

Article

Water Clarity Assessment Through Satellite Imagery and Machine Learning

Joaquín Salas ^{1,*} , Rodrigo Sepúlveda ²  and Pablo Vera ¹ 

¹ CICATA Querétaro, Instituto Politécnico Nacional, Cerro Blanco 141, Colinas del Cimatarío, Querétaro 76030, Mexico; pvera@ipn.mx

² Facultad de Ingeniería, Universidad Nacional Autónoma de México, Ciudad Universitaria CDMX 04510, Mexico; rtsh@unam.mx

* Correspondence: jsalasr@ipn.mx

Abstract: Leveraging satellite monitoring and machine learning (ML) techniques for water clarity assessment addresses the critical need for sustainable water management. This study aims to assess water clarity by predicting the Secchi disk depth (SDD) using satellite images and ML techniques. The primary methods involve data preparation and SSD inference. During data preparation, AquaSat samples, originally from the L1TP collection, were updated with the Landsat 8 satellite's latest postprocessing, L2SP, which includes atmospheric corrections, resulting in 33,261 multispectral observations and corresponding SSD measurements. For inferring the SSD, regressors such as SVR, NN, and XGB, along with an ensemble of them, were trained. The ensemble demonstrated performance with an average determination coefficient of R^2 of around 0.76 and a standard deviation of around 0.03. Field data validation achieved an R^2 of 0.80. Furthermore, we show that the regressors trained with L1TP imagery for predicting SSD result in a favorable performance with respect to their counterparts trained on the L2SP collection. This document contributes to the transition from semi-analytical to data-driven methods in water clarity research, using an ML ensemble to assess the clarity of water bodies through satellite imagery.

Keywords: water clarity; Secchi disk depth; machine learning; satellite imagery



Academic Editors: Anuj Tiwari and Anukesh Krishnan Kutty Ambika

Received: 14 November 2024

Revised: 10 January 2025

Accepted: 14 January 2025

Published: 17 January 2025

Citation: Salas, J.; Sepúlveda, R.; Vera, P. Water Clarity Assessment Through Satellite Imagery and Machine Learning. *Water* **2025**, *17*, 253. <https://doi.org/10.3390/w17020253>

Copyright: © 2025 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 (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Clean water is indispensable for maintaining public health, ensuring environmental sustainability, and fostering economic prosperity, underscored by its inclusion in the UN Sustainable Development Goals [1]. In 2020, an estimated 3.6 billion people, or approximately 46% of the global population at that time, were reported to lack access to managed drinking water services [2], a figure projected to rise to 6 billion by 2050 [3].

The contamination of water resources is often attributed to insufficient waste management in industrial activities [4] and the runoff from agricultural practices [5], posing significant risks to water quality. Moreover, climate change intensifies these water-related challenges by altering precipitation patterns [6] and triggering extreme weather events like droughts and floods [7], compromising water resources' availability and quality. The scarcity and degradation of clean water resources have a profound economic impact [8] and also highlight the urgent need to develop monitoring techniques that are cost-effective, rapid, and robust to ensure sustainable water management and access.

Water quality monitoring at various points within a water body is costly. It requires mobilizing material and human resources, ensuring accessibility to each sampling point, and transporting the samples to a laboratory. Performing this task regularly is economically

unfeasible for most water bodies. Additionally, scientific studies are still lacking on how machine learning (ML) algorithms can be effectively utilized and how the latest remote sensing post-processing algorithms improve existing methods for assessing water clarity in aquatic environments. A promising solution lies in leveraging satellite monitoring capabilities for remote sensing combined with machine learning (ML) techniques. These data-driven approaches offer a pathway to achieving faster, more robust, and economically viable water clarity assessments. Water clarity indicates the presence or absence of contaminants in the water, such as suspended solids, dissolved solids, algae, and other pollutants. These factors influence how light is reflected and transmitted in the water. Consequently, measuring water turbidity using the Secchi disk depth (SDD) remains a widely used parameter for assessing water quality.

Using satellites for environmental monitoring encompasses a comprehensive scope that enables the collection of global observations with detailed local specificity. As reported by the United Nations Office for Space Affairs [9], by February 2024, the total number of satellites in orbit reached 12,568, with a significant 61.02% of these launched in the short span from 2020 to 2023. These satellites have various instruments, including optical cameras, radar systems, and LiDAR sensors, facilitating remote observations across a broad spectrum of variables [10]. Such capabilities allow for data collection on Earth's surface coverage and detailed ocean metrics like temperature, height, salinity, and color [11]. Beyond terrestrial and marine environments, satellites offer insights into atmospheric conditions, including humidity and precipitation [12]. They are instrumental in monitoring natural disasters and their impacts, such as volcanic eruptions, fires, hurricanes, earthquakes, and floods [13]. Furthermore, they are important in tracking long-term environmental changes, including sea-level rise, ice mass fluctuations, and global temperature trends [14]. Among the most relevant applications is the monitoring of ice sheets, glaciers, and various bodies of water, such as rivers, wetlands, lakes, and dams, which is vital for assessing changes in water clarity and overall environmental health [15]. Some available satellites for water clarity monitoring include the Sentinel and Landsat constellations [16] and geostationary satellites [17].

This study combines satellite imagery with ML techniques for water clarity assessment by predicting SDD. The methodology leverages the observable changes in water's spectral signature caused by pollutant loads, which affect the reflectance across various electromagnetic spectrum bands. Such alterations facilitate the development of predictive models linking these spectral changes to SDD, a measure significantly influenced by the concentration of suspended matter in water [18]. The study employs multifaceted modeling to operationalize this approach, creating an array of regressors that include decision trees, kernel-based methods, and neural networks. The optimization of hyperparameters for these models ensures that the predictive accuracy is maximized. By repeatedly training these models on various data splits, the study validates their robustness and facilitates the selection of the most effective model configurations. The culmination of this process is the formation of an ensemble regressor, combining the strengths of individual models to provide superior inferences of SDD. This ensemble model is then applied to new satellite images to generate inferences indicating water clarity, offering a scalable and dynamic environmental monitoring and management tool.

This document outlines utilizing ML-based regression techniques for estimating water clarity, specifically by estimating SDD. Its contributions include the following:

- A thorough evaluation of machine learning algorithms to estimate the SDD using the L1TP (Level-1 Terrain Precision) and L2SP (Level-2 Science Product) Landsat image collections, aimed at assessing the atmospheric processing results of the latter collection.

- Spatiotemporal evaluation of the enriched dataset within the context of the Valle de Bravo Dam's waterbody. This dam is an integral part of the Cutzamala system, which is crucial for supplying water to large areas of Mexico's population.
- The public release of the dataset and code encourages external validation of the findings and facilitates future advancements in water clarity research through remote sensing.

The subsequent sections of the article are dedicated to elaborating on the related literature, the dataset preparation process, the methodologies adopted, the outcomes achieved, and their broader implications for the study of water clarity using remote sensing technologies. We conclude by summarizing the findings and outlining prospective avenues for further investigation in this area of environmental research.

2. Literature Review

Secchi's pioneering experiments aboard the *Immacolata Concezione* in 1864 laid the groundwork for understanding light and color behavior in marine environments by measuring sea transparency using various disks [19]. This historical method has evolved with technology, as exemplified by Yu et al. [20], who mapped MODIS aqua satellite observations to SDD values in the Yellow and East China seas using a polynomial model, achieving a significant correlation. Similarly, advancements in theoretical approaches and empirical algorithms have been developed to improve SDD estimation accuracy. Lee et al. [21] revised the Law of Contrast Reduction, introducing a new theoretical model based on the diffuse attenuation coefficient. This shift towards more accurate models is echoed in studies across various regions, including the Korean Peninsula by Kim et al. [22], which highlighted spatial and temporal variations in MODIS-derived SDD, and the Arabian Gulf by Kaabi et al. [23], utilizing a regionally calibrated algorithm to assess water clarity through empirical correlations.

Further research efforts have extended the utility and accuracy of SDD estimations in diverse aquatic environments. Alikas and Kratzer [24] developed empirical and semi-analytical algorithms for lakes and coastal waters with high concentrations of organic matter. At the same time, Rodrigues et al. [25] applied the mechanistic model by Lee et al. [21] to Brazilian waters, leading to the QAAR17 model's development for improved SDD mapping. Jin et al. [26] mapped the spatial extent of surface water resources and evaluated their water quality over time using remote sensing and models of SDD, chlorophyll-a (Chl-a), and suspended solids (SS) concentration. These advancements signify a continuous effort to refine SDD estimation methods, addressing challenges such as high variability in water qualities and the need for model adaptations based on specific water types, as highlighted by recent studies like those by Qing et al. [27], Guo et al. [28], and Zhang et al. [29], which strive for increased accuracy in SDD estimation through semi-analytical models and enhanced data analysis techniques.

Recent advancements in remote sensing and ML have significantly improved the estimation and monitoring of water clarity, particularly in measuring SDD across various water bodies. Studies by Alsahli and Nazeer [30], Zhou et al. [31], and Zhang et al. [32] have utilized different atmospheric correction methods, classification-based approaches, and ML techniques such as generalized regression neural network (GRNN), sparse spectrum Gaussian process regression (SSGPR), extreme gradient boosting (XGBoost), and random forest (RF) to enhance SDD estimations. These methods, applied to data from satellites like Sentinel-2 and Landsat, demonstrate a promising potential for remote sensing in water clarity monitoring.

Efforts to refine the estimation of water clarity from remote observations include Golubknov and Golubkov [33], who employ Principal Component Analysis to select

features and variance analysis for clustering. Shamloo and Sima [34] model the SDD and Landsat multispectral response with Artificial Neural Networks. Wang et al. [35] harmonize different remote-sensing satellite platforms, such as SPP, MODIS, and MERIS, using the minimization of a physics-based cost function, which relates previous field observations with satellite images. Zheng et al. [36] use the correlation coefficient between specific satellite wavebands and SDD field measurements to develop an inversion model that accurately estimates water transparency in Poyang Lake, achieving a high correlation for bands such as blue and red. Feng et al. [37] emphasize the importance of satellite remote sensing, interdisciplinary research, and data-sharing frameworks to address the problem of harmful algal blooms (HABs) in inland waters. Youssef et al. [38] employ Landsat and Grace satellite imagery and Geographic Information System (GIS) tools to study the impact of climate change and human activities, such as urbanization and agriculture, on land cover and water resources in the Eastern Nile Delta.

