On the Spatial-Temporal Behavior, and on the Relationship Between Water Quality and Hydrometeorological Information to Predict Dissolved Oxygen in Tropical Reservoirs. Case Study: La Miel, Hydropower Dam

Authors: Juliana-Andrea, Alzate-Gómez, Cesar, Aguirre-Duran, Jorge Alberto, Escobar-Vargas, Luis-Javier, Montoya-Jaramillo, and Carlos-César, Piedrahita-Escobar

Source: Air, Soil and Water Research, 16(1)

Published By: SAGE Publishing

URL: https://doi.org/10.1177/11786221221150189

On the Spatial-Temporal Behavior, and on the Relationship Between Water Quality and Hydrometeorological Information to Predict Dissolved Oxygen in Tropical Reservoirs. Case Study: La Miel, Hydropower Dam

Air, Soil and Water Research
Volume 16: 1–14

DOI: 10.1177/11786221221150189

Alzate-Gómez Juliana-Andrea[1], Aguirre-Duran Cesar[2], Escobar-Vargas Jorge Alberto[2], Montoya-Jaramillo Luis-Javier[1] and Piedrahita-Escobar Carlos-César[1]

[1]Universidad de Medellín, Colombia. and [2]Pontificia Universidad Javeriana, Bogotá, Colombia.

ABSTRACT: Hydropower is currently one of the leading renewable energy sources in developing countries. Despite the benefits that it can provide, it also triggers significant environmental impacts, such as changes in the reservoirs' water quality. In quantifying those changes, dissolved oxygen (DO) is used as one of the water quality indicators and is the most used variable to quantify water quality and analyze water pollution. This paper aims to establish a relationship between water quality and hydrometeorological variables in tropical reservoirs to better estimate dissolved oxygen. Univariate and multivariate techniques were used to analyze temporal and spatial changes in watersheds to better select vital variables for the forecast model, such as Vector Autoregression (VAR). The results show that, for all monitoring stations, the water quality variables associated with the DO process are COD, BOD, and PO_4. Likewise, precipitation and flow discharge were the hydrometeorological parameters that had the most significant impact on DO. Also, the principal component analysis (PCA) allowed us to identify that the strength of the relationships between water quality and hydrometeorology changes depending on the location of the monitoring site. Finally, the implementation of a VAR model showed good performance metrics for dissolved oxygen predictions based on all analyses.

KEYWORDS: Tropical reservoir, water quality, hydrometeorology, hydroinformatics, hydrological time series

RECEIVED: June 15, 2022. **ACCEPTED:** December 14, 2022.

TYPE: Original Research

CORRESPONDING AUTHORS: Montoya-Jaramillo Luis-Javier, Faculty of Enginering, Universidad de Medellín, Cra. 87 #30–65, Medellin, 050026, Colombia. Email: ljmontoya@udemedellin.edu.co

Alzate-Gómez Juliana-Andrea, Universidad de Medellín, Cra. 87 #30-65, Medellin, 050026, Colombia. Email: jalzate321@soyudemedellin.edu.co

Introduction

Hydropower is a significant source of renewable electricity, with a share of 16%–17% of the total world electricity generation (Killingtveit, 2019), and currently it is the main renewable energy source in most countries in Asia (Vaidya et al., 2021; Li, Chen et al., 2018), Europe (Alsaleh & Abdul-Rahim, 2021, 2022), Africa (Gyimah et al., 2021; Woldeab et al., 2018), and South America (de Oliveira et al., 2021; Semensatto et al., 2021). Furthermore, future hydropower development is primarily concentrated in developing countries and emerging economies of Southeast Asia, South America, and Africa, also with the Balkans, Anatolia, and the Caucasus being additional centers of future dam construction (Zarfl et al., 2015).

Despite the benefits that hydropower can provide, these structures can also trigger substantial environmental impacts like the, such as water supply to communities, flood control, and greenhouse gases reduction (Silva & Castillo, 2021), these structures can also trigger substantial environmental impacts like the disruption of aquatic ecosystems, reduction of riparian biodiversity, modification of stream morphology, or water quality degradation, among others (Barbarossa et al., 2020; Zarfl et al., 2019). In the case of water quality inside the reservoir, this factor is affected principally by meteorological, hydrological, and geological factors, as well as land use (Dalu & Wasserman, 2018; Marcé et al., 2010; Jerves-Cobo et al., 2020; Vega et al., 2018).

In the case of effects on reservoir water quality caused by meteorological and hydrological factors, these are described mainly based on geographic location. This is particularly the case for high and low-latitude systems also called temperate systems, where the four seasons have an impact on the system (Hwang et al., 2016; Weirich et al., 2019). Most of the studies carried out focus mainly on the assessment of water quality and its relationship with meteorological, and anthropic variables in eutrophic reservoirs (Winton et al., 2019). About the foregoing, it has been learned that nutrients and organic matter are responsible for most of the variation in reservoir water quality related to anthropogenic activities that directly impact the reservoir, followed by suspended solids related to both anthropogenic and natural processes, but not directly with dissolved oxygen (Mamun et al., 2021). In the same way, water quality parameters exhibited a seasonal fluctuation, with predominantly higher concentrations during the dry season than the wet season (Woldeab et al., 2018). Additionally, most of the best-documented examples of impacts—which stem from oligotrophication and water quality behavior, under this trophic state—come from temperate catchments with important and carefully monitored fisheries (Winton et al., 2019). In temperate systems, seasonal patterns, hydrology, and watershed morphology are the main regulatory factor for the nutrient concentration in reservoirs (Mamun et al., 2018; Nadarajah et al., 2019). Similarly, water quality variables, except total nitrogen, show vast seasonal differences as a result of high seasonality in water temperature and water flow (Varol, 2020b). However, all those conclusions cannot be directly applied to tropical case studies where the hydrological regime is different (Winton et al., 2019). Nevertheless, based on the studies mentioned above, it is necessary to go deeper into studies that directly relate water quality to hydrological and meteorological factors in hydropower systems in tropical latitudes (Calamita, 2020).

Regarding tropical systems, water quality analysis has been applied, and the hydrological cycle was shown to affect reservoir water quality (Lobato et al., 2015) directly. However, the application of multivariate statistical analysis in tropical reservoirs is limited and concentrated in eutrophic systems impacted by intense anthropic activities (Ling et al., 2017; Marques et al., 2019).

Statistical analysis and prediction play an essential role in processing surface water quality time series, with tools such as outliers detection, normality tests, and trend detection, among others, granting an excellent first approach (Fu & Gan Wang, 2012). To perform an integrated statistical analysis of water quality is necessary to consider hydrological, meteorological, and anthropic activities. Multivariate statistical analysis, such as principal component analysis and correlation analysis, facilitates integrated water quality data analysis since it allows the identification of factors that influence water quality (Chen et al., 2015; Varol, 2020a). These methods exhibit a practical approach to assessing and forecasting the water quality of reservoirs and can be used as a tool for water quality management (Varol, 2020b).

Analyzing water quality is possible through dissolved oxygen (DO) because is used as one of the water quality indicators and is the most used variable to quantify water quality and analyze water pollution. Since it plays a substantial role in aquatic environment characterization and shows the equilibrium between the processes that produce or consume oxygen in that environment, predicting its concentration could be advantageous for the environmental custodians. Accurate predictions of DO concentrations can help better manage tropical ecosystems. Low DO concentrations can lead to the mortality of aquatic organisms and the release of nutrients from the sediments, among others (Vilas et al., 2018), which is why we should consider studies to predict it.

Concerning dissolved oxygen forecasting, previous studies have explored the prediction. Some of the statistical methods are multiple linear regression models, artificial neural networks, classification tree, principal component/factor analysis, discriminant analysis, and Normal hidden Markov models. They are used independently or jointly to predict the temporal evolution of water, and the results are of outstanding quality but do not take into account spatial and hydrological factors (Liu et al., 2021). Regression models are most widely used for modeling the stochastic behavior of DO concentrations and artificial intelligence (Yaseen et al., 2018). Even so, regression models and neural networks require more data preprocessing, complex relationships cannot be modeled without transforming the input, and non-linear relationships cannot be captured, being very sensitive to different scales of variables. They usually require larger amounts of data for model training and require a lot of computing resources. Finally, knowing the rules or reasons why the artificial network returns those results is not usually easy and needs other analyses.

This study's main objective is to establish relations between water quality and hydrometeorological variables to predict and estimate the dissolved oxygen concentration in a tropical reservoir. The results obtained using the vector autoregressive (VAR) with different approaches focus on critical parameters and be efficient in predicting.

In this same sense, it is analyzed that changes in the dissolved oxygen are given by water temperature, and the intensity of biological processes such as photosynthesis, respiration, and decomposition of organic matter (Rajwa-Kuligiewicz et al., 2015), due to changing hydrometeorological conditions (Rajwa et al., 2014), in the case of the reservoir it has been identified that DO have changes due seasonal and vertical dynamics (Lliev & Hadjinikolova, 2013), air temperature and nutrients (Dordoni et al., 2022) and physical process the atmosphere (Liqoarobby et al., 2021). Following this, we hypothesized that in tropical reservoirs the dissolved oxygen dynamics depend on the air temperature, Sunshine duration, flow discharge, and precipitation, and internal chemical process as water temperature and decomposition of organic matter and Chemical oxygen demand. Identifying the key variables will allow for a more accurate prediction of dissolved oxygen.

