Prepublished copy

Nature Environment & Pollution Technology

This is a peer-reviewed prepublished version of the paper

A REVIEW ON ARTIFICIAL INTELLIGENCE FOR WATER QUALITY PREDICTION IN AMAZONIAN COUNTRIES

J.E. Cruz de La Cruz*, W.A. Mamani**, F. Pineda***, V. Yana-Mamani****, R. Santa Cruz*****, Í. Maldonado-Ramírez*****, R. Pérez-Astonitas*****, E. Morales-Rojas*****.

*Facultad de Mecánica Eléctrica, Electrónica y Sistemas, Universidad Nacional del Altiplano, Puno 21001, Perú;

**Universidad de Alicante, España;

***Pontificia Universidad Católica del Perú;

****Universidad Nacional de Moquegua, Perú,

*****Facultad de Ingeniería y Sistemas y Mecánica Eléctrica, Universidad Nacional Toribio Rodríguez de Mendoza de Amazonas; Calle Agropecuaria N°520, Bagua, Amazonas. Perú;

† Corresponding author: E. Morales-Rojas; eli.morales@untrm.edu.pe

ABSTRACT

Water quality prediction plays an important role in environmental monitoring, ecosystem sustainability in the Amazon. Therefore, this review focuses on determining the advances of the scientific production of artificial intelligence in water quality prediction in the Amazon, as well as the limitations and perspectives compared to water quality indexes (WQI). In this sense, Boolean operators were applied, using the following terms: “artificial intelligence”, “machine learning”, “water quality” and “Amazonia”, the databases were through Scopus, web of science, Springer and IEEE. In this study, 14 scientific articles published during the period 2000-2024 focused on Amazonian countries were evaluated. Although in the Amazon low scientific production was evidenced and is led by Brazil, the highest scientific growth was for 2021 and 93% belongs to the Scopus database, with a compound annual rate of 12.16%. The IA is characterized by using data from governmental institutions and is only limited to parameters such as Total Suspended Solids (TSS), Total Organic Carbon, Turbidity and Chlorophyll, using satellite imaging techniques and the most commonly used algorithm was the Clustering Algorithms. In this context, AI application is still very low in Amazonian countries compared to other European countries. Its limitations are in the accuracy and the limited amount of physicochemical and microbiological data used for predictions. However, AI is a tool that will replace the water quality indexes used manually.

Key Words	Artificial intelligence, water quality, amazon
DOI	https://doi.org/10.46488/NEPT.2025.v24i02.D1705 (DOI will be active only after the final publication of the paper)

Prepublished copy

Nature Environment & Pollution Technology

This is a peer-reviewed prepublished version of the paper

Citation of the Paper	J.E. Cruz de La Cruz, W.A. Mamani, F. Pineda, V. Yana-Mamani, R. Santa Cruz, Í. Maldonado-Ramírez, R. Pérez-Astonitas, E. Morales-Rojas , 2025. Spatial Analyses of Reliability of Solar Power in the Western Part of Iraq. *Nature Environment and Pollution Technology*, 24(2), D11705. https://doi.org/10.46488/NEPT.2025.v24i02.D1705

INTRODUCTION

Water quality is an essential ecological value for health and economic development (Villena Chávez, 2018), it is the most abundant component of the human body, so it must have proper management (Salas-Salvadó et al., 2020; Veneros et al., 2024). However, water quality has been threatened by various sources, wastewater treatment plants, mining activities (heavy metals), food processing wastes (Morales-Rojas, 2020), agricultural runoff (Kaur & Sinha, 2019), chemical and household disposal. Water sources are divided into surface water and groundwater sources, of this rivers are most used water source by the population living near river Banks (S. Mustafa et al., 2017). In that sense, artificial intelligence (AI) plays an important role in living beings (Romero Lopez & Vargas Mato, 2017) and is divided into biological, hydromorphological and physicochemical quality (Nikolaou et al., 2008;Gad et al., 2020).

Therefore, a productive and cost-effective methodology for water quality estimation is required, but they are difficult to design due to their nonlinearity, non-stationary characteristics and constant unpredictable natural changes (Lintern et al., 2018; Perona et al., 1999). Therefore, reliable methodologies such as artificial intelligence (AI) models and their advanced tools are needed to determine water quality and achieve sustainable development (Chaudhary, 2023; Tiyasha et al., 2020). Therefore, AI is a significantly powerful resource for water monitoring (Rajaee et al., 2020).

Studies have applied artificial neural network and adaptive neuro-fuzzy inference system to calculate dissolved oxygen (DO) levels and organic matter parameters (BOD_5 and COD), these results were optimal with the values measured in the river water (Emamgholizadeh et al., 2014). Neural network models are studied, to examine and mimic the relationship of water quality index (WQI) with water quality variables in a tropical environment, being an encouraging alternative to predict WQI, as opposed to manual calculation methods of WQI that involve long calculations and transformations (Hameed et al., 2017; Limon & Webb, 1964), water quality indices have a quantitative description, it is qualitatively interpreted with the help of classification schemes (Egbueri et al., 2023). However, this process has been hampered by labor and testing costs (Agbasi & Egbueri, 2023).

The adoption of AI in water has led to major revolutions and innovations with respect to water quality assessment and monitoring and high quality research should be prioritized

(Myllyaho et al., 2021). Other studies have analyzed a variety of modeling methods, such as deep learning (DL), which have been shown to increase efficiency compared to traditional machine learning (ML) models (Irwan et al., 2023).

Water in the Amazon plays a crucial role in climate regulation, biodiversity and ecosystem sustainability, as well as in the life and livelihood of local communities and population (Marengo et al., 2018). The Amazon basin covers an area of approximately 7 million km2. Amazon forests cover about 5.3 million km2 , representing 40% of the global tropical forest area (Laurance et al., 2001). However, water quality in this region has been affected by a number of factors, including human activities such as mining, deforestation, and land-use change as well as by the effects of climate change (Weng et al., 2018).

In this context, the application of advanced technologies such as AI offers new opportunities to more effectively monitor and manage water resources in the Amazon. To this end, the following questions need to be addressed:

What will be the evolution of scientific production over time through the assessment and monitoring of water quality with AI techniques in the Amazon?

What are the AI algorithms used in Amazon?

What are the limitations and perspectives of AI versus water quality indices in the Amazon?

Based on the above, the objective of this study was to determine the advances of the scientific production of artificial intelligence in the prediction of water quality in the Amazon, as well as the limitations and perspectives of its use in the prediction of water quality in the Amazon.

