

Article

Monitoring Water Quality of Valle de Bravo Reservoir, Mexico, Using Entire Lifespan of MERIS Data and Machine Learning Approaches

Leonardo F. Arias-Rodriguez ¹, Zheng Duan ^{1,2,*}, Rodrigo Sepúlveda ³,
Sergio I. Martinez-Martinez ⁴ and Markus Disse ¹

¹ Hydrology and River Basin Management, Technical University of Munich, Arcisstrasse 21, 80333 Munich, Germany; leonardo.arias@tum.de (L.F.A.-R.); markus.disse@tum.de (M.D.)

² Department of Physical Geography and Ecosystem Science, Lund University, Sölvegatan 12, S-223 62 Lund, Sweden

³ Department of Sanitary and Environmental Engineering, National Autonomous University of Mexico, Ciudad Universitaria, Mexico City 04510, Mexico; rtsh@unam.mx

⁴ Center of Design and Construction Sciences, Autonomous University of Aguascalientes, Av. Universidad 940, 20131 Aguascalientes, Mexico; simartin@correo.uaa.mx

* Correspondence: zheng.duan@tum.de or zheng.duan@nateko.lu.se; Tel.: +49-89-289-23258

Received: 17 April 2020; Accepted: 14 May 2020; Published: 16 May 2020



Abstract: Remote-sensing-based machine learning approaches for water quality parameters estimation, Secchi Disk Depth (SDD) and Turbidity, were developed for the Valle de Bravo reservoir in central Mexico. This waterbody is a multipurpose reservoir, which provides drinking water to the metropolitan area of Mexico City. To reveal the water quality status of inland waters in the last decade, evaluation of MERIS imagery is a substantial approach. This study incorporated in-situ collected measurements across the reservoir and remote sensing reflectance data from the Medium Resolution Imaging Spectrometer (MERIS). Machine learning approaches with varying complexities were tested, and the optimal model for SDD and Turbidity was determined. Cross-validation demonstrated that the satellite-based estimates are consistent with the in-situ measurements for both SDD and Turbidity, with R^2 values of 0.81 to 0.86 and RMSE of 0.15 m and 0.95 nephelometric turbidity units (NTU). The best model was applied to time series of MERIS images to analyze the spatial and temporal variations of the reservoir's water quality from 2002 to 2012. Derived analysis revealed yearly patterns caused by dry and rainy seasons and several disruptions were identified. The reservoir varied from trophic to intermittent hypertrophic status, while SDD ranged from 0–1.93 m and Turbidity up to 23.70 NTU. Results suggest the effects of drought events in the years 2006 and 2009 on water quality were correlated with water quality detriment. The water quality displayed slow recovery through 2011–2012. This study demonstrates the usefulness of satellite observations for supporting inland water quality monitoring and water management in this region.

Keywords: turbidity; secchi disk depth; trophic state; remote sensing; gaussian processes regression; support vector machines; random forest regression; inland waters

1. Introduction

Inland waters such as lakes, reservoirs, and rivers are important water resources; they regulate climate and hydrological flows; support soil formation, nutrient cycling, and pollination; enable food production and water supply; and provide aesthetic conditions, cultural services, and recreation [1]. Therefore, their protection is vital and their water quality must be assured. Based on a water body's intended purpose, the parameters of water quality must achieve certain standards. Monitoring these

parameters allows for the detection of sudden harmful changes and establishes opportunities for implementing preventive and restorative measures to recover healthy conditions.

Conventional methods for monitoring water quality, so-called point sampling methods, determine water quality indicators by collection of samples directly from the field and their analysis in a laboratory. However, these traditional methods have important constraints. First, in-situ sampling is laborious and requires extensive time to cover large areas, which increases costs. Furthermore, investigating the spatial and temporal trends of water quality parameters in large waterbodies is not feasible due to limited sample points, which do not accurately represent the complete status of the water surface [2]. Moreover, the topography can play an important role in restricting access to some areas of water bodies, and errors may still exist in field and laboratory measurements.

From the conventional parameters, Secchi Disk Depth (SDD) is a common measurement of water transparency, which can be evaluated using the approach developed by Pietro Angelo Secchi [3]. In this method, a white and black disk disappears inside a water column at a certain water depth; therefore, SDD is commonly measured as a numerical variable for distance as meters (m). Additionally, SDD is inversely related to the average amount of organic and inorganic materials along the water column [3] and is a practical indicator of trophic conditions [4–7]. SDD is employed to study relative nutrient loads and particle contents as well as visually track the flow of suspended detritus and the displacement of sediment influxes from tributary streams and rivers. In a eutrophication process, the water is affected by algae saturation and other aquatic plants due to excess nutrients. The remaining matter of such aquatic plants depletes oxygen from the water, causing oxygen-dependent life to deplete. Fertilizers from fields, human sewage and animal wastes are the main sources of such nutrient loads. Inland waters with high eutrophication are characterized by poor water quality, which can potentially threaten human health and constrain usage [8]. Turbidity in a water body is caused by suspended chemical and biological particles via scattering and absorption of light. This water quality parameter has implications for both water safety and aesthetics regarding drinking-water supplies [9]. To measure Turbidity, an electronic turbidimeter in nephelometric turbidity units (NTU) is employed, which requires water samples.

Through Remote Sensing (RS), it is possible to acquire information from the Earth's surface. This can be achieved over different scales, regions, and periods of time. Information concerning inland waters can be utilized to retrieve physical and biochemical parameters of the water using the spectral reflectance measured by RS sensors in several bands of the electromagnetic spectrum. This procedure has helped to develop water quality monitoring with RS in the recent decades. The successful history of water quality monitoring applications has been detailed in studies by Dekker [10], Cheng [6], Odematt [11], Matthews [12] and Hansen [13]. During its period of operation (2002–2012) and beyond, the Medium Resolution Imaging Spectrometer (MERIS) provided by the European Space Agency (ESA) has been successfully employed to monitor inland waters [11,14,15], and its archives are considered a rich source of data for water research [16]. MERIS has outstanding advantages for monitoring water quality, including full spatial resolution of 260×290 m, 15 visible (VIS) and near-infrared (NIR) bands, as well as an extensive web-enabled image archive (2002–2012) [17]. MERIS also enables temporal analysis applications with its three-day temporal resolution.

The current ESA satellites, the Sentinel-2 and 3 are under operation since 2015 and research of former events must be addressed using archived imagery from prior sensors. Increased use of MERIS for inland water quality analysis was visible during the years before the launch of the Sentinels (2010–2015) with a decrease after 2015. During those years, MERIS was an important source of data for inland water quality research considering the quantity of available sensors (Figure 1).

Currently, archived data from MERIS contain valuable information in many fields that has yet to be processed, as is the case of applications in inland water quality. This situation opens the need to further research and increases the limited number of studies analyzing the complete MERIS imagery, which leads to a broader scope and better representation of study cases. Despite the advantages of monitoring water quality using RS techniques, these methodologies are not yet broadly applied by

water resources and policy managers [18]. Further research using RS with field measurements is therefore necessary to evidence its benefits and potential in protecting water resources and promptly detecting potential hazards.

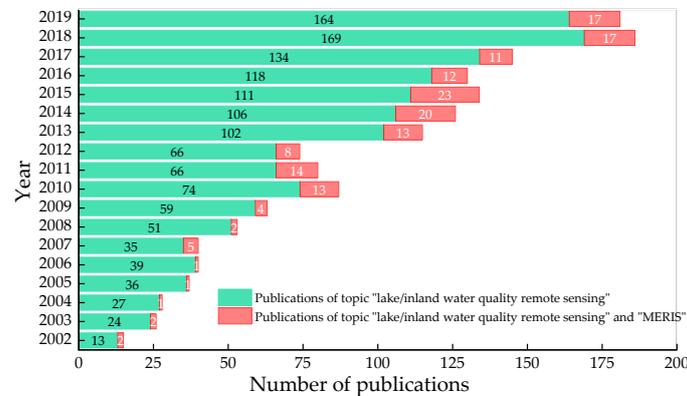


Figure 1. Number of publications (2002–2019) listed in Web of Science for the terms “lake water quality remote sensing” and “inland water quality remote sensing” and the further addition of the word “MERIS” to both terms (literature published January 2020).

To estimate water quality parameters, various approaches have been developed based on the relationship between RS reflectance and optical characteristics of water constituents. In general, these methods can be broadly divided in empirical, semi-analytical and machine learning methods [16,19,20] with further sub-classifications among them. Empirical methods employ band and band ratio as coefficients to establish relationships. Frequently, several combinations of input values are evaluated through comparison of error metrics looking for the best fit. The result is a regression algorithm that can be applied to the images of the study area and dates of interest to estimate spatial and temporal variations in water quality parameters. This approach is, to some extent, easily applicable when there are enough in-situ and RS data; however, its application is limited to the studied water body and cannot be generalized to other regions due the variations of atmosphere and water composition [21]. If an empirical method selects bands or band ratios based on the knowledge of the physical characteristics of water components that may affect specific wavelengths, then it is classified as a semi-empirical method. On the other hand, analytical approaches use the knowledge of physics of light. They define the specific and necessary parameters of a model on the base of the optical properties of the water and atmosphere also known as inherent optical properties (IOPs). The modelling process produces theoretical absorption and backscattering values which can be separated to estimate optically active water quality constituents using an inverse equation [22]. The semi-analytical approaches implement in addition in-situ measurements, to define the parameters of the inverse equation and to reduce the difficulty of modelling complex waters. These models can derive several water quality parameters simultaneously [23] and they can be applicable to other regions different from the original study area. However, their use require various large spectral datasets for training and computing, as well as considerable fieldwork in the regional context to develop robust algorithms [16,24–26].

The machine learning (ML) techniques in the RS field were introduced to overcome the complex association between the RS data and the water constituents present in the parametric regression models as least-squares or multiple regression [27,28]. A standard procedure of regression approaches is the linear regression (LR) which is a statistical method that allows to observe the relationship between two constant numerical variables. It can be classified as an empirical approach in the water quality modelling field or as a ML basic algorithm for data analysts. During this paper we will define LR as a ML approach for further comparison. Another widely applied algorithm is the Support Vector Regression (SVR) [29–31] which is a supervised learning method trained with labeled data. As the support vector machines (SVM) used for classification, SVR algorithm includes the C hyperparameter and the kernel trick. It is useful with a limited number of samples because of its good generalization

ability. Also common, the random forest is an adaptable procedure useful for classification and regression (RFR). It employs subsets of the data which are averaged for enhancement of predictive capacity, control of over-fitting and handling of large datasets. RFR has been implemented to several RS applications including water resources [32,33]. A more recent method for estimation of biophysical parameters, the Gaussian Processes Regression (GPR) [34,35] provides a Bayesian approach to learn regression problems using kernels [34]. It has lately been applied for water quality parameters retrieval from remotely sensed data with high performance in its estimations [36–38]. When lacking spectral field measurements, the modelling process in ML algorithms can be implemented with less data and different assumptions for their training stage in comparison with radiative transfer models [39]. For water quality studies, the ML approaches analyzing completely and intensively the MERIS imagery of lakes and reservoirs are sparse due to their recent development and the previous operating timeframe of MERIS. Thus, these studies using novel algorithms could take a greater advantage of the legacy of this sensor increasing the usage of such rich source of data.

The Valle de Bravo reservoir in central Mexico is a multipurpose waterbody that provides drinking water to the metropolitan area of Mexico City. It is also the most important reservoir in the country for recreational activities such as tourism, fishing, and sailing [40]. Most of the previous research in Valle de Bravo is limited due to the use of conventional measuring methods. These constraints are in temporal and spatial scale due to scarce measuring stations or impossibility of continuous sampling campaigns due to the time and costs demands. In the last two decades, studies by Olvera-Viascan [41,42], Ramirez-Garcia [43], Nandini [44] and Figueroa-Sanchez [45] analyzed the reservoir and expressed concern about its trophic state. Some authors ultimately offered strategies for improving the reservoir's water quality and reducing the presence of toxic cyanobacteria, pointing as main contributors of the degradation of water quality the scarce wastewater management, the agricultural runoff and the surrounding ecosystems factors. In Mexico, there is a national monitoring water quality program under the "Sistema Nacional de Información del Agua" (SINA) with measurement stations (around 5000) distributed across the inland waters of the country, with five fixed stations located in Valle de Bravo. However, these five stations and the measured water quality parameters can likely be insufficient for accurately representing the spatial and temporal scale of harmful events in the water, especially in cases of eutrophication or harmful algae. Moreover, the measured parameters are limited to control the pollution from wastewaters as biochemical oxygen demand (BOD), chemical oxygen demand (COD), total suspended solids (TSS) and fecal coliforms.

The major installation of monitoring stations began in 2012, which indicates there is no comprehensive water quality data about the reservoir prior to this time. As a result of the limited monitoring capacity in the reservoir, there is an increasing demand for continuous monitoring of water quality parameters in the region, especially for such important reservoirs which supply drinking water to great urban areas where millions of people reside. Furthermore, a lack of knowledge of the water quality conditions may persist in the years prior the establishment of monitoring programs. Similar limitations can likely be present in transition and developing economies either because they lack extensive survey networks or because these networks are of recent implementation and therefore no previous data can be acquired. Standard procedures which may help to overcome these limitations are needed and they are of particular benefit for such regions to improve their water quality monitoring capacity. One way to overcome these restrictions is using available resources in combination with current analysis techniques. This leads to clarification of the variations of inland water quality in recent years, together with the implications of natural and anthropogenic hazards in water quality detriment. In this way, overall conclusions of the water quality could be achieved even in lack of extensive field or surface spectral data measurements.