Furthermore, the integration of neural networks (NNs) and Mixture Density Networks (MDNs) in studies by Sun [39], Gan et al. [40], and Maciel et al. [41], along with the employment of a convolutional neural network (CNN) regressor by Schatz et al. [42], indicates a shift towards more sophisticated analytical models. These models have successfully mapped satellite data to SDD observations with high accuracy, reflected in R^2 values up to 0.93, and addressing challenges such as sensor-specific errors and data harmonization, particularly for scenarios with open waters.

3. Data Resources

This section describes the datasets and resources used to analyze water clarity, focusing on multispectral satellite observations and derived indices. We detail the AquaSat database, including its composition, preprocessing levels, and the selected Landsat 8 records used in this study. Additionally, we present the multispectral indices employed to enrich the feature set and the geographical context of the study site within the Cutzamala System.

3.1. Landsat Image Collections

Our analysis begins with AquaSat [43], a database for water clarity measurement. This database comprises 603,432 records, including intensity values for various multispectral bands from the Landsat 5, 7, and 8 platforms and depths associated with the SDD. The data in AquaSat were obtained from the L1TP (Level-1 Terrain Precision) collection, which corrects for radiometric and geometric issues, including sensor irregularities and distortions due to the Earth's rotation. In 2017, the United States Geological Survey (USGS) introduced the L2SP (Level-2 Science Products) processing, which accounts for atmospheric effects, such as absorption and scattering phenomena. Consequently, we extract the L2SP collection values for the geographic locations in the AquaSat database through queries to Google Earth Engine (GEE) collections. We also deemed it interesting to explore the result of the regressors applied to the 3×3 neighborhood centered at the AquaSat sampling geolocation. Thus, we also extracted the multispectral intensity values from the GEE collections corresponding to L1TP and L2SP. Due to the presence of stripes in some Landsat 7 images, the aim to maximize the probability of obtaining high-quality pixels, and the period covered by the AquaSat water body sampling, our focus is exclusively on the Landsat 8 platform, yielding 33,261 observations. From there, we gather information on pixel quality and specific bands: coastal aerosol (0.43–0.45 μm), blue (0.450–0.51 μm), green (0.53–0.59 μm), red (0.64–0.67 μm), near-infrared (NIR) (0.85–0.88 μm), short-wave infrared 1 (SWIR1) (1.65–2.07 μm), and SWIR 2 (SWIR2) (2.11–2.29 μm), all with a resolution of 30 m per pixel.

3.2. Multispectral Indices

The characteristics provided by the satellite multispectral bands are enriched with an additional set of descriptors derived from studies on water quality, turbidity, suspended particle studies, and SDD. Partially inspired by the work of Avdan et al. [44], we identified the following descriptors:

Normalized Difference Water Index (NDWI) [45]: The NDWI is designed to enhance the presence of water bodies by contrasting the reflectance in the green and near-infrared (NIR) bands. Water strongly absorbs NIR wavelengths while reflecting green light. Therefore, a high NDWI value corresponds to the presence of water. The NDWI can be obtained using the following expression:

$$\text{NDWI} = \frac{\text{Green} - \text{NIR}}{\text{Green} + \text{NIR} + \epsilon} \quad (1)$$

where ϵ is a very small number used to avoid indetermination. A higher NDWI generally indicates clearer water with fewer suspended particles, which often corresponds to greater SDD.

Normalized Difference Suspended Sediment Index (NDSSI) [46]: NDSSI is tailored to detect suspended sediments. Blue light is reflected more by suspended particles such as sand, clay, and organic particles, while clear water absorbs it. NIR, on the other hand, is absorbed by water and reflected by sediments. NDSSI is computed using:

$$\text{NDSSI} = \frac{\text{Blue} - \text{NIR}}{\text{Blue} + \text{NIR} + \epsilon} \quad (2)$$

A high NDSSI indicates higher concentrations of suspended sediments, corresponding to a lower SDD.

Normalized Multi-band Drought Index (NMDI) [47]: The NMDI is used to detect water content in vegetation and soil. SWIR bands capture the absorption due to water content, while NIR reflects the overall moisture presence. The NMDI can be obtained using the expression:

$$\text{NMDI} = \frac{\text{NIR} - (\text{SWIR1} - \text{SWIR2})}{\text{NIR} + (\text{SWIR1} - \text{SWIR2}) + \epsilon} \quad (3)$$

This index helps detect turbidity and suspended matter in water bodies since clearer water has different reflectance in SWIR bands from turbid water.

Normalized Difference Turbidity Index (NDTI) [48]: The NDTI quantifies water turbidity by contrasting green and red reflectance. Turbid waters with high sediment levels reflect more red light and absorb green light. The expression to compute NDTI is given by:

$$\text{NDTI} = \frac{\text{Red} - \text{Green}}{\text{Red} + \text{Green} + \epsilon} \quad (4)$$

Higher turbidity corresponds to lower SDD values. Therefore, NDTI is directly related to water clarity estimation.

These indices capture the interactions of light with water, sediment, and other particulates across different wavelengths. While NDWI and NDSSI focus on water presence and suspended solids, NMDI emphasizes moisture and particulate matter. Meanwhile, NDTI relates to optical water properties.

In this methodology, the geolocation provided by AquaSat is utilized to identify observations corresponding to the Landsat 8 satellite. The image identifier is then used to formulate a query in Google Earth Engine, retrieving the image associated with that geographic location and extracting the values of its bands. This process results in a database of 47,405 samples. From this database, we select those samples whose pixel quality indicates

cloudless locations [49]. Records that do not contain the value of the SDD, which serves as the response variable, are discarded, leaving us with a final database of 33,621 records.

3.3. Experimental Field of Interest

Given its significance as a water source affecting the livelihood of millions, we choose the Cutzamala System as the site for applying the inferences made by the ML regressors. The Cutzamala system, inaugurated in 1978, is crucial for supplying water to Mexico City (CDMX) and the State of Mexico, contributing $12 \text{ m}^3/\text{s}$ of water in 2023. This system, along with the Lerma System, which provides an additional $5.724 \text{ m}^3/\text{s}$ and groundwater extraction, totaling $58.322 \text{ m}^3/\text{s}$, meets various needs: urban (67.746%), agricultural (4.068%), and industrial (23.804%) [50]. The Valle de Bravo Dam, a key component within the Cutzamala System, spans 1700 ha with a maximum depth of 35 m. Initially designed for hydroelectric generation, it has evolved into a multipurpose facility, contributing $6 \text{ m}^3/\text{s}$ to the Cutzamala System's flow, serving as a flood regulation basin, and becoming a prime tourist destination as the country's most significant recreational dam [51]. Fed by seven streams, including the Amanalco and Molino Rivers, the Valle de Bravo Reservoir plays a crucial role in the region's hydrology and environmental sustainability [52]. Analyzing water clarity within the Valle de Bravo Dam is a starting point for identifying pollution sources, understanding pollutant behavior and distribution, and potentially optimizing the operations of potabilization plants that supply drinking water to CDMX.

Specifically, the Valle de Bravo Dam is located in the hydrological basin of the Balsas River, located at geographic coordinates $19^\circ 21' 30''$ North and $100^\circ 11' 00''$ West [51] (see Figure 1). It was completed in 1954 and has a maximum height of 35 m. Along with the El Bosque and Villa Victoria dams, it is an essential component of the Cutzamala System [50], collectively supplying $8.313 \text{ m}^3/\text{s}$ and $5.362 \text{ m}^3/\text{s}$ of potable water to CDMX and the State of Mexico, respectively [51].

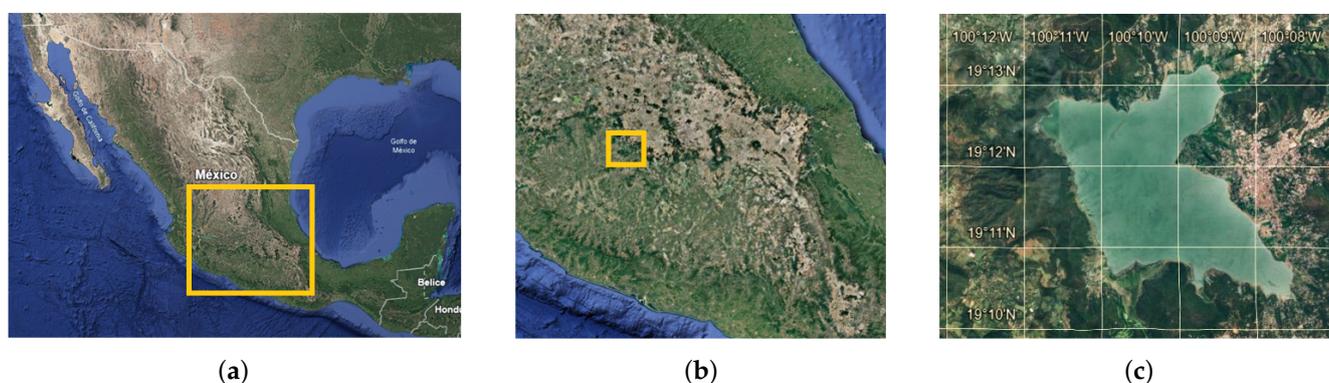


Figure 1. Google Earth maps showing the Valle de Bravo Dam location in Mexico (a), in central Mexico (b), and locally (c) [53].

4. Methods

We aim to infer the SDD over water bodies from Earth observations over time. To that end, the inference problem contains two stages. The first focuses on inferring the SDD on static images, which is solved by constructing an ML-based ensemble regressor. The second stage pursues the assessment of the SDD over time. Sometimes, the pixels are not usable in satellite images because clouds cover the view. Thus, we take a state estimation approach. This approach will serve us well later in cases where the water body is wholly or partially covered with clouds. Figure 2 illustrates a schematic representation of the proposed methodology.