MATERIALS AND METHODS

The bibliography consulted goes back to the year 2000, up to 2024. Boolean operators were applied, using the following terms: “artificial intelligence”, “machine learning”, “water quality” and “Amazonia”. All the research was carried out through a search in Google Scholar, due to its capacity to compile free access texts. As well as the use of the main databases Scopus, web of science, Springer and IEEE as a support of all the findings on this topic, by passing after a rigorous peer review (Bakhmat et al., 2022) and 95 scientific articles were found.

Inclusion criteria

Publications from 2000 to January 2024 in all languages were considered. Titles, abstracts, methodology (to determine the amount of data used in the study) and main results were examined to select the articles of interest. As the geographic scope was Amazon-wide, only 14 studies were rescued.

Exclusion conditions

Book and encyclopedia chapters, conferences and reviews that did not consider artificial intelligence and others that did not fit the study topic were excluded. We also excluded, gray literature for not passing peer review (Valentine et al., 2019). In addition, inconclusive studies and duplicates were not taken into account.

Compound annual growth rate (CAGR)

CAGR is a factor to measure economic growth (Castillo & Powell, 2019), with the purpose of describing the evolution of scientific studies, the last 23 years were studied. Calculations were performed using a CAGR calculator. An open calculator was applied because it is easier and faster to use (Calcuvio, 2022).

Limitations and prospects

The selected research was related to the physicochemical parameters analyzed in each study and how these are contributing to the closing of gaps. The benefits of AI compared to traditional methods for determining water quality (water quality indices) were addressed. Likewise, each selected article was analyzed to determine the limitations and future perspectives for the development of research.

Data analysis

The data were downloaded in CSV format and processed in Microsoft Office Excel 2016. This facilitated the determination of the distribution of studies by year and country. The analysis was performed with VOSviewer software version 1.6.19, a tool widely used in the scientific community to represent and visualize bibliometric networks. VOSviewer uses several colors to help understand and discover keyword relationships (Eck, Ludo Waltman, 2013).

RESULTS AND DISCUSSION

Research behavior according to countries

Table 2 shows the distribution of scientific articles according to the affiliation and country of origin of the first author, showing that Brazil (6 articles) and the USA (3 articles) lead with studies related to the application of artificial intelligence in the Amazon. These results may be related to the participation in the development and dissemination of knowledge in water resources management led by the Brazilian Association of Water Resources (ABRhidro) (de Paiva et al., 2020). Another country that leads in verified water quality prediction studies is the USA. These results may be related to the fact that it is the second country after China with the highest amount of organic pollutant emissions to water in the world, with an impressive value of 1,850,753 kg/day (Paraschiv et al., 2015).

Table 2. Distribution of scientific production by country

Country	Publications
Brazil	6
USA	3

Prepublished copy

Nature Environment & Pollution Technology

This is a peer-reviewed prepublished version of the paper

Canada	1
China	2
United Kingdom	1

Behavior of scientific production by year

Fig. 1 shows the evolution of publications by year, highlighting 6 articles in 2021 and only one article found for 2024 (Fig 1A), which is related to the CAGR of 12.16%. Most publications are in the Scopus database (95%) and only 7% are in Web of Science (Fig 1 B). These databases are world leaders in terms of scientific content (Zhu & Liu, 2020).

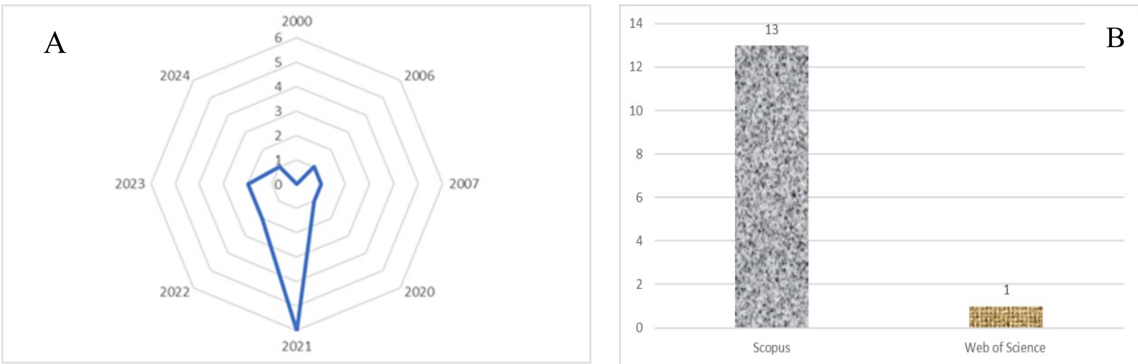

Figure. 1: Behavior of scientific production by year (A); Scientific production by database (B).

Fig 2 shows the amount of data per study, with three studies using the largest amount of data, including the studies by Valdes & Pou, (2021a) with 99473 data, while Liu et al., (2023) reached 15204 data and Strobl et al. (2007) used 8097 data. Generally, the studies that used large amounts of data were obtained from public data from organizations dedicated to water quality monitoring. In that sense, the importance of the amount of data in water quality predictions is affirmed and aims to improve decision making to improve quality indexes and public health (De Fortuny et al., 2013), however, the states, devote little attention to water management in the Amazon, although for example in the Amazon basin of Peru 94 percent transport it manually to their homes from rivers and only 50% boil the water for consumption (McClain et al., 2001). Therefore, improving water quality through efficient techniques while avoiding the smallest margin of error (James et al., 2015), is through artificial intelligence that uses methods to generate highly accurate results. To counteract problems in the amount of data, studies demonstrate the application of data imputation techniques (process of replacing missing values) (Ghapor et al., 2017) and synthetic data generation which is widely used for exploratory data analysis (Dankar & Ibrahim, 2021). These methods can be justified because of the high cost of physicochemical and microbiological analyses in water, preventing more continuous and prolonged sampling in the Amazon basin and the world.