Study Area Description and Data Source

Study area

La Miel river basin is in the Central Mountain range of the Colombian Andes. The river reaches being studied have a length of 62 km and drain an area of 712 km^2. In this area, the La Miel river has a variety of tributaries, including Tenerife, Salado, Manso, Moro, Pennsylvania, and Samaná, among other rivers.

"La Miel I" is a gravity dam located between 5°15′ N–5°35′ N and 74°53′ W–75°15′ on the riverbed of "La Miel" river in Caldas Department, Colombia, as is shown in Figure 1. Between 1997 and 2002, the dam was built for the primary purpose of hydroelectric power generation. It is 188 m high, giving it a storage capacity of 571 million m^3 and a surface area of 12.2 km^2, with an installed generation capacity of 396 MW in three turbine units. Its commercial operation began in December 2002. In 2010, the Guarinó diversion dam on the Guarinó River was opened, and the Manso diversion dam began operating in 2013. Both divert water into the Amaní Reservoir through a tunnel.

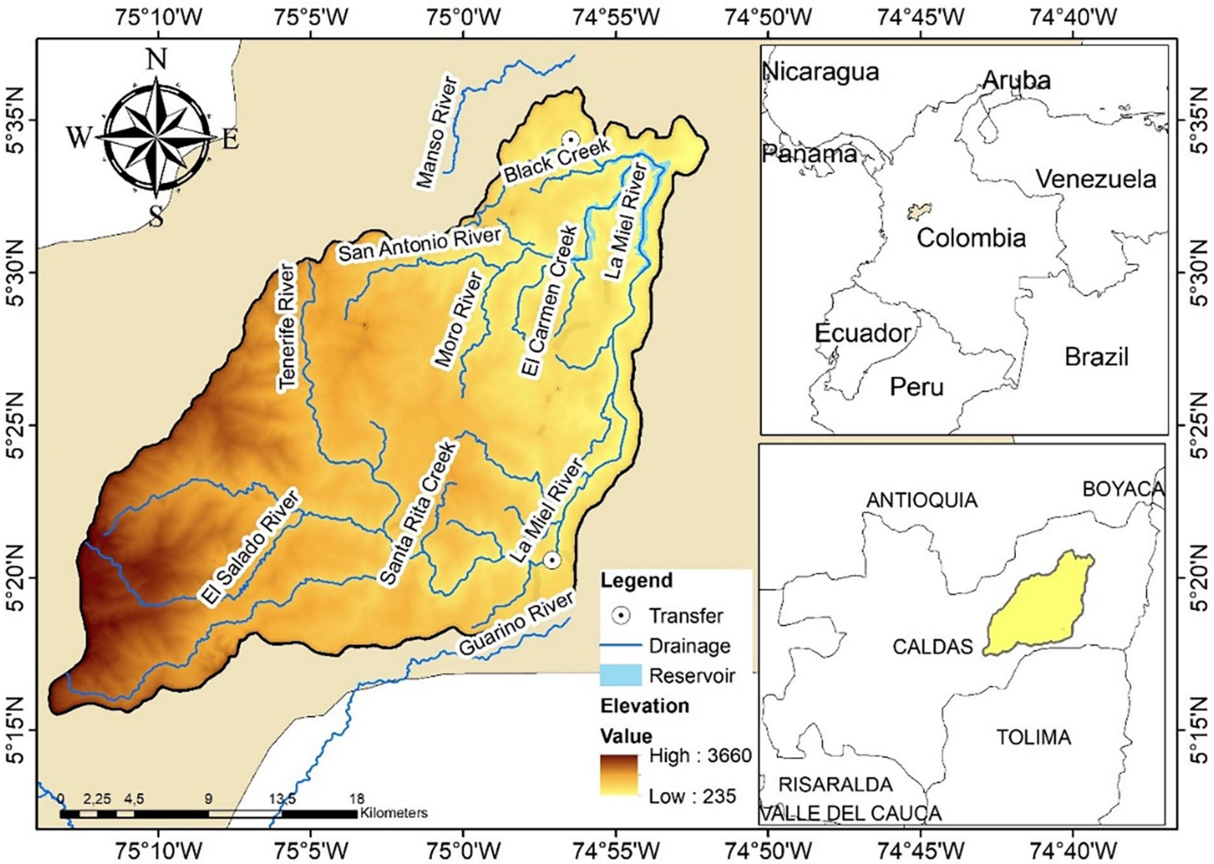

Figure 1. Location area study of "La Miel" hydropower dam and its basin.

In terms of hydrology, the watershed has a bimodal regime, composed of two wet and dry seasons. The average annual precipitation in the reservoir catchment is about 4,428 mm, with the heaviest rainfall occurring approximately in November with little precipitation between June and August. The dry season occurs from June to August, and the wet season—which is broken by dry periods—spans from May to November. Furthermore, the highest monthly evaporation rates occur in the dry season, with 165 mm/day rates, while the lowest evaporation rates occur in the wet season with an average of 46 mm/day. La Miel river basin experiences rainfall and a humid atmosphere throughout the year.

Hydro-meteorological and water quality data

Monthly data of water quality parameters, such as Temperature (T °C), pH, total suspended solids (TSS, mg/L), dissolved oxygen (DO, mg/L), and conductivity (µS/cm) were measured in situ, at the reservoir. In addition, other monthly data were taken to perform laboratory analysis to measure concentrations of chemical oxygen demand (COD mg/L), biochemical oxygen demand (BOD mg/L), ammonia (NH_4 mg/L), nitrite (NO_2 mg/L), nitrate (NO_3 mg/L), total Kjeldahl nitrogen (TKN mg/L), phosphate (PO_4 mg/L), Total phosphorus (TP mg/L). Inside the reservoir, there are five monitoring stations. These are distributed in two tributaries and the dam. The monitoring stations E4 and E6 are located in the Moro River. The monitoring stations E3 and E5 are in the La Miel river. Finally, the E7 monitoring station is at the dam, as shown in Figure 2. The measurements and samplings were carried out between January 2002 and December 2015.

For the same period, seven hydrological parameters (discharge, precipitation, solar brightness, evaporation, relative humidity, cloudiness, and air temperature) were collected, covering all basin areas. These data were obtained from hydrological and meteorological gaging stations. Specifically, nine monitoring sites were used for measuring precipitation, four sites for discharge, and two sites for the other hydro-meteorological parameters. In addition, hydro-meteorological data were obtained from the DHIME website of the Institute of Hydrology, Meteorology, and Environmental Studies (IDEAM, acronym in Spanish).

Methods

Data treatment and analysis

Firstly, visual inspections such as scatterplots and boxplots were performed to test for outliers in water quality and hydro-meteorological time series, then anomalous values from the time series were identified and removed. Based on the hydrology described in the study area section, the time series were divided into the dry and rainy seasons for analysis, considering tropical seasonality.

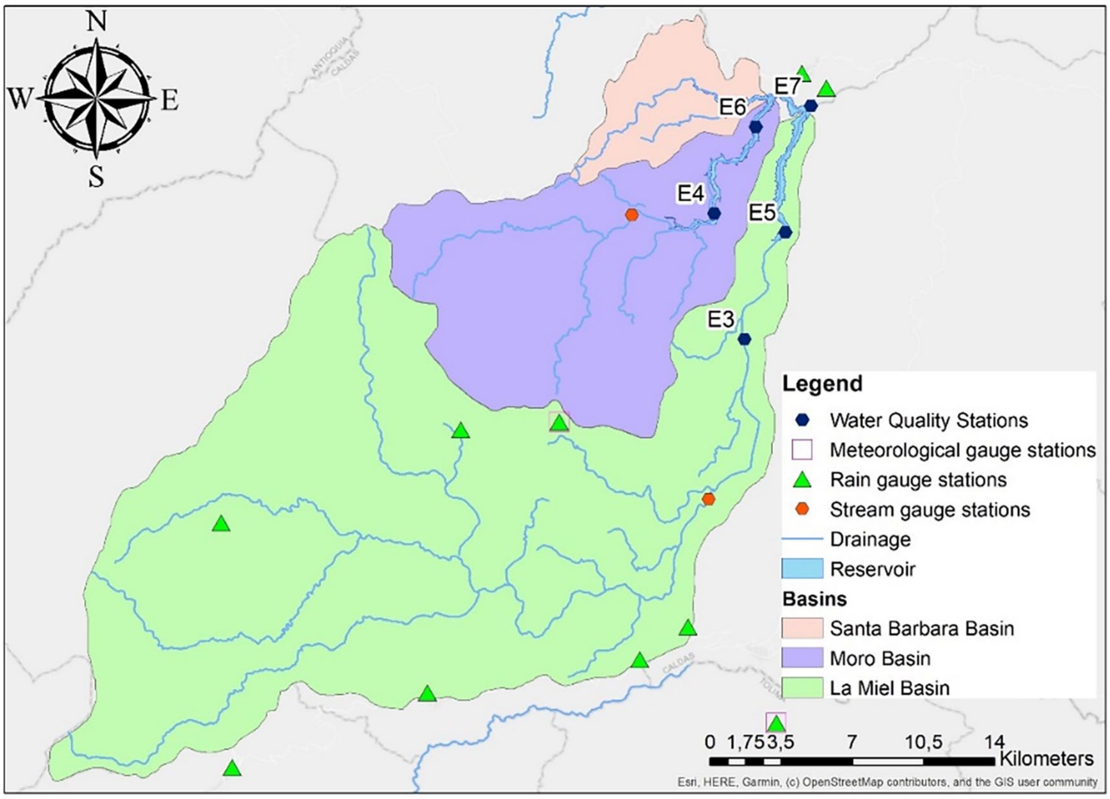

Figure 2. Location of water quality monitoring and hydrometeorological gaging stations.