Concerns about the water quality conditions and quantity of the water supply raised for the urban region of Mexico City during the previous decade [46–48] and until today regulation in the supply is commonly applied. As the most important drinking water source for the region, the protection and continuous monitoring of Valle de Bravo reservoir is an essential duty. The understanding of disruptive

events that occurred in previous years may lead to a clear comprehension of the current situation and to avoid formerly occurred threats. To contribute to such needs in the region, this paper analyzes the water quality parameters variations in the Valle de Bravo reservoir for a period of 11 years, prior to the launch of current sensors used for water quality monitoring. Water quality measurements from sampling campaigns conducted in 2010 and RS data from matchup MERIS imagery are used as input for ML algorithms. From the analysis, the best model is selected and applied to the complete MERIS data archives (2002–2012) to examine the spatial and temporal variations of water quality. This could contribute to future research on water quality of lakes and reservoirs where limited monitoring is implemented but the resources to increase its investigation exist. The main objectives of this research are focused firstly, on the development and evaluation of a methodology based on ML approaches using MERIS spectral data and physically water quality data measured in Valle de Bravo. Secondly, on the analysis of the spatial and temporal dynamics of the water quality in the reservoir during the entire MERIS operation timeframe (11 years), which also complements the scarce number of studies taking advantage of the complete MERIS imagery. Also, as the ML techniques are commonly based on different assumptions, a further and continuous evaluation of their predicting capacity is necessary to determine which approach may be better to evaluate and map water quality in the region using MERIS data. Finally, this study also contributes to increase the use of ML techniques in the analysis of water quality parameters in lakes and reservoirs, which are of recent implementation. The results of this work will complement the existing literature for water quality evaluation in the reservoir. To our best knowledge, no comprehensive integration of in-situ water quality measurements and RS techniques has yet been implemented to monitor water quality in this region for such amount of time or using ML approaches. This study aims to fill this research gap for the intended water quality parameters. The findings of this work are expected to provide guidance to policy makers on incorporating satellite RS into national in-situ water quality control program.

2. Study Area

Valle de Bravo (Figure 2) is a tropical ($19^{\circ}11'N$, $100^{\circ}09'W$) and high-altitude (1780 masl) reservoir. It has a surface area of 18.55 km^2 and an average depth of 20 m, with a storage capacity of $418.25 \times 10^6 \text{ m}^3$ and a drainage basin of 547 km^2 [49–51]. The precipitation measures 836 mm year^{-1} , while the mean annual evaporation measures $1620 \text{ mm year}^{-1}$ [52].

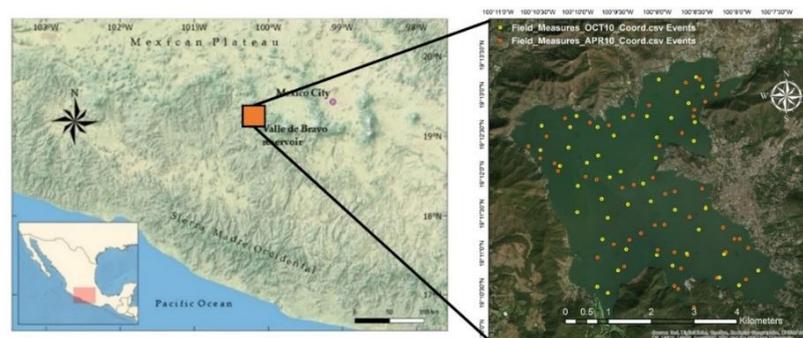


Figure 2. Location of the study area and water measurement locations in Valle de Bravo. Orange points indicate field measurements taken on 25 April 2010; yellow dots indicate samples taken on 02 October 2010.

The reservoir receives water discharge primarily from the Amanalco River and also from smaller tributaries (Molino, González and Carrizal rivers), as well as sewage outlets from adjacent towns (Valle de Bravo and Avándaro). The Amanalco and Carrizal rivers were formerly detected as the main sources of physical and chemical pollution causing bacterial presence due to incoming nutrient loads of phosphorus and nitrogen from their discharges [41]. The reservoir provides most of the drinking water to Mexico City (21 million inhabitants) through the Cutzamala System—a 330 km network

of open channels, tunnels, and aqueducts that brings the drinking water of neighboring reservoirs toward the capital. This network system supplies 25% (19 m³/s) of the city's drinking water demand by pumping it a total of 1100 m from its lowest to highest point in Mexico City (2250 masl) [53]. The water balance of such system depends on extractions and injections of water from other neighbor reservoirs, such as Los Colorines, El Bosque, and Tuxpan, located at the east of the reservoir. Critical periods of volume storage occurred during 2006, 2009 and again in 2013, where the reservoir lost 50% of its maximum capacity due to water scarcity. These situations posed serious issues regarding tourism, water quality [46], and supply of drinking water in Mexico City [47], which escalated the establishment of extraordinary tariffs to control demand.

3. Materials

3.1. Field Campaigns

The in-situ data were acquired as part of the research program IN107710 “Water quality monitoring using remote sensing” funded by the National Autonomous University of Mexico (UNAM) through the “Support program to research and innovation technology”. Sampling campaigns were performed on 25 April and 2 October 2010. Weather conditions were considered optimal with no rain or cloud coverage in the study area. A total of 96 samples (50 on April and 46 on October 2010) were collected and analyzed. SDD was measured in the field campaign and Turbidity under lab conditions. The SDD was measured using a standard 20-cm-diameter acrylic disk divided in black and white quarters. For the Turbidity measurements, the collected water samples were kept in containers with ice and transported to the lab facilities of the Sanitary Engineering Department at the UNAM. The following day, the Turbidity was measured in a Hach® 2100N device.

3.2. MERIS Satellite Data

MERIS full-resolution Level 1P products were collected from 2002–2012 with the MERIS FRS extraction tool (<https://merisfrs-merci-ds.eo.esa.int/merci/welcome.do>) with the selection criteria of low cloud coverage and period of time (image at the mid or end of each month) to ensure a fixed time reference in the analysis and allow significant changes to be visible. For model development, remote sensing reflectance (Rrs) was obtained from the processing of Level 1P MERIS products with acquisition dates on 27 April 27 and 3 October 2010 with ±2 days to the respective field measurement dates (25 April and 2 October 2010) (Table 1), ensuring the same water quality conditions in the reservoir.

Table 1. MERIS imagery corresponding to the field campaign data.

Product Name	Acquisition	Field Campaign
MER_FRS_1PPBCM20100427_171020_000000172089_00012_42651_0001	27 April 2010	25 April 2010
MER_FRS_1PPBCM20101003_171308_000000142093_00284_44927_0001	3 October 2010	2 October 2010

The two images present low cloud coverage (9–11%) and were processed using ESA's SNAP© software for Sentinel products, which is suitable for MERIS image analysis. The MERIS Level 1 Radiometric Processor was applied for SMILE correction, equalization and radiometric recalibration. Geometrical correction was applied using the MERIS Orthorectification Processor. Adjacency effect was corrected using the Improve Contrast over Ocean and Land processor (ICOL) [54]. For atmospheric correction and retrieval of Rrs the Case-2 Regional CoastColour (C2RCC) Processor was implemented which is based on inversion of radiative transfer and bio-optical models using neural networks [55]. The remote sensing reflectance was selected as output instead of the water leaving reflectance. With training ranges of 0.016–43.18 mg m⁻³ of chl-a, C2RCC stands as an adequate processor for the MERIS products used in this study which provides Rrs values and retrieval of inherent optical properties concentrations. The processor is described in detail in Doerffer and Schiller [55]. For model development, to test the effect that different processing levels have in the final retrievals, two different

datasets (DS1 and DS2) of Rrs were produced. The DS1 avoids the adjacency processor, as it was seen it modifies considerably the reflectance values, then the Rrs was taken directly from the C2RCC after atmospheric correction; the DS2 does include all the above described corrections. The entire image processing procedure is shown in Figure 3.

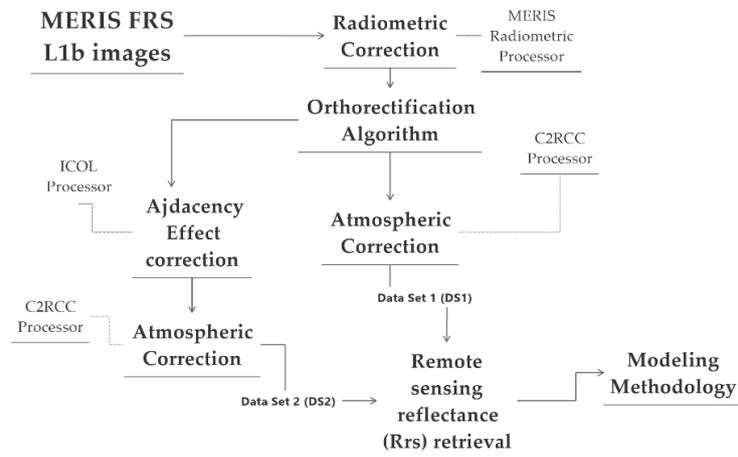


Figure 3. Image processing flowchart, including atmospheric correction, optional adjacency effect correction and remote sensing reflectance (Rrs) retrieval.

The Valle de Bravo reservoir (and most inland waters) was masked as land for the processor pixel-expression detection, thus the default-pixel expression was removed allowing the algorithm to process the complete scene. Additionally, the atmospherically corrected reflectance was retrieved as Rrs. All other processing parameters were used as default. The retrieved 12 Rrs bands with wavelengths (in nm) are: b1(412.69), b2(442.56), b3(489.88), b4(509.81), b5(559.69), b6(619.60), b7(664.57), b8(680.82), b9(708.32), b10(753.37), b12(778.40), b13(864.87). The Rrs data of each measurement location and the in-situ measurements of SDD and Turbidity were used as input base for model development. The reflectance values taken from each pixel correspond to one coordinate in the image, which is already an average value of the pixel area (260 × 290 m).

4. Methods

Different regression algorithms were evaluated to develop the predictive model: linear regression (LR), random forest regression (RFR), support vector regression (SVR) and Gaussian processes regression (GPR). Their accuracy was further compared with cross-validation to retrieve R^2 and RMSE, selecting the best model accordingly. All the regression analysis was produced using the open-source resources of the Scikit-learn library in a Python environment. Hyperparameter tuning results for each algorithm are shown in Section 4.5.

4.1. Linear Regression (LR)

LR is a standard procedure used in many studies since decades [6,7,56] and until recently [57]. It fits a linear model with coefficients to minimize the residual sum of squares between the observed targets in the dataset and the targets predicted by the linear approximation [58]. Its procedure allows relative straight-forward predictions and has been utilized in absence of spectral field measurements as is the case of this study. The regression analysis followed the general form expressed as:

$$y = \beta_0 + \beta_1x_1 + \beta_2x_2 + \dots + \beta_nx_n \quad (1)$$

where y is the selected water quality parameter, x refers to the value of the reflectance of a MERIS band, and β is the coefficient band obtained from the multiple regression. Similar approaches can be observed in the studies of Härmä and Kloiber [59,60] and more recently from Garaba, Toming or Alikas [61–63].

4.2. Random Forest Regression (RFR)

The Random Forest algorithm has been proved as an effective ML algorithm for classification and regression in many fields, including water quality monitoring [64,65]. Regression trees model non-linear relationships in data between predictors and response variables but with likely problems of overfitting. Random forest introduces randomness into individual regression trees to solve this problem [66]. The forest is composed of decision trees with different subset features and added flexibility, as bootstrap sampling from the dataset. Each tree is therefore trained with a random vector sampled independently with the same distribution, leading to a generalization of the error for the forest. The result is an increased accuracy using the mean of individual predictions of trees who acted as learners [67]. The algorithm of random forest for regression [68] is constructed for $b = 1$ up to $b = B$ trees, then a bootstrap sample of size N from the training data is selected. After, m random variables from the initial p should be selected as well, the best variable/split-point among the m chosen, and the node split into two daughter nodes for each terminal node of the tree (T_b). The ensemble of trees is retrieved as the computed average of such B trees to make predictions in the form:

$$\hat{Y}(x_i) = \frac{1}{B} \sum_{b=1}^B T_b(x_i) \quad (2)$$

Hyperparameters needed to be optimized for RFR are the $n_estimators$ that specifies the number of trees in the forest, the max_depth which sets the maximum depth of the tree, the $min_samples_split$ which is the minimum number of samples required to split an internal node and the $min_samples_leaf$ which is the minimum number of samples required to be at a leaf node.