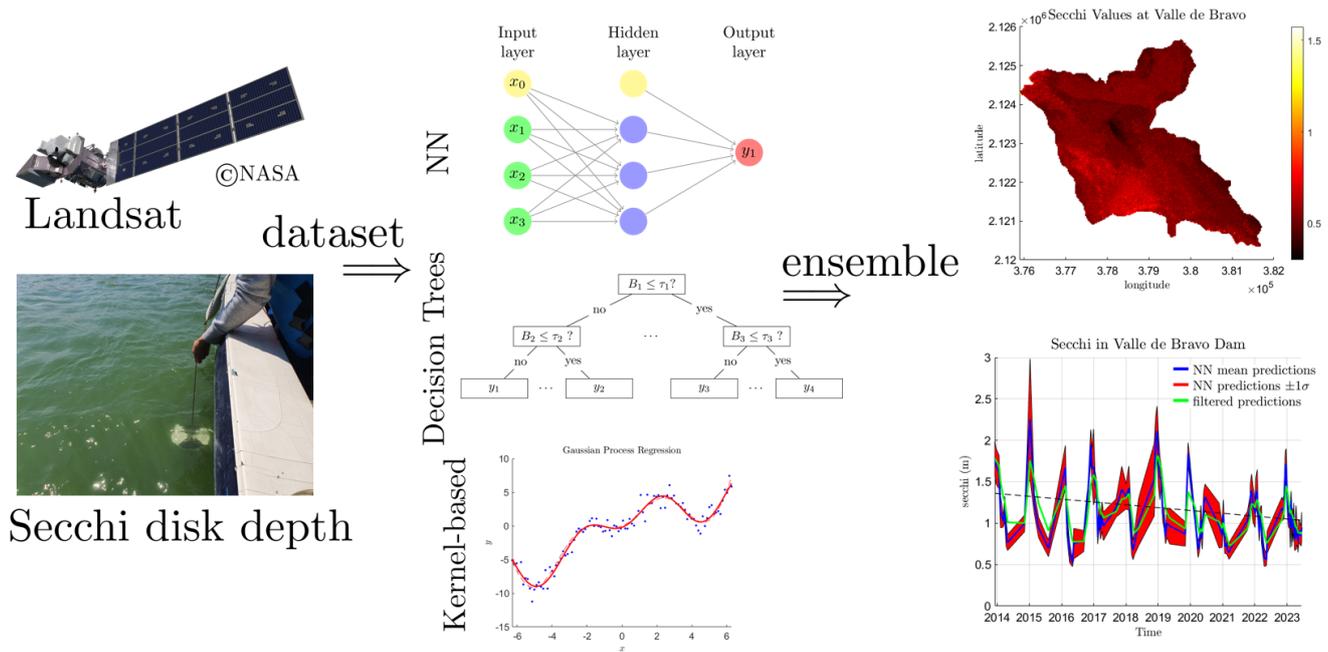


Figure 2. A schematic representation of the methodology for assessing water clarity using machine learning-based regression techniques.

4.1. Setting Up the Regressors

Theoretical and empirical research shows that a good ensemble selection should include individual, accurate regressors that make mistakes in different regions of the input data distribution [54]. Among the many regressors, we select representatives of decision trees, kernel-based methods, and graphical models. Decision trees establish their discrimination boundaries parallel to the main axis (although some variations lift this restriction [55]), and kernel-based methods excel in establishing similarity between observations non-linearly projected into multidimensional space. In contrast, neural networks project data into non-linear space in stages, generating increasingly abstract semantic representations. By broadening the spectrum of approaches, we aim to reveal the interpretability of different regions of the dataset. We deliberately dismissed techniques requiring large spatial support, such as convolutional neural networks (CNNs) or Transformers [56]. Given the 30 m/pixel resolution of Landsat observations, applying these techniques to such data might not yield meaningful spatial feature extraction, as even small image patches represent relatively large areas (e.g., 10 × 10 pixels cover 300 × 300 m). This spatial scale may be too coarse for the localized nature of SDD measurements, making them less suitable for SDD assessments.

In the present approach, a set of regressors is evaluated, including support vector regression (SVR), neural networks (NNs), and extreme gradient boosting (XGB) [57]. Subsequently, an ensemble of regressors is constructed through an NN.

XGB [58]. We optimized the hyperparameters related to the learning rate ($0.01 \leq \eta \leq 0.3$), the percentage of columns sampled in each tree ($0.01 \leq cs \leq 0.3$), the maximum depth ($1 \leq mp \leq 11$), the percentage of data sampled to build a tree ($30 \leq ss \leq 100$), and the regularization term ($0 \leq \gamma \leq 100$). The hyperparameter search is performed randomly using a uniform distribution with 1000 samples over the specified interval. The evaluation is carried out by cross-validation. During hyperparameter selection, 50 decision trees are generated. With the chosen parameters, a model with 100 trees is trained.

SVR [59]. The goal is to regress the objective curve using a kernel on the input data that projects them into a non-linear space. We aim to approximate them in this space by

a straight line within a tolerance margin ϵ . For SVR, we optimize the hyperparameters related to the constant C in the range $[1, 1000]$, allowing some flexibility in crossing the defined margin; γ in the range $[0.01, 1]$, determining the flexibility of the decision boundary; and ϵ in the range $[0.01, 0.1]$, defining the margin within which errors are not penalized.

NN [60]. Using gridded hyperparameter search, the selected neural network model is chosen from architectures that include between 32 and 512 units per layer, with the number of layers ranging from one to five. A ReLU activation function and an L_1 regularizer on the layer weights are employed, searching for the optimal regularization constant per layer. The optimizer is Adam, and the loss function is Mean Square Error (MSE). The regressor is trained during up to 500 epochs with *early stopping* based on the loss value in the validation partition, having patience during ten epochs. The model that achieves the best loss value during training is retained.

Ensemble. An NN was designed as the underlying structure of the ensemble. Using gridded search to look for the best hyperparameters for the number of layers (between one and five) and the number of units per layer (between three and one hundred in steps of five). The Adam optimizer and an MSE-based loss function were used for parameter optimization. The training process was carried out for up to 500 epochs, using *early stopping* with the patience of ten epochs. The model that showed the best performance on the validation set was retrained. For training, the predictors are scaled according to the inference models related to NN, XGB, and SVR before being introduced into the ensemble. The predictions are then normalized using the training partition before parameter optimization.

Following Afendras and Markatou [61], who offer theoretical foundations, the data were divided into 50% for training, 20% for validation, and 30% for testing. Using the training partition, the predictors were normalized while retaining the mean value and standard deviation across all bands. These parameters were later applied to normalize the validation and testing splits. Performance is measured using the coefficient of determination R^2 [62]. The procedure was repeated 20 times, conducting the learning process for each of the 20 random data partitions. This process resulted in a mean value of R^2 and a standard deviation of the performance for each regressor.

4.2. Estimating the SDD over Time

The permanence of remote sensing platforms makes it possible to explore the temporal assessment of SDD over water bodies. For instance, the first Landsat platform was launched on 23 July 1972, making it the longest-serving satellite platform for Earth observation analysis. Although it was only on 1 October 2008 that the full images platform was publicly available, the length of its records made it possible to revisit historical observations for analysis. NASA, the USGS, and the Landsat program agencies, have tried to maintain band compatibility over time. In particular, they have added the Quality Assessment Band (BQA) to check for the presence of clouds, cloud shadows, snow, and ice, all of which prevent the correct assessment of SDD by distorting observations intended to be made of the surface reflectance.

Additionally, Landsat images undergo several transformations from when an image is captured until the data are delivered to users. These transformations include radiometric, geometric, and atmospheric corrections, notwithstanding the corresponding data compression. In order to obtain a robust estimate of the SDD over time, we have devised the use of a Kalman filter. The state estimation of the Kalman filter will marginalize the complex effects of the different data processing stages. However, it will serve as an imputation strategy for missing values in the presence of bad BQA values.

Each evaluation of the SDD in the water body estimates its depth for each image pixel. Added value can be obtained by considering the observations over time z_k and the

corresponding measurement noise covariance \mathbf{R}_k . We model the estimation of the state of the system $\hat{\mathbf{x}}_k = [z_k, \dot{z}_k]$ and the corresponding transition noise covariance \mathbf{Q}_k using a Kalman filter. The prediction equations for the state $\hat{\mathbf{x}}_{k|k-1}$ and corresponding state covariance $\mathbf{P}_{k|k-1}$ at time k given the measurements up to time $k-1$ can be expressed by [63]:

$$\begin{aligned}\hat{\mathbf{x}}_{k|k-1} &= \mathbf{A}\hat{\mathbf{x}}_{k-1|k-1}, \\ \mathbf{P}_{k|k-1} &= \mathbf{A}\mathbf{P}_{k-1|k-1}\mathbf{A}^T + \mathbf{Q}_k,\end{aligned}\quad (5)$$

where $\mathbf{A} = \begin{bmatrix} 1 & \Delta t \\ 0 & 1 \end{bmatrix}$ is the state transition matrix, and \mathbf{Q}_k is the system noise covariance matrix. On the other hand, the equations for update are given by:

$$\begin{aligned}\mathbf{K}_k &= \mathbf{P}_{k|k-1}\mathbf{H}^T(\mathbf{H}\mathbf{P}_{k|k-1}\mathbf{H}^T + \mathbf{R}_k)^{-1}, \\ \hat{\mathbf{x}}_{k|k} &= \hat{\mathbf{x}}_{k|k-1} + \mathbf{K}_k(z_k - \mathbf{H}\hat{\mathbf{x}}_{k|k-1}), \\ \mathbf{P}_{k|k} &= (\mathbf{I} - \mathbf{K}_k\mathbf{H})\mathbf{P}_{k|k-1},\end{aligned}\quad (6)$$

where the observation model \mathbf{H} is represented as $\mathbf{H} = \begin{bmatrix} 1 & 0 \end{bmatrix}$, \mathbf{K}_k stands for the Kalman filter gain at time k , and \mathbf{I} is the identity matrix. In this formulation, $\hat{\mathbf{x}}_{k|k}$ contains the best estimate for the SDD at time k . In contrast, $\hat{\mathbf{x}}_{k|k-1}$ contains the best prediction for the measurement at time k , particularly useful when clouds occlude the surface covered by a specific pixel or the BQA value discards its employment.