Descriptive statistics were calculated with mean, maximum, minimum, and standard deviation values to identify the order of magnitude of water quality time series in each season. The dispersion of the data sets in each monitoring station assesses preliminary spatial and seasonal variation.

Univariate statistical analysis

The Kolmogorov-Smirnov one-sample test assessed the normality of the time series and the Shapiro-Wilk test with a significance level of 0.05, was applied to each parameter at each monitoring station. These tests determine whether parametric or nonparametric tests are more reliable for the following analyses. Afterward, a Pettitt and Mann-Kendall nonparametric test was performed to assess the homogeneity of the time series and the existence of a monotonic trend. In the case of Pettit's test, it identifies points where stepwise shifts (breaks) occur when inhomogeneity is extracted.

In addition, Kruskal-Wallis H tests were performed for each water quality parameter between the tail and mid sites and between the tail and dam sites.

Correspondingly, Levene and Mann-Whitney tests were applied to each parameter, assessing homoscedasticity and stationarity of time series between dry and wet seasons. A detrend of time series was applied for both, knowing that these analyses are affected by trends.

Multivariate statistical analysis with missing values

Correlation analysis was performed to measure the strength of the association between parameters from each monitoring site and between water quality parameters, hydro-meteorological parameters, and mixtures of water quality and hydro-meteorological parameters. The analysis was evaluated using the Pearson, Kendall, and Spearman correlation coefficients. The Pearson correlation coefficient is often used for jointly customarily distributed data (data that follows a bivariate normal distribution). For the latter case, the time series is normalized for better test results. For non-normally distributed continuous data, ordinal data, or data with correlated outliers, Spearman's rank correlation and Kendall's rank correlation can be used as measures of monotonic correlation (Schober et al., 2018).

Data imputation

Once univariate and correlation function analysis has been performed on all-time series, imputation techniques are applied to the time series to fill in the gaps.

Due to the fact that the next analyses require continuously sampled data, it is necessary to obtain the estimates of the missing values. The data produced by this method is used to define the most important variables, then to build possible modeling scenarios, and the analysis is performed using PCA.

In the first approximation, linear, quadratic, and cubic padding is applied. In the case of a large number of consecutive out-of-stocks, use the average value calculated before filling in the data.

Since these are univariate techniques, they do not obtain process variance. Hence, Multiple Imputation of Chain Equations (MICE) is applied as a multivariate technique, as this technique uses the time series of other variables to obtain their variances to complete the series of missing data. Consequently, the relationships and correlations discovered in the previous stages are considered to select time series to fill in the gaps.

In this study, the linear regression model of the MICE technique was changed by machine learning models, specifically Bayesian Ridge, Decision Tree Regressor, Extra Tree Regressor, and K-Nearest Neighbors. In addition, all search engines were evaluated, and the method that best preserved the variance of the existing data was selected.

To define the imputation technique that best fits the time series gaps to carry out the following statistical tests, a normality test was performed to compare the results before and after imputation, and it was found that all filling techniques maintained the results after imputation. Therefore, another validation for evaluating padding adjustments is to compare the mean and standard deviation of the time series before and after applying the imputation technique.

Principal component analysis (PCA)

This method is used to define the most representative climate and quality variables and to define modeling scenarios that allow the prediction of dissolved oxygen using the variables identified as significant for each site in the previous analyses. PAC helps you interpret your data, in this case, simplifying the complexities of high-dimensional data while preserving trends and patterns. Considering that there were at first 24 variables, it was necessary to determine, which of these were the variables that dominated the dynamics of the dissolved oxygen interaction. For this, multivariate analysis is required to analyze the arrangement in which several interrelated quantitative dependent variables describe the observations.

Its goal is to extract the vital information from the arrangement, to represent it as a set of new orthogonal variables called principal components, and to display the pattern of similarity of the observations, and the variables as points (Abdi & Williams, 2010). To ensure that each variable contributed the same proportion to the analysis, the time series were normalized. After that, PCA analysis was applied.

The first two components are extracted and analyzed to complete and define the scenarios for the prediction, considering that they are the components that contribute the most to the variability of the system. Since they explain more than 70% of the variability of the data.

Forecasting model vector autoregression (VAR)

Dissolved oxygen is the most used variable to quantify water quality and analyze water pollution. Since it plays a substantial role in aquatic environment characterization and shows the equilibrium between the processes that produce or consume oxygen in that environment.

In the previous phases are identified the main climatic and water quality processes that determine the OD behavior, then these processes are linked through univariate and multivariate analyses until reaching the definition of the main Spatio-temporal processes of water quality and climatology.

Dissolved oxygen is predicted using vector autoregression (VAR), which generalizes univariate autoregressive models and allows the modeling of multivariate time series systems. Each variable is modeled through a linear equation, including its lagged values, the lagged values of the other variables, and an error term.

To forecast dissolved oxygen in the E3 and E7 stations, the VAR model is used in two scenarios. The first scenario consists of variables associated with the first component found in the PCA analysis, and the second scenario consists of variables associated with the first two components produced by the PCA analysis.

The selected time series were divided into training and testing sequences; this separation was done using the last 2 years (of the time series) for testing and the rest for training. The stationarity test was again applied to the training series to evaluate whether all the time series were stationary. If any series is non-stationary, according to the test, then the time series is differential in this case. The Akaike Information Criterion (AIC) is used to determine the order of the model (Hurvich & Tsai, 1991) and the predictability of the developed model is evaluated and validated using statistical indicators such as root mean squared errors (RMSE) and mean absolute errors (MAE). All statistical analyses and methods were performed using Python 3.8.

Results and Discussion

Data analysis a general overview of the reservoir water quality

The mean and standard deviation of the Time series for the dam was plotted in Figure 3. These visualization techniques show that outliers are negligible. The order of magnitude corresponds to the expected value for such a system, considering the case of the water quality series.

In the case of water temperature, for both seasons, the lowest temperature was 23.4°C. However, the highest temperature varied depending on the season, finding that the highest temperature was 32.8°C for the dry season, and in the wet season, the highest temperature was 31.6°C. It is also worth mentioning that the standard deviation does not exceed 1.2°C. From these findings, it can be inferred that water temperature depends on

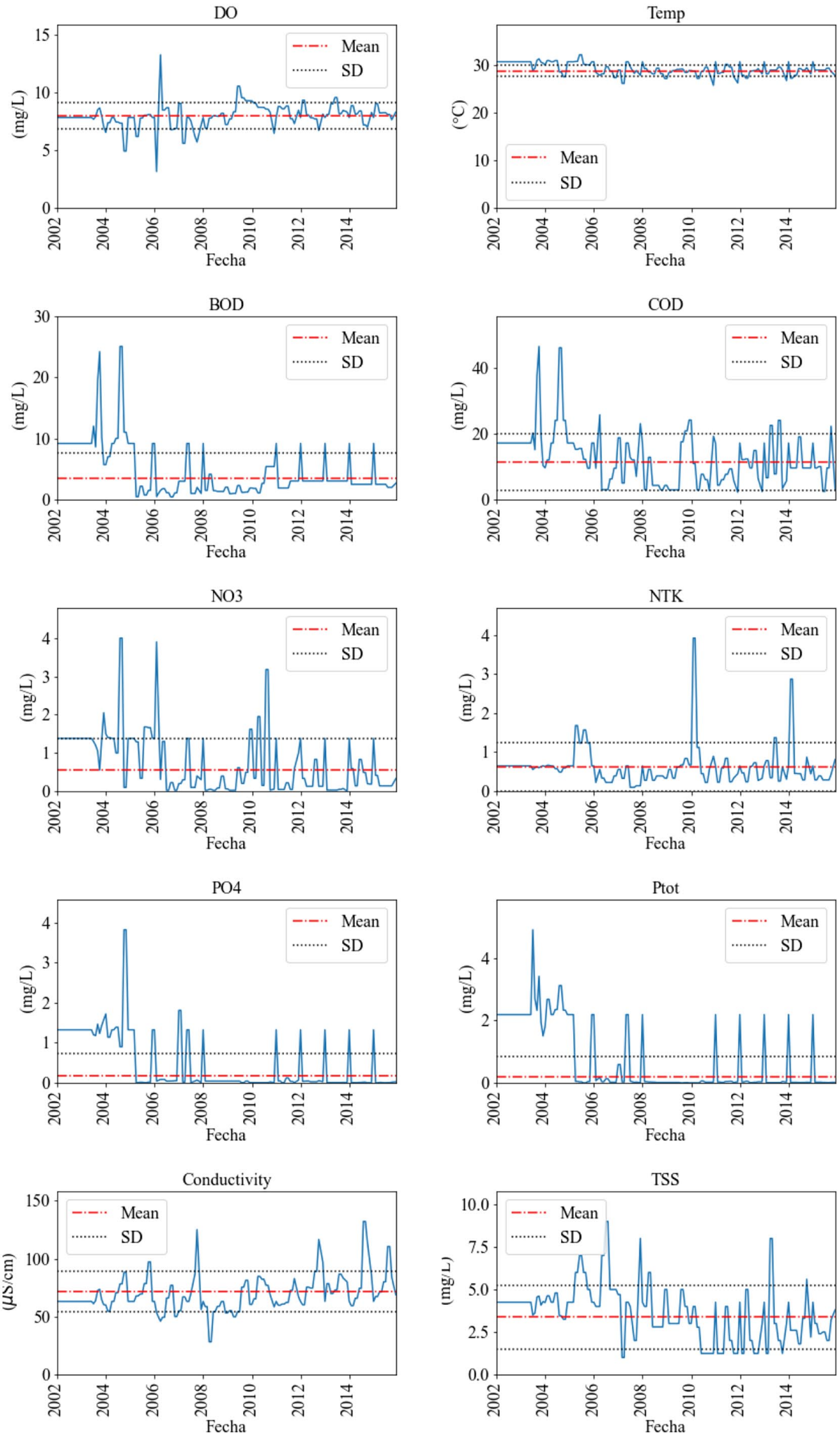

Figure 3. Mean and standard deviation of water quality time series dam.

water retention. And solar radiation plays a decisive role in surface water conditions and is related to dissolved oxygen.