4.3. Support Vector Regression (SVR)

The support vector regression (SVR), the regression version of the support vector machine (SVM) algorithm, is a well-positioned ML algorithm that has been applied for water quality studies [39] and its use is considered a standard procedure in ML evaluations. It is known for its good generalization capability, particularly with a limited number of samples [69]. The algorithm looks at the extremes of the datasets and draws a decision boundary defined as a hyperplane near those extreme points, establishing a frontier which best segregates between the classes of data. This is done with the aid of separation lines known as support vectors which are defined as data points that the margin pushes up against all points that are close to the opposing class. The SVR algorithm gives greater importance to the support vectors. With the hyperplanes, the SVR can be used in multidimensional datasets. For multidimensional data, a function is used to overcome linearity and transform the data into a high dimensional space but at a higher computational demand. Therefore, a Kernel function, a function that takes vectors as inputs in the original space and returns the dot product of the vectors in the feature space, is implemented. For SVR, linear, polynomial, gaussian radial basis, or hyperbolic sigmoid functions are common. For the regression formulation, consider a set of training points, $\{(x_1, z_1), \dots, (x_l, z_l)\}$, where $x_i \in R^n$ is a feature vector and $z_i \in R^1$ is the target output. Under given parameters $C > 0$ and $\varepsilon > 0$, the standard form of support vector regression is:

$$\min_{\omega, b, \xi, \xi^*} \frac{1}{2} \omega^T \omega + C \sum_{i=1}^l \xi_i + C \sum_{i=1}^l \xi_i^* \quad (3)$$

subject to

$$\omega^T \phi(x_i) + b - z_i \leq \varepsilon + \xi_i \quad (4)$$

$$z_i - \omega^T \phi(x_i) - b \leq \varepsilon + \xi_i^* \quad (5)$$

$$\xi_i, \xi_i^* \geq 0, \quad i = 1, \dots, l \quad (6)$$

where $C > 0$ is the regularization parameter [58,70]. In this study the radial basis function (rbf) kernel function was adopted. Hyperparameters needed to be optimized for SVR are the C parameter that acts as a penalty measure of the term and the $gamma$ parameter which is the kernel coefficient for types rbf, poly, and sigmoid.

4.4. Gaussian Processes Regression (GPR)

The GPR is a non-linear kernel method that establishes a relation between the input and the output variables, in this case, the spectral bands of MERIS and the field-measured water quality parameters, respectively. It has been applied in water quality parameters predictions with successful results and its use starts to be common in evaluation of ML approaches [36,37,71]. The main objective is to describe a distribution over functions using a Gaussian process specified by its mean and covariance function. The mean $m(x)$ and the covariance function $k(x, x')$ of a real process $f(x)$ are defined as:

$$m(x) = \mathbb{E}[f(x)], \quad (7)$$

$$k(x, x') = \mathbb{E}[(f(x) - m(x))(f(x') - m(x')))] \quad (8)$$

Defining the Gaussian process as:

$$f(x) \sim \mathcal{GP}(m(x), k(x, x')) \quad (9)$$

where usually the mean function is considered to be zero [34].

To produce predictions in a multidimensional space, the GPR uses diverse kernel types. In this study we selected the *rbf* kernel together with a noise white kernel function. Hyperparameters needed to be optimized for GPR are the $alpha$ which is a value added to the diagonal of the kernel matrix during fitting; larger values correspond to increased noise level in the observations and the $n_starts_optimizer$ which is the number of restarts of the optimizer for finding the kernel's parameters which maximize the log-marginal likelihood.

4.5. Hyperparameter Tuning

The hyperparameters used on the ML algorithms play a vital role in the performance of the developing models. Fitness and error behavior are affected depending on the assigned values and thus, hyperparameter tuning is a critical and challenging step in the development of the models. In this study, a 12-fold cross-validation was implemented using a GridSearch on the relevant hyperparameters. The optimal values are selected from the dataset with better performance in error metrics and presented in Table 2. The remaining hyperparameters of each model used default values.

Table 2. Hyperparameters setting and results of cross-validation with GridSearch for the ML algorithms.

Method	Hyperparameter	GridSearch Values	SDD Result	Turb. Result
LR	-	-	-	-
RFR	n_estimators	1, 10, 50, 100, 200, 500, 1000, 1500, 2000	1	10
	min_samples_leaf	0.1, 0.5, 1, 5, 10	1	1
	min_samples_split	2, 5, 10, 50, 100	10	2
	bootstrap	True, False	True	True
	max_depth	2, 4, 10, 20, 50, 100, None	50	20
SVR	C	0.0001, 0.001, 0.005, 0.0075, 0.1, 0.5, 1, 5, 10, 15, 20, 50, 100, 1000	1000	1000
	gamma	0.0001, 0.001, 0.01, 0.1, 1, 5, 10, 100, 1000	1000	1000
GRP	alpha	0.0001, 0.001, 0.0045, 0.0055, 0.0080, 0.01, 0.1, 1, 10	0.0045	1
	n_restarts_optimizer	0, 1, 2, 4, 8, 10, 12, 16, 20, 32, 64	2	0

4.6. Model Evaluation

To determine the most relevant MERIS bands as input for the algorithms, all possible combinations of the 12 MERIS bands were determined through the implementation of a power set (PS) as follows:

$$PS(b) = 2^b \quad (10)$$

where b is 12, the number of MERIS bands, with a total of 4096 possible band combinations per each dataset. The evaluation of each possible combination was assessed with a 12-fold cross-validation as matching number of folds for the available data (96 samples). Conventional proportions for training and validating are in the range of 70–30% or 80–20% when having enough data. With limited data, a leave one out cross-validation (LOOCV) could be applied, which evaluates all the available data except for one value. In this work we set an intermediate proportion between both above approaches, with a 12-fold cross validation, we settle for the middle between 25–20% and the extreme case of LOOCV; which brings us to a 12.5–10% of validation size. We did not consider evaluating further cross-validation proportions to consider it out of the scope of this work. Finally, the dataset was divided into training (88 samples, 91.67% of the total) and validation sets (8 samples, 8.33% of the total) for the cross-validation. This procedure ensured extensive validation of the dataset and avoided skew results due to random sampling. The error metrics controlling the performance were the R^2 and RMSE:

$$RMSE(y, \hat{y}) = \sqrt{\frac{1}{n_{samples}} \sum_{i=0}^{n_{samples}-1} (y_i - \hat{y}_i)^2} \quad (11)$$

$$R^2(y, \hat{y}) = 1 - \frac{\sum_{i=0}^{n_{samples}-1} (y_i - \hat{y}_i)^2}{\sum_{i=0}^{n_{samples}-1} (y_i - \bar{y})^2} \quad (12)$$

The validations were dependent on the number of bands used as input. To allow an analysis of the spectral sensitivity, all the possible combinations using only 1 band were evaluated and the band with best error metrics determined; similarly, this process was repeated for all the combinations of 2 bands, 3 bands, until 12 bands. On each validation the R^2 and RMSE were calculated to determine the optimal number of bands and its specific wavelength. The entire approach was applied to all the ML algorithms. When the best conditions (type and number of bands) were found for each model, a further comparison among them was performed using its best resources with several cross-validations. The model with the best metrics was selected for a posterior multitemporal analysis of MERIS data. The workflow diagram of this methodology is shown in Figure 4.

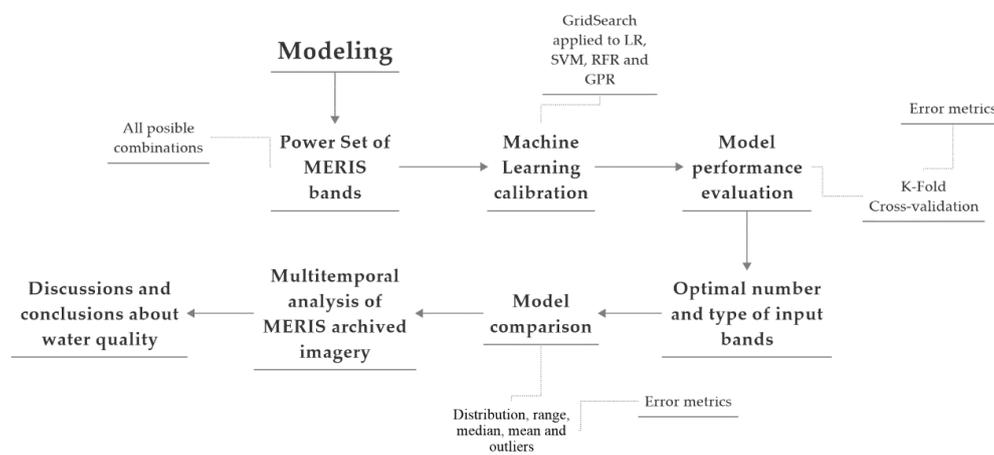


Figure 4. Overview of the methodology used in this study.

5. Results

5.1. In-Situ Measurements

The two sampling campaigns were performed during rainy (April) and dry (October) seasons. With this, the two predominant conditions of the seasons of the year were acquired. This contributed to the enrichment of the developed models. The statistical properties of the data for the SDD exhibit a mean of 1.36 m with a maximum of 2.03 m and minimum of 0.72 m. The Turbidity mean was 8.2 NTU and the maximum and minimum values were 13 NTU and 4.5 NTU, respectively. The standard deviation measured 0.38 m for the SDD and 3.1 NTU for the Turbidity.

5.2. Spectral Sensitivity

The feature engineering is of major importance in ML model development. The main objective is to provide higher accuracy and robust results. The spectral sensitivity behavior of the tested algorithms is shown in Figure 5.

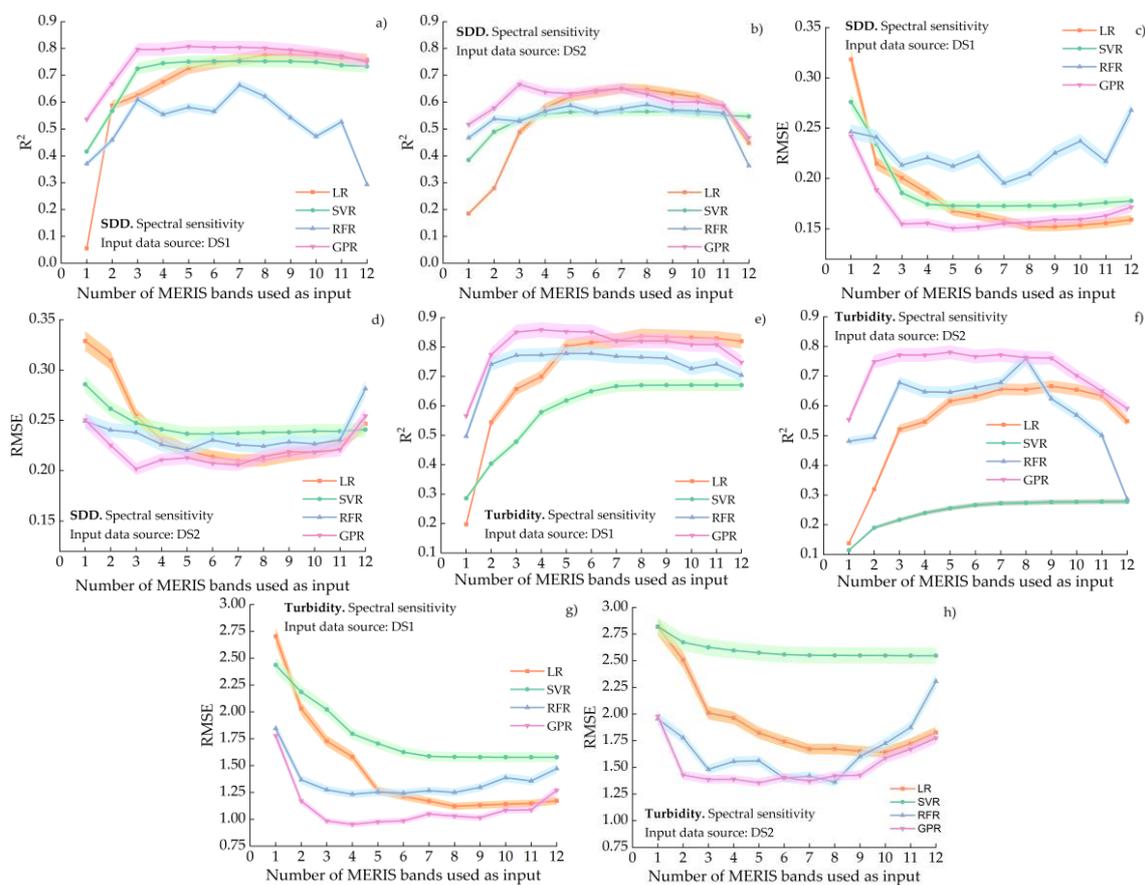


Figure 5. Spectral sensitivity of the ML models for different number of MERIS bands and Rrs datasets. R^2 and RMSE are displayed as error metrics. The horizontal axis represents the number of used MERIS bands, not the specific MERIS channel. Dataset origin is indicated in the graph. (a,b) show R^2 and (c,d) RMSE for SDD. Similarly, (e–h) for Turbidity. The filling area of each model stands for the error bars.