5. Results

In this section, we analyze the performance of regressors in the spatiotemporal inference of the SDD. The regressors are programmed in Python. For the neural networks, we used TensorFlow 2.12.1. For XGBoost, we used XGBoost 2.0.0; for SVR, we used scikit-learn 1.5.1.

5.1. Predictors

Upon verifying the input data, we found that while the predictors can range between 0 and 2^{16} , they predominantly fall within the range of 5000 to 13,000. The response variable is generally between 0 and 10 m but can reach values higher than 60 m. The normalized mutual information (NMI) between variables associated with ultra-blue and blue colors registers a value of 0.350. Similarly, the NMI between SWR1 and SWR2 is 0.484. However, the NMI concerning the response variable is generally low, oscillating between 0.008 for SWR2 and 0.046 for the red band. Figure 3 illustrates the relationship between the multispectral bands and the SDD response variable. The linear correlations between the coastal aerosol, blue, red, green, NIR, SWIR1, and SWIR2 bands and the SDD variable are -0.06 , -0.09 , -0.18 , -0.18 , -0.11 , -0.11 , and -0.10 , respectively. Please note that we multiplied the integer intensity values in the Landsat images by 0.0000275, as recommended by the USGS, to compute the band reflectance.

Derived from the multispectral bands, we investigated the incorporation of multi-spectral indices, including the NDWI, NMDI, NDTI, and NDSSI (see Section 3.2). These indices allow for the establishment of non-linear relationships between bands to enrich the feature set. The multispectral indices were constrained to values between -1 and 1 for subsequent use.

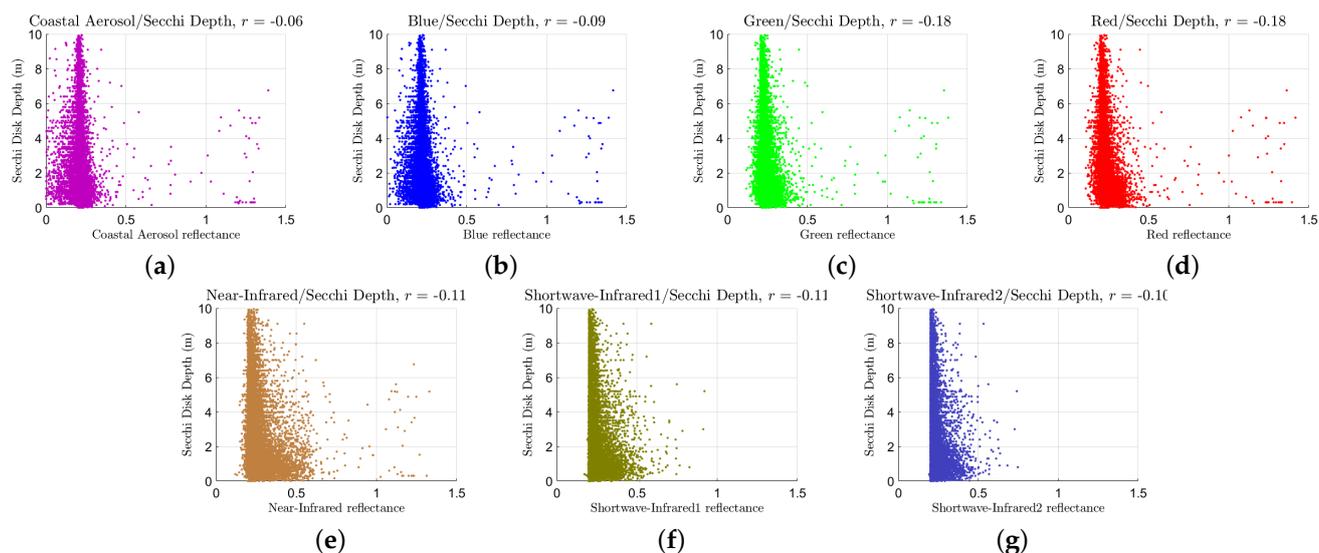


Figure 3. Relationship between the multispectral bands and the SDD. The linear correlation between the coastal aerosol (a), blue (b), green (c), red (d), NIR (e), SWIR1 (f), and SWIR2 (g) with the SDD variable is -0.06 , -0.09 , -0.18 , -0.18 , -0.11 , -0.11 , and -0.10 .

5.2. Regressors Training

For the XGB and SVR regressors, the best hyperparameters for models with the lowest loss on the validation partition were identified through a random search in the parameter space. The XGB regressors included values for the learning rate (η), column sample by tree (cs), maximum depth (md), subsample ratio (ss), regularization term (γ), and number of estimators. For SVR regressors, the optimal parameters comprised the penalty term (C), epsilon (ϵ), and kernel coefficient (γ). The best parameters for the NN and ensemble regressors were obtained through a grid search across the architectural space. For the NN regressors, the optimal architectures varied in terms of the number of layers and neurons per layer, with layer counts ranging from two to four and neurons per layer ranging from 32 to 512. The last layer in each architecture contained a single neuron, which is not shown in the table. For the ensemble regressors, the architectures consisted of two to four layers, with neuron counts tailored for each method to achieve the best performance. The specific hyperparameters for all regressors are summarized in Table 1. The table provides details about each regressor type, including whether the L1TP (T) or L2SP (S) collection was used, whether only bands (b) or both bands and spectral indices (s) were included, and whether the central pixel (c) or the 3×3 neighborhood (n) was utilized. The results were obtained using the best regressor for each method and repeating the data partition 20 times.

The results obtained are described and illustrated in Table 2 and Figure 4, showing that XGB and SVR perform similarly across all options. Interestingly, the L1TP collection appears to offer marginally better results. Superior performance is achieved with the neural network, improving the determination coefficient by approximately 0.09 consistently, except for the L2SP collection when using multispectral indices. However, the ensemble appears to outperform the different options, offering R^2 values around 0.75 for almost all cases, except when using the central pixel in the L2SP collection, where its performance drops to around 0.66–0.69.

Table 1. Regressor hyperparameters. The table shows the best hyperparameters for the constructed regressors. The regressors include those where the image source corresponds to the L1TP (T) or L2SP (S) collection, where either the bands (b) or the bands and spectral indices (s) were employed, and whether the central (c) pixel or its 3 × 3 neighborhood (n) was used. “n.l.” stands for neurons per layer. The last layer of the NN and the ensemble (Ens) contains one neuron that is not shown.

Method	XGB						SVR			NN n. l.	Ens n. l.
	η	cs	md	ss	γ	n	C	ϵ	γ		
Tbc	0.022	0.900	10	0.418	0.131	430	50.2	0.984	0.101	480, 448, 32, 32	23, 39, 90
Tbn	0	0.418	9	0.491	0.828	490	13.359	0.301	0.107	64, 288	3, 15
Tsc	0	0.900	8	0.791	0.111	420	6.564	0.579	0.100	96, 256	13, 18
Tsn	0	0.455	8	0.700	0.343	370	11.107	0.050	0.109	128, 416, 32	3, 81, 51
Sbc	0	0.955	7	0.427	0.899	380	26.93	0.988	0.104	256, 192, 96	48, 6, 3
Sbn	0	0.709	9	0.518	0.313	240	10.924	0.252	0.101	416, 416	98, 63
Ssc	0	0.900	7	0.573	0.253	350	5.563	0.681	0.107	512, 160	13, 66
Ssn	0	0.600	9	0.318	0.667	480	10.58	0.028	0.107	32, 416	43, 90, 96

Table 2. Regressor performance evaluation. The table summarizes the regressors’ results for the problem of determining the SDD (see also Figure 4). As expected, the ensemble proved to be the best regressor (highlighted in bold). In the column names, T indicates that the data come from the L1TP collection and S from L2SP, b signifies that the satellite bands were used or s that spectral indices were computed, and c denotes that only the central pixel was employed or n that a 3 × 3 neighborhood was used.

Method	Tbc	Tbn	Tsc	Tsn	Sbc	Sbn	Ssc	Ssn
XGB	0.51 ± 0.02	0.54 ± 0.03	0.56 ± 0.02	0.59 ± 0.03	0.51 ± 0.02	0.54 ± 0.02	0.53 ± 0.03	0.58 ± 0.02
SVR	0.52 ± 0.01	0.56 ± 0.01	0.57 ± 0.02	0.59 ± 0.02	0.50 ± 0.02	0.54 ± 0.02	0.55 ± 0.01	0.57 ± 0.02
NN	0.61 ± 0.02	0.65 ± 0.02	0.63 ± 0.02	0.64 ± 0.02	0.59 ± 0.02	0.62 ± 0.02	0.60 ± 0.02	0.60 ± 0.02
Ensemble	0.76 ± 0.03	0.76 ± 0.03	0.74 ± 0.03	0.75 ± 0.03	0.66 ± 0.01	0.75 ± 0.03	0.69 ± 0.02	0.76 ± 0.03

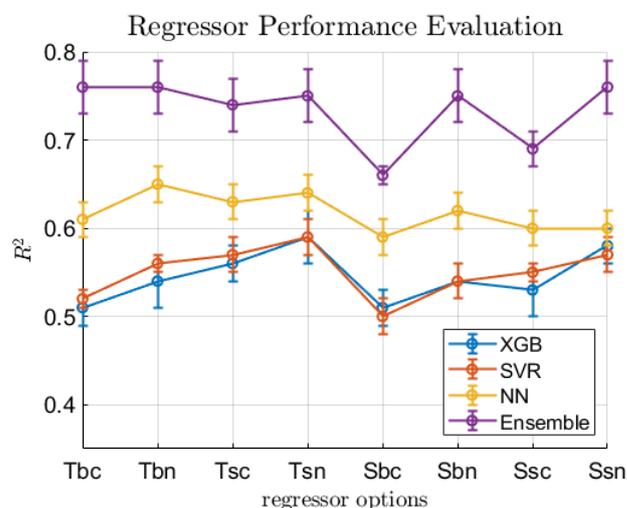


Figure 4. Regressor performance evaluation. The figure illustrates the performance of the regressors for the problem of determining the SDD (see also Table 2). The ensemble was the best-performing regressor. In the column names, T indicates that the data comes from the L1TP collection, and S denotes data from L2SP. b signifies that the satellite bands were used, while s indicates that spectral indices were computed. Additionally, c denotes that only the central pixel was used, whereas n indicates that a 3 × 3 neighborhood was employed.