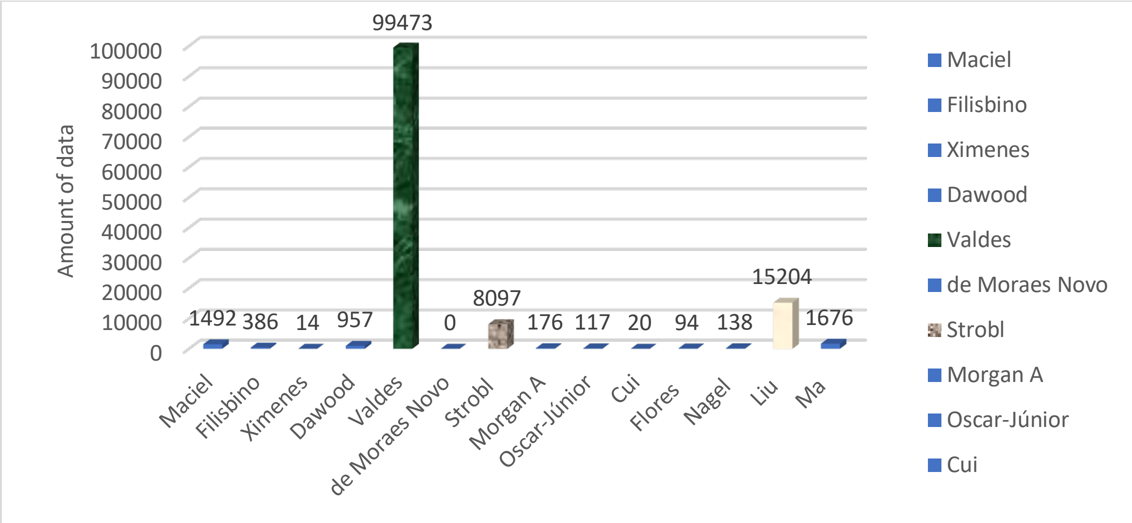

Figure .2: Amount of data used in the studies.

Table 3, the explored studies are evidenced, the minimum female contribution as first author (3 articles) in this subject, in spite of this is evidencing the increase in the number of women working in the information technology industry and even now they prefer careers related to Information and Communications Technology-ICT (Akter et al., 2021) regarding the AI algorithms used are Support Vector Machines, Ensemble Learning, Support Vector Machines, Ensemble Learning, Correlation Maps.

With respect to data collection, it was evident that 8 studies were from the public sector, influenced by the large amount of data. This is important because the data are analyzed and these are visualized in open access in this case obtained from government institutions (Hossain et al., 2016), with this an analysis is achieved from a holistic point of view, which propose future directions and the role that can be played by organizations of scientific work has the potential to advance and transform the common knowledge from which the science-based sectors are nourished (Perkmann & Schildt, 2015). While only 3 studies were with data collected by same authors, taking into account that this type of studies is costly reagents and instruments for the analysis of water samples and 3 studies used the public-public combination. Therefore, the promotion and access to water quality monitoring data by governments should be free access and artificial intelligence should be a tool for analysis and decision making by public and private institutions.

The validation methods used in the studies were the validity and accuracy index, which allows inferences to be made and hypotheses to be evaluated (Lamprea et al., 2007). Among these studies we have the R2, standard error of the estimation, Student's t-test analysis widely used in the studies. Prediction errors (value that quantifies the uncertainty of a prediction), as well as the coefficient of determination, which is a function of the proportion of variability of the variable and is attributed to the linear relationship with X (Roy-García et al., 2019).

The physicochemical parameters found was the “total phosphorus” parameter that is found in water in the form of phosphates, this can be dangerous, if it is not properly controlled, for example if the water is used for irrigation purposes (Morales-Rojas, 2020) and total suspended solids in water is important to monitor because they carry toxic substances absorbed and limit the availability of light and photosynthesis (Park, 2007). This parameter is used in predictions by using empirical models that use surface reflectance from satellite images to estimate Total Suspended Solids (TSS) concentrations (Isidro et al., 2018). Likewise, the studies analyzed are related to the evaluation of chlorophyll and dissolved organic carbon (DOC). Regarding chlorophyll in water, they have distinctive spectral characteristics, therefore, they can be measured by spectral indices in various water bodies, including lakes, ponds, rivers and streams (Yang et al., 2017). However, it is still challenging to accurately estimate chlorophyll concentrations by remote sensing in various water bodies, because the presence of suspended sediments with high inorganic content and colored dissolved organic matter can nullify the chlorophyll spectral signal (Mouw et al., 2015; Odermatt et al., 2012; Sun et al., 2014; Yacobi et al., 2011). On the other hand, from the studies the evaluation of dissolved organic carbon in water is evidenced, this is because they carry pollutants and can negatively affect water treatment processes (Ledesma et al., 2012).

Another of the main results is the use of advances in remote sensing in landscape ecology, highlighting its contribution to the structure and function of the landscape, analysis of historical changes and simulation of future changes, an important topic for the Amazon, which represents interdisciplinarity and will help ensure that landscape ecology can benefit the most from remote sensing (Foody, 2023).

Table 3. Characteristics of the studies on water quality prediction

Cita	Quartile	Number of appointments	Database	Genre of first author	AI algorithms used	Where data were collected	Data sources	Validation methods	Physical-chemical parameters in water quality prediction /Source.	Main results
(de Moraes Novo et al., 2006)	Q1	106	Scopus	F	Other techniques	Brazil	Own and public/ historical Landsat-TM imagery	Validity and accuracy index (R2, standard error of estimation, Student's t-test)	No mention/Lago=concentration of chlorophyll, dissolved organic carbon dissolved organic carbon (DOC)	Evaluated seasonal changes in chlorophyll distribution in lakes. They used a linear mixing model on spectroradiometer reflectance data. It turned out that phytoplankton patches are concentrated in lakes close to land with clear water influence.
(Strobl et al., 2007)	Q2	10	Scopus	M	Other techniques	World Cup	Own	-	Total phosphorus/-*	Describes a methodology called Critical Sampling Points (CSP) to identify crucial sampling locations in rural watersheds for total phosphorus contamination in water.
(Crowley & Cardille, 2020)	SC	37	Scopus	F	Support Vector Machines	World Cup	Public	-	No mention of-	It highlights the contribution in landscape structure and function, analysis of historical changes and simulation of future changes. It underlines the importance of the continued accessibility of free imagery from satellite sources and open access data analysis software in landscape ecology. It also emphasizes the future opportunities these advances offer for broader and more detailed studies in areas such as landscape connectivity, metapopulation dynamics, and social-ecological systems.

