

Monitoring Water Quality in Small Reservoirs Using Sentinel-2 Imagery and Machine Learning [†]

Victoria Amores-Chaparro ¹, Fernando Broncano-Morgado ^{2,*} , Pablo Fernández-González ¹, Aurora Cuartero ³ 
and Jesús Torrecilla-Pinero ⁴ 

¹ Department of Informatics and Telematics Systems, University of Extremadura, 10003 Cáceres, Spain; vicamoresc@unex.es (V.A.-C.); pablofg@unex.es (P.F.-G.)

² Department of Informatics and Telematics Systems, University of Extremadura, 06006 Badajoz, Spain

³ Department of Graphics Expression, University of Extremadura, 10003 Cáceres, Spain; acuartero@unex.es

⁴ Department of Construction, University of Extremadura, 10003 Cáceres, Spain; jtorreci@unex.es

* Correspondence: fbroncano@unex.es

[†] Presented at the First Summer School on Artificial Intelligence in Cybersecurity, Cancún, Mexico, 3–7 November 2025.

Abstract

This article investigates the estimation of water quality parameters, specifically chlorophyll-a, applying machine learning techniques to Sentinel-2 images. This study focuses on five small reservoirs located in the Extremadura region (Spain), as these are the ones for which continuous daily records from automatic in situ sensors are available. Chlorophyll-a estimates are obtained from two sources: (1) From the C2RCC atmospheric correction of Sentinel-2 images using Sen2Cor and radiometric calibration to ensure temporal consistency, and (2) from in situ data obtained from the official website of the Guadiana Basin Automatic Network Information System. The machine learning (ML)-based methodology significantly improves the predicted results for inland water bodies, enabling enhanced continuous assessment of water quality in small reservoirs.

Keywords: remote sensing; water quality monitoring; Sentinel-2 imagery; machine learning; continuous monitoring

1. Introduction

Access to water for human consumption tends to be a recurrent source of conflict across multiple regions. In addition, hydro-intensive production practices have contributed to a decline in the availability of fresh water due to contamination and unhealthy conditions. The continuous measurement of reservoir variables can contribute to assessing and implementing actions aimed at maintaining stored water under sanitary conditions. The concentration of chlorophyll-a is commonly used as an indicator of the trophic state of the water.

The Sentinel satellite network, operated by the Copernicus Programme of the European Commission, enables the monitoring of reservoir conditions. One of the parameters that can be measured through hyperspectral imaging is the concentration of chlorophyll-a present in the water [1]. Several studies have proposed suitable methods for these techniques [2–4], which are based on the use of the SNAP tool [5] and the C2-NET model [6]. Nevertheless, the extrapolation of these tools to small reservoirs has proven to be unstable since the training was performed using imagery from large terrestrial and oceanic water bodies.



Academic Editors: Luis Javier García Villalba, Héctor Manuel Pérez Meana, Gabriel Sánchez Pérez and Ana Lucila Sandoval Orozco

Published: 2 February 2026

Copyright: © 2026 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the [Creative Commons Attribution \(CC BY\)](https://creativecommons.org/licenses/by/4.0/) license.

In this work, we propose a machine learning model to adjust the chlorophyll-a levels obtained from satellite imagery of small reservoirs. Sentinel-2 images were processed with the C2-NET, specifically C2RCC model, to infer chlorophyll-a concentration, and the results were calibrated against real-time in situ measurements to improve model accuracy.

2. Materials and Methods

2.1. Study Area

The study area consists of five different reservoirs within the Guadiana River Basin, managed by the Guadiana River Basin Authority (Confederación Hidrográfica del Guadiana). This institution provides continuous water quality data for several reservoirs, including the Alange, Orellana, Tentudía, Peña del Águila, and Zújar reservoirs. The geographical locations of these reservoirs are shown in Figure 1.