The 12 bands of MERIS imply a rigorous assessment due to the many possibilities of different combinations as input data for intended algorithms, which also implies high computational demand. The process is challenging due to the different nature and characteristics of the ML algorithms. To identify the optimal type and number of MERIS spectral bands required for both SDD and Turbidity, we rely on the error metrics evaluated with a rigorous 12-fold cross-validation combined with the

appraisal given by the PS. For DS1, differences exist within the models with the addition of spectral bands, especially for the RFR and LR (Figure 5a,c,e,g). But most important, the high collinearity and correlation of the spectral bands is present in all the ML algorithms. This is visible for the interval of 1 to 3 bands, with the high increase of accuracy and minimization of the error. This behavior is present in all the models but more constant in the GPR and less in the RFR. For both water quality parameters, SVR and GPR require only 3 bands to perform satisfactory and constant in R^2 (Figure 5a,e), with no major improvement with the addition of more bands. LR performs satisfactorily ($R^2 > 0.70$) when using 6 bands or more. On the other hand, the RFR behaves inconsistently depending on the number of features added. Its maximum performance is reached with 8 bands. The turning point occurring when using 3 bands is similar for SVR and GPR, however, for SDD the SVR and the GPR start a small decrease in accuracy and error after using 5 bands. For Turbidity, the improvement is high from 1 to 3 bands and then the tendency is constant for GPR. RFR behaves constant after using 2 bands but its performance remains constant at $R^2 \approx 0.75$. SVM does not reach values of the R^2 higher than 0.7 and LR shows high performance when using more than 5 bands with its peak at 8 bands. Accordingly, the RMSE is lower for GPR in all cases (Figure 5c,g). For DS2, a remarkably similar behavior is seen in the spectral sensitivity but with an important decrease in the performance of all models in R^2 and RMSE (Figure 5b,d,f,g). High collinearity is also present in high degree in all the ML models but in lesser extend in RFR. Satisfactory results are not achieved for SDD nor R^2 or RMSE and the GPR stands as the best model for both water quality parameters. For Turbidity GPR and RFR perform with better error metrics (Figure 5f,h) than SDD (Figure 5b,d) but not improving the use of DS1 (Figure 5e,g). The best results of each algorithm with different DS and water quality parameters are shown in Table 3. In Table 4 the highest error metrics of each algorithm and their optimal number and type of bands are presented and they all are product of the DS1 as result of the evaluation.

Table 3. Best combination and number of MERIS bands results of cross-validation. Coefficient of determination (R^2) and root mean square error (RMSE) are shown for the ML methods.

SDD								
Model	LR		SVR		RFR		GPR	
Dataset	DS1	DS2	DS1	DS2	DS1	DS2	DS1	DS2
R^2	0.78	0.65	0.75	0.57	0.66	0.59	0.81	0.67
RMSE (m)	0.15	0.21	0.17	0.24	0.2	0.22	0.15	0.2
Turbidity								
Model	LR		SVR		RFR		GPR	
Dataset	DS1	DS2	DS1	DS2	DS1	DS2	DS1	DS2
R^2	0.84	0.67	0.67	0.28	0.78	0.76	0.86	0.78
RMSE (NTU)	1.12	1.64	1.58	2.55	1.24	1.36	0.95	1.35

Table 4. Best combination and number of MERIS bands results of cross-validation. Coefficient of determination (R^2) and root mean square error (RMSE) are shown for the ML methods. These combinations belong to the DS1 as result of the evaluation.

Model	SDD			Turbidity		
	Band Combination	R^2	RMSE	Band Combination	R^2	RMSE
LR	b1, b3, b4, b5, b6, b7, b8, b9, b10	0.78	0.15	b1, b2, b3, b4, b5, b6, b7, b8	0.84	1.12
RFR	b1, b2, b4, b5, b6, b8, b10	0.66	0.20	b2, b5, b8, b9, b10, b13	0.78	1.24
SVR	b3, b4, b5, b6, b8	0.75	0.17	All Bands	0.67	1.58
GPR	b4, b5, b6, b7, b8	0.81	0.15	b2, b5, b12, b13, b14	0.86	0.95

The minimum number of bands required for SDD and Turbidity retrieval with relatively good accuracy (however not the best) are recognized as 2 for GPR, 3 for SVR (SDD) and RFR, and 5 for LR. It is visible that increasing the input bands of MERIS does not significantly improve the fitness or minimize the error of the prediction. As computational demand often increases when adding more bands to validation procedures and evaluation of further data, this result provides valuable information on how to improve the efficiency of modelling.

5.3. Model Performance

The results from spectral sensitivity allow a standardized evaluation among the ML models using the optimum number and type of bands determined for each algorithm (Table 4), ensuring each model performs under its best conditions. The random sampling of training and testing datasets has a strong influence in the results retrieved for each individual training. Thus, to retrieve a representative set of results and avoid atypical responses, we executed random runs of the models to yield a dataset of 120 predicted values. The results of this process are displayed in Figure 6.

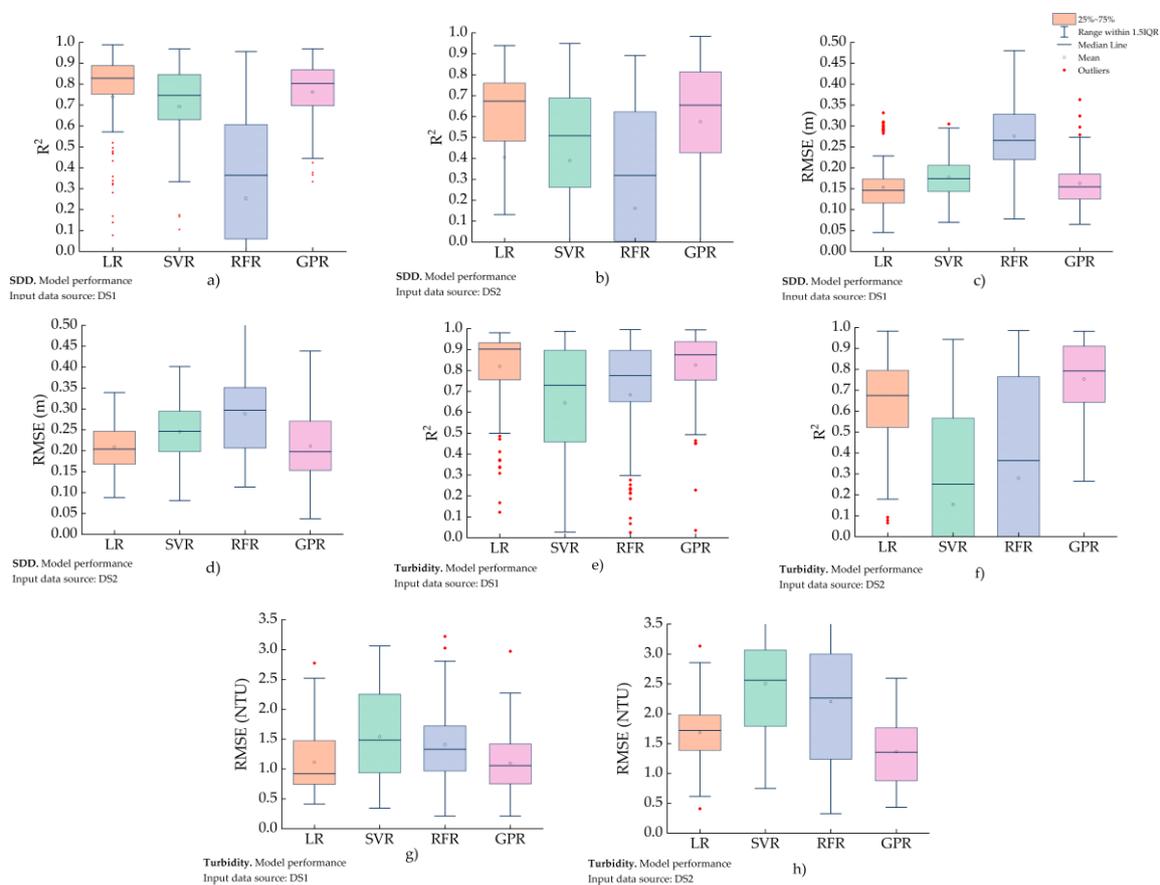


Figure 6. R^2 and RMSE distribution of random repetitions (120 values) between the predicted and measured SDD and Turbidity. The dataset origin is indicated in the graph. (a,b) show R^2 and (c,d) RMSE for SDD, respectively. Similarly, (e–h) for Turbidity.

Similar to spectral sensitivity results, the DS1 outperforms the DS2 in all the models and error metrics. In DS1, GPR ($R^2 = 0.762$, RMSE = 0.163) and LR ($R^2 = 0.739$, RMSE = 0.153) perform better and more robust for SDD followed by SVM ($R^2 = 0.693$, RMSE = 0.177) and RFR ($R^2 = 0.253$, RMSE = 0.276) (Figure 6a,c). For Turbidity, GPR performs better and surpasses the other models ($R^2 = 0.826$, RMSE = 1.099). It is important to note that RFR acts extremely poor in robustness for SDD and SVM for Turbidity. GPR and LR perform similarly for both water quality parameters. The main differences are seen on the RMSE where GPR acts more robust (Figure 6e,g). From these results, it is

clear that GPR is more stable to the random sampling processes. For DS2, R^2 and RMSE models produced values that are more spread and less robust, following the tendency of the spectral sensitivity (Figure 6b,d,f,h). The best results for SDD are produced using the GPR ($R^2 = 0.58$, $RMSE = 0.21$) as well as Turbidity ($R^2 = 0.75$, $RMSE = 1.36$), however, these metrics are still lower than the ones of DS1. In Table 5 a summary of the mean results of error metrics is shown. In general, using the DS1, GPR and LR achieved satisfactory and similar performances, which indicates they are good options for water quality parameters retrieval. SVM performed better for SDD than Turbidity and the opposite behavior is seen in RFR. From these results, GPR and LR are the potential methods for retrieval of SDD and Turbidity using MERIS spectral data in this study.

Table 5. Cross-validation mean results from 120 samples produced with the best combinations and number of MERIS bands. Coefficient of determination (R^2) and root mean square error (RMSE) are shown for the ML methods.

SDD								
Model	LR		SVR		RFR		GPR	
Dataset	DS1	DS2	DS1	DS2	DS1	DS2	DS1	DS2
R^2	0.74	0.41	0.69	0.39	0.25	0.16	0.76	0.58
RMSE (m)	0.15	0.21	0.18	0.25	0.28	0.29	0.16	0.21
Turbidity								
Model	LR		SVR		RFR		GPR	
Dataset	DS1	DS2	DS1	DS2	DS1	DS2	DS1	DS2
R^2	0.82	0.63	0.64	0.15	0.68	0.28	0.83	0.75
RMSE (NTU)	1.11	1.69	1.54	2.50	1.41	2.20	1.10	1.36

5.4. Processing Efficiency

It is important to consider the processing demand during training and validation on computational resources; depending on the desired application this can play a crucial role. All the models were implemented using Python version 3.7.4. The hardware used was an Intel(R) Core(TM) i7-8665U CPU processor @ 1.90 GHz and 2.11 GHz, 32.0 GB of installed memory (RAM) and system type 64-bit, x64-based processor. The results of the processing performance are illustrated in Figure 7.

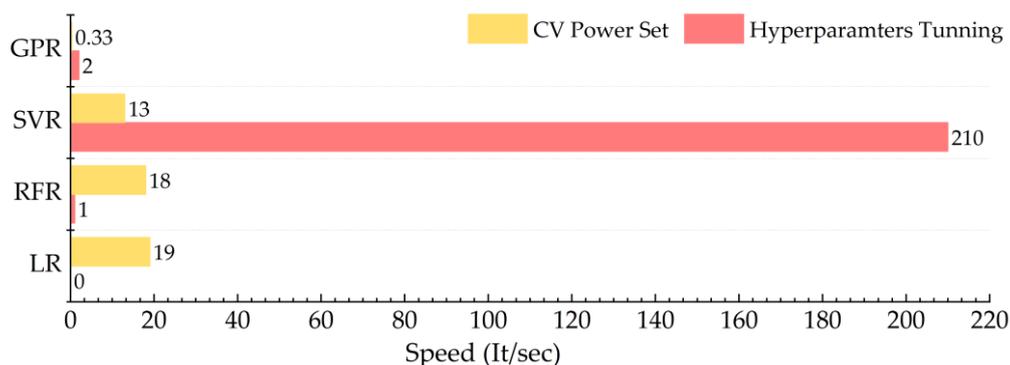


Figure 7. Processing performance of the ML models.

The process of cross-validation on a power set of 12 predictors (bands) which includes GridSearch of several possible hyperparameters in more than 4000 cases is highly demanding. A major influencing factor is the number of iterations required on the hyperparameter tuning process and the number of predictors needed to be evaluated. A larger number of terms and type of kernel also increases the required processing time in these methods. The settings are as described in Section 4.5 and no increment was used for the *cache size*.

SVR immediately stands out, performing at 210 it/sec in the hyperparameter tuning process, far away from the similar results of RFR and GRP. The LR has the advantage here of no tuning need. During the cross validation of the power set SVR, RFR and LR perform similar with 18–19 it/sec. However, GPR performs the lowest at this point with 0.33 it/sec.

5.5. SDD and Turbidity Maps

MERIS images from the sampling campaign dates were used to produce spatial distribution maps for both water quality parameters. The GPR was selected according to previous results of model performance and processing efficiency. Figure 8 displays the SDD and Turbidity spatial distribution over the water surface.