The dissolved oxygen levels measured at the water surface changed depending on the season for the maximum and the minimum values. For the dry season, the extreme values were 3.16 and 12.55 mg/L. For the wet season, the extreme values were 4.7 and 13.3 mg/L. The latter can be because the organic matter load is low, and turbulence is present during the rainy season. These findings are consistent with the low chemical oxygen demand (COD) observed, however there were not significant differences between seasons (p value = .16). For chemical Oxygen Demand (COD) revealed values between 1.0 and 11.8 mg/L for the dry season, and between 1.9 and 14.8 mg/L for the wet season.

For Biochemical Oxygen Demand (BOD) values were low, the measurements for the dry season and between 0.4 and 4.2 mg/L for the wet season, 0.6 and 8.9 mg/L, show higher loads for the rainy season, possibly due to sediment transport. All this, considering that COD and BOD are indicators of organic pollution of water bodies, in this case, low (Rangel-Peraza et al., 2009).

The study area has a low impact on anthropic activities. One possible reason that explains the observed loads can be the geological formation of the area. Finally, according to (Chapman, 2021), the observed COD concentrations in uncontaminated surface water are typically below 20 mg/L. Similar values for BOD are less than 2 mg/L for uncontaminated water. Based on these criteria, it implies that the water of "La Miel" hydropower has a low content of biodegradable organic matter and, in general, high dissolved oxygen levels.

The same behavior is observed for nitrogenous and phosphorous compounds for TKN. Measurements in the dry season are higher than in the rainy season. Average values between 0.26 and 0.28 mg/L and between 0.57 and 0.78 mg/L are observed for the dry season. The average values for the wet season are between 0.18 and 0.22 mg/L and between 0.52 and 0.57 mg/L. NO_3, PO_4, and TP describe similar behaviors. All mentioned nutrient values may be associated with good water quality and low nutrient content, in the reservoir, due to low TSS levels below 8.9 mg/L on average. Thus, it can be concluded that it is an oligotrophic reservoir.

These results showed that the year's season has a clear impact on water quality and DO throughout the reservoir. The latter indicates that the amount of rain or solar brightness that the system receives is vital to the reservoir's water quality and plays a vital role in dissolved oxygen behavior. On the other hand, the water quality variables that seem to have the greatest impact on dissolved oxygen are BOD, COD, and water temperature. In a smaller proportion, the nutrient is the most influential phosphate.

Temporal variation of water quality data and hydrometeorology

The normality test results showed that none of the water quality time series described a normal distribution, which is why subsequent analyses were performed using nonparametric tests.

Four of the five monitoring stations showed changes points in water temperature. However, none of these changes coincided with the stations. In the case of nitrogenous and phosphorous compounds, all their chemical forms showed change points, but they did not coincide between compounds. On the other hand, dissolved oxygen and BOD showed change points in all stations and coincided with four of the five stations. From the latter, it can be inferred that changes in the processes, that affect them, affect the entire system.

On the contrary, COD showed change points in three out of five stations, but these changes did not coincide when the changes occurred. In addition, conductivity values showed that the change point at all sites and the date the change occurred were consistent with other water quality parameters, which, as expected, were previously influenced by other variables throughout the system. Finally, concerning water quality parameters, it is worth noting that the TKN did not show mean changes at any site. Therefore, this parameter was not affected by the processes that occurred in the reservoir throughout the years of measurement. In summary, the mass variables have a point of change and are related to changes in dissolved oxygen: water temperature, nutrients, BOD, and, a lesser extent, TSS.

Changes in hydro-meteorological variables, precipitation, flow, cloud cover, and evaporation occurred around the first half of 2005 and overlapped with the changes indicated by the water temperature at the dam site. Again, air temperature and relative humidity describe a change point, they are related to changes in dissolved oxygen.

Monitoring point E4 for water temperature showed an upward trend, and other monitoring points showed a downward trend.

Regarding nutrients, phosphorus compounds in all monitoring stations showed a downward trend, and nitrite in the study area also showed a downward trend. On the other hand, one station showed an increasing trend for nitrates (E3), and the rest of the monitoring site results showed a decreasing trend. On the contrary, ammonia showed a tendency in four stations (E4, E5, E6, and E7), describing an incremental tendency. All this allows us to infer that, although the amount of nutrients is low, the form of nitrogen available in water increasing is ammoniacal nitrogen, which, since it is not converted into nitrite to continue the nitrogen cycle, becomes a substance that generates pollution in the aquatic system.

COD and BOD showed decreasing trend effects in four stations (E4, E5, E6, and E7) and increasing trends in E3. However, dissolved oxygen showed an upward trend at all sampling stations. These results allowed us to presume that the surface water of the reservoir is evolving toward good aeration conditions. Conductivity shows similar behavior, as both parameters described dissolved solids. Conductivity shows an increasing trend in four sampling stations (E4, E5, E6, and E7). The latter behavior possibly occurs because this station is at the tail of the reservoir r(E3), so as they belong to different basins, the quality characteristics come from different conditions.

In contrast, TSS showed a trend in four (E3, E5, E6, and E7) of the five sampling points, and this trend was decreasing, so this fact allowed us to see more clearly that the increase in free ions, that cause the increase in conductivity, is mainly related to the basin geology and not to inputs from suspended solids.

In brief, the trend analysis revealed that, in all monitoring stations, there is a decreasing trend for the parameters related to the nutrients of the system. Likewise, the fact that dissolved oxygen shows an increasing trend for all monitoring stations allows us to infer that the evolution of the study system tends to remain as a simple oligotrophic system. The water quality variables showed the greatest change over time, climate variables will not change significantly.

In addition, it is worth mentioning that there is no trend in evaporation, air temperature, flow, and precipitation. The sun's brightness and relative humidity showed a downward trend. These variables correlate in time with some slight changes in dissolved oxygen.

To determine the significance of seasonality on water quality and dissolved oxygen time series. Seasonality plays a significant role in the water quality state within reservoirs (Rangel-Peraza et al., 2009). The results of the Mann-Whitney U test denote that described seasonality, with intense periods every 3 months for water temperature ammonia and SST, Total Nitrogen, Phosphorus, BOD, and PO_4, 6 months for dissolved oxygen DO and COD, and 12 months for conductivity. The Levene test results contrast with the findings made before, where pH, TSS, NO_3, and TP were significantly different among seasons.

For all hydroclimate variables, the cycles show 6-month and 1-year cycles, which vary between dry and rainy seasons. This shows its main relationship with nutrients and TSS.

At this point, it was possible to establish to temporally the processes affecting dissolved oxygen fluctuations. Considering BOD, COD, water temperature, air temperature, solar brightness, relative humidity, and precipitation are key parameters of DO.

Spatial variation of water quality e hydrometeorological data

The results of the Kruskal-Wallis H-test, spatial analysis considering the arms of the reservoir. Significant differences were measured in DO, conductivity, NO_3, TSS, and water temperature for the principal arm between stations E3 and E5. Likewise, significant differences were measured on the other arm of the reservoir between stations E4 and E6 for DO, TP, and Water Temperature. At last, tail monitoring points were assessed at the dam site. It was come to know that: between stations E3 and E7, significant differences in conductivity, NO_3, pH, TSS, and WT were observed. On the other hand, between stations E4 and E7, significant differences in ammonia, DO, TP, and water temperature were noted. These findings confirm that Water Temperature is significantly different among all monitoring stations. Furthermore, conductivity, NO_3, and TSS are significantly different through the "La Miel river" arm and dam site, as DO and TP are significantly different through the "Moro river" arm.

The results of the Kruskal-Wallis H-test, hydroclimatological spatial analysis considering the arms of the reservoir. Significant differences were not founded in the flow discharge, precipitation, solar brightness, humidity, and air temperature for the principal arm between stations E3 and E5. Moreover, no significant differences were measured on the other arm of the reservoir between stations E4 and E6 for all hydroclimatic variables. Tail monitoring points were assessed at the dam site. Between stations E3 and E7, significant differences in precipitation. And between E4 and E7 flow discharge is significantly different.

Using this technique, it was determined that the arm (E3) and the dam (E7) were the sites that better represented changes in water quality and DO. This implies that spatially important processes occur at E3 y E8.