(Oscar-Júnior, 2021)	Q1	9	Scopus	M	Other techniques	Brazil	Public	Validity and accuracy index (Shapiro-Wilk)	Not mentioned/Rio	They found a correlation between precipitation and sea surface temperatures in the Pacific and Atlantic, highlighting the importance of the South Atlantic Convergence Zone in precipitation variability. These results are relevant for development planning and water resource management in the region.
(Maciel et al., 2021)	Q1	40	Scopus	M	Ensemble Learning, Support Vector Machines	Brazil	Own and public	Prediction errors	SST	Machine Learning methods outperformed Semi-Analytical methods, with Random Forest obtaining the best results (errors less than 22%). In addition, it was observed that Semi-Analytics could be useful for Zsd retrieval. The application of Random Forest to Sentinel-2 atmospherically corrected images showed reasonable feasibility with errors of 28%.
(Filisbino Freire da Silva et al., 2021)	Q1	7	Scopus	M	Support Vector Machines	Brazil	Own and public	Coefficient of Determination	Total Suspended Sediments (TSM)	Method for monitoring optical water types in Brazil using Sentinel-2 MSI data

(Ximenes et al., 2021)	Q1	7	Scopus	M	Clustering Algorithms	Brazil	Own	Validity and accuracy rates	NC	They mapped 14 ecoregions in the Purus-Madeira interfluve using machine learning techniques, highlighting the environmental heterogeneity of the region. This study can help in systematic conservation planning actions to maintain a greater number of protected ecoregions (as ecoregions represent environmental heterogeneity).
(Dawood et al., 2021)	Q1	8	Scopus	F	Other techniques	Peru	Own	Coefficient of Determination	NC	The study presents an integrated framework using Adaptive Neuro-Fuzzy Inference System (ANFIS) and Fuzzy Inference System (FIS) to predict water network condition index. The performance of the framework, superior to the multiple linear regression (MLR) model with an R^2 of 0.9145, demonstrates its effectiveness in optimizing prediction results and supporting sustainable urban water management.
(Valdes & Pou, 2021b)	SI	2	Scopus	M	Other techniques	World Cup	Public	validity and accuracy index	NC	The results highlight temporal changes in water vapor patterns. In the conclusions, it is highlighted that accurate seasonal prediction is achieved with a reduced set of features, and the association of water vapor patterns with specific events in hemispheric regions is demonstrated.

(Cui et al., 2022)	SI	0	Scopus	M	Support Vector Machines	Peru	Public	validity and accuracy index	NC	Results indicate reasonable performance, even with limited annotations, and the effectiveness of incorporating Lab color space is evaluated. Spatial regularity was observed and active learning is proposed as a future approach, along with testing unsupervised clustering algorithms on artisanal gold mining datasets.
(Flores Júnior et al., 2022)	SI	4	WOS	M	Other techniques	Brazil	Public	Prediction errors	NC	The results showed a significant improvement in the accuracy of Chl-a recovery with the hybrid semi-analytical algorithm (HSAA) compared to empirical models, with a mean absolute percent error (MAPE) of less than 37%. Furthermore, the applicability of the calibrated model for estimating Chl-a concentration in OLCI images was demonstrated, thus supporting the usefulness of semi-analytical models in highly turbid waters and their importance for monitoring Amazonian aquatic ecosystems.
(Nagel et al., 2023)	Q1	0	Scopus	M		World Cup	Public	-	NC	This paper highlights the importance of satellite data in environmental science, providing information on river migration patterns and their implications for ecosystem management and conservation.

(Liu et al., 2023)	Q1	8	Scopus	M	Correlation maps	World Cup	Public	correlation index	NC	Synergistic relationships between global nature contributions to people (PNC) were found to prevail over trade-offs, particularly in low-latitude areas vulnerable to climate change impacts. This research underscores the critical need for strategies that foster landscape multifunctionality and regional collaboration to enhance human well-being, despite challenges in selection.
(Ma et al., 2024)	Q1	3	Scopus	M		World Cup	Public	-	NC	The systematic review highlights the growing application and success of Transfer Learning (TL) in environmental remote sensing in various domains, including land cover mapping and disaster management. Future directions emphasize the need for reference datasets for TL evaluation, improving model interpretability, and leveraging basic models for remote sensing tasks, indicating the crucial role of TL in improving the efficiency of environmental monitoring.

NC=No analysis of physico-chemical or microbiological parameters; SC=no quartile

Fig 3 shows the co-occurrence of keywords where the word climate change is highlighted and it is the node that is related to the other keywords such as Computational intelligence. In addition, it is observed that the green nodes are the emerging words such as water resources, south Atlantic convergence zone. The visualization of the co-occurrence of keywords is important in recent years, since they exploit the mapping of knowledge related to each topic (Radhakrishnan et al., 2017).

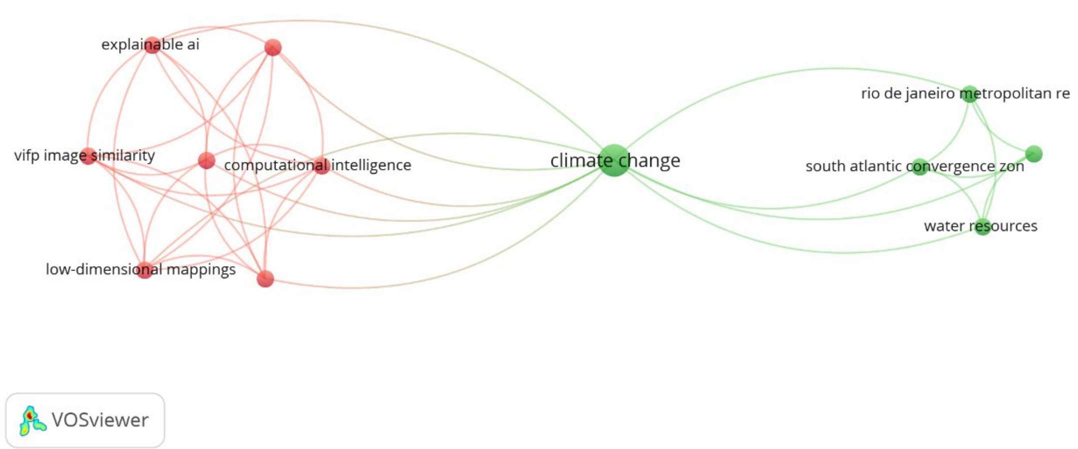

Figure .3: Keyword analysis of the studies

Limitations and future prospects

The limitations of the studies lie in insufficient accuracy, in the face of this, researchers are working on how to improve the accuracy of water quality prediction (Chen et al., 2023). The explored studies are based on spatial analysis of images and address few water quality parameters such as chlorophyll and turbidity, TSS (Feyisa et al., 2014; Su, 2017). Therefore, systematic methods to estimate error should be developed and the images should be of high resolution (Burns & Nolin, 2014). On the other hand, it is evident the need for future work to assess water quality in the Amazon, incorporating in the analysis several parameters (physicochemical-microbiological), especially the parameters of organic matter and heavy metals through in situ monitoring, associated with biological indicators of water quality and incorporating sensors for real-time evaluation.