5.3. Verification Through Fieldwork

Landsat images are downloaded through the USGS platform during operation using their Python API (note that we employed GEE-downloaded images to construct the regressors). The polygon defining the dam area is set to obtain the multispectral values that feed

the ensemble regressor. Typical examples are shown in Figure 5. Despite the observation resolution of Landsat being 30 m/pixel, the presence of clustered zones with similar depths along the dam is noticeable. The pixels in the images were filtered, considering only those without clouds or cloud shadows.

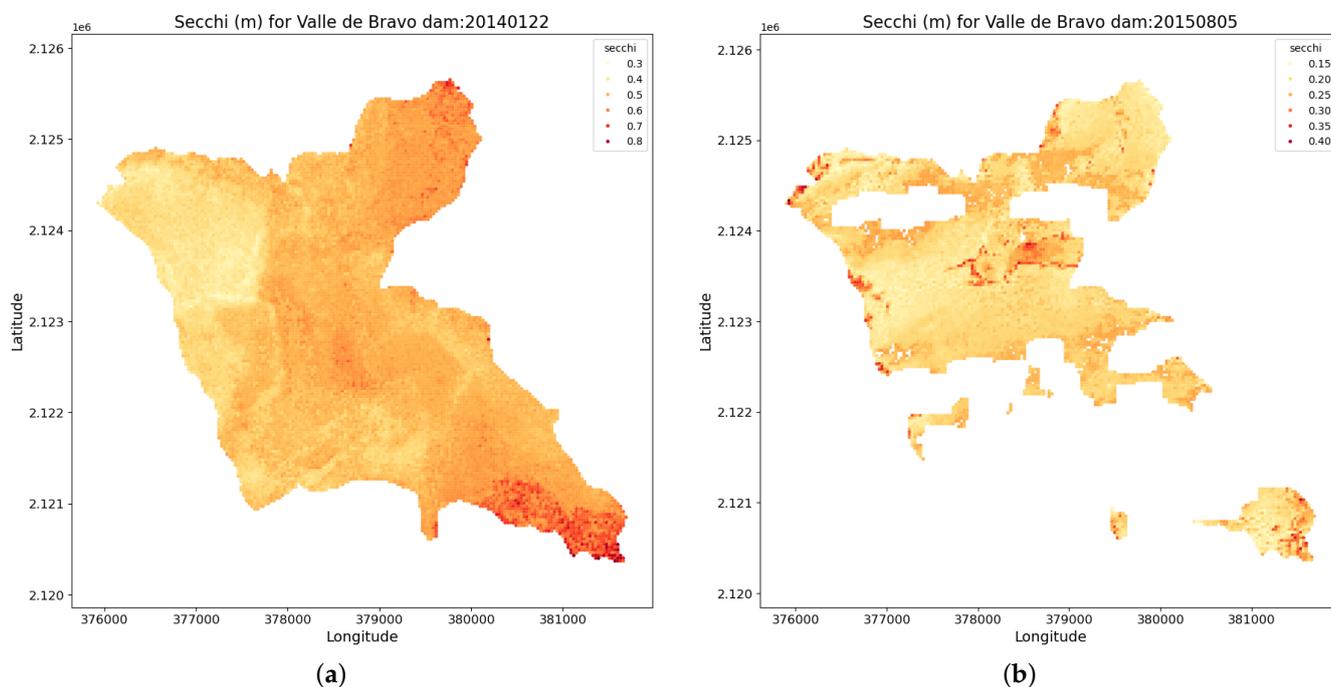


Figure 5. Predictions of the SDD in the Valle de Bravo Dam. (a) Even with a resolution of 30 m/pixel, characteristic of Landsat, clusters with similar depths can be observed. (b) The pixel quality was examined to determine which portions contained reliable information for processing.

The availability of satellite data from Landsat 8 facilitates temporal analysis based on the sequence of observations. For this purpose, the mean value of the SDD predictions for the entire dam area and its standard deviation are used to monitor their evolution over time. To achieve a robust estimation, the Kalman filter is applied. Experimentally, we define a process Q variance with a value of 0.2. From 5 December 2013 to 6 July 2023, satellite observations were collected, obtaining 110 images. Figure 6 illustrates the temporal variations in the SDD, with notable peaks at the end or beginning of each year, followed by a decline a few months later. Additionally, there is a slight but perceptible reduction in the SDD over time, with an average decrease of 3.42 cm per year.

We analyzed the contribution of neighboring pixels by incorporating a 3×3 pixel neighborhood ($90 \text{ m} \times 90 \text{ m}$) into the model. As shown in Table 2, this approach led to improved performance. However, the area covered by 3×3 pixels in a Landsat image is substantial, and special care must be taken during sampling to account for the effects of shallow waters and the presence of soil. These factors can compromise remote sensing observations, particularly in inland water bodies, as highlighted by the field sampling. For our application in the Valle de Bravo Dam, we ultimately employed the 1×1 pixel Tbc regressor, which provided more localized and reliable results.

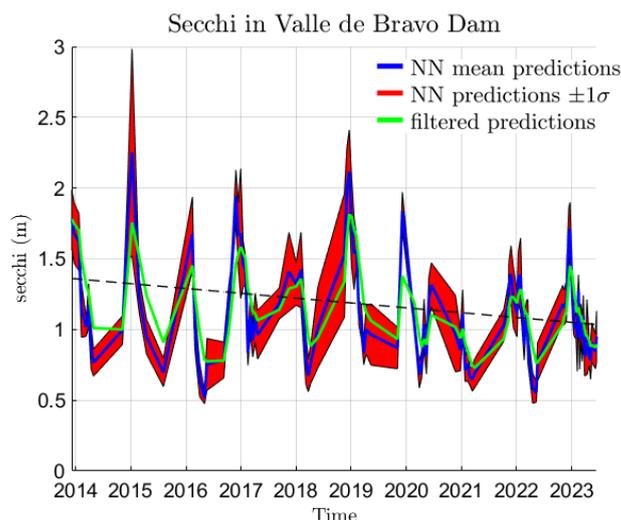


Figure 6. Observations and robust estimation of the average SDD. The data indicate peaks at the end of/beginning of the year, followed by valleys a few months later. A gradual decrease in the SDD is observed throughout the observation interval.

5.4. Field Data Validation

A field visit to the Valle de Bravo Reservoir was conducted to obtain field-measured values. Direct sampling is a fundamental tool for constructing models using satellite images, as it provides ground truth. The reference values, in turn, enable the construction and refinement of predictions made by these models.

The field visit necessitates careful planning, determination of sampling sites, and accurate collection and processing of quality parameters during sampling and laboratory work. These steps are important for obtaining and calibrating models with reliable data. Typically, selecting sites for representative reservoir sampling is based on assumptions regarding its discharges. Even in water clarity studies utilizing remote sensing, the determination of these sites can be somewhat arbitrary.

For this study, we followed a stratified sampling approach to maximize the multispectral diversity of the study area. Specifically, the intensity values were first normalized by evenly dividing the range of each band into 12 levels. This means that each group was represented as a vector of values between 0 and 11, with as many positions as spectral bands in the satellite images. Subsequently, all possible groups were identified, and the percentage of pixels belonging to each was calculated. We retained groups containing more than 5% of the pixels, while the remaining groups were assigned to the most similar one using the smallest Euclidean distance. This approach ensures that the samples are distributed across areas with the highest potential diversity of results.

On 7 October 2022, Valle de Bravo Dam measurements coincided with a Landsat 8 satellite overflight. During our fieldwork, observations were taken from forty points, with geolocation and measurement of the SDD y at each site. Subsequently, we downloaded the image set corresponding to this observation and performed inference calculations using the described procedure. Figure 7a illustrates the portion of the dam that was not obscured by clouds or over which a cloud was not projected (Figure 7b), highlighting the location of the points and the difference between the reference value and the prediction colored. To obtain the prediction \hat{y} , an interpolation function was created to evaluate the sampling geo-coordinates. Figure 7c displays the difference between the field measurements and the predictions. The coefficient of determination was found to be $R^2 = 0.80$.