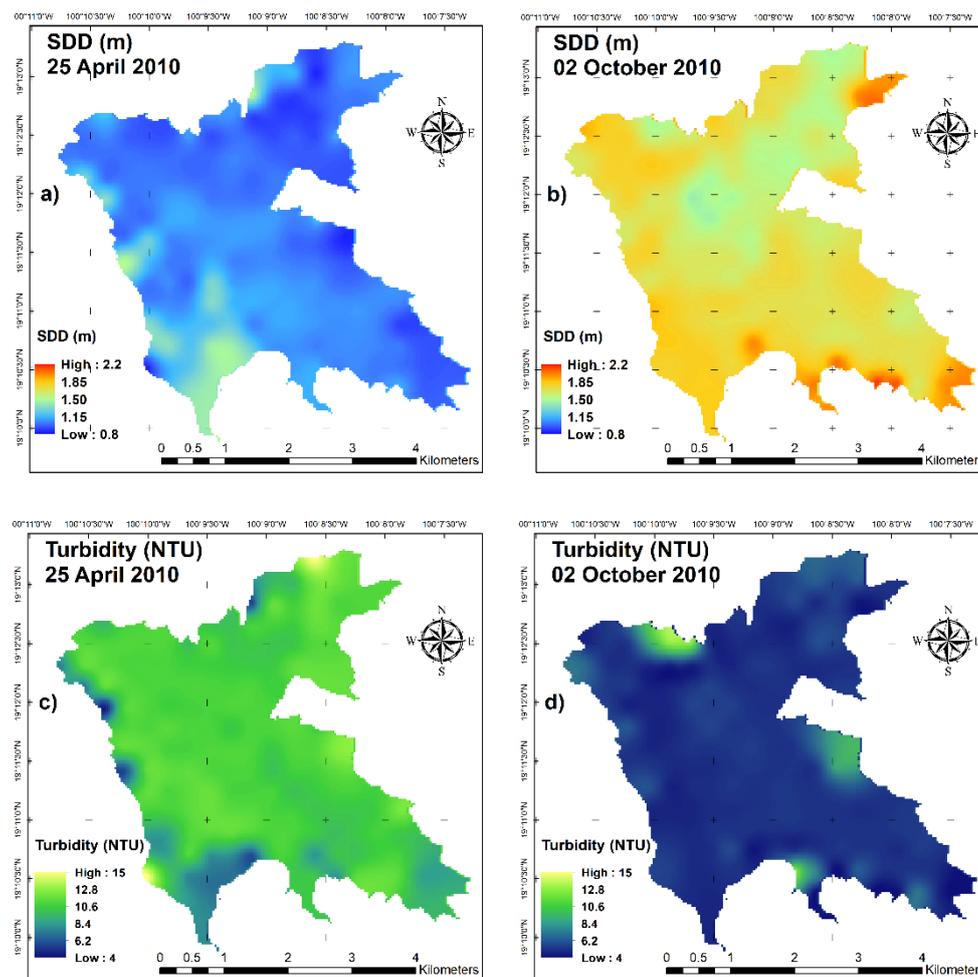


Figure 8. SDD and Turbidity spatial distribution maps for the sampling campaigns days in Valle de Bravo. SDD upper (a,b), and Turbidity lower (c,d).

For the spatial resolution, different levels of pixel size were tested to increase the 300 m resolution of MERIS. Interpolation with Spline with Barriers technique was applied in ArcMap© GIS software which interpolates a raster surface using barriers from points with a minimum curvature spline technique. The final resolution of the maps stands in 3 m. The maps revealed higher values of SDD during April and lower in October, the opposite pattern was observed for Turbidity in agreement with the negative correlation of the two water quality parameters. From the maps, it is visible that the higher levels of transparency and lower turbidity are present during October. In April, the southern part of the reservoir presents the higher values of SDD (Figure 8a, green color) and lower Turbidity (Figure 8c, clear blue color), particularly in the entrance of the incoming rivers (Carrizal, González

and Molino) and the wastewater discharges from neighbor towns. In October, the SDD (Figure 8b, blue color) and Turbidity (Figure 8d, clear blue color) acquire the opposite values and they are more homogenous all around the water surface.

5.6. Multitemporal Analysis of MERIS Imagery

Figure 9 exhibits the average monthly value for each parameter in the entire reservoir. Retrievals indicated that SDD ranged from 0 to 1.92 m and Turbidity up to 23 NTU. The correspondence between the SDD and Turbidity is clearly visible, as well as the seasonal dependence of both parameters.

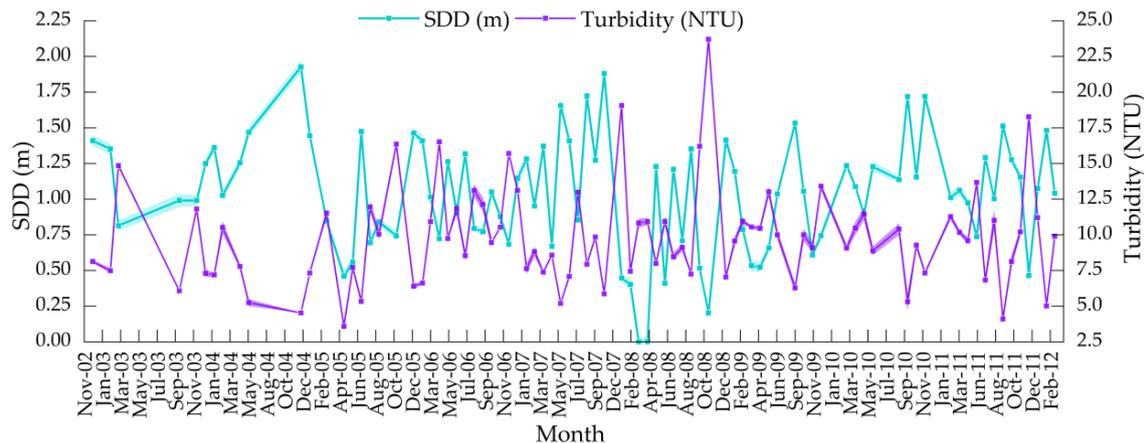


Figure 9. Estimated values for SDD and Turbidity in Valle de Bravo for the period 2002–2012.

The significant correlation between the two parameters indicates robustness in the model, since the GPR was trained for SDD and Turbidity independently. In general, high turbidity periods (Nov 2005 (16 NTU), Apr and Dec 2006 (16 NTU), Jan, Oct and Nov 2008 (10, 16 and 23 NTU) and Dec 2012 (18 NTU)) correspond with low transparency (SDD = 0.63 m). During 2008, SDD experienced exceptionally low levels during Mar (≈ 0 m), Apr (≈ 0 m) and Nov (0.20 m). Furthermore, constant tendency of low SDD and high Turbidity were found during a long period in Apr 2005–Mar 2006 and intermittently in Feb–Oct 2008 and Apr–Dec 2009.

In contrast, the estimations for the rest of the years have recognizable patterns with high clarity in dry seasons and low clarity during rainy periods. Valle de Bravo experienced the highest transparency during the years of 2004, end of 2007, 2010 and 2012 throughout the period of November reaching peak values in Dec 2004, Sep and Nov 2007 and Dec 2010. During Jan and Nov 2008 high Turbidity is observed, atypical even for other records of years. Recovery was visible from May 2010 and continued with higher SDD during 2012. A further analysis of SDD values per each month during the complete timeframe is shown in Figure 10a. Each year is also displayed with the corresponding months in Figure 10b. Similarly, Figure 10c,d show the Turbidity values per month and year.

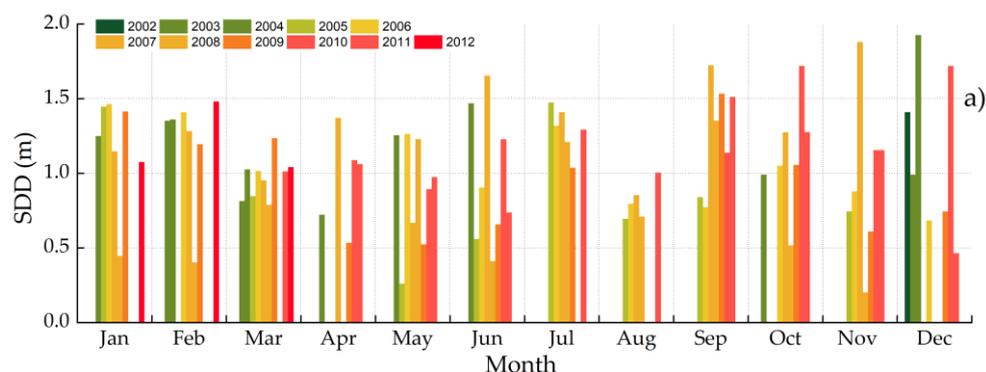


Figure 10. Cont.

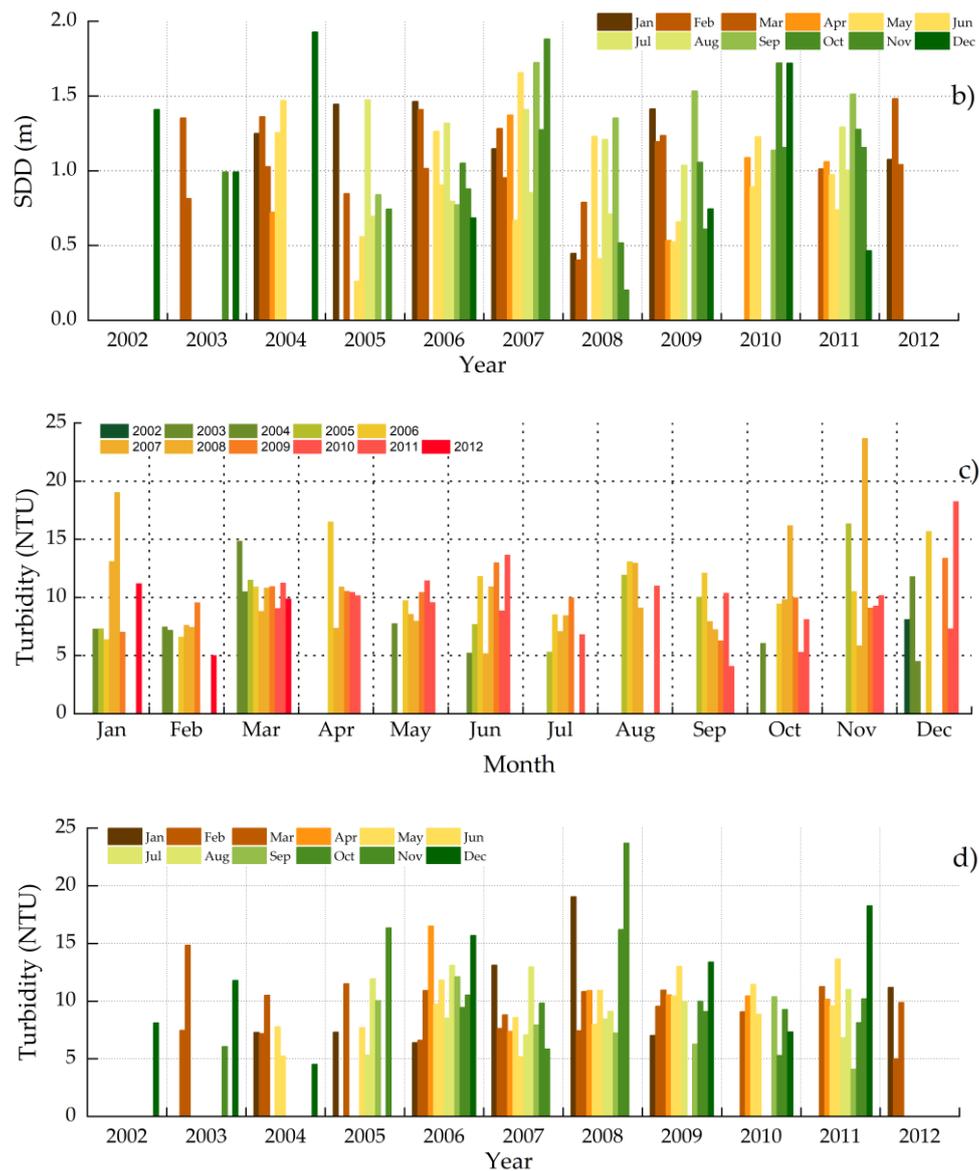


Figure 10. SDD and Turbidity analysis per month and year during 2002–2012. Figure 10a,b correspond to monthly and yearly analysis of SDD, similarly Figure 10c,d for Turbidity. Number of images used per year (year: number): 2002: 1, 2003: 4, 2004: 6, 2005: 8, 2006: 12, 2007: 11, 2008: 11, 2009: 10, 2010: 7, 2011: 10, 2012: 3. Number of images used per month (month: number): Jan: 7, Feb: 7, Mar: 10, Apr: 6, May: 8, Jun: 8, Jul: 6, Aug: 5, Sep: 7, Oct: 6, Nov: 7, Dec: 7.