Correlation of water quality and hydrometeorological seasonal data

In previous analyses, attempts were made to interpret the time series results and analyze spatial-temporal behavior that was important to predicting dissolved oxygen. A correlation analysis was performed to determine the direct relationship between water quality and hydro-meteorological data. Relationships which not demonstrated in the previous steps were done using this statistical technique, and for this analysis, the series was split between dry or wet seasons.

To gain an initial understanding of the complete monthly dataset relationships, Pearson, Kendall, and Spearman's methods were applied. Considering that the Spearman test is an appropriate tool for this type of assessment, as it targets non-normal data, allowing us to identify variables with better associations between them. The wet and dry season correlation matrices for all monitoring sites were produced.

For the analysis of the results, the correlations discovered between absolute values of 0.0 and 0.39 were considered weak or negligible. In contrast, the correlation between absolute values .4 and 1 is considered moderate or intense so the latter will be the correlation considered in the following analysis (Schober et al., 2018).

The study revealed that whether it is the dry season or rainy season, the variables with the highest correlation coefficient are water quality variables among themselves. So then, there is a specific relationship between hydro-meteorological variables and water quality variables. In addition, the correlation in the dry season is more significant than that in the rainy season.

Table 1. Delta Mean and Delta Standard Deviation of Imputation Techniques for Each Station.

METHOD	MEAN DELTA					STANDARD DEVIATION DELTA				
	E3	E4	E5	E6	E7	E3	E4	E5	E6	E7
Linear	0.9	0.11	0.3	0.4	0.2	1.24	1.64	1.1	1.46	1.28
Bayesian	0.1	0.18	0.25	0.04	0.14	1.34	1.66	0.89	1.48	1.18
KNN	0.56	0.5	0.5	0.26	0.45	1.16	1.45	0.95	1.37	1.13
Extra Tree	0.22	0.24	0.16	0.33	0.97	0.98	1.42	1.03	1.41	0.99
DTree	0.8	0.29	0.21	0.5	0.7	0.59	0.46	0.33	0.26	0.33

Moving on to a more specific analysis of the relationship between water quality and hydrometeorological variables, it was found that the most recurrent relationship in the sampling sites was between precipitation and conductivity for the dry and wet seasons, lacking only at the E3 site. This relationship is also noticed by Ricardo et al. (2016) and Zhang et al. (2018), described as a possible consequence of runoff and soil type around the reservoir. Precipitation is also associated with nitrogen and phosphorus compounds, especially during the dry season, as studied by Branco et al. (2019), showing that lower precipitation leads to increased nutrient concentrations due to lower dilution. In the same way, precipitation and air temperature were correlated with the water temperature at sites E6 and E7 behavior. These can be explained because these sites have the lowest surface water movement and the largest surface area in contact with the atmosphere. Lastly, another striking correlation was the one between solar brightness and PO_4 found for the dry season of stations E3 and E5, a link that was also found in Li et al. (2019), Yang et al. (2021) highlighting the impact of solar radiation on phosphorus concentration in continental water systems.

The results described in the previous analyses showed correlations between water quality variables and hydro-meteorological variables. It asserts at this point and with these statistical methods and taking into consideration seasonal variations that the hydro-meteorological variables to be considered for accurate predictions of dissolved oxygen data are precipitation, flow discharge, relative humidity, air temperature, and solar brightness. And water quality parameters: water temperature, BOD, COD, and PO_4.

Data imputation, filling the water quality series with missing data

The results of the comparison of the different gap-filling methods at the five stations are presented in Table 1. These results are shown, where the best is near zero.

Comparing the means between the initial series and the filled series, 0.06 for linear regression, 0.14 for Bayesian regression, 0.38 for ExtraTree regression, 0.45 for K-Nearest Neighbors, and 0.5 for Decision Tree regression. Shows better results for linear regression and Bayesian.

Comparing the standard deviation values between the initial series and the filled series, it is 1.34 for linear regression, 1.31 for Bayesian regression, 1.17 for Extra Tree regression, 1.21 for K-Nearest Neighbors, and 0.3 for Decision Tree regression. Shows better results for Decision Trees and Extra Trees.

Evaluation of the results of imputation techniques showed that MICE techniques estimated by machine learning methods achieved better results than mean, linear interpolation, or linear regression methods. And the estimator A Decision Tree regression was found to be a better estimator for the imputation process for missing data on water quality according to the metrics.

Principal component analysis (PCA) of water quality e hydrometeorological data

The results of applying the PCA in all water quality sampling stations revealed that it is possible to explain more than 70% of the accumulated variance of the data with only six principal components, which allowed for reducing the dimensionality from nine selected in the previous steps to six dimensions.

In addition, it was learned that the more significant variance at each monitoring point of the time series is due to the first principal component, which changed as a function of the location of the station in the reservoir, showing that for E3 (First tributary La Miel river), which is the most upstream monitoring point of the reservoir, the first principal component was composed of hydro-meteorological variables like precipitation, flow, and air temperature, for points E4 (First tributary Moro river) and E5 (second tributary La Miel river) The first component is a mixture of climate and water quality variables, such as precipitation, air temperature, PO_4, and water temperature.

On the other hand, for points E6 (Second tributary Moro River) and E7 (dam), the first principal component consists of water quality variables such as PO_4, NO_3, BOD, COD, and water temperature, as shown for stations E3 and E7 in Figure 4 Where, which displays the differences.

These results allow us to infer that the main drivers of the processes change depending on the location within the dam, being the upstream reservoir processes driven by meteorological variables, and as the location of the processes approaches

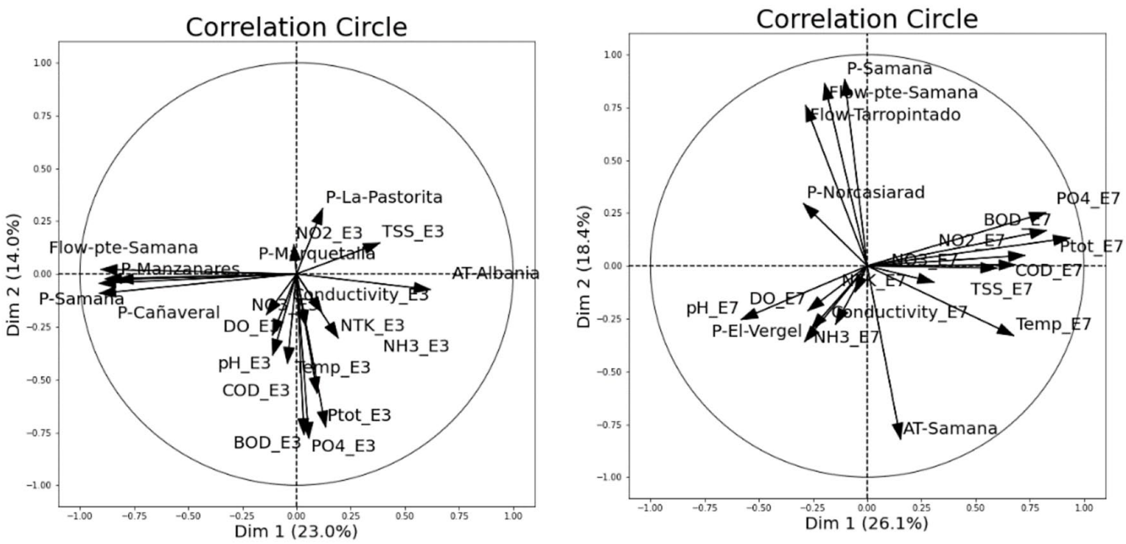

Figure 4. PCA circle plot for stations E3 and E7.

the dam point, water quality variables take center stage as the main drivers of the processes in the reservoir.

The results also show that, for all monitoring stations, the water quality variables associated with the first or second principal component, depending on the location of the monitoring point in the reservoir, are PO_4, BOD, COD, and NO_3. Likewise, precipitation and currents at the Samaná sampling were the hydro-meteorological parameters that had the most significant impact on the first or second principal component, depending on the location of the monitoring points in the reservoir. Another important aspect highlighted from the PCA results is the TSS itself, as a single principal component. Therefore, it can be concluded that the variance associated with SST is not related to any other climatic or water quality variables. Finally, it is observed (Figure 4) that the decrease in dissolved oxygen depends on the parameters of the first component for E3 and E7.

Dissolved oxygen prediction model based on vector autoregressive (VAR) technique

To perform VAR analysis to forecast dissolved oxygen in monitoring sites, E3 and E7, the PCA results were used as input to create the predictive model case, knowing that, at this point, all previous analyses allowed us to understand which variables were the main drivers of the reservoir process in Spatio-temporal context.

Two cases were implemented for the two monitoring sites, applying the variables from the first component of the PCA analysis and, in the other case, applying the variables from the first and second components. For site E3, Precipitation and Flow for Case 1 were selected, and for Case 2, the variables mentioned and BOD, COD, and PO_4 were selected. A site E7, the water temperature, BOD, COD, and PO_4 was selected for case 1, and for case 2, the variables mentioned as well as precipitation and air temperature were selected. Finally, after selecting the time series, splitting them into training and test

Table 2. Performance Indicators of VAR Predictive Model for Each Case.