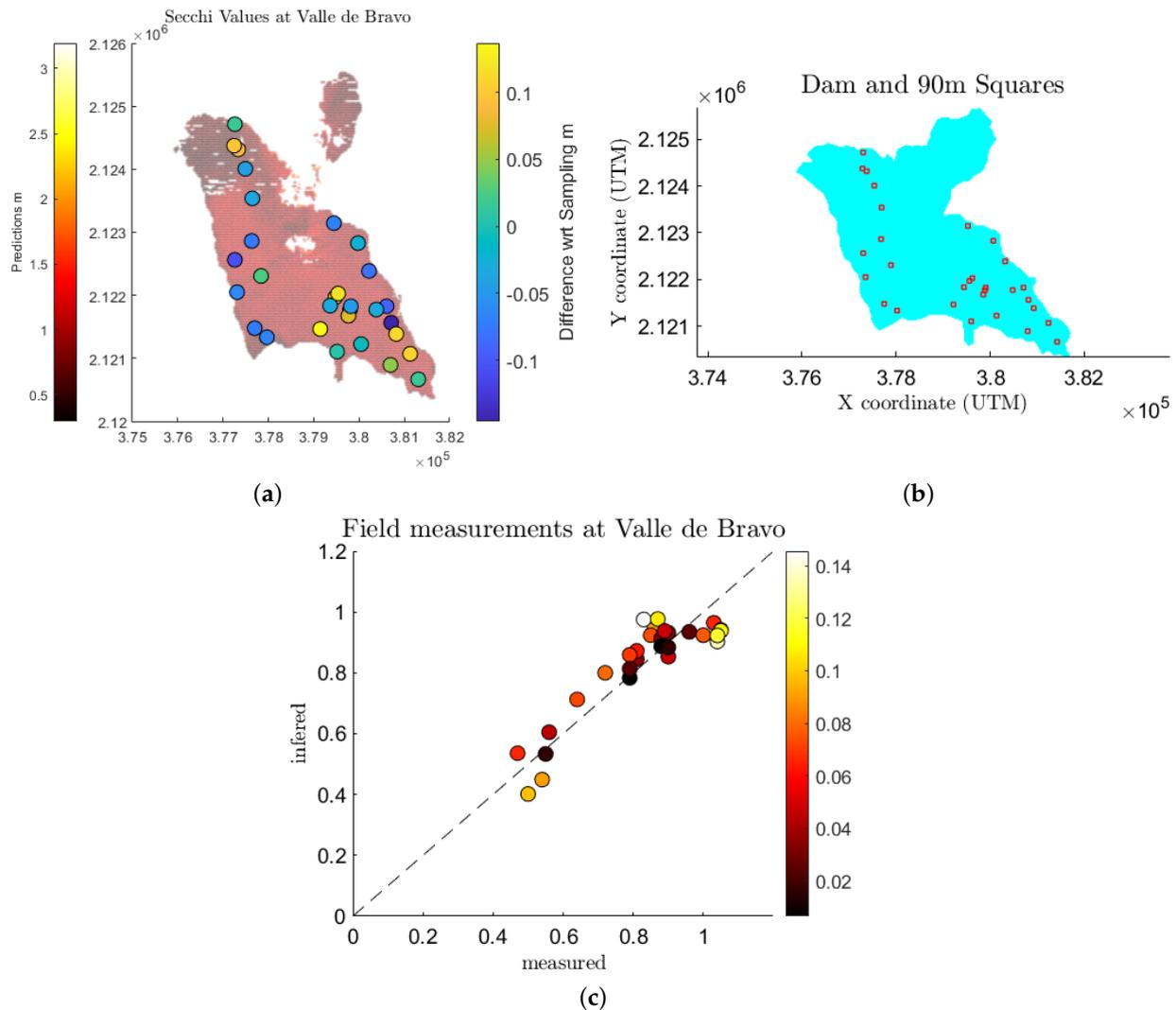


Figure 7. Comparison and error analysis of SDD predictions. Correlation between observed and predicted values shows a strong positive relationship. (a) The grayscale intensity of the sample point reflects the difference from the reference in meters. (b) illustrates some difficulties in extending the support area as some outbound regions may lie on shallow water or soil. (c) Relationship between the measurements and inferences. The system tends to underestimate the SDD. The coefficient of determination has a value of $R^2 = 0.80$.

The present approach consists of regressors for XGB, SVR, NN, and an ensemble, which require fine-tuning of their hyperparameters. In addition, XGB, SVR, and NN require preparing the regressors before the ensemble. In this case, we repeated the training stage 20 times with different data splits for training, validation, and testing. Note that SVR requires the most time, while XGB is the fastest. Also, the time taken to fine-tune the hyperparameters for a 3×3 neighborhood does not increase linearly with the number of predictors.

6. Discussion

With the increasing availability of the data corpus [43], research efforts to determine water clarity have shifted from semi-analytical approaches [21] to data-driven methods. Likewise, new insights into surface reflectance scattering and absorbing effects [16], due to temporal, spatial, and spectral variations in the presence of gases, aerosols, and water vapor, underscore the importance of evaluating available datasets. Unlike past efforts, we updated the Aquasat dataset predictors, obtained with L1TP processing, with intensity

values corresponding to the newly developed L2SP image post-processing algorithms [64]. While the L1TP processing includes geophysical corrections related to ground control point (GCP) and digital elevated map (DEM) corrections, L2SP adds additional corrections for atmospheric effects related to light absorption and scattering. Since L2SP attempts to remove atmospheric interference, the observations estimate the best surface reflectance, aiming to make the inferences more trustworthy [65]. Note that the cloud-corrected L2C2 interpolates over clouds, whereas L2SP flags the presence of clouds. We have preferred the latter option to reduce misinterpretations of the surface reflectance values.

To assess the capacity of this new dataset to infer manually obtained SSD observations, we constructed a baseline machine learning scheme with kernel, decision trees, and neural network-based ML schemes. The resulting ML ensemble, trained with 33,261 measurements originally part of Aquasat, has been updated and tested in the field on a water body crucial for millions of people [50]. Note that the aggregated analysis has allowed obtaining a general trend for water clarity at the Valle de Bravo Dam, as illustrated in Figure 7, which highlights potential applications of the present approach.

The techniques employed to solve the SDD problem have evolved significantly since Secchi's pioneering work [19]. These methods include polynomial models [20], empirical and semi-analytical algorithms [24], mechanistic models [25,27–29], and ML techniques such as generalized regression neural networks (GRNNs), sparse spectrum Gaussian-process regression (SSGPR), XGBoost, and random forest (RF). The present approach contributes by exploring various ML techniques to benchmark an updated dataset and apply it to a novel water body.

The selected regressors represent the most commonly used non-probabilistic approaches discussed in the literature [54], purposely chosen to offer a wide spectrum of criteria and ensembled to cover different dataset regions. Particularly for this application with inland water bodies, where the coarse resolution of satellite images and the punctual nature of SDD sampling pose challenges, the chosen approach proved appropriate. In our case, employing other techniques, such as CNNs [42], which are more suitable for open-water scenarios, is less feasible.

Our results show that for the case of calculating SDD, the L1TP collection offers acceptable results compared to those obtained from the L2SP collection. This result is consistent with Li et al. [66], who showed that the former provided better performance when calculating land surface temperature and emissivity from L1TP and L2SP. Other results, such as those by Sun et al. [67], show that L2SP outperforms L1TP in reflectance consistency, particularly in applications such as NDVI calculation. However, when atmospheric conditions are minimal, such as where the effects of scattering, absorption, and refraction are reduced or negligible, L1TP is sufficiently good. Some of the reasons for these results may be related to overcorrection in the L2SP collection, particularly concerning water turbidity, interpolation, or smoothing, especially along water–land transitions. In this sense, the L1TP collection may more faithfully preserve the original radiance values, particularly useful in complex environments such as water bodies. These ideas will be of interest to future research.

The results suggest that atmospheric correction in the L2SP does not improve the results obtained with L1TP. A possible way to gain intuition about this situation is by implementing atmospheric correction mechanisms on the L1TP collection, such as ACOLITE [68], 6S (Second Simulation of a Satellite Signal in the Solar Spectrum) [69], and iCOR [70], followed by a comparison with the results of the regressors built on top of the L2SP. This comparison will be the subject of future research.

This manuscript studies the construction of machine learning regression schemes from satellite images. Nonetheless, the number of available remote sensing platforms

useful for satellite-based observations has steadily increased over recent years [9], which could potentially enhance spatiotemporal analysis, particularly in dealing with seasonal or episodic events. It is important to note that the in situ observations and the satellite overflight dates must align for the construction of the regressors. In our case, AquaSat [43] was constructed using Landsat satellite passes. Nonetheless, several strategies have been proposed to harmonize remote sensing observations across different platforms [71,72], including Sentinel, PlanetScope, MODIS, and others. These approaches offer an exciting opportunity to enhance the spatiotemporal analysis of water bodies. However, this requires obtaining reference values from the newer platform to assess performance properly.

7. Conclusions

As pressure on water resources increases, continuous monitoring of water bodies becomes more important. While this is especially true for water bodies intended for human consumption and agriculture, it is essential to remember that water bodies are an integral and essential part of ecosystems on which the life and well-being of plants and animals depend. This article evaluates the feasibility of using machine learning algorithms and remote sensing from Landsat satellite images to estimate water clarity by determining SDD. In particular, we compare decision tree-based schemes, kernel-based approaches, neural networks, and regressor ensembles. Additionally, we compare the effectiveness of Collection 2 at its processing levels: Level-1 Terrain Precision (L1TP) and Level-2 Surface Product (L2SP). Moreover, we examine the performance obtained when using the central pixel, which indicates the geographical position of the reference sample and the use of pixels surrounding that position. The results indicate that machine learning-based approaches can effectively estimate SDD.

To demonstrate the proposed scheme, we reviewed a key water body that supplies water resources to CDMX. We conducted a historical analysis of water clarity, highlighting its temporal trends. During field visits, we collected samples, which were subsequently verified using the developed system, yielding satisfactory results. This study provides insights into the spatiotemporal distribution of water quality in strategic reservoirs, such as the Valle de Bravo Dam, which supplies Mexico's most populated city. The model was incorporated into a Geographical Information System used by CDMX authorities to monitor the water clarity of the Valle de Bravo Dam [73]. Consistent monitoring will support informed decision-making regarding water management and treatment and effective communication with citizens.

This research may be enriched with data from precipitation measurements, water runoff flows, demographic density analysis, and research into water currents within the dam, which would improve the understanding of water movement and augment the current system's efficacy. Additionally, it will be interesting to explore why the L1TP collection continues to provide competitive performance compared to the post-processed L2SP collection. Perhaps this can be supported by the implementation of alternative atmospheric correction methods. Another interesting direction for future research may involve harmonizing across multiple remote-sensing platforms to increase monitoring resilience in the face of seasonal or episodic events.

Author Contributions: Conceptualization, J.S. and R.S.; methodology, J.S. and R.S.; software, J.S.; validation, J.S., R.S. and P.V.; formal analysis, J.S., R.S. and P.V.; investigation, J.S., R.S. and P.V.; resources, J.S. and R.S.; data curation, J.S. and R.S.; writing—original draft preparation, J.S.; writing—review and editing, J.S., R.S. and P.V.; visualization, J.S. and R.S.; supervision, J.S.; project administration, J.S.; funding acquisition, J.S. and R.S. All authors have read and agreed to the published version of the manuscript.

Funding: Work partially funded by SIP-IPN 20240619 and SECTEI under grant 201/2021 for Joaquín Salas.

Data Availability Statement: Publicly available datasets were analyzed in this study. This data and code can be found at <https://github.com/joquinsal/as/waterClarity>, accessed on 13 January 2025.

Acknowledgments: Thanks to Mathew Ross and Francisco Vaca for insightful discussions concerning the topics of this paper.

Conflicts of Interest: The authors declare no conflicts of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript; or in the decision to publish the results.

