Estimation of Water Quality Parameters through a Combination of Deep Learning and Remote Sensing Techniques in a Lake in Southern Chile

Lien Rodríguez-López, David Bustos Usta, Iongel Duran-Llacer, Lisandra Bravo Alvarez, Santiago Yépez, Luc Bourrel, Frederic Frappart, Roberto Urrutia

Article

Estimation of Water Quality Parameters through a Combination of Deep Learning and Remote Sensing Techniques in a Lake in Southern Chile

Lien Rodríguez-López [1,*], **David Bustos Usta** [2], **Iongel Duran-Llacer** [3], **Lisandra Bravo Alvarez** [4], **Santiago Yépez** [5], **Luc Bourrel** [6], **Frederic Frappart** [7] and **Roberto Urrutia** [8]

[1] Facultad de Ingeniería, Arquitectura y Diseño, Universidad San Sebastián, Lientur 1457, Concepción 4030000, Chile

[2] Facultad de Oceanografía, Universidad de Concepción, Concepción 4030000, Chile; davidbustos@udec.cl

[3] Hémera Centro de Observación de la Tierra, Facultad de Ciencias, Ingeniería y Tecnología, Universidad Mayor, Camino La Pirámide 5750, Huechuraba 8580745, Chile; iongel.duran@umayor.cl

[4] Department of Electrical Engineering, Universidad de Concepción, Edmundo Larenas 219, Concepción 4030000, Chile; lisanbravo@udec.cl

[5] Department of Forest Management and Environment, Faculty of Forestry, Universidad de Concepcion, Calle Victoria, Concepción 4030000, Chile; syepez@udec.cl

[6] Géosciences Environnement Toulouse, UMR 5563, Université de Toulouse, CNRS-IRD-OMP-CNES, 31000 Toulouse, France; luc.bourrel@ird.fr

[7] INRAE, Bordeaux Sciences Agro, UMR 1391 ISPA, Université de Bordeaux, 33604 Talence, France; frederic.frappart@inrae.fr

[8] Facultad de Ciencias Ambientales, Universidad de Concepción, Concepción 4030000, Chile; rurrutia@udec.cl

* Correspondence: lien.rodriguez@uss.cl; Tel.: +56-999168115

Abstract: In this study, we combined machine learning and remote sensing techniques to estimate the value of chlorophyll-a concentration in a freshwater ecosystem in the South American continent (lake in Southern Chile). In a previous study, nine artificial intelligence (AI) algorithms were tested to predict water quality data from measurements during monitoring campaigns. In this study, in addition to field data (Case A), meteorological variables (Case B) and satellite data (Case C) were used to predict chlorophyll-a in Lake Llanquihue. The models used were SARIMAX, LSTM, and RNN, all of which showed generally good statistics for the prediction of the chlorophyll-a variable. Model validation metrics showed that all three models effectively predicted chlorophyll as an indicator of the presence of algae in water bodies. Coefficient of determination values ranging from 0.64 to 0.93 were obtained, with the LSTM model showing the best statistics in any of the cases tested. The LSTM model generally performed well across most stations, with lower values for MSE (<0.260 (μg/L)^2), RMSE (<0.510 ug/L), MaxError (<0.730 μg/L), and MAE (<0.442 μg/L). This model, which combines machine learning and remote sensing techniques, is applicable to other Chilean and world lakes that have similar characteristics. In addition, it is a starting point for decision-makers in the protection and conservation of water resource quality.

Keywords: water quality; chlorophyll; remote sensing; deep learning; Chile; lakes

1. Introduction

Eutrophication is a phenomenon that occurs in lakes when excess nutrients such as phosphorus and nitrogen accumulate in the water [1,2]. This leads to the growth of algae and other aquatic plants, which can deplete oxygen levels in water and create harmful algal blooms [3,4]. The effects of eutrophication can be devastating to lake ecosystems, leading to the death of fish and other aquatic species [5,6]. Interventions to reduce eutrophication in lake watersheds include decreasing fertilizer use in nearby agriculture, limiting the discharge of raw sewage into the lake, and introducing species that consume excess nutrients, such as carp and tilapia [7,8]. In addition, other practices carried out by lake managers such

as dredging and aeration are causing an increase in oxygen levels and reducing nutrient concentrations [9,10].