	CASE 1	CASE 2	CASE 1	CASE 2
0	17.06	20.58	5.104	13.29
1	16.79	20.65	5.274	13.35
2	16.23	18.94	4.438	12.09
3	16.3	19.22	4.53	12.29
4	16.0	18.57	4.348	11.9
5	15.98	18.71	4.553	12.24
6	15.82	18.52	4.402	12.03
7	15.76	18.44	4.583	12.06
8	15.6	18.12	4.489	11.79
9	15.63	18.27	4.403	11.19
10	15.48*	18.01	4.411	10.49
11	15.56	18.15	4.3	10.17
12	15.67	18.05	3.806*	8.494*
13	15.63	17.79*	4.009	9.088
14	15.67	17.9	3.936	9.432

data, leaving 5 years as test data and the rest as training data, the optimal model order was selected using the Akaike Information Criterion (AIC) by equating the maximum lag order to 14, to select the right order of the VAR model, we iteratively fit increasing orders of VAR model and pick the order that gives a model with least AIC, as is shown in Table 2.

The Table 3 shows the performance metrics for the two cases at each site. It is shown that for station E3, and its case 1, an RMSE of 1.03 was found. For case 2, an RMSE of 2.04 was

Table 3. Performance Indicators of VAR Predictive Model for Each Case.

CASE 1 STATION E3		CASE 2 STATION E3		CASE 1 STATION E7		CASE 2 STATION E7	
FORECAST ACCURACY	DO	FORECAST ACCURACY	DO	FORECAST ACCURACY	DO	FORECAST ACCURACY	DO
MAPE	0.10	MAPE	0.21	MAPE	0.08	MAPE	0.21
ME	0.11	ME	0.52	ME	0.51	ME	0.32
MAE	0.79	MAE	1.66	MAE	0.68	MAE	1.69
MPE	0.03	MPE	0.04	MPE	0.06	MPE	0.03
RMSE	1.03	RMSE	2.04	RMSE	0.84	RMSE	2.09

obtained, which is why it is considered that both models had a good performance in predicting dissolved oxygen. However, it could be inferred that for this station, a model based only on hydro-meteorological variables obtains a better approximation than a model where, in addition to the hydro-meteorological variables, water quality variables are included. Additionally, monitoring site E7, Table 3. shows that for case 1, an RMSE of 0.84 was found, and for case 2, an RMSE of 2.09 was obtained.

As a result, although the metric results show that the model produces a good approximation, in this monitoring site, there is a significant difference between the performance metrics, indicating that Case 1 provides better prediction results than Case 2.

The behavior described by the performance metrics is also reflected in the graphs as presented in Figure 5, where for the monitoring site E3, the two cases present a similar forecast between them with minor differences. In case 2, the model better reproduces the variability of the system because it has more input variables but fails to improve the forecast compared to the results of case 1. Conversely, for monitoring site E7, it is observed that case 1 presents a better fit than case 2, showing for case 2 that the forecast values are above the expected mean values. Even though it well reproduces the variability of the series, in case 1, the variability also approaches the expected variability and, at the same time, yields values that are within what is expected.

The model evaluation demonstrated that DO can be predicted as a function of precipitation and flow discharge in the E3 station and terms of air temperature, water temperature BOD, COD, and PO_4 in the E7 station.

Discussion

This study used various statistical tests to analyze water quality and hydroclimatological aspects and to forecast dissolved oxygen on a monthly basis. Since the selection of an appropriate set of input variables from all possible input variables during The VAR model development is important for obtaining a high-quality model. Many of the described methods for input variable selection are based on heuristics, expert knowledge, statistical analysis, or a combination of these. However, although there is a well-justified need to consider input variable selection carefully, there is currently no consensus on how to accomplish this task (Ranković et al., 2012). For this reason, it is imperative to correctly predict using multivariate statistical techniques and to assess spatiotemporal water quality and hydroclimate variables to select them appropriately.

In this study, exploratory data analysis was used to aid what variables should be used. Doing so, allowed a better understanding of the underlying process occurring in the basin that affects reservoir water quality and DO.

The input variables analyzed in this paper were: Temperature, pH, total suspended solids, dissolved oxygen, conductivity, chemical oxygen demand, biochemical oxygen demand, ammonia, nitrite, nitrate, total Kjeldahl nitrogen, phosphate, Total phosphorus, as well discharge, precipitation, solar brightness, evaporation, relative humidity, cloudiness, and air temperature. The number of total variables was 20 for five water quality stations in the reservoir. Throughout the process, variables were refined, and concluding that only six were the most appropriate for the prediction of dissolved oxygen and that a monthly period in two (E3 and E7) of the five stations in the reservoir could well represent the spatial and temporal dynamics of dissolved oxygen in the tropical reservoir. It was concluded that DO can be predicted as a function of precipitation, flow, air temperature, water temperature, BOD, COD, and PO_4.

Furthermore, as mentioned in Chen et al. (2020) and Yaseen et al. (2018), while the selection of variable key parameters for prediction is essential, there is clearly a need for reservoir zoning in addition to this. The latter allows refinements to improve the performance of metrics and forecasts, as shown in case 2, where the performance did not improve. Identifying two dissolved oxygen dynamics related to the location of the reservoir stations, one in the tail E3 and the other in the reservoir vessel E7.

Other studies have found that water temperature, photosynthesis, respiration, and BOD (Rajwa et al., 2014; Rajwa-Kuligiewicz et al., 2015) seasonal variation (Lliev & Hadjinikolova, 2013), air temperature and nutrition (Dordoni et al., 2022), are variables that affect dissolved oxygen.

Referring to the original hypothesis, it was found that variables such as water temperature, organic matter decomposition, and COD were indeed important succeeded by phosphates to

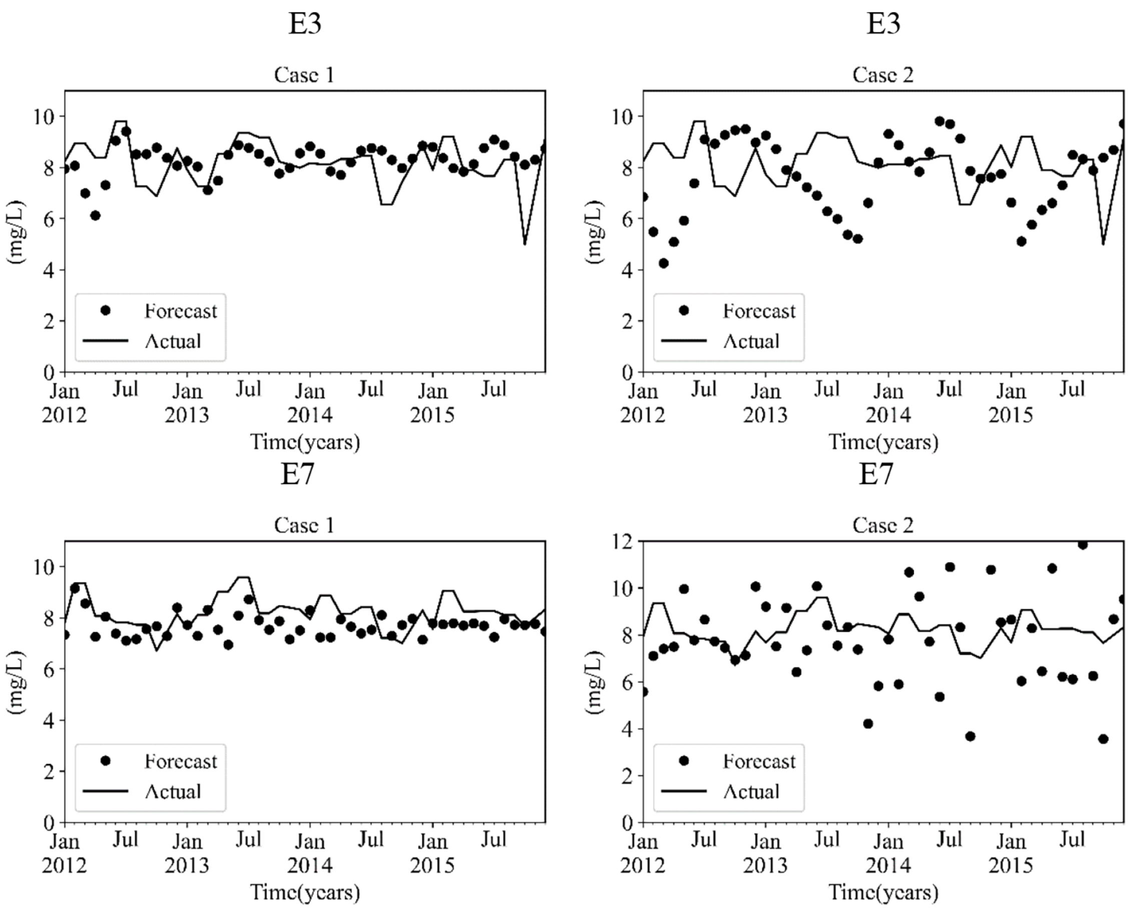

Figure 5. Case 1 and Case 2 VAR models for stations E3 and E7.

changes in dissolved oxygen concentrations. And that these are the correct inputs for the prediction of dissolved oxygen in a tropical reservoir. However, nitrogen compounds, conductivity, pH, and suspended solids did not show much correlation in this investigation.

It was previously known that different climatic conditions and temperatures were the basis for determining DO dynamics, and this study also involved precipitation and flow discharge as fundamental variables for predicting DO.

Although climatic variables and chemical interactions should indeed be considered together for the prediction of dissolved oxygen, they should also be explored for future work, other variables such as water levels, wind speed and direction, relationships with geologic conditions, soils, and biological processes, since in this study due to the scarcity of information in these areas it was not possible to carry them out. Likewise, consider changes at finer scales than monthly, for example, consider the dynamics of oxygen at an hourly resolution.