References

1. Henriksen, H.Z.; Thapa, D.; Elbanna, A. Sustainable development goals in IS research. *Scand. J. Inf. Syst.* **2021**, *33*, 3.
2. Koncagül, E.; Connor, R.; UNESCO World Water Assessment Programme. *The United Nations World Water Development Report 2023: Partnerships and Cooperation for Water; Facts, Figures and Action Examples*; UNESCO: Paris, France, 2023.
3. Boretti, A.; Rosa, L. Reassessing the projections of the world water development report. *NPJ Clean Water* **2019**, *2*, 15. [[CrossRef](#)]
4. Wang, R.; Lu, Y.; Song, S.; Yang, S.; Wu, Y.; Cui, H. Industrial source discharge estimation for pharmaceutical and personal care products in China. *J. Clean. Prod.* **2022**, *381*, 135129. [[CrossRef](#)]
5. Majsztrik, J.C.; Fernandez, R.T.; Fisher, P.R.; Hitchcock, D.R.; Lea-Cox, J.; Owen, J.S.; Oki, L.R.; White, S.A. Water use and treatment in container-grown specialty crop production: A review. *Water Air Soil Pollut.* **2017**, *228*, 1–27. [[CrossRef](#)]
6. Benestad, R.E. Implications of a decrease in the precipitation area for the past and the future. *Environ. Res. Lett.* **2018**, *13*, 044022. [[CrossRef](#)]
7. Sánchez-García, C.; Francos, M. Human-environmental interaction with extreme hydrological events and climate change scenarios as background. *Geogr. Sustain.* **2022**, *3*, 232–236. [[CrossRef](#)]
8. García-López, M.; Cuadrado-Quesada, G.; Montano, B. Untangling the vicious cycle around water and poverty. *Sustain. Dev.* **2023**, *32*, 1845–1860. [[CrossRef](#)]
9. Kojima, A.; Yárnoz, D.G.; Di Pippo, S. Access to space: A new approach by the united nations office for outer space affairs. *Acta Astronaut.* **2018**, *152*, 201–207. [[CrossRef](#)]
10. Chuvieco, E. *Fundamentals of Satellite Remote Sensing: An Environmental Approach*; CRC Press: Boca Raton, FL, USA, 2020.
11. Ho, C.R.; Liu, A.K.; Li, X. *Remote Sensing Applications in Ocean Observation*; MDPI-Multidisciplinary Digital Publishing Institute: Basel, Switzerland, 2023.
12. Kikuchi, M.; Braun, S.A.; Suzuki, K.; Liu, G.; Battaglia, A. Satellite Precipitation Measurements: What Have We Learnt About Cloud-Precipitation Processes From Space? In *Clouds and Their Climatic Impacts: Radiation, Circulation, and Precipitation. 2023*; pp. 303–324. Available online: <https://agupubs.onlinelibrary.wiley.com/doi/10.1002/9781119700357.ch15> (accessed on 13 January 2025).
13. Schumann, G.; Giustarini, L.; Tarpanelli, A.; Jarihani, B.; Martinis, S. Flood modeling and prediction using earth observation data. *Surv. Geophys.* **2023**, *44*, 1553–1578. [[CrossRef](#)]
14. Minnett, P.; Alvera-Azcárate, A.; Chin, T.; Corlett, G.; Gentemann, C.; Karagali, I.; Li, X.; Marsouin, A.; Marullo, S.; Maturi, E.; et al. Half a century of satellite remote sensing of sea-surface temperature. *Remote Sens. Environ.* **2019**, *233*, 111366. [[CrossRef](#)]
15. Cretaux, J.F.; Calmant, S.; Papa, F.; Frappart, F.; Paris, A.; Berge-Nguyen, M. Inland surface waters quantity monitored from remote sensing. *Surv. Geophys.* **2023**, *44*, 1519–1552. [[CrossRef](#)]
16. Tottrup, C.; Druce, D.; Meyer, R.P.; Christensen, M.; Riffler, M.; Dulleck, B.; Rastner, P.; Jupova, K.; Sokoup, T.; Haag, A.; et al. Surface water dynamics from space: A round robin intercomparison of using optical and sar high-resolution satellite observations for regional surface water detection. *Remote Sens.* **2022**, *14*, 2410. [[CrossRef](#)]
17. Portela, C.F.; Martins, V.S.; Novo, E.M.; Paulino, R.S.; Barbosa, C.C. Recent advances in geostationary satellites for inland and coastal aquatic systems: Scientific research and applications. *Int. J. Remote Sens.* **2024**, *45*, 1574–1607. [[CrossRef](#)]
18. Cui, Z.; Huang, Q.; Sun, J.; Wan, B.; Zhang, S.; Shen, J.; Wu, J.; Li, J.; Yang, C. The Secchi disk depth to water depth ratio affects morphological traits of submerged macrophytes: Development patterns and ecological implications. *Sci. Total Environ.* **2024**, *907*, 167882. [[CrossRef](#)]
19. Secchi, P.A. Relazione delle esperienze fatte a bordo della pontificia pirocorvetta Imacolata Concezione per determinare la trasparenza del mare. *Il Nuovo C.* **1864**, *20*, 205–238.
20. Yu, D.; Xing, Q.; Lou, M.; Shi, P. Retrieval of Secchi disk depth in the Yellow Sea and East China Sea using 8-day MODIS data. In *Proceedings of the IOP Conference Series: Earth and Environmental Science*; IOP Publishing: Bristol, UK, 2014; Volume 17, p. 012112. [[CrossRef](#)]

21. Lee, Z.; Shang, S.; Hu, C.; Du, K.; Weidemann, A.; Hou, W.; Lin, J.; Lin, G. Secchi disk depth: A new theory and mechanistic model for underwater visibility. *Remote Sens. Environ.* **2015**, *169*, 139–149. [[CrossRef](#)]
22. Kim, S.H.; Yang, C.S.; Ouchi, K. Spatio-temporal patterns of Secchi depth in the waters around the Korean Peninsula using MODIS data. *Estuarine Coast. Shelf Sci.* **2015**, *164*, 172–182. [[CrossRef](#)]
23. Al Kaabi, M.R.; Zhao, J.; Ghedira, H. MODIS-based mapping of Secchi disk depth using a qualitative algorithm in the shallow Arabian Gulf. *Remote Sens.* **2016**, *8*, 423. [[CrossRef](#)]
24. Alikas, K.; Kratzer, S. Improved retrieval of Secchi depth for optically-complex waters using remote sensing data. *Ecol. Indic.* **2017**, *77*, 218–227. [[CrossRef](#)]
25. Rodrigues, T.; Alcântara, E.; Watanabe, F.; Imai, N. Retrieval of Secchi disk depth from a reservoir using a semi-analytical scheme. *Remote Sens. Environ.* **2017**, *198*, 213–228. [[CrossRef](#)]
26. Jin, H.; Fang, S.; Chen, C. Mapping of the Spatial Scope and Water Quality of Surface Water Based on the Google Earth Engine Cloud Platform and Landsat Time Series. *Remote Sens.* **2023**, *15*, 4986. [[CrossRef](#)]
27. Qing, S.; Cui, T.; Lai, Q.; Bao, Y.; Diao, R.; Yue, Y.; Hao, Y. Improving remote sensing retrieval of water clarity in complex coastal and inland waters with modified absorption estimation and optical water classification using Sentinel-2 MSI. *Int. J. Appl. Earth Obs. Geoinf.* **2021**, *102*, 102377. [[CrossRef](#)]
28. Guo, J.; Lu, J.; Zhang, Y.; Zhou, C.; Zhang, S.; Wang, D.; Lv, X. Variability of chlorophyll-a and secchi disk depth (1997–2019) in the Bohai Sea based on monthly cloud-free satellite data reconstructions. *Remote Sens.* **2022**, *14*, 639. [[CrossRef](#)]
29. Zhang, X.; Li, C.; Zhou, W.; Zheng, Y.; Cao, W.; Liu, C.; Xu, Z.; Yang, Y.; Yang, Z.; Chen, F. Study of the Profile Distribution of the Diffuse Attenuation Coefficient and Secchi Disk Depth in the Northwestern South China Sea. *Remote Sens.* **2023**, *15*, 1533. [[CrossRef](#)]
30. Alsahli, M.M.; Nazeer, M. Modeling Secchi Disk Depth Over the North Arabian Gulf Waters Using MODIS and MERIS Images. *PFG–J. Photogramm. Remote Sens. Geoinf. Sci.* **2022**, *90*, 177–189. [[CrossRef](#)]
31. Zhou, Y.; Liu, H.; He, B.; Yang, X.; Feng, Q.; Kutser, T.; Chen, F.; Zhou, X.; Xiao, F.; Kou, J. Secchi Depth estimation for optically-complex waters based on spectral angle mapping-derived water classification using Sentinel-2 data. *Int. J. Remote Sens.* **2021**, *42*, 3123–3145. [[CrossRef](#)]
32. Zhang, Y.; Shi, K.; Sun, X.; Zhang, Y.; Li, N.; Wang, W.; Zhou, Y.; Zhi, W.; Liu, M.; Li, Y.; et al. Improving remote sensing estimation of Secchi disk depth for global lakes and reservoirs using machine learning methods. *GIScience Remote Sens.* **2022**, *59*, 1367–1383. [[CrossRef](#)]
33. Golubkov, M.; Golubkov, S. Patterns of the relationship between the Secchi disk depth and the optical characteristics of water in the Neva Estuary (Baltic Sea): The influence of environmental variables. *Front. Mar. Sci.* **2024**, *11*, 1265382. [[CrossRef](#)]
34. Shamloo, A.; Sima, S. Investigating the potential of remote sensing-based machine-learning algorithms to model Secchi-disk depth, total phosphorus, and chlorophyll-a in Lake Urmia. *J. Great Lakes Res.* **2024**, *50*, 102370. [[CrossRef](#)]
35. Wang, Y.; Xiang, J.; Zhou, S. A variational optimization algorithm for Secchi disk depth based on Multi-satellite data. In *Proceedings of the Journal of Physics: Conference Series*; IOP Publishing: Bristol, UK, 2024; Volume 2718, p. 012008.
36. Zheng, P.; Xu, X.; Xu, X.; Huang, P.; Zhou, X. Study on multi-annual and simultaneous satellite-ground remote sensing retrieval of water transparency (Secchi disk depth) in Poyang lake. In *Proceedings of the International Conference on Image, Signal Processing, and Pattern Recognition (ISPP 2024)*, SPIE, Guangzhou, China, 1–3 March 2024; Volume 13180, pp. 539–545.
37. Feng, L.; Wang, Y.; Hou, X.; Qin, B.; Kuster, T.; Qu, F.; Chen, N.; Paerl, H.W.; Zheng, C. Harmful algal blooms in inland waters. *Nat. Rev. Earth Environ.* **2024**, *5*, 631–644. [[CrossRef](#)]
38. Youssef, Y.M.; Gemal, K.S.; Atia, H.M.; Mahdy, M. Insight into land cover dynamics and water challenges under anthropogenic and climatic changes in the eastern Nile Delta: Inference from remote sensing and GIS data. *Sci. Total Environ.* **2024**, *913*, 169690. [[CrossRef](#)] [[PubMed](#)]
39. Sun, J. Estimations of Secchi depth from the turbid Eastern China Coastal Seas using MODIS data. *Int. J. Remote Sens.* **2023**, *44*, 3209–3226. [[CrossRef](#)]
40. Gan, T.; Kalinga, O.; Ohgushi, K.; Araki, H. Retrieving seawater turbidity from Landsat TM data by regressions and an artificial neural network. *Int. J. Remote Sens.* **2004**, *25*, 4593–4615. [[CrossRef](#)]
41. Maciel, D.A.; Pahlevan, N.; Barbosa, C.C.; Martins, V.S.; Smith, B.; O’Shea, R.E.; Balasubramanian, S.V.; Saranathan, A.M.; Novo, E.M. Towards global long-term water transparency products from the Landsat archive. *Remote Sens. Environ.* **2023**, *299*, 113889. [[CrossRef](#)]
42. Schatz, J.; Morse, C.; Wanner, B.; Agarwal, T. LakeNet: Water Quality Monitoring with Satellite Images and CNNs. 2022. Available online: <https://christheissmorse.github.io/files/publications/lakenet.pdf> (accessed on 13 January 2025).
43. Ross, M.; Topp, S.; Appling, A.; Yang, X.; Kuhn, C.; Butman, D.; Simard, M.; Pavelsky, T. AquaSat: A data set to enable remote sensing of water quality for inland waters. *Water Resour. Res.* **2019**, *55*, 10012–10025. [[CrossRef](#)]
44. Yigit Avdan, Z.; Kaplan, G.; Goncu, S.; Avdan, U. Monitoring the water quality of small water bodies using high-resolution remote sensing data. *ISPRS Int. J. Geo-Inf.* **2019**, *8*, 553. [[CrossRef](#)]

