines (i.e., aquatic ecology and data science), environmental companies, and public water managers to address this increasingly severe issue. Susqueda (Ter basin, Catalonia) is a eutrophic, large and deep reservoir (Z_{max} = 110 m) where monitoring efforts in recent years have focused on a monthly measurement in more than 30 physico-chemical, hydrological, meteorological and biological parameters, some of them involving expert intervention and costly efforts that could not be held at a higher temporary frequency. Cuerda del Pozo (Duero basin, Castilla y León) is a deep reservoir (Z_{max} = 30 m) where monitoring efforts have focused on daily data collection through probes mounted in automatic profilers. This strategy allowed a higher monitoring frequency for fewer parameters and a narrower time span. In both cases, the parameter chosen as a proxy of cyanobacterial proliferation (output of models) is fluorometric measurements of chlorophyll-a and phycocyanin. The results of our machine-learning-based analyses suggest that the selected modelling path mainly depends on two aspects: (1) the time span where data are collected, and (2) the frequency and type of data measured (i.e., one discrete measurement at the surface vs. many measurements along the water column). Thus, a Susqueda dataset analysis led to more interpretative results, allowing for a better understanding of the system and the main factors related to CyanoHABs with limited predictive capacity. Meanwhile, the Cuerda del Pozo dataset is treated as a time series where autoregressive forecasting techniques, combined with information of exogenous parameters, are applied to foresee cyanobacterial blooms before they occur, losing part of the interpretability in the process. The results from this work are expected to provide an effective tool to boost smart and goal-orientated sampling planning, while improving data-driven decision-making processes essential for the water management of cyanobacterial blooms.

Keywords: CyanoHABs; monitoring; machine learning; predictive modeling; water management

Citation: Fournier, C.; Quesada, A.; Justel, A.; Monteoliva, A.P.; Cirera, J.; Solà, C.; Munné, A.; Garcia, J.C.; Rodríguez, J.J.; Cirés, S. Machine Learning Approach towards the Early Warning of Cyanobacterial Blooms in Drinking Water Reservoirs. *Biol. Life Sci. Forum* **2022**, *14*, 38. https://doi.org/10.3390/blsf2022014038

Academic Editor: Vitor Gonçalves

Published: 26 July 2022

Author Contributions: Conceptualization, C.F., S.C., A.Q. and A.J.; methodology, C.F., S.C., A.Q. and A.J.; formal analysis, C.F. and A.J.; investigation, C.F., S.C., A.Q. and A.J.; resources, A.P.M., J.C., C.S., A.M., J.C.G., J.J.R.; data curation, C.F.; writing—original draft preparation, C.F.; writing—review and editing, C.F., S.C., A.Q., A.J., A.P.M., J.C., C.S., A.M., J.C.G. and J.J.R. All authors have read and agreed to the published version of the manuscript.

Funding: Part of this research was funded by Catalan Water Agency, Spain, grant number CTN20C0533 and by project "AIHABs-Artificial Intelligence-powered forecast for Harmful Algal Blooms" (reference PCI2021-121915) funded by Spanish State Research Agency (AEI) and the Joint Programming Initiatives (JPIs) on Water with the contribution of the European Commission.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Conflicts of Interest: The authors declare no conflict of interest.