It is evident that artificial intelligence can be used to monitor water quality in the Amazon. This in order to have more accurate data and to replace the evaluation of quality indices applied manually, and their results are expressed in categories (Maroneze et al., 2014) (Table 4), these indices are currently classified according to values, as shown in the ICAs of Colombia and Brazil. In the future, it is expected that AI will help organizations in monitoring, decision making and achieving safe and improved water quality for users to support a sustainable future (Chauhan & Sahoo, 2024; Mustafa et al., 2021).

Table 4. Categories of water quality indices commonly used in the Amazon countries

Categories of ICA			
	Colombia		Brazil
Code	ICA Rojas	ICAUCA	IAP
1	Very Bad	Lousy	Lousy
2	Inadequate	Inadequate	Mala
3	Acceptable	Acceptable	Regular
4	Good	Good	Good
5	Optima	Optima	Optima

Table 5 shows the physicochemical and microbiological parameters considered by the Colombian ICAs, which are based on weighted values; however, there are few of them, in this sense, AI has the great challenge of improving and establishing evaluation patterns.
Table 5. Physicochemical and microbiological parameters used by ICA and IA in Amazonia.

Parameter	Colombia		Amazonia
	Index		IA
	ICA Rojas 1991	ICAUCA 2004	Parameters used
Dissolved oxygen (DO)	0.25	0.21	-
pH	0.17	0.08	-
Biochemical oxygen demand (BOD5)	0.15	0.15	-
Nitrates			-
Fecal Coliforms	0.21	0.16	-
Temperature			-
Turbidity	0.11	0.07	-
Total Dissolved Solids	0.11	0.07	-
Total Phosphorus	-	0.08	X
Cadmium	-	-	-
Mercury	-	-	-
Electrical conductivity	-	-	-
Suspended Solids	-	0.05	X
Color	-	0.05	-
Total Nitrogen	-	0.08	-
Chlorides	-	-	-

Arsenic	-	-	-
Fluoride	-	-	-
Total Coliforms	-	-	-
Chemical Oxygen Demand (COD)	-	-	-
Alkalinity	-	-	-
Hardness	-	-	-
Phosphates	-	-	-
Cyanide	-	-	-
Selenium	-	-	-

- = Parameters not contemplated; X= Parameters used in AI studies in Amazonia; Adapted from: Torres et al.(2009)

CONCLUSIONS

The IA is characterized by using data from governmental institutions and it was evidenced that it is limited to parameters such as Total Suspended Solids (TSS), Turbidity, total organic carbon and Chlorophyll, limiting the analysis of water quality. It is expected that this will be overcome and the IA will use monitoring with the use of a greater number of parameters. Brazil is the leading Amazonian country with AI studies, however, its application is still very low compared to other European countries. The compound annual growth rate showed a value of 12.16% and 95% of the selected articles are hosted in Scopus 95% and only 7% are in Web of Science. This shows that these databases are the leaders in the scientific academic world due to the high standard of review during their editorial processes.

AI is expected to increase its potential to solve the challenges faced by water monitoring in the Amazon, providing technical support for more efficient management and operation.

REFERENCES

Agbasi, J. C., & Egbueri, J. C. (2023). Intelligent soft computational models integrated for the prediction of potentially toxic elements and groundwater quality indicators: a case study. *Journal of Sedimentary Environments*, *8*(1), 57–79. https://doi.org/10.1007/s43217-023-00124-y

Akter, S., Siddique, S., Jahan, I., Rabeya, T., & Mim, M. R. (2021). People Thoughts Prediction using Machine Learning on Women's Contribution in ICT in Bangladesh. *2021 IEEE 6th International Conference on Computing, Communication and Automation, ICCCA 2021*, *January*, 626–631. https://doi.org/10.1109/ICCCA52192.2021.9666268

Bakhmat, N., Kolosova, O., Demchenko, O., Ivashchenko, I., & Strelchuk, V. (2022). Application of International Scientometric Databases in the Process of Training Competitive Research and Teaching Staff: Opportunities of Web of Science (Wos), Scopus, Google Scholar. *Journal of Theoretical and Applied Information Technology*, *100*(13), 4914–4924.

Burns, P., & Nolin, A. (2014). Using atmospherically-corrected Landsat imagery to measure glacier area change in the Cordillera Blanca, Peru from 1987 to 2010. *Remote Sensing of Environment*, *140*, 165–178. https://doi.org/10.1016/j.rse.2013.08.026

Calcuvio. (2022). *Calculadora de tasa de crecimiento anual compuesto o CAGR | Calcuvio*. https://www.calcuvio.com/crecimiento-anual

Castillo, J. A., & Powell, M. A. (2019). Análisis de la producción científica del Ecuador e impacto de la colaboración internacional en el periodo 2006-2015. *Revista española de Documentación Científica*, *42*(1), 225. https://doi.org/10.3989/redc.2019.1.1567

Chaudhary, G. (2023). Environmental Sustainability: Can Artificial Intelligence be an Enabler for SDGs? *Nature Environment and Pollution Technology*, *22*(3), 1411–1420. https://doi.org/10.46488/NEPT.2023.v22i03.027

Chauhan, M., & Sahoo, D. R. (2024). Towards a Greener Tomorrow: Exploring the Potential of AI, Blockchain, and IoT in Sustainable Development. *Nature Environment and Pollution Technology*, *23*(2), 1105–1113. https://doi.org/10.46488/nept.2024.v23i02.044

Chen, Z., Liu, L., Wang, Y., & Gao, J. (2023). Review of Water Quality Prediction Methods. *Lecture Notes in Civil Engineering*, *341 LNCE*, 237–265. https://doi.org/10.1007/978-981-99-1919-2_17

Crowley, M. A., & Cardille, J. A. (2020). Remote Sensing's Recent and Future Contributions to Landscape Ecology. *Current Landscape Ecology Reports*, *5*(3), 45–57. https://doi.org/10.1007/s40823-020-00054-9

Cui, K., Camalan, S., Li, R., Pauca, V. P., Alqahtani, S., Plemmons, R., Silman, M., Dethier, E. N., Lutz, D., & Chan, R. (2022). Semi-Supervised Change Detection of Small Water Bodies Using Rgb and Multispectral Images in Peruvian Rainforests. *Workshop on Hyperspectral Image and Signal Processing, Evolution in Remote Sensing*, *2022-Septe*, 2–5. https://doi.org/10.1109/WHISPERS56178.2022.9955140