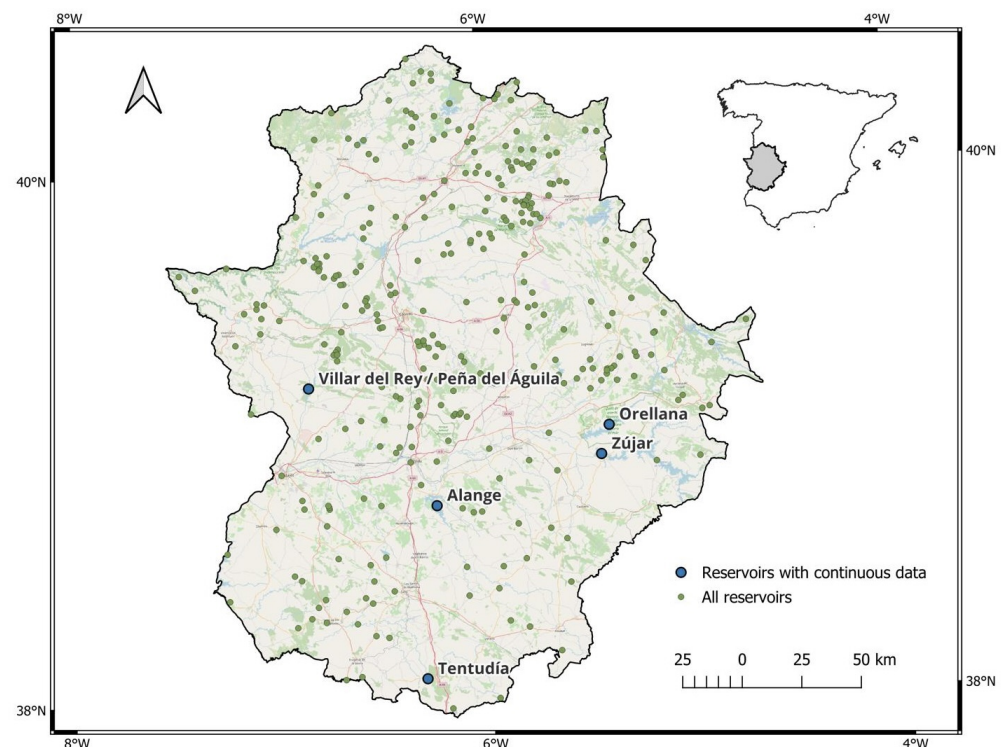


Figure 1. Study area in southwestern Spain, a region of the Iberian Peninsula. The five reservoirs with continuous in situ chlorophyll-a data are marked with a blue circle.

2.2. Data Used

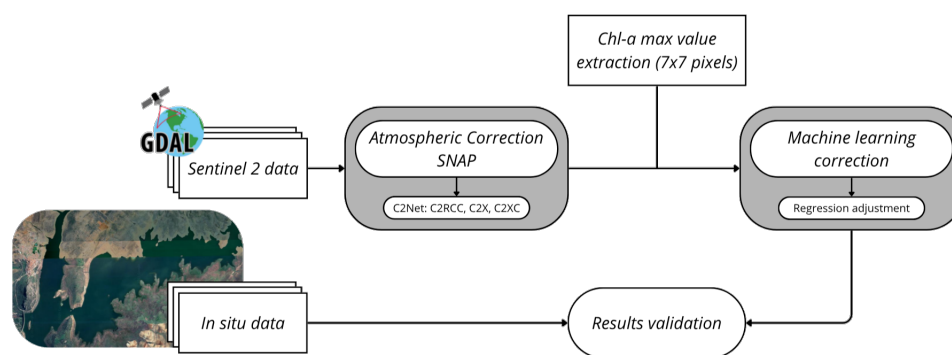
For this study, measurement data obtained directly from the reservoirs and images from the Sentinel-2 satellites were used. The in situ measurement data from the reservoirs were obtained from SIRA (<https://siraguadiana.com>, accessed on 30 January 2026), while the hyperspectral images were retrieved from the Copernicus Browser (<https://browser.dataspace.copernicus.eu/>, accessed on 30 January 2026) and processed using SNAP with C2RCC correction in a maximum 7×7 window [7,8]. All available satellite images and in situ measurements from the years 2023 and 2024 were collected. Subsequently, the data were cross-referenced so that each fully visible (cloud-free) hyperspectral image of the reservoir was assigned the nearest daily measured value of chlorophyll. Table 1 presents several statistics of measurements, broken down by reservoirs and the overall total.

Table 1. Identification of the data used. The measured data are shown on the left, and the data from satellite images using SNAP are shown on the right. Chlorophyll-a concentration in (mg/m³).