The MERIS coverage was poor for the months of August and during the years of 2002, 2003 and 2004. In Figure 10a–d, the patterns of SDD and Turbidity are shown. The higher values of SDD are recognizable during the mid-part of the year (rainy) and the lower values during the last part (dry). The Turbidity shows a corresponding behavior. The years of 2005 (May–Jun), 2008 (Jan–Mar, Oct–Nov), and 2009 (Apr–Jun) show lower values of SDD, Figure 10a,b, and higher Turbidity (2005, 2006, 2008 and 2011), Figure 10c,d. An important missing part of the analysis is the record from the years of 2003–2004 where no major droughts or other emergencies were reported. Due this lack of images, the comparison of common patterns gains uncertainty and the degree of disruption of events during the remaining years remains partially unclear.

As said before, the peaks in values of SDD and Turbidity were associated with the years of 2005 and 2008 and they likely represent special cases with consequences during 2006 and 2010. Therefore,

these events which are not common in a period of 11 years, should be treated as serious incidents which might pose health threats from suspended solids accumulations and highly turbid water.

6. Discussion

6.1. Performance of Machine Learning Algorithms

The great variety of ML algorithms offers great potential for development of robust models that could predict water quality parameters. This variety, however, also challenges the selection among many candidates and possible train and validation DS. Firstly, the selection of diverse DS of RS data for training is an important step which will contribute to the goodness in performance of the models. In this case, the exception of the Adjacency Effect correction using the ICOL processor in the DS1 implies an unaccounted error in the developed model that was chosen to perform the multitemporal analysis. The adjacency effect affects mainly overestimating or under-correcting the NIR bands. Nevertheless its influence may be lesser significant, thanks to the relatively good error metrics of the models here developed when validating against in-situ measurements. The validation procedure demonstrated that DS1 produced accurate predictions with substantial improvement between 30–90% in R^2 for SDD and up to >100% for Turbidity over predictions retrieved with DS2 with similar dedicated training time conditions (Table 5). Furthermore, the chosen model for multitemporal analysis, the GPR, only uses NIR bands for Turbidity retrieval (Table 4) and the error metrics associated with it remain the highest ($R^2 = 0.83$, RMSE = 1.05) (Table 5). It is recommended, however, to complement the methodology with alternative correction approaches to evaluate further the adjacency effect and its associated errors when using RS in small reservoirs as this case [72,73]. Secondly, feature selection of representative bands is an important stage that should be addressed with a proper evaluation, to contrast the contributions of different spectral regions to the model. Random selection of the training dataset has a considerable influence in the performance predictions that can be reduced via cross-validation. Furthermore, the tuning of hyperparameters requires rigorous analysis due to the multiple values of a GridSearch; its proper validation should be assessed as well with a cross-validation. Finally, much of the goodness of the models will come from the quantity and quality of the field data gathered, which, in many cases, represents an important limitation. This study contributes to solve the above-mentioned constrains by applying a comprehensive methodology with state-of-the-art ML algorithms using MERIS data. Furthermore, the tuning process results of the hyperparameters are also provided, which could give more insights of typical ranges used for applications in water quality retrievals using RS data. The GPR and the LR models here developed performed with relatively good results and a similar behavior. SVR and RFR performed relatively good on only one of the predicted parameters (SDD for RFR and Turbidity for SVR). The reason for this behavior in RFR could be due to the complexity on the tuning process of the algorithm and the many hyperparameters required to be adjusted as well as the limited number of field data used. In the SVR most likely an extensive search of GridSearch values would be required. In both cases, this also implies a higher computational demand for a proper tuning process, increasing the complexity when apply it in regression applications like this study.

From the results, it is clear that there is a high correlation between the MERIS spectral bands; thus, a specific region of the electromagnetic spectrum cannot be pointed as a clear dominant for the development of the ML algorithms using MERIS due that many combinations produce similar performance (Figure 6). In the case of Valle de Bravo, the addition of the blue band could have had an important influence in the GPR model for the estimation of SDD when measuring VIS reflectance, however the green and red regions produce the best model. On the other hand, for Turbidity, the red and NIR bands agrees with the reflection of electromagnetic energy from suspended solids present in the water, which reflect in those regions. However, there is need of only 2–3 bands to produce relatively good performance models and the addition of bands only improves slightly the initial results for GPR. On the contrary, LR and SVM (for Turbidity) require additional bands for a better performance. Although research may tend to use only visible and NIR spectral regions that are known

to contribute significantly to the absorption of water, in this study and as part of the limited data, preconceived ideas of former methodologies were not considered and all the available bands were evaluated in correlation with the field data. The result from the mathematical point of view is that, for the developed ML algorithms, some MERIS bands had a strong correlation for this study case and a limited number of bands are likely to produce relatively good results for both water quality parameters (Figure 6). This work contributes therefore to open a discussion about the introduction of ML, empirical and semi-empirical methods and their further integration with other existing approaches, as semi-analytical algorithms.

6.2. Dynamics of Water Quality Parameters and Its Influencing Factors

The reservoir has marked seasons with dry autumns and winters, along with rainy springs and summers (Figure 9). According to the results, in rainy seasons water transparency decreases and Turbidity increases. This could be explained by runoff, suspended matter, and dissolved solids carried in rainy months, enhanced by the possible growth of phytoplankton from incoming nutrients. In contrast, the autumns and winters are characterized by low rainfalls and thus experience less runoff, avoiding resuspension and instead allowing settlement of the suspended matter and dissolved solids, particularly in deeper and less turbid waters. These observations bolster findings from previous studies [45,74]. However, it would be expected that these factors are only causing the regular patterns and not the anomalies seen in a deeper analysis for the years 2006–2008 and 2010 (Figures 9 and 10b,d). In these years some rainy months exhibited high transparency (> 1 m) and lower Turbidity (≈ 10 NTU or lower) (2006: May, Jun; 2007: Jun–July, Sep; 2010: Jun, Sep) (Figure 10a,b). This fluctuation affected the regular patterns also in dry months with low transparency (≈ 1 m or lower) and high Turbidity (≈ 10 NTU or higher) (2006: Oct–Dec; 2007: Jan, Mar; 2010: Mar) (Figure 10c,d). The inconsistency in these values could also affected the observed behavior in years of recovered water volume storage (2010–2012) where SDD and Turbidity variations were clearly correlated with an inverse relationship. As said before, different aspects such as resuspension, shoreline erosion, loads from river inputs, wind, and water depth are considered important influencers of transparency (therefore SDD) and Turbidity in reservoirs [75]. Furthermore, the reservoir is surrounded by neighboring hills with important altitude differences to the reservoir surface (up to 2100 masl compared to 1780 masl) on the reservoir's west and northwest side. The runoff produced by rainfall in rainy seasons carries loads of suspended materials and dissolved solids into the reservoir, causing reduction in SDD and increase in Turbidity because of light attenuation. Therefore, low SDD and high Turbidity are expected from river inflows to the east and southeast during rainy seasons. In dry seasons, the suspended materials and dissolved solids tend to form sediment, and the penetration of light is higher for deep waters. However, these events could not explain completely the change in patterns seen during 11 years of monitoring. Consequently, there is a need of further clarification for the major events affecting the water in the reservoir during this period. Researchers have studied the decrease of water availability in Valle de Bravo, extractions, droughts, and climate change [48].

In this sense, a correlation between the water scarcity and the disruption in patterns of water quality parameters may exist. According to local records [76], in the period of 2006–2007 and 2009–2010, the reservoir lost up to 50% of its storage (Figure 11).

The years of critical volume storage observed in Figure 11 (2005–2006, 2008–2009) coincide with the periods of disturbance of SDD and Turbidity parameters, clearly recognizable in the years 2006–2009. The findings in this study suggest that there is a possible correlation between the water quality behavior and the decrease of water volume caused by low precipitation (Figure 12), which could lead to increased Turbidity and low SDD, particularly in the dry season (Nov–May).

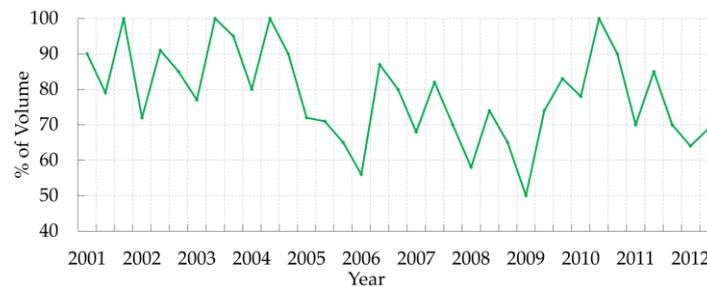


Figure 11. Water volume variation in Valle de Bravo (% of its maximum capacity) during the period of 2001–2012. Adapted from ProValle [77].

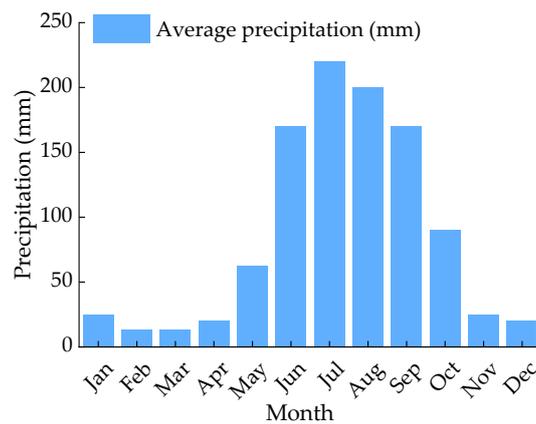


Figure 12. Average monthly precipitation in Valle de Bravo river basin. Adapted from CONAGUA [76].

6.3. Water Quality Status in the Reservoir

The water quality status in the reservoir during the studied period is difficult to infer from physical type parameters as SDD and Turbidity. For an appraisal approach, biological and chemical measures as chlorophyll-a or total phosphorus would be required. However, with the available data, some correlations could be established, and useful insights of biological and chemical parameters obtained. Regarding this, a classification based on collected lake data from the international program on eutrophication of the Organization for Economic Cooperation and Development (OECD) [78] is a useful resource for further categorization and an estimation can be then inferred for nutrient and load-eutrophication responses in Valle de Bravo. Applying its classification according to SDD, during most of the 2002–2012 Valle de Bravo was under a hypertrophic status with intermittent recovery in trophic conditions. The hypertrophic conditions are more evident for the period of 2005–2009 with a slight recovery in 2007 (Figure 13).

Valle de Bravo experimented Eutrophic status ($3 < \text{SDD} < 1.5 \text{ m}$) for 7 months (8% of the period) and Hypertrophic for 77 months (92%). For this period, total phosphorus and chlorophyll-a concentrations could be present with amounts up to 100 mg/m^3 and 25 mg/m^3 respectively according the OECD classification. High levels of phosphate and nitrate could serve as an indicator of the presence of cyanobacteria blooms, which reduces potability of drinking water. This study gathers limited data to obtain strict conclusions about the water quality conditions of the reservoir; however, the findings suggest that there is a clear correlation between the water quality behavior and the decrease of water volume caused by low precipitation.

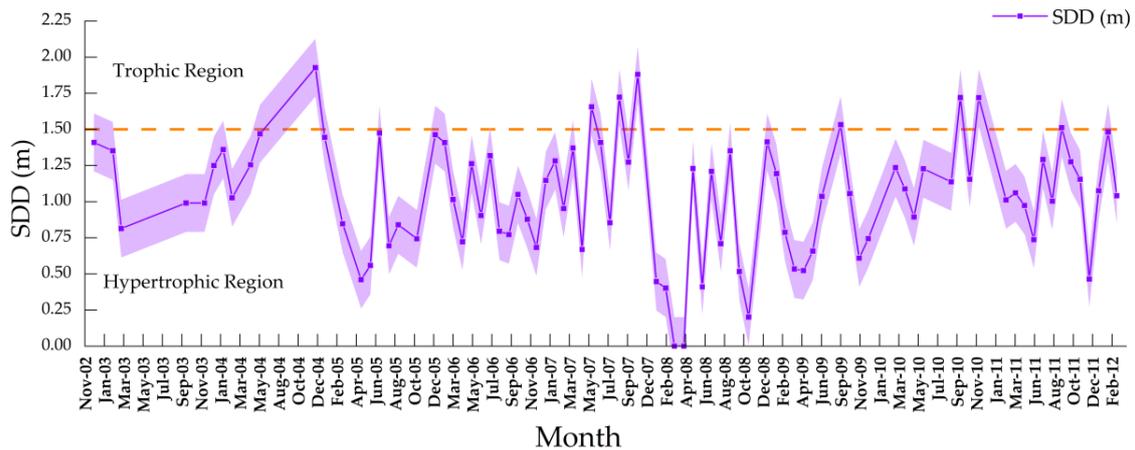


Figure 13. SDD (m) values and standard deviation (0.50 scaling factor) of Valle de Bravo during 2002–2012 with the corresponding regions of classification of the OECD [78].

7. Conclusions

Utilizing the remote sensing reflectance from MERIS data and in-situ collected samples, this study developed and validated machine learning algorithms to estimate the water quality parameters of Secchi Disk Depth (SDD) and Turbidity for Valle de Bravo reservoir in central Mexico from 2002 to 2012. Using the dataset 1 (DS1), the models performed well for estimation of both water quality parameters with satisfactory cross-validation results and with a slightly outperformance for GPR (SDD: $R^2 = 0.81$; RMSE = 0.15, Turbidity: $R^2 = 0.86$; RMSE = 0.95) followed by LR, SVM and RFR. With this, the contribution to the continuous analysis of MERIS imagery stored is reinforced. The results obtained confirm that ML algorithms are current useful approaches to retrieve water quality parameters from RS data.