Dankar, F. K., & Ibrahim, M. (2021). Fake it till you make it: Guidelines for effective synthetic data generation. *Applied Sciences (Switzerland)*, *11*(5), 1–18. https://doi.org/10.3390/app11052158

Dawood, T., Elwakil, E., Novoa, H. M., & Delgado, J. F. G. (2021). Ensemble intelligent systems for predicting water network condition index. *Sustainable Cities and Society*, *73*(May). https://doi.org/10.1016/j.scs.2021.103104

De Fortuny, E. J., Martens, D., & Provost, F. (2013). Predictive modeling with big data: Is bigger really better? *Big Data*, *1*(4), 215–226. https://doi.org/10.1089/big.2013.0037

de Moraes Novo, E. M. L., de Farias Barbosa, C. C., Freitas, R. M., Shimabukuro, Y. E., Melack, J. M., & Filho, W. P. (2006). Seasonal changes in chlorophyll distributions in Amazon floodplain lakes derived from MODIS images. *Limnology*, *7*(3), 153–161. https://doi.org/10.1007/s10201-006-0179-8

de Paiva, R. C. D., Chaffe, P. L. B., Anache, J. A. A., Fontes, A. S., de Araujo, L. M. N., de Araujo, A. N., Bartiko, D., Bleninger, T., de Amorim, P. B., Buarque, D. C., Carlotto, T., Collischonn, W., Detzel, D. H. M., Fan, F. M., Formiga-Johnsson, R. M., Kobiyama, M., Mannich, M., Marques, G., Michel, G. P., … Zanandrea, F. (2020). Advances and challenges in the water sciences in brazil: A community synthesis of the xxiii brazilian water resources symposium. *Revista Brasileira de Recursos Hidricos*, *25*, 1–28. https://doi.org/10.1590/2318-0331.252020200136

Eck, Ludo Waltman, N. J. van. (2013). Full-Text Citation Analysis : A New Method to Enhance. *Journal of the American Society for Information Science and Technology*, *64*(July), 1852–1863. https://doi.org/10.1002/asi

Egbueri, J. C., Mgbenu, C. N., Digwo, D. C., & Nnyigide, C. S. (2023). A multi-criteria water quality evaluation for human consumption, irrigation and industrial purposes in Umunya area, southeastern Nigeria. *International Journal of Environmental Analytical Chemistry*, *103*(14), 3351–3375. https://doi.org/10.1080/03067319.2021.1907360

Emamgholizadeh, S., Kashi, H., Marofpoor, I., & Zalaghi, E. (2014). Prediction of water quality parameters of Karoon River (Iran) by artificial intelligence-based models. *International Journal of Environmental Science and Technology*, *11*(3), 645–656. https://doi.org/10.1007/s13762-013-0378-x

Feyisa, G. L., Meilby, H., Fensholt, R., & Proud, S. R. (2014). Automated Water Extraction Index: A new technique for surface water mapping using Landsat imagery. *Remote Sensing of Environment*, *140*, 23–35. https://doi.org/10.1016/j.rse.2013.08.029

Filisbino Freire da Silva, E., Márcia Leão de Moraes Novo, E., de Lucia Lobo, F., Clemente Faria Barbosa, C., Tressmann Cairo, C., Almeida Noernberg, M., & Henrique da Silva Rotta, L. (2021). A machine learning approach for monitoring Brazilian optical water types using Sentinel-2 MSI. *Remote Sensing Applications: Society and Environment*, *23*(July). https://doi.org/10.1016/j.rsase.2021.100577

Flores Júnior, R., Barbosa, C. C. F., Maciel, D. A., Novo, E. M. L. de M., Martins, V. S., Lobo, F. de L., Sander de Carvalho, L. A., & Carlos, F. M. (2022). Hybrid Semi-Analytical Algorithm for Estimating Chlorophyll-A Concentration in Lower Amazon Floodplain Waters. *Frontiers in Remote Sensing*, *3*(April), 1–20.

https://doi.org/10.3389/frsen.2022.834576

Foody, G. M. (2023). Remote sensing in landscape ecology. *Landscape Ecology*, *38*(11), 2711–2716. https://doi.org/10.1007/s10980-023-01753-4

Gad, M., Elsayed, S., Moghanm, F. S., Almarshadi, M. H., Alshammari, A. S., Khedher, K. M., Eid, E. M., & Hussein, H. (2020). Combining water quality indices and multivariate modeling to assess surface water quality in the Northern Nile Delta, Egypt. *Water (Switzerland)*, *12*(8). https://doi.org/10.3390/W12082142

Ghapor, A. A., Zubairi, Y. Z., & Imon, A. H. M. R. (2017). Missing value estimation methods for data in linear functional relationship model. *Sains Malaysiana*, *46*(2), 317–326. https://doi.org/10.17576/jsm-2017-4602-17

Hameed, M., Sharqi, S. S., Yaseen, Z. M., Afan, H. A., Hussain, A., & Elshafie, A. (2017). Application of artificial intelligence (AI) techniques in water quality index prediction: a case study in tropical region, Malaysia. *Neural Computing and Applications*, *28*, 893–905. https://doi.org/10.1007/s00521-016-2404-7

Hossain, M. A., Dwivedi, Y. K., & Rana, N. P. (2016). State-of-the-art in open data research: Insights from existing literature and a research agenda. *Journal of Organizational Computing and Electronic Commerce*, *26*(1–2), 14–40. https://doi.org/10.1080/10919392.2015.1124007

Irwan, D., Ali, M., Ahmed, A. N., Jacky, G., Nurhakim, A., Ping Han, M. C., AlDahoul, N., & El-Shafie, A. (2023). Predicting Water Quality with Artificial Intelligence: A Review of Methods and Applications. *Archives of Computational Methods in Engineering*, *30*(8), 4633–4652. https://doi.org/10.1007/s11831-023-09947-4

Isidro, C. M., McIntyre, N., Lechner, A. M., & Callow, I. (2018). Quantifying suspended solids in small rivers using satellite data. *Science of the Total Environment*, *634*, 1554–1562. https://doi.org/10.1016/j.scitotenv.2018.04.006

James, D. E., Schraw, G., & Kuch, F. (2015). Using the sampling margin of error to assess the interpretative validity of student evaluations of teaching. *Assessment and Evaluation in Higher Education*, *40*(8), 1123–1141. https://doi.org/10.1080/02602938.2014.972338

Kaur, T., & Sinha, A. K. (2019). Pesticides in Agricultural Run Offs Affecting Water Resources: A Study of Punjab (India). *Agricultural Sciences*, *10*(10), 1381–1395. https://doi.org/10.4236/as.2019.1010101

Lamprea, J., M, J. A. L., & Gómez-restrepo, C. (2007). Validez en la evaluación de escalas. *Revista Colombiana de Psiquiatría*, *36*(2), 340–348.