Compared to current models, the power of VAR is a systematic yet flexible approach to capturing complex real-world behavior and exhibits better predictive performance as well as the ability to capture the interwoven dynamics of time-series data (Aptech, 2021). For the VAR model, splitting the series into two groups, was sufficient to obtain high prediction rates.

In this study, it was evident that the VAR statistical tool has great potential, although its use is incipient in water quality analysis in reservoirs. For this reason, more research on this topic should be done. The results of this study can be compared with the results reported in the literature.

Future works will be important not only to the link between parameters but taking into consideration other statistical methods and do a comparison between knowing the Spatio-temporal processes that affect dissolved oxygen behavior in tropical reservoirs.

After all, accurate prediction of dissolved oxygen may provide a cost-effective solution to prevent water quality crises in a tropical reservoir. The models and methods presented here can be applied to dissolved oxygen prediction in other tropical ecosystems such as lakes and rivers.

Conclusion

This study applied a methodology that allows a determination of the behavior of hydro-meteorological and water quality variables in tropical reservoirs and establishes the relationship between water quality parameters and hydrometeorology to predict dissolved oxygen. Statistical tests and analysis showed a statistically significant influence of hydro-meteorological variables like precipitation, flow discharge, relative humidity, solar brightness, and air temperature over water quality parameters, especially with nutrients, and dissolved oxygen. These analyses also showed that, as the relationship between water quality and hydro-meteorological variables varied from site to site, therefore the behavior of water quality parameters is influenced by the area within the reservoir.

When applied to water quality data analysis and prediction, nonparametric procedures have several advantages over parametric procedures. Some of these advantages are: (1) prior transformations are not required, even when approximate normality could be achieved, (2) normality for water-quality data is not required, (3) comparisons are made between central values such as the median rather than the mean; and finally, (4) data below the detection limit can be incorporated without fabrication of values or bias. So, the information contained in less-than-values is accurately used, not misrepresenting the state of that information. Furthermore, its ability to handle missing values, outliers, and its capacity to update, makes it ideal for prediction.

Descriptive statistical analyses are performed by separating the time series according to the climatic seasons showing significant differences between the dry and wet seasons. Similarly, the correlation analysis results of each season show that the reservoir is affected by seasons, and the influence of meteorological variables in the dry season is more evident than that in the rainy season. Therefore, this funding increases the metrics for DO predictions.

Only the variables that are part of the first principal component are recommended for DO forecasting. The combination of more components creates noise that reduces the quality of forecasting.

Due to the spatial and temporal heterogeneity of water quality in a tropical reservoir, water quality monitoring should be designed to capture the temporal dynamics close to dams' inlets to predict dam water quality. This finding suggests that designing and maintaining effective reservoir water quality monitoring is key to sustainable management and prediction.

It is essential to begin analyzing and modeling the environmental impacts of large dams more holistically to better inform stakeholders and decision-makers on the balance between exploiting hydropower potential and maintaining critical natural resources.

Author Contributions

A.G.J.A designed this study, have made a substantial contribution to the concept of the article; analysis, or interpretation of data for the article; A.D.C. wrote the original draft, analysis, or interpretation of data for the article; E.V.J.A Revised it critically for important intellectual content; approved the version to be published; M.J.L.J and P.E.C.C reviewed the several versions of the manuscript.

Data Availability Statement

The datasets generated during and/or analyzed during the current study are available from the corresponding author on reasonable request.

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) received no financial support for the research, authorship, and/or publication of this article.

ORCID iDs

Alzate-Gómez Juliana Andrea https://orcid.org/0000-0002-9029-6989

Montoya-Jaramillo Luis Javier https://orcid.org/0000-0002-5832-2219

REFERENCES

Abdi, H., & Williams, L. J. (2010). Principal component analysis. *Wiley Interdisciplinary Reviews: Computational Statistics*, *2*, 433–459.

Alsaleh, M., & Abdul-Rahim, A. (2022). Moving toward sustainable environment: The effects of hydropower industry on water quality in EU economies. *Energy & Environment*, *33*, 1304–1325. https://doi.org/10.1177/0958305x211039452

Alsaleh, M., & Abdul-Rahim, A. S. (2021). Do global competitiveness factors effects the industry sustainability practices? Evidence from European hydropower industry. *Journal of Cleaner Production*, *310*, 127492. https://doi.org/10.1016/j.jclepro.2021.127492

Aptech. (2021). https://www.aptech.com/blog/introduction-to-the-fundamentals-of-vector-autoregressive-models/

Barbarossa, V., Schmitt, R. J. P., Huijbregts, M. A. J., Zarfl, C., King, H., & Schipper, A. M. (2020). Impacts of current and future large dams on the geographic range connectivity of freshwater fish worldwide. *Proceedings of the National Academy of Sciences of the United States of America*, *117*(7), 3648–3655. https://doi.org/10.1073/pnas.1912776117

Branco, C. W. C., Leal, J. J. F., Huszar, V. L. D. M., Farias, D. D. S., Saint'Pierre, T. D., Sousa-Filho, I. F., de Palermo, E. F. D. A., Guarino, A. W. S., Gomes, A. R., & Kozlowsky-Suzuki, B. (2019). New lake in a changing world: The construction and filling of a small hydropower reservoir in the tropics (Rio de Janeiro, Brazil). *Environmental Science and Pollution Research*, *26*(35), 36007–36022. https://doi.org/10.1007/s11356-019-06665-y

Calamita, E. (2020). *Modelling the effects of large dams on water quality in tropical rivers*. ETH Zürich.

Chapman, D. (2021). *Water quality assessments*. CRC Press.

Chen, K., Chen, H., Zhou, C., Huang, Y., Qi, X., Shen, R., Liu, F., Zuo, M., Zou, X., Wang, J., Zhang, Y., Chen, D., Chen, X., Deng, Y., & Ren, H. (2020). Comparative analysis of surface water quality prediction performance and identification of key water parameters using different machine learning models based on big data. *Water Research*, *171*, 115454. https://doi.org/10.1016/j.watres.2019.115454

Chen, P., Li, L., & Zhang, H. (2015). Spatio-temporal variations and source apportionment of water pollution in Danjiangkou Reservoir Basin, Central China. *Water*, *7*(12), 2591–2611. https://doi.org/10.3390/w7062591

Dalu, T., & Wasserman, R. J. (2018). Cyanobacteria dynamics in a small tropical reservoir: Understanding spatio-temporal variability and influence of environmental variables. *The Science of the Total Environment*, *643*, 835–841. https://doi.org/10.1016/j.scitotenv.2018.06.256

de Oliveira, K. L., Ramos, R. L., Oliveira, S. C., & Christofaro, C. (2021). Spatial variability of surface water quality in a large Brazilian semiarid reservoir and its main tributaries. *Environmental Monitoring and Assessment*, *193*(7), 1–15. https://doi.org/10.1007/s10661-021-09194-9

Dordoni, M., Seewald, M., Rinke, K., Schmidmeier, J., & Barth, J. A. C. (2022). Novel evaluations of sources and sinks of dissolved oxygen via stable isotopes in lentic water bodies. *The Science of the Total Environment*, *838*, 156541. https://doi.org/10.1016/j.scitotenv.2022.156541

Fu, L., & Gan Wang, Y. (2012). Statistical tools for analyzing water quality data. *Water quality monitoring and assessment*. (pp. 143–168). InTech Open.

Gyimah, R. A. A., Gyamfi, C., Anornu, G. K., Karikari, A. Y., & Tsyawo, F. W. (2021). Multivariate statistical analysis of water quality of the Densu River, Ghana. *International Journal of River Basin Management*, *19*(2), 189–199. https://doi.org/10.1080/15715124.2020.1803337

Hurvich, C. M., & Tsai, C. L. (1991). Bias of the corrected AIC criterion for underfitted regression and time series models. *Biometrika*, *78*(3), 499–509. https://doi.org/10.1093/biomet/78.3.499

Hwang, S.-J., Kim, K., Park, C., Seo, W., Choi, B.-G., Eum, H. S., Park, M.-H., Noh, H. R., Sim, Y. B., & Shin, J.-K. (2016). Hydro-meteorological effects on water quality variability in Paldang Reservoir, confluent area of the South-Han River-North-Han River-Gyeongan Stream, Korea. *Korean Journal of Ecology and Environment*, *49*(4), 354–374. https://doi.org/10.11614/ksl.2016.49.4.354

Jerves-Cobo, R., Forio, M. A. E., Lock, K., Van Butsel, J., Pauta, G., Cisneros, F., Nopens, I., & Goethals, P. L. M. (2020). Biological water quality in tropical rivers during dry and rainy seasons: A model-based analysis. *Ecological Indicators*, *108*, 105769. https://doi.org/10.1016/j.ecolind.2019.105769

Killingtveit, Å. (2019). Hydropower. In T. Letcher (Ed.), *Managing Global Warming* (pp. 265–315). Academic Press.