From the temporal analysis it is suggested that the droughts of 2006 and 2009 acted in detriment of the water quality of the reservoir. The seasonal fluctuations were affected with unusual behaviors during 2006–2009 and contributed to lower values in 2010. The water transparency measured with SDD retrieved low values (≈ 1 m) during these periods. The Turbidity estimations confirmed this behavior with high values (≈ 12 NTU) during the same years. The suggested classification indicated an evolution from an initial trophic stage in 2002–2005 to an intermittent hypertrophic one during 2006–2008 and 2010, before a slight recovery to trophic status during 2011–2012. The water patterns also suggest that periods with low SDD and high Turbidity coincide with the rainy months (June–October) and thus, runoff of surrounding areas could have had influence on transparency owing to the loads of suspended materials and dissolved solids. On the contrary, opposite behavior, high SDD and low Turbidity was observed in dry seasons.

The methodology applied in this study yielded results that were consistent with independent evaluations, confirming the idea that RS techniques are powerful tools for overcoming limited resources when planning monitoring programs of water quality, even across long time periods. The synchrony of field measurements and the acquisitions of the sensor is of major importance. Ideally, same day in situ data is preferred for validation of satellite products. Regarding this study, it is necessary to consider that greater uncertainties in the results may be present due the variation of ± 2 days between field and satellite data collection. To avoid this, it is recommended to conduct continuous field measurements and use sensors with enough temporal resolution.

Local water quality monitoring systems are present in different countries of the world to periodically analyze the state of inland waters. Such systems have great potential for integration with RS techniques. This combination could allow extensive spatial and temporal analysis on a greater scale. Scheduled field campaigns paired with the date of image acquisition by respective satellites could be useful for data calibration, training, and validation. The continuous measurements of water quality parameters could serve as a constant source of field data. The RS resources, as the MERIS archives,

offer valuable data and an important opportunity to contribute to the understanding of how diverse events influence inland waters. The study of data acquired from sensors such as MERIS is essential for the understanding of the water quality of lakes and reservoirs in the last two decades. The current operational satellites, particularly the Sentinel-2 and Sentinel-3, are the natural successors of ENVISAT with MERIS sensors; however, extensive analysis of periods of time of any inland water compelling the first 20 years of the 2000 years would require the contribution from MERIS for a wider monitoring. For the case of Valle de Bravo, this study could serve as a base for further monitoring using Sentinel data and investigate its evolution during the remaining 8 years of the decade. The full exploration of the usefulness and performance of MERIS in monitoring inland water quality would be beneficial to the development and improvement in utilization of successor satellite missions/sensors, i.e., the Sentinel-3 with OLCI instrument, that is continuity of the MERIS instrument capability. The validated good performance of estimated water quality parameters using MERIS data in this study provides confidence in combining MERIS and successor satellite missions to extend a longer-term monitoring of inland water quality.

ML regression models are useful methods to retrieve water quality parameters for the first decade of the century using MERIS imagery, particularly in inland waters with special importance for human health, as seen in the encouraging accuracies retrieved. Future work will focus on (i) gather in-situ national water quality monitoring system datasets, (ii) process spectral datasets of current sensors like Landsat 8 OLCI or Sentinel satellites, (iii) extend the analysis of inland waters of the region where the most of the water quality remains uncertain at long-time period scale, (iv) assess new approaches like variations of linear regression (ridge linear regression, radius neighbor regression, elastic net regression) or trees (gradient boost regression trees) and (v) estimate other important water quality parameters as Chl, CDOM, TSS or nutrients. All the above with the aim to contribute to the knowledge of water quality status and trophic state of inland waters in regions which have not been previously studied using remote sensing techniques.

Author Contributions: Z.D. and L.F.A.-R. conceived this study. L.F.A.-R. and Z.D. conducted data processing and analysis. L.F.A.-R wrote the original version of the manuscript with extensive guidance from Z.D. The in-situ measurements of water quality parameters for this study were provided by R.S. Constructive comments and improvements of the manuscript were provided by Z.D., S.I.M.-M. and M.D. through extensive discussion. All authors have read and agreed to the published version of the manuscript.

Funding: This article was accomplished with the financial support for research of the Mexican National Council for Science and Technology (CONACYT) and the Federal Department of Energy (SENER) through its funding “CONACYT-SENER Sustentabilidad Energética” CVU 678957 to L.F.A.R. Furthermore, this work was supported by the German Research Foundation (DFG) and the Technical University of Munich (TUM) in the framework of the Open Access Publishing Program.

Acknowledgments: The authors would like to thank the persons involved in the field campaigns in Valle de Bravo, with the participation of R.S. and the support of the Sanitary and Environmental Engineering Department (DISA) from the National Autonomous University of Mexico (UNAM). They would also like to thank the Technical University of Munich (TUM) and its Graduate School (TUM-GS) for providing all the institutional services and facilities used in this study. They are further grateful to the ESA for providing the necessary MERIS data and software. Additionally, they would like to acknowledge the TUM fellows A.G. Padilla, Jaime Vigil, and Maria Galli. Many thanks as well to Thomas Schneider from the TUM Chair of Aquatic Systems Biology and Marco Körner from the TUM Chair of Remote Sensing Technology for their valuable advice and opinion about this paper. We would like finally to thank the peer reviewers for providing constructive comments, which extensively improved this manuscript.

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

References

1. MEA. *Ecosystems and Human Well Being: Synthesis*; Millennium Ecosystem Assessment: Washington, DC, USA, 2005.
2. Ritchie, J.C.; Zimba, P.V.; Everitt, J.H. Remote sensing techniques to assess water quality. *Photogramm. Eng. Remote Sens.* **2003**, *69*, 695–704. [[CrossRef](#)]

3. Preisendorfer, R.W. Secchi disk science: Visual optics of natural waters. *Limnol. Oceanogr.* **1986**, *31*, 909–926. [[CrossRef](#)]
4. Luhtala, H.; Tolvanen, H. Optimizing the Use of Secchi Depth as a Proxy for Euphotic Depth in Coastal Waters: An Empirical Study from the Baltic Sea. *ISPRS Int. J. Geo-Inf.* **2013**, *2*, 1153–1168. [[CrossRef](#)]
5. Carlson, R.E. A trophic state index for lakes. *Limnol. Oceanogr.* **1977**, *22*, 361–369. [[CrossRef](#)]
6. Cheng, K.-S.; Lei, T.-C. Reservoir Trophic State evaluation using Landsat TM Images. *J. Am. Water Resour. Assoc.* **2001**, *37*, 1321–1334. [[CrossRef](#)]
7. Lathrop, R. Landsat Thematic Mapper monitoring of turbid inland water quality. *Photogramm. Eng. Remote Sens.* **1992**, *58*, 465–470.
8. Khan, F.A.; Ansari, A.A. Eutrophication: An ecological vision. *Bot. Rev.* **2005**, *71*, 449–482. [[CrossRef](#)]
9. WHO. *Water Quality and Health—Review of Turbidity: INFORMATION for Regulators and Water Suppliers*; World Health Organization: Geneva, Switzerland, 2017.
10. Dekker, A.G.; Peters, S.W.M. The use of the Thematic Mapper for the analysis of eutrophic lakes: A case study in the Netherlands. *Int. J. Remote Sens.* **1993**, *14*, 799–821. [[CrossRef](#)]
11. Odermatt, D.; Heege, T.; Nieke, J.; Kneubühler, M.; Itten, K. Water quality monitoring for lake constance with a physically based algorithm for MERIS data. *Sensors* **2008**, *8*, 4582–4599. [[CrossRef](#)]
12. Matthews, M.W. Eutrophication and cyanobacterial blooms in South African inland waters: 10 years of MERIS observations. *Remote Sens. Environ.* **2014**, *155*, 161–177. [[CrossRef](#)]
13. Hansen, C.; Burian, S.J.; Dennison, P.E.; Williams, G. Spatiotemporal Variability of Lake Water Quality in the Context of Remote Sensing Models. *Remote Sens.* **2017**, *9*, 409. [[CrossRef](#)]
14. Giardino, C.; Candiani, G.; Zilioli, E. Detecting Chlorophyll-a in Lake Garda using TOA MERIS radiances. *Photogramm. Eng. Remote Sens.* **2005**, *71*, 1045–1051. [[CrossRef](#)]
15. Kratzer, S.; Brockmann, C.; Moore, G. Using MERIS full resolution data to monitor coastal waters—A case of study from Himmerfjärden, a fjord-like bay in the northwestern Baltic Sea. *Remote Sens. Environ.* **2008**, *112*, 2284–2300. [[CrossRef](#)]
16. Matthews, M.W. A current review of empirical procedures of remote sensing in inland and near-coastal transitional waters. *J. Remote Sens.* **2011**, *32*, 6855–6899. [[CrossRef](#)]
17. Sepulveda, R. Diseño de modelos de calidad del agua mediante el uso de percepción remota. In *Master and Doctoral Program in Engineering*; National Autonomous University of Mexico: Mexico City, Mexico, 2011.
18. Schaeffer, B.A.; Schaeffer, K.G.; Keith, D.; Lunetta, R.S.; Conmy, R.; Gould, R.W. Barriers to adopting satellite remote sensing for water quality management. *Int. J. Remote Sens.* **2013**, *34*, 7534–7544. [[CrossRef](#)]
19. Topp, S.N.; Pavelsky, T.M.; Jensen, D.; Simard, M.; Ross, M.R. Research Trends in the Use of Remote Sensing for Inland Water Quality Science: Moving Towards Multidisciplinary Applications. *Water* **2020**, *12*, 169. [[CrossRef](#)]
20. Odermatt, D.; Giardino, C.; Heege, T. Chlorophyll retrieval with MERIS Case-2-Regional in perialpine lakes. *Remote Sens. Environ.* **2010**, *114*, 607–617. [[CrossRef](#)]
21. Giardino, C.; Oggioni, A.; Bresciani, M.; Yan, H. Remote sensing of suspended particulate matter in Himalayan lakes. *Mt. Res. Dev.* **2010**, *30*, 157–168. [[CrossRef](#)]
22. Hedley, J.; Roelfsema, C.; Chollett, I.; Harborne, A.; Heron, S.; Weeks, S.; Skirving, W.; Strong, A.; Eakin, C.; Christensen, T.; et al. Remote Sensing of Coral Reefs for Monitoring and Management: A Review. *Remote Sens.* **2016**, *8*, 118. [[CrossRef](#)]
23. Giardino, C.; Brando, V.; Gege, P.; Pinnel, N.; Hochberg, E.; Knaeps, E.; Reusen, I.; Doerffer, R.; Bresciani, M.; Braga, F.; et al. Imaging Spectrometry of Inland and Coastal Waters: State of the Art, Achievements and Perspectives. *Surv. Geophys.* **2019**, *40*, 401–429. [[CrossRef](#)]
24. Le, C.; Hu, C.; Cannizzaro, J.; English, D.; Muller-Karger, F.; Lee, Z. Evaluation of chlorophyll-a remote sensing algorithms for an optically complex estuary. *Remote Sens. Environ.* **2013**, *129*, 75–89. [[CrossRef](#)]
25. Feng, L.; Hu, C.; Han, X.; Chen, X.; Qi, L. Long-term distribution patterns of chlorophyll-a concentration in China's largest freshwater lake: MERIS full-resolution observations with a practical approach. *Remote Sens. Environ.* **2014**, *7*, 275–299. [[CrossRef](#)]
26. Kallio, K.; Koponen, S.; Ylöstalo, P.; Kervinen, M.; Pyhälähti, T.; Attila, J. Validation of MERIS spectral inversion processors using reflectance, IOP and water quality measurements in boreal lakes. *Remote Sens. Environ.* **2015**, *157*, 147–157. [[CrossRef](#)]