Laurance, W. F., Cochrane, M. A., Bergen, S., Fearnside, P. M., Delamônica, P., Barber, C., D’Angelo, S., & Fernandes, T. (2001). The future of the Brazilian Amazon. *Science*, *291*(5503), 438–439.

https://doi.org/10.1126/science.291.5503.438

Ledesma, J. L. J., Köhler, S. J., & Futter, M. N. (2012). Long-term dynamics of dissolved organic carbon: Implications for drinking water supply. *Science of the Total Environment*, *432*, 1–11. https://doi.org/10.1016/j.scitotenv.2012.05.071

Limon, P. J., & Webb, R. H. (1964). A Magnetic Resonance Experiment for the Undergraduate Laboratory. *American Journal of Physics*, *32*(5), 361–364. https://doi.org/10.1119/1.1970348

Lintern, A., Webb, J. A., Ryu, D., Liu, S., Bende-Michl, U., Waters, D., Leahy, P., Wilson, P., & Western, A. W. (2018). Key factors influencing differences in stream water quality across space. *Wiley Interdisciplinary Reviews: Water*, *5*(1), 1–31. https://doi.org/10.1002/WAT2.1260

Liu, Y., Fu, B., Wang, S., Rhodes, J. R., Li, Y., Zhao, W., Li, C., Zhou, S., & Wang, C. (2023). Global assessment of nature's contributions to people. *Science Bulletin*, *68*(4), 424–435. https://doi.org/10.1016/j.scib.2023.01.027

Ma, Y., Chen, S., Ermon, S., & Lobell, D. B. (2024). Transfer learning in environmental remote sensing. *Remote Sensing of Environment*, *301*(August 2023), 113924. https://doi.org/10.1016/j.rse.2023.113924

Maciel, D. A., Barbosa, C. C. F., Novo, E. M. L. de M., Flores Júnior, R., & Begliomini, F. N. (2021). Water clarity in Brazilian water assessed using Sentinel-2 and machine learning methods. *ISPRS Journal of Photogrammetry and Remote Sensing*, *182*(October), 134–152. https://doi.org/10.1016/j.isprsjprs.2021.10.009

Marengo, J. A., Souza, C. M., Thonicke, K., Burton, C., Halladay, K., Betts, R. A., Alves, L. M., & Soares, W. R. (2018). Changes in Climate and Land Use Over the Amazon Region: Current and Future Variability and Trends. *Frontiers in Earth Science*, *6*(December), 1–21. https://doi.org/10.3389/feart.2018.00228

Maroneze, M. M., Zepka, L. Q., Vieira, J. G., Queiroz, M. I., & Jacob-Lopes, E. (2014). A tecnologia de remoção de fósforo: Gerenciamento do elemento em resíduos industriais. *Revista Ambiente e Agua*, *9*(3), 445–458. https://doi.org/10.4136/1980-993X

McClain, M. E., Aparicio, L. M., & Llerena, C. A. (2001). Water use and protection in rural communities of the Peruvian Amazon Basin. *Water International*, *26*(3), 400–410. https://doi.org/10.1080/02508060108686932

Morales-Rojas, E. (2020). Mixed greywater treatment for irrigation uses. *Revista Ambiente e Agua*, *15*(6), 445–458. https://doi.org/https://doi.org/10.4136/ambi-agua.2599

Mouw, C. B., Greb, S., Aurin, D., DiGiacomo, P. M., Lee, Z., Twardowski, M., Binding, C., Hu, C., Ma, R., Moore, T., Moses, W., & Craig, S. E. (2015). Aquatic color radiometry remote sensing of coastal and inland waters: Challenges and recommendations for future satellite missions. *Remote Sensing of Environment*, *160*, 15–30. https://doi.org/10.1016/j.rse.2015.02.001

Mustafa, H. M., Mustapha, A., Hayder, G., & Salisu, A. (2021). Applications of IoT and Artificial Intelligence in Water Quality Monitoring and Prediction: A Review. *Proceedings of the 6th International Conference on Inventive Computation Technologies, ICICT 2021*, *February*, 968–975. https://doi.org/10.1109/ICICT50816.2021.9358675

Myllyaho, L., Raatikainen, M., Männistö, T., Mikkonen, T., & Nurminen, J. K. (2021). Systematic literature review of validation methods for AI systems. *Journal of Systems and Software*, *181*, 111050. https://doi.org/10.1016/j.jss.2021.111050

Nagel, G. W., Darby, S. E., & Leyland, J. (2023). The use of satellite remote sensing for exploring river meander migration. *Earth-Science Reviews*, *247*(October), 104607. https://doi.org/10.1016/j.earscirev.2023.104607

Nikolaou, A. D., Meric, S., Lekkas, D. F., Naddeo, V., Belgiorno, V., Groudev, S., & Tanik, A. (2008). Multi-parametric water quality monitoring approach according to the WFD application in Evros trans-boundary river basin: priority pollutants. *Desalination*, *226*(1–3), 306–320. https://doi.org/10.1016/j.desal.2007.02.113

Odermatt, D., Gitelson, A., Brando, V. E., & Schaepman, M. (2012). Review of constituent retrieval in optically deep and complex waters from satellite imagery. *Remote Sensing of Environment*, *118*, 116–126. https://doi.org/10.1016/j.rse.2011.11.013

Oscar-Júnior, A. C. (2021). Precipitation Trends and Variability in River Basins in Urban Expansion Areas. *Water Resources Management*, *35*(2), 661–674. https://doi.org/10.1007/s11269-020-02749-4

Paraschiv, D., Tudor, C., & Petrariu, R. (2015). The textile industry and sustainable development: A holt-winters forecasting investigation for the Eastern European area. *Sustainability (Switzerland)*, *7*(2), 1280–1291. https://doi.org/10.3390/su7021280

Park, G. S. (2007). The role and distribution of total suspended solids in the macrotidal coastal waters of Korea. *Environmental Monitoring and Assessment*, *135*(1–3), 153–162. https://doi.org/10.1007/s10661-007-9640-3

Perkmann, M., & Schildt, H. (2015). Open data partnerships between firms and universities: The role of boundary organizations. *Research Policy*, *44*(5), 1133–1143. https://doi.org/10.1016/j.respol.2014.12.006

Perona, E., Bonilla, I., & Mateo, P. (1999). Spatial and temporal changes in water quality in a Spanish river. *Science of the Total Environment*, *241*(1–3), 75–90. https://doi.org/10.1016/S0048-9697(99)00334-4

Radhakrishnan, S., Erbis, S., Isaacs, J. A., & Kamarthi, S. (2017). Correction: Novel keyword co-occurrence network-based methods to foster systematic reviews of scientific literature. *PLoS ONE*, *12*(9), 1–16. https://doi.org/10.1371/journal.pone.0185771

Rajaee, T., Khani, S., & Ravansalar, M. (2020). Artificial intelligence-based single and hybrid models for prediction of water quality in rivers: A review. *Chemometrics and Intelligent Laboratory Systems*, *200*(August 2019), 103978. https://doi.org/10.1016/j.chemolab.2020.103978

Romero Lopez, T. de J., & Vargas Mato, D. (2017). Uso de microorganismos eficientes para tratar aguas contaminadas. *Iha*, *XXXVIII*(3), 88–100.