45. McFeeters, S.K. The use of the Normalized Difference Water Index (NDWI) in the delineation of open water features. *Int. J. Remote Sens.* **1996**, *17*, 1425–1432. [CrossRef]
46. Hossain, A.; Jia, Y.; Chao, X. Development of remote sensing based index for estimating/mapping suspended sediment concentration in river and lake environments. In Proceedings of the 8th International Symposium on ECOHYDRAULICS, Seoul, Republic of Korea, 12–16 September 2010; Volume 435, pp. 578–585.
47. Wang, L.; Qu, J.J. NMDI: A normalized multi-band drought index for monitoring soil and vegetation moisture with satellite remote sensing. *Geophys. Res. Lett.* **2007**, *34*. [CrossRef]
48. Lacaux, J.; Tourre, Y.; Vignolles, C.; Ndione, J.; Lafaye, M. Classification of ponds from high-spatial resolution remote sensing: Application to Rift Valley Fever epidemics in Senegal. *Remote Sens. Environ.* **2007**, *106*, 66–74. [CrossRef]
49. Short, N.M. *The Landsat Tutorial Workbook: Basics of Satellite Remote Sensing*; National Aeronautics and Space Administration, Scientific and Technical Information Branch: Washington, DC, USA, 1982; Volume 1078.
50. del Agua (México). Gerencia Regional de Aguas del Valle de México y Sistema Cutzamala, C.N. In *Sistema Cutzamala: Agua Para Millones de Mexicanos*; Comisión Nacional del Agua, Gerencia Regional de Aguas del Valle de México: Ciudad de México, México, 2005.
51. Olvera Viascán, V. Estudio de Eutroficación de la Presa Valle de Bravo, México. 1992. Available online: <https://www.revistatyca.org.mx/index.php/tyca/article/view/687> (accessed on 13 January 2025).
52. Valencia, S.E.D. Modelación de las Cargas de Carbono en la Cuenca Hidrológica de la Presa de Valle de Bravo. Ph.D. Thesis, Universidad Nacional Autónoma de México, Mexico City, Mexico, 2020.
53. Google Earth. Available online: <https://earth.google.com/web/@19.2249,-100.1953,1789.9684a,24022.6242d,35y,0h,0t,0r> (accessed on 1 October 2024).
54. Hastie, T.; Tibshirani, R.; Friedman, J. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*; Springer: Berlin/Heidelberg, Germany, 2009; Volume 2.
55. Carreira-Perpiñán, M.Á.; Hada, S.S. Counterfactual explanations for oblique decision trees: Exact, efficient algorithms. In Proceedings of the AAAI Conference on Artificial Intelligence, Online, 2–9 February 2021; Volume 35, pp. 6903–6911.
56. Bengio, Y.; Lecun, Y.; Hinton, G. Deep learning for AI. *Commun. ACM* **2021**, *64*, 58–65. [CrossRef]
57. Fernández-Delgado, M.; Sirsat, M.S.; Cernadas, E.; Alawadi, S.; Barro, S.; Febrero-Bande, M. An extensive experimental survey of regression methods. *Neural Netw.* **2019**, *111*, 11–34. [CrossRef] [PubMed]
58. Chen, T.; He, T.; Benesty, M.; Khotilovich, V.; Tang, Y.; Cho, H.; Chen, K.; Mitchell, R.; Cano, I.; Zhou, T.; et al. *Xgboost: Extreme Gradient Boosting, R Package Version 0.4-2*; 2015; Volume 1, pp. 1–4. Available online: <https://cran.r-project.org/web/packages/xgboost/vignettes/xgboost.pdf> (accessed on 13 January 2025)
59. Awad, M.; Khanna, R.; Awad, M.; Khanna, R. Support vector regression. In *Efficient Learning Machines: Theories, Concepts, And Applications for Engineers and System Designers*; Springer Nature: Berlin/Heidelberg, Germany, 2015; pp. 67–80. [CrossRef]
60. Specht, D.F. A general regression neural network. *IEEE Trans. Neural Netw.* **1991**, *2*, 568–576. [CrossRef] [PubMed]
61. Afendras, G.; Markatou, M. Optimality of training/test size and resampling effectiveness in cross-validation. *J. Stat. Plan. Inference* **2019**, *199*, 286–301. [CrossRef]
62. Chicco, D.; Warrens, M.; Jurman, G. The coefficient of determination R-squared is more informative than SMAPE, MAE, MAPE, MSE and RMSE in regression analysis evaluation. *PeerJ Comput. Sci.* **2021**, *7*, e623. [CrossRef]
63. Prince, S. *Computer Vision: Models, Learning, and Inference*; Cambridge University Press: Cambridge, UK, 2012. [CrossRef]
64. USGS. *Landsat 8-9 Collection 2 (C2) Level 2 Science Product (L2SP) Guide*; Technical report; United States Geological Survey: Asheville, NC, USA, 2022.
65. KC, M.; Leigh, L.; Pinto, C.T.; Kaewmanee, M. Method of Validating Satellite Surface Reflectance Product Using Empirical Line Method. *Remote Sens.* **2023**, *15*, 2240. [CrossRef]
66. Li, X.J.; Wu, H.; Ni, L.; Cheng, Y.L.; Zhang, X.X. A General Framework for Retrieving Land Surface Emissivity and Temperature Using Sensors with Split-Window Thermal Infrared Channels: A Case Study with Landsat 9. *IEEE Trans. Geosci. Remote Sens.* **2024**, *62*, 5008012. [CrossRef]
67. Sun, Y.; Wang, B.; Teng, S.; Liu, B.; Zhang, Z.; Li, Y. Continuity of Top-of-Atmosphere, Surface, and Nadir BRDF-Adjusted Reflectance and NDVI between Landsat-8 and Landsat-9 OLI over China Landscape. *Remote Sens.* **2023**, *15*, 4948. [CrossRef]
68. Vanhellemont, Q.; Ruddick, K. Acolite for Sentinel-2: Aquatic applications of MSI imagery. In Proceedings of the 2016 ESA Living Planet Symposium, Prague, Czech Republic, 9–13 May 2016; Volume 9.
69. Vermote, E. 6S User Guide. In *Second Simulation of the Satellite Signal in the Solar Spectrum*. 1996. Available online: https://ltdri.org/files/6S/6S_Manual_Part_1.pdf (accessed on 13 January 2025).
70. De Keukelaere, L.; Sterckx, S.; Adriaensen, S.; Knaeps, E.; Reusen, I.; Giardino, C.; Bresciani, M.; Hunter, P.; Neil, C.; Van der Zande, D.; et al. Atmospheric correction of Landsat-8/OLI and Sentinel-2/MSI data using iCOR algorithm: Validation for coastal and inland waters. *Eur. J. Remote Sens.* **2018**, *51*, 525–542. [CrossRef]

71. Shao, Z.; Cai, J.; Fu, P.; Hu, L.; Liu, T. Deep learning-based fusion of Landsat-8 and Sentinel-2 images for a harmonized surface reflectance product. *Remote Sens. Environ.* **2019**, *235*, 111425. [[CrossRef](#)]
72. Saunier, S.; Louis, J.; Debaecker, V.; Beaton, T.; Cadau, E.G.; Boccia, V.; Gascon, F. Sen2like, a tool to generate Sentinel-2 harmonised surface reflectance products—first results with Landsat-8. In Proceedings of the IGARSS 2019-2019 IEEE International Geoscience and Remote Sensing Symposium, Yokohama, Japan, 28 July–2 August 2019; pp. 5650–5653.
73. Salas, J.; Arenas, J.C.; Briseño, A. Monitoreo de fenómenos sociales y ambientales mediante observaciones de la Tierra. *Ciencia* **2025**, *forthcoming*.

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