Reservoir	Alange		Orellana		Tentudía		Peña del Águila		Zújar		Total	
Start date	31 January 2023		22 March 2023		31 January 2023		24 January 2023		22 March 2023		31 January 2023	
End date	24 December 2024		21 December 2024		6 December 2024		20 October 2024		21 December 2024		24 December 2024	
N	23		22		22		17		22		106	
Maximum	15.92	32.61	26.35	32.81	155.36	33.14	200.00	32.87	163.80	30.33	200.00	33.14
Minimum	1.69	7.66	2.91	6.50	1.11	0.17	2.55	8.48	0.00	2.78	0.00	0.17
Mean	5.63	17.19	8.56	20.81	31.09	21.69	26.51	24.30	16.43	17.16	17.11	20.01
Std. Dev.	3.57	6.17	5.39	7.34	35.32	7.76	45.06	6.96	33.60	7.60	30.34	7.87

2.3. Machine Learning Algorithms

The proposed methodology aims to correct the presented data through a machine learning framework that enables the inference of a more accurate value of chlorophyll-a concentration in the reservoirs of the study area, compared with the measured data. The application of the methodology is illustrated graphically in Figure 2.

**Figure 2.** Workflow used in the analysis of water quality (chlorophyll-a) in the reservoir.

The proposed machine learning model uses a Voting Ensemble with five Gradient Boosting regressors to aggregate reservoir data. Extreme values from the Peña del Águila (1), Tentudia (1) and Zújar (5) reservoirs were removed to improve statistical reliability.

3. Results

The model was trained with 66% of the data and tested with 34%, using a Voting Regression with five Gradient Boosting estimators. Table 2 presents the resulting metrics, while Figure 3 illustrates the model performance on the test dataset versus the actual data.

Table 2. Model Fitting Metrics. Chlorophyll-a concentration (mg/m³).

RSME	MAE	R ²	r (Pearson)	p-Value
9.94	8.64	−0.65	0.32	0.067

Although the R² value is negative, the predicted data generally follow the observed trends. This calibration enables the correction of chlorophyll-a concentrations estimated from satellite imagery and the C2-NET model, ensuring consistency with in situ sensor measurements. Consequently, the proposed approach improves the reliability of chlorophyll-a estimates from remote sensing and supports reservoir water quality assessment.

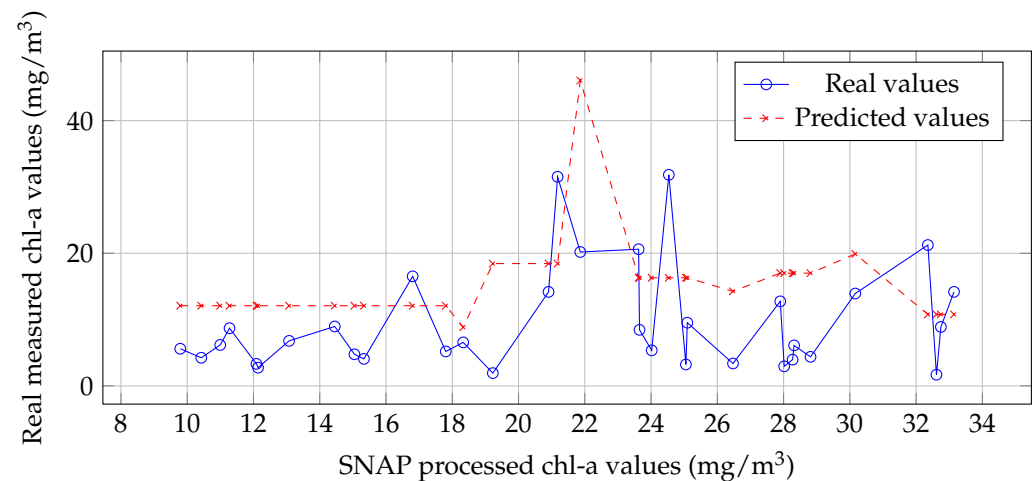


Figure 3. Comparison between real data and model behavior on test data.