Roy-García, I., Rivas-Ruiz, R., Pérez-Rodríguez, M., & Palacios-Cruz, L. (2019). Correlation: Not all correlation entails causality. *Revista Alergia Mexico*, *66*(3), 354–360. https://doi.org/10.29262/ram.v66i3.651

S. Mustafa, A., O. Sulaiman, S., & H. Shahooth, S. (2017). Application of QUAL2K for Water Quality Modeling and Management in the lower reach of the Diyala river. *Iraqi Journal of Civil Engineering*, *11*(2), 66–80. https://doi.org/10.37650/ijce.2017.134910

Salas-Salvadó, J., Maraver, F., Rodríguez-Mañas, L., de Pipaon, M. S., Vitoria, I., & Moreno, L. A. (2020). The importance of water consumption in health and disease prevention: The current situation. *Nutricion Hospitalaria*, *37*(5), 1072–1086. https://doi.org/10.20960/nh.03160

Strobl, R. O., Robillard, P. D., & Debels, P. (2007). Critical sampling points methodology: Case studies of geographically diverse watersheds. *Environmental Monitoring and Assessment*, *129*(1–3), 115–131. https://doi.org/10.1007/s10661-006-9346-y

Su, T. C. (2017). A study of a matching pixel by pixel (MPP) algorithm to establish an empirical model of water quality mapping, as based on unmanned aerial vehicle (UAV) images. *International Journal of Applied Earth Observation and Geoinformation*, *58*, 213–224. https://doi.org/10.1016/j.jag.2017.02.011

Sun, D., Hu, C., Qiu, Z., Cannizzaro, J. P., & Barnes, B. B. (2014). Influence of a red band-based water classification approach on chlorophyll algorithms for optically complex estuaries. *Remote Sensing of Environment*, *155*, 289–302. https://doi.org/10.1016/j.rse.2014.08.035

Tiyasha, Tung, T. M., & Yaseen, Z. M. (2020). A survey on river water quality modelling using artificial intelligence models: 2000–2020. *Journal of Hydrology*, *585*(June), 2000–2020. https://doi.org/10.1016/j.jhydrol.2020.124670

Torres, P., Cruz, C. H., & Patiño, P. (2009). Índices De Calidad De Agua En Fuentes Superficiales Utilizadas En La Producción De Agua Para Consumo Humano. Una

Revisión Crítica. *Revista Ingenierías Universidad de Medellín*, *8*(15), 79–94.

Valdes, J. J., & Pou, A. (2021a). A Machine Learning - Explainable AI approach to tropospheric dynamics analysis using Water Vapor Meteosat images. *2021 IEEE Symposium Series on Computational Intelligence, SSCI 2021 - Proceedings*, 8–11. https://doi.org/10.1109/SSCI50451.2021.9660188

Valdes, J. J., & Pou, A. (2021b). A Machine Learning - Explainable AI approach to tropospheric dynamics analysis using Water Vapor Meteosat images. *2021 IEEE Symposium Series on Computational Intelligence, SSCI 2021 - Proceedings*, 2–5. https://doi.org/10.1109/SSCI50451.2021.9660188

Valentine, J. C., Hedges, L. V., & Cooper, H. M. (2019). Handbook of Research Synthesis and Meta-Analysis 2nd Edition. In *The Lancet* (Vol 389, Number 10082).

Veneros, J., Ramos, L. C., Goñas, M., Morales, E., Auquiñivín-Silva, E., Oliva, M., & García, L. (2024). Seasonal Variability of Water Quality for Human Consumption in the Tilacancha Conduction System, Amazonas, Peru. *Nature Environment and Pollution Technology*, *23*(2), 899–909. https://doi.org/10.46488/nept.2024.v23i02.025

Villena Chávez, J. A. (2018). Calidad del agua y desarrollo sostenible. *Revista Peruana de Medicina Experimental y Salud Pública*, *35*(2), 304. https://doi.org/10.17843/rpmesp.2018.352.3719

Weng, W., Luedeke, M., Zemp, D., Lakes, T., & Kropp, J. (2018). Aerial and surface rivers: Downwind impacts on water availability from land use changes in Amazonia. *Hydrology and Earth System Sciences*, *22*(1), 911–927. https://doi.org/10.5194/hess-22-911-2018

Ximenes, A. C., Amaral, S., Monteiro, A. M. V., Almeida, R. M., & Valeriano, D. M. (2021). Mapping the terrestrial ecoregions of the Purus-Madeira interfluve in the Amazon Forest using machine learning techniques. *Forest Ecology and Management*, *488*(February). https://doi.org/10.1016/j.foreco.2021.118960

Yacobi, Y. Z., Moses, W. J., Kaganovsky, S., Sulimani, B., Leavitt, B. C., & Gitelson, A. A. (2011). NIR-red reflectance-based algorithms for chlorophyll-a estimation in mesotrophic inland and coastal waters: Lake Kinneret case study. *Water Research*, *45*(7), 2428–2436. https://doi.org/10.1016/j.watres.2011.02.002

Yang, Z., Reiter, M., & Munyei, N. (2017). Estimation of chlorophyll-a concentrations in diverse water bodies using ratio-based NIR/Red indices. *Remote Sensing Applications: Society and Environment*, *6*(February), 52–58. https://doi.org/10.1016/j.rsase.2017.04.004

Zhu, J., & Liu, W. (2020). A tale of two databases: the use of Web of Science and Scopus in academic papers. *Scientometrics*, *123*(1), 321–335.

Prepublished copy
Nature Environment & Pollution Technology
This is a peer-reviewed prepublished version of the paper
https://doi.org/10.1007/s11192-020-03387-8