Ling, T. Y., Gerunsin, N., Soo, C. L., Nyanti, L., Sim, S. F., & Grinang, J. (2017). Seasonal changes and spatial variation in water quality of a large young tropical reservoir and its downstream river. *Journal of Chemistry*, *2017*, 1–16. https://doi.org/10.1155/2017/8153246

Liqoarobby, R., Suparman, S., & Fadilah, Y. K., (2021). Aqueous systems of dissolved oxygen in reservoir. *E3S Web of Conferences*, *249*, 03015. https://doi.org/10.1051/e3sconf/202124903015

Liu, C., Pan, C., Chang, Y., & Luo, M. (2021). An integrated autoregressive model for predicting water quality dynamics and its application in Yongding River. *Ecological Indicators*, 133, 108354. https://doi.org/10.1016/j.ecolind.2021.108354

Li, X., Guo, M., Duan, X., Zhao, J., Hua, Y., Zhou, Y., Liu, G., & Dionysiou, D. D. (2019). Distribution of organic phosphorus species in sediment profiles of shallow lakes and its effect on photo-release of phosphate during sediment resuspension. *Environment International, 130, 104916*. https://doi.org/10.1016/j.envint.2019.104916

Li, X. Z., Chen, Z. J., Fan, X. C., & Cheng, Z. J. (2018). Hydropower development situation and prospects in China. *Renewable and Sustainable Energy Reviews*, *82*, 232–239. https://doi.org/10.1016/j.rser.2017.08.090 (August 2017.

Lliev, I., & Hadjinikolova, L. (2013). *Seasonal and vertical dynamics of the water temperature and oxygen content in Kardzhali reservoir.* Agricultural Science and Technology. https://www.researchgate.net/publication/305781610

Lobato, T. C., Hauser-Davis, R. A., Oliveira, T. F., Silveira, A. M., Silva, H. A., Tavares, M. R., & Saraiva, A. C. (2015). Construction of a novel water quality index and quality indicator for reservoir water quality evaluation: A case study in the Amazon region. *Hydrology Journal*, *522*, 674–683. https://doi.org/10.1016/j.jhydrol.2015.01.021

Mamun, M., Kim, J. Y., & An, K. G. (2021). Multivariate statistical analysis of water quality and trophic state in an artificial dam reservoir. *Water*, *13*(2), 186. https://doi.org/10.3390/w13020186

Mamun, M., Lee, S. J., & An, K. G. (2018). Temporal and spatial variation of nutrients, suspended solids, and chlorophyll in Yeongsan watershed. *Journal of Asia-Pacific Biodiversity*, *11*(2), 206–216. https://doi.org/10.1016/j.japb.2018.02.006

Marcé, R., Rodríguez-Arias, M. À., García, J. C., & Armengol, J. (2010). El Niño Southern Oscillation and climate trends impact reservoir water quality. *Global Change Biology*, *16*(10), 2857–2865. https://doi.org/10.1111/j.1365-2486.2010.02163.x

Marques, É. T., Gunkel, G., & Sobral, M. C. (2019). Management of tropical river basins and reservoirs under water stress: Experiences from northeast Brazil. *Environments*, *6*(6), 62. https://doi.org/10.3390/environments6060062

Nadarajah, S., Wijenayake, W. M. H. K., & Amarasinghe, U. S. (2019). Influence of hydrology on water quality and trophic state of irrigation reservoirs in Sri Lanka. *Lakes and Reservoirs: Research and Management*, *24*(3), 287–298. https://doi.org/10.1111/lre.12283

Rajwa, A., Bialik, R. J., Karpiński, M., & Luks, B. (2014). Dissolved oxygen in rivers: Concepts and measuring techniques. In R. Bialik, M. Majdański, & M. Moskalik (Eds.), *Achievements* (pp. 337–350). Springer.

Rajwa-Kuligiewicz, A., Bialik, R. J., & Rowiński, P. M. (2015). Dissolved oxygen and water temperature dynamics in lowland rivers over various timescales. *Journal of Hydrology and Hydromechanics*, *63*(4), 353–363. https://doi.org/10.1515/johh-2015-0041

Rangel-Peraza, J. G., De Anda, J., González-Farias, F., & Erickson, D. (2009). Statistical assessment of water quality seasonality in large tropical reservoirs. *Lakes and Reservoirs: Research and Management*, *14*(4), 315–323. https://doi.org/10.1111/j.1440-1770.2009.00412.x

Ranković, V., Radulović, J., Radojević, I., Ostojić, A., & Čomić, L. (2012). Prediction of dissolved oxygen in reservoirs using adaptive network-based fuzzy inference system. *Journal of Hydroinformatics*, *14*(1), 167–179. https://doi.org/10.2166/hydro.2011.084

Ricardo, P., Medeiros, P., Cavalcante, G. H., Brandini, N., & Knoppers, B. A. (2016). Inter-annual variability on the water quality in the lower São Francisco River (NE-Brazil). *Acta Limnologica Brasiliensia*, *28*, 5. https://doi.org/10.1590/s2179-975x3515

Schober, P., Boer, C., & Schwarte, L. A. (2018). Correlation coefficients: Appropriate use and interpretation. *Anesthesia-Analgesia Research Society, 126*(5), 1763–1768.

Semensatto, D., Labuto, G., Zorzal-Almeida, S., & McRae, D. V. (2021). Spatio-temporal changes in water quality in the Guarapiranga reservoir (São Paulo, Brazil): Insights from a long-term monitoring data series. *Environmental Monitoring and Assessment*, *193*(7), 1–15. https://doi.org/10.1007/s10661-021-09167-y

Silva, S. N., & Castillo, J. Á. D. (2021). An approach of the hydropower: Advantages and impacts. A review. *Journal of Energy Research and Reviews, June*, 10–20. https://doi.org/10.9734/jenrr/2021/v8i130201 del.

Vaidya, R. A., Molden, D. J., Shrestha, A. B., Wagle, N., & Tortajada, C. (2021). The role of hydropower in South Asia's energy future. *International Journal of Water Resources Development*, *37*(3), 367–391. https://doi.org/10.1080/07900627.2021.1875809

Varol, M. (2020a). Spatio-temporal changes in surface water quality and sediment phosphorus content of a large reservoir in Turkey. *Environmental Pollution*, *259*, 113860. https://doi.org/10.1016/j.envpol.2019.113860

Varol, M. (2020b). Use of water quality index and multivariate statistical methods for the evaluation of water quality of a stream affected by multiple stressors: A case study. *Environmental Pollution*, *266*, 115417. https://doi.org/10.1016/j.envpol.2020.115417

Vega, A. S., Lizama, K., & Pastén, P. A. (2018). Water quality: Trends and challenges. In *G. Donoso (Ed.),* Water policy in Chile (pp. 25–51). Springer.

Vilas, M. P., Marti, C. L., Oldham, C. E., & Hipsey, M. R. (2018). Macrophyte-induced thermal stratification in a shallow urban lake promotes conditions suitable for nitrogen-fixing cyanobacteria. *Hydrobiologia*, *806*(1), 411–426. https://doi.org/10.1007/s10750-017-3376-z

Weirich, C. A., Robertson, D. M., & Miller, T. R. (2019). Physical, biogeochemical, and meteorological factors responsible for interannual changes in cyanobacterial community composition and biovolume over two decades in a eutrophic lake. *Hydrobiologia*, *828*(1), 165–182. https://doi.org/10.1007/s10750-018-3810-x

Winton, R. S., Calamita, E., & Wehrli, B. (2019). Reviews and syntheses: Dams, water quality and tropical reservoir stratification. *Biogeosciences*, *16*(8), 1657–1671. https://doi.org/10.5194/bg-16-1657-2019

Woldeab, B., Beyene, A., Ambelu, A., Buffam, I., & Mereta, S. T. (2018). Seasonal and spatial variation of reservoir water quality in the southwest of Ethiopia. *Environmental Monitoring and Assessment*, *190*(3), 163. https://doi.org/10.1007/s10661-018-6527-4

Yang, X., Yuan, J., Yue, F.-J., Li, S.-L., Wang, B., Mohinuzzaman, M., Liu, Y., Senesi, N., Lao, X., Li, L., Liu, C.-Q., Ellam, R. M., Vione, D., & Mostofa, K. M. G. (2021). New insights into mechanisms of sunlight- and dark-mediated high-temperature accelerated diurnal production-degradation of fluorescent DOM in lake waters. *Science of The Total Environment, 760, 143377.* https://doi.org/10.1016/j.scitotenv.2020.143377

Yaseen, Z., Ehteram, M., Sharafati, A., Shahid, S., Al-Ansari, N., & El-Shafie, A. (2018). The integration of nature-inspired algorithms with least square support vector regression models: Application to modeling river dissolved oxygen concentration. *Water*, *10*(9), 1124. https://doi.org/10.3390/w10091124

Zarfl, C., Berlekamp, J., He, F., Jähnig, S. C., Darwall, W., & Tockner, K. (2019). Future large hydropower dams impact global freshwater megafauna. *Scientific Reports*, *9*(1), 1–10. https://doi.org/10.1038/s41598-019-54980-8

Zarfl, C., Lumsdon, A. E., Berlekamp, J., Tydecks, L., & Tockner, K. (2015). A global boom in hydropower dam construction. *Aquatic Sciences*, *77*(1), 161–170. https://doi.org/10.1007/s00027-014-0377-0

Zhang, R., Cuartas, L. A., de Castro Carvalho, L. V., Reis Deusdará Leal, K., Mendiondo, E. M., Abe, N., Birkinshaw, S., Samprogna Mohor, G., Seluchi, M. E., & Nobre, C. A. (2018). Season-based rainfall-runoff modelling using the probability-distributed model (PDM) for large basins in southeastern Brazil. *Hydrological Processes*, *32*, 2217–2230. https://doi.org/10.1002/hyp.13154