4. Conclusions

The resulting model, based on regression algorithms, provides a more comprehensive and accurate fit for the reservoirs within the study area. Through this adjustment, corrections can be applied to the chlorophyll-a concentration estimated from satellite imagery and C2-NET. This approach ensures that the chlorophyll-a concentration trends obtained from satellite images are consistent with those measured by sensors, making it a useful method for estimating the health and quality of stored water.

Author Contributions: Conceptualization, A.C., P.F.-G. and J.T.-P.; methodology, F.B.-M. and V.A.-C.; software, F.B.-M. and V.A.-C.; validation, A.C. and P.F.-G.; formal analysis, A.C. and F.B.-M.; investigation, A.C. and J.T.-P.; resources, V.A.-C.; data curation, V.A.-C.; writing—original draft preparation, F.B.-M.; writing—review and editing, A.C. and P.F.-G.; supervision, J.T.-P.; project administration, A.C.; funding acquisition, A.C. All authors have read and agreed to the published version of the manuscript.

Funding: This initiative was carried out within the framework of the funds of the Recovery, Transformation and Resilience Plan, financed by the European Union (Next Generation)—National Cybersecurity Institute (INCIBE) in the project C107/23 “Artificial Intelligence Applied to Cybersecurity in Critical Water and Sanitation Infrastructures”.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The model and data are available at GitHub (<https://github.com/fbroncano/C2RCC-ML>, accessed on 30 January 2026).

Acknowledgments: The authors would like to thank the Guadiana River Basin Authority (Confederación Hidrográfica del Guadiana) for providing the real-time data available at SIRA (<https://siraguadiana.com>, accessed on 30 January 2026). The authors would also like to thank the Copernicus Programme (<https://www.copernicus.eu>, accessed on 30 January 2026) for providing access to Sentinel satellite imagery used in this study.

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

References

1. Delegido, J.; Tenjo, C.; Ruiz-Verdú, A.; Peña, R.; Moreno, J. Empirical model for chlorophyll-a determination in inland waters from the forthcoming Sentinel-2 and 3. Validation from HICO images. *Rev. Teledetec.* **2014**, *41*, 37–47. [[CrossRef](#)]

2. Soriano-González, J.; Urrego, E.P.; Sòria-Perpinyà, X.; Angelats, E.; Alcaraz, C.; Delegido, J.; Ruíz-Verdú, A.; Tenjo, C.; Vicente, E.; Moreno, J. Towards the Combination of C2RCC Processors for Improving Water Quality Retrieval in Inland and Coastal Areas. *Remote Sens.* **2022**, *14*, 1124. [[CrossRef](#)]
3. Zhang, J.; Fu, P.; Meng, F.; Yang, X.; Xu, J.; Cui, Y. Estimation algorithm for chlorophyll-a concentrations in water from hyperspectral images based on feature derivation and ensemble learning. *Ecol. Inform.* **2022**, *71*, 101783. [[CrossRef](#)]
4. Homaei, M.; Di Bartolo, A.J.; Molano Gómez, R.; Rodríguez, P.G.; Caro, A. Enabling RPL on the Internet of Underwater Things. *J. Netw. Syst. Manag.* **2025**, *33*, 55. [[CrossRef](#)]
5. Leskovec, J.; Sosič, R. SNAP: A General-Purpose Network Analysis and Graph-Mining Library. *ACM Trans. Intell. Syst. Technol. (TIST)* **2016**, *8*, 1. [[CrossRef](#)] [[PubMed](#)]
6. Brockmann, C.; Doerffer, R.; Peters, M.; Kerstin, S.; Embacher, S.; Ruescas, A. Evolution of the C2RCC neural network for Sentinel 2 and 3 for the retrieval of ocean colour products in normal and extreme optically complex waters. In Proceedings of the Living Planet Symposium, Prague, Czech Republic, 9–13 May 2016.
7. Cuartero, A.; Cáceres-Merino, J.; Torrecilla-Pinero, J.A. An application of C2-Net atmospheric corrections for chlorophyll-a estimation in small reservoirs. *Remote Sens. Appl. Soc. Environ.* **2023**, *32*, 101021. [[CrossRef](#)]
8. Cáceres-Merino, J.; Cuartero, A.; Torrecilla-Pinero, J.A. Finding optimal spatial window size: The influence of methodology on remote-sensing based Chl-a prediction in small reservoirs. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2024**, *17*, 1234–1248.

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.