27. Brezonik, P.; Menken, K.D.; Bauer, M. Landsat-based remote sensing of lake water quality characteristics, including chlorophyll and colored dissolved organic matter (CDOM). *Lake Reserv. Manag.* **2005**, *21*, 373–382. [[CrossRef](#)]
28. Zheng, Z.; Li, Y.; Guo, Y.; Xu, Y.; Liu, G.; Du, C. Landsat-Based Long-Term Monitoring of Total Suspended Matter Concentration Pattern Change in the Wet Season for Dongting Lake, China. *Remote Sens.* **2015**, *7*, 13975–13999. [[CrossRef](#)]
29. Noori, R.K.; Karbassi, A.R.; Moghaddamnia, A.; Han, D.; Zokaei-Ashtiani, M.H.; Farokhnia, A.; Ghafari Gousheh, M. Assessment of input variables determination on the SVM model performance using PCA, Gamma test, and forward selection techniques for monthly stream flow prediction. *J. Hydrol.* **2011**, *401*, 177–189. [[CrossRef](#)]
30. Mountrakis, G.; Im, J.; Ogole, C. Support vector machines in remote sensing: A review. *ISPRS J. Photogramm. Remote Sens.* **2011**, *66*, 247–259. [[CrossRef](#)]
31. Vapnik, V.; Golowich, S.E.; Smola, A. Support vector method for function approximation, regression estimation and signal processing. In *Advances in Neural Information Processing Systems*; Mozer, M.C., Jordan, M.I., Petsche, T., Eds.; MIT Press: Cambridge, MA, USA, 1997; pp. 281–287.
32. Liu, M.; Liu, X.; Liu, D.; Ding, C.; Jiang, J.L. Multivariable integration method for estimating sea surface salinity in coastal waters from in situ data and remotely sensed data using random forest algorithm. *Comput. Geosci.* **2015**, *75*, 44–56. [[CrossRef](#)]
33. Rodriguez-Galiano, V.; Mendes, M.P.; Garcia-Soldado, M.J.; Chica-Olmo, M.; Ribeiro, L. Predictive modeling of groundwater nitrate pollution using random forest and multisource variables related to intrinsic and specific vulnerability: A case study in an agricultural setting (southern Spain). *Sci. Total Environ.* **2014**, *476–477*, 189–206.
34. Rasmussen, C.E.; Williams, C.K.I. *Gaussian Processes for Machine Learning*; The MIT Press: Cambridge, MA, USA, 2005.
35. Verrelst, J.; Rivera, J.P.; Moreno, J.; Camps-Valls, G. Gaussian processes uncertainty estimates in experimental Sentinel-2 LAI and leaf chlorophyll content retrieval. *ISPRS J. Photogramm. Remote Sens.* **2013**, *86*, 157–167. [[CrossRef](#)]
36. Blix, K.; Pálffy, K.R.; Tóth, V.; Eltoft, T. Remote Sensing of Water Quality Parameters over Lake Balaton by Using Sentinel-3 OLCI. *Water* **2018**, *10*, 1428. [[CrossRef](#)]
37. Pasolli, L.; Melgani, F.; Blanzieri, E. Gaussian Process Regression for Estimating Chlorophyll Concentration in Subsurface Waters from Remote Sensing Data. *IEEE Geosci. Remote Sens. Lett.* **2010**, *7*, 464–468. [[CrossRef](#)]
38. Verrelst, J.; Muñoz, J.; Alonso, L.; Rivera, J.P.; Camps-Valls, G.; Moreno, J. Machine learning regression algorithms for biophysical parameter retrieval: Opportunities for Sentinel-2 and -3. *Remote Sens. Environ.* **2012**, *118*, 127–139. [[CrossRef](#)]
39. Kim, Y.H.; Im, J.; Ha, H.K.; Choi, J.-K.; Ha, S. Machine learning approaches to coastal water quality monitoring using GOCI satellite data. *Geosci. Remote Sens.* **2014**, *51*, 158–174. [[CrossRef](#)]
40. Ipomex. *Diagnostico en Materia de Turismo Valle de Bravo*; Ayuntamiento Constitucional de Valle de Bravo: Municipio de Valle de Bravo, Estado de México, México, 2014. Available online: https://www.ipomex.mx/recursos/ipo/files_ipo/2014/8/8/2ed859f540454faa56eba99a59eedb19.pdf (accessed on 10 September 2018).
41. Olvera-Viascan, V. Estudio de la Eutrofication del Embalse Valle de Bravo, Mexico. Master's Thesis, Facultad de Ciencias, Universidad Nacional Autonoma de Mexico, Mexico City, Mexico, 1990.
42. Olvera-Viascan, V.; Bravo-Inclan, L.; Sanchez-Chavez, J. Aquatic ecology management assessment in Valle de Bravo reservoir and its watershed. *Aquat. Ecosyst. Health Manag.* **1998**, *1*, 277–290. [[CrossRef](#)]
43. García, P.R.; Nandini, S.; Sarma, S.S.S.; Valderrama, E.R.; Cuesta, I.; Hurtado, M.D. Seasonal variations of zooplankton abundance in the freshwater reservoir Valle de Bravo. *Hydrobiologia* **2002**, *467*, 99–108. [[CrossRef](#)]
44. Nandini, S.; Merino-Ibarra, M.; Sarma, S.S.S. Seasonal changes in the zooplankton abundances of the reservoir Valle de Bravo (State of Mexico, Mexico). *Lake Reserv. Manag.* **2008**, *24*, 321–330. [[CrossRef](#)]
45. Figueroa-Sanchez, M.A.; Nandini, S.; Sarma, S.S.S. Zooplankton community structure in the presence of low levels of cyanotoxins: A case study in a high altitude tropical reservoir (Valle de Bravo, Mexico). *J. Limnol.* **2014**, *73*, 157–166. [[CrossRef](#)]
46. CNN. La Ciudad de México, en Crisis de Agua, in Expansión in Alliance with CNN. 2010. Available online: <https://expansion.mx> (accessed on 18 October 2018).

47. Fondo para la Comunicación y la Educación Ambiental A.C. Recorte en el Suministro de Agua del Sistema Cutzamala. 2009. Available online: <https://agua.org.mx> (accessed on 10 September 2018).
48. Escolero, Ó.; Martínez, S.; Kralisch, S.; Perevochtchikova, M. *Vulnerabilidad de las Fuentes de Abastecimiento de Agua Potable de la Ciudad de México en el Contexto de Cambio Climático*; Centro Virtual de Cambio Climático de la Ciudad de México-UNAM: Ciudad de México, México, 2009.
49. Ramirez, P.; Olvera, V.; Pulido, M.; Duran, A. Presence of *Vibrio cholerae* in a fresh water Reservoir of Valle de Bravo (México State, México). *Int. Rev. Hydrobiol.* **1998**, *83*, 647–650.
50. Merino-Ibarra, M.; Monroy-Ríos, E.; Vilaclara, G.; Castillo, F.S.; Gallegos, M.E.; Ramírez-Zierold, J. Physical and chemical limnology of a wind-swept tropical highland reservoir. *Aquat. Ecol.* **2008**, *42*, 335–345. [[CrossRef](#)]
51. Gaytan-Herrera, M.L.; Martinez-Almeida, V.; Oliva-Martinez, M.G.; Duran-Diaz, Á.; Ramirez-Garcia, P. Temporal variation of phytoplankton from the tropical reservoir Valle de Bravo, Mexico. *J. Environ. Biol.* **2011**, *32*, 117–126. [[PubMed](#)]
52. Ramírez-Zierold, J.A.; Merino-Ibarra, M.; Monroy-Ríos, E.; Olson, M.; Castillo, F.S.; Gallegos, M.E.; Vilaclara, G. Changing water, phosphorus and nitrogen budgets for Valle de Bravo reservoir, water supply for Mexico City Metropolitan Area. *Lake Reserv. Manag.* **2010**, *26*, 23–34. [[CrossRef](#)]
53. Gobierno de México. Sistema Cutzamala, la Llave de Agua del Valle de México. Available online: <https://www.gob.mx/temas/archivo/articulos/agua?page=123&post=articulos&query%5Btopics%5D=agua> (accessed on 10 September 2018).
54. Santer, R.; Zagolski, F.; Gilson, M. *ICOL—Improve Contrast between Ocean and Land*; MERIS: Los Angeles, CA, USA, 2009.
55. Schiller, R.D.H. The MERIS Case 2 water algorithm. *Int. J. Remote Sens.* **2007**, *28*, 517–535.
56. Duan, H.; Ma, R.; Zhang, Y.; Zhang, B. Remote-sensing assessment of regional inland lake water clarity in northeast China. *Limnology* **2009**, *10*, 135–141. [[CrossRef](#)]
57. Bonansea, M.; Ledesma, M.; Rodriguez, C.; Pinotti, L. Using new remote sensing satellites for assessing water quality in a reservoir. *Hydrol. Sci. J.* **2018**, *64*, 34–44. [[CrossRef](#)]
58. Pedregosa, F.; Varoquaux, G.; Gramfort, A.; Michel, V.; Thirion, B.; Grisel, O.; Blondel, M.; Prettenhofer, P.; Weiss, R.; Dubourg, V.; et al. Scikit-learn: Machine Learning in Python. *J. Mach. Learn. Res.* **2011**, *12*, 2825–2830.
59. Härmä, P.; Vepsäläinen, J.; Hannonen, T.; Pyhälähti, T.; Kämäri, J.; Kallio, K.; Eloheimo, K.; Koponen, S. Detection of water quality using simulated satellite data and semi-empirical algorithms in Finland. *Sci. Total Environ.* **2001**, *268*, 107–121. [[CrossRef](#)]
60. Kloiber, S.M.; Brezonik, P.L.; Olmanson, L.G.; Bauer, M.E. A procedure for regional lake water clarity assessment using Landsat multispectral data. *Remote Sens. Environ.* **2002**, *82*, 38–47. [[CrossRef](#)]
61. Garaba, S.P.; Badewien, T.H.; Braun, A.; Schulz, A.-C.; Zielinski, O. Using ocean colour remote sensing products to estimate turbidity at the Wadden Sea time series station Spiekeroog. *J. Eur. Opt. Soc.-Rapid* **2014**, *9*. [[CrossRef](#)]
62. Toming, K.; Kutser, T.; Uiboupin, R.; Arikas, A.; Vahter, K.; Paavel, B. Mapping Water Quality Parameters with Sentinel-3 Ocean and Land Colour Instrument imagery in the Baltic Sea. *Remote Sens.* **2017**, *9*, 1070. [[CrossRef](#)]
63. Alikas, K.; Kratzer, S. Improved retrieval of Secchi depth for optically-complex waters using remote sensing data. *Ecol. Indic.* **2017**, *77*, 218–227. [[CrossRef](#)]
64. Ruescas, A.B.; Hieronymi, M.; Mateo-Garcia, G.; Koponen, S.; Kallio, K.; Camps-Valls, G. Machine Learning Regression Approaches for Colored Dissolved Organic Matter (CDOM) Retrieval with S2-MSI and S3-OLCI Simulated Data. *Remote Sens.* **2018**, *10*, 786. [[CrossRef](#)]
65. Maier, P.M.; Keller, S. Application of Different Simulated Spectral Data and Machine Learning to Estimate the Chlorophyll a Concentration of Several Inland Waters. In Proceedings of the 2019 10th Workshop on Hyperspectral Imaging and Signal Processing: Evolution in Remote Sensing (WHISPERS), Amsterdam, The Netherlands, 24–26 September 2019.
66. James, G.; Witten, D.; Hastie, T.; Tibshirani, R. *An Introduction to Statistical Learning with Applications in R*; Springer: Berlin/Heidelberg, Germany, 2013.
67. Breiman, L. Random Forests. In *Machine Learning*; Kluwer Academic Publishers: Berlin, Germany, 2001; Volume 45, pp. 5–32.

68. Hastie, T.; Tibshirani, R.; Friedman, J. *The Elements of Statistical Learning*, 2nd ed.; Springer: Berlin/Heidelberg, Germany, 2009.
69. Batur, E.; Maktav, D. Assessment of Surface Water Quality by Using Satellite Images Fusion Based on PCA Method in the Lake Gala, Turkey. *IEEE Trans. Geosci. Remote Sens.* **2019**, *57*, 2983–2989. [[CrossRef](#)]
70. Lin, C.-J.; Chang, C.-C. LIBSVM: A Library for Support Vector Machines. 2001. Available online: <https://www.csie.ntu.edu.tw/~cjlin/papers/libsvm.pdf> (accessed on 10 September 2018).
71. Blix, K.; Camps-Valls, G.; Jenssen, R. Gaussian Process Sensitivity Analysis for Oceanic Chlorophyll Estimation. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2017**, *10*, 1265–1277. [[CrossRef](#)]
72. Candiani, G.; Giardino, C.; Brando, V.E. Adjacency effects and bio-optical model regionalisation: MERIS data to assess lake water quality in the subalpine ecoregion. In Proceedings of the Envisat Symposium, Montreux, Switzerland, 23–27 April 2007.
73. Brando, V.E.; Dekker, A.G. Satellite hyperspectral remote sensing for estimating estuarine and coastal water quality. *IEEE Trans. Geosci. Remote Sens.* **2003**, *41*, 1378–1387. [[CrossRef](#)]
74. Carnero-Bravo, V.; Merino-Ibarra, M.; Ruiz-Fernández, A.C.; Sanchez-Cabeza, J.A.; Sanchez-Cabeza, B. Sedimentary record of water column trophic conditions and sediment carbon fluxes in a tropical water reservoir (Valle de Bravo, Mexico). *Environ. Sci. Pollut. Res.* **2015**, *22*, 4680–4694. [[CrossRef](#)]
75. Sokoletsky, L.G. MERIS Retrieval of Water Quality Components in the Turbid Albemarle-Pamlico Sound Estuary, USA. *Remote Sens.* **2011**, *3*, 684–707. [[CrossRef](#)]
76. CONAGUA. Diagnóstico para el manejo integral de las subcuencas Tuxpan, El Bosque, Ixtapan del Oro, Valle de Bravo, Colorines-Chilesdo y Villa Victoria pertenecientes al Sistema Cutzamala. *World Bank Group* **2015**, *104*, 36–51.
77. ProValle. El Valor del monitoreo. In *Boletín del Patronato ProValle A.C.*; Municipio de Valle de Bravo: Estado de México, México, 2013.
78. OECD. *Eutrofication of Waters: Monitoring, Assessment and Control*; Organization for Economic Co-Operation and Development: Paris, France, 1982; p. 154.



© 2020 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).