Chlorophyll-a (Chl-a) is a green pigment found in the chloroplasts of algae, plants, and some bacteria that is responsible for capturing light energy during photosynthesis [11,12]. It is often used as an indicator of algae presence because it is a primary photosynthetic pigment found in all types of algae [13,14]. The concentration of Chl-a is directly proportional to the number of algae present in a water sample, making it a useful tool for monitoring algal growth and detecting harmful algal blooms [15,16]. Additionally, Chl-a can be used to estimate the primary productivity of aquatic ecosystems, which is an important factor for understanding the health and functioning of these systems [17,18].

Remote sensing techniques, combined with artificial intelligence models, have revolutionized the way scientists study, and manage the Earth's natural resources [19,20]. These techniques involve the use of sensors to collect data remotely, often from satellites, aircraft or drones [21–23]. By using artificial intelligence algorithms to analyze the collected data, researchers can gain insights into environmental patterns and make predictions about future trends [24,25]. With the help of machine learning models, scientists can develop early warning systems to detect harmful algal blooms, helping to mitigate the negative effects of eutrophication on lake ecosystems [26–28]. Moreover, multiple remote sensing studies have been used to monitor water quality in lakes and identify changes in nutrient concentrations that could lead to eutrophication [29–32]. The combination of remote sensing and artificial intelligence provides a powerful tool for understanding and managing complex environmental systems.

Chile has several lake districts, from north to south: the district of Altiplanic Lakes, Nabuelbutan Lakes, Araucanian Lakes, Chiloe Lakes, and Nordpatagonian Lakes or Paine Towers. Araucanian lakes stand out for their economic, social, and environmental importance. Lake Llanquihue is the largest lake in this chain and there is little scientific knowledge about aspects of its water quality, which is why it was selected for study in a previous investigation and specifically in the present one. The objective of this study is to contribute through combined techniques of remote sensing and machine learning to develop early tools for monitoring lakes in the follow-up of algal bloom phenomena. For this, we will follow the following specific objectives: (i) analyze the behavior of the physicochemical and biological variables best related to algal bloom events during the period 1989–2021; (ii) train artificial intelligence models with real in situ data of limnological and meteorological variables and data from Landsat satellite image sources and (iii) estimate the concentration of chlorophyll-a in the lake for seasons of the year where monitoring data are not available and validate these results with data from monitoring campaigns.

2. Materials and Methods

2.1. Site Description

Lake Llanquihue (41°08′S and 72°47′W) is a large freshwater lake located in southern Chile, in the Los Lagos Region [33]. The lake is located at an altitude of 51 m above sea level (m.a.s.l.) and has an area of approximately 870 km^2 (Figure 1), making it one of the largest lakes in Chile [34]. It is also one of the most emblematic natural landmarks in the region and is surrounded by the impressive backdrop of the Andes Mountains. The lake is fed by several rivers and streams, such as the Maullín and Petrohué rivers, which flow into the lake from east and west, respectively. The mouth of Lake Llanquihue is the Maullín River that flows into the Pacific Ocean. The lake has a maximum depth of 317 m and an average depth of 207 m, making it one of the deepest lakes in South America [35]. Lake Llanquihue is surrounded by several active volcanoes, such as Osorno, Calbuco and Puntiagudo, which contribute to the region's rugged and dramatic landscape [36]. The lakeshore is characterized by beaches, cliffs, and rocky outcrops, with a variety of flora and fauna present in the surrounding forests and wetlands. The climate around Lake Llanquihue is classified as temperate oceanic, with mild temperatures and high humidity throughout the year. The average annual temperature is around 11 °C, with averages ranging from

5 °C in winter to 16 °C in summer [37]. The region is also known for its frequent rainfall, especially during the winter months, which contributes to the lush vegetation and fertile soil of the surrounding area. Overall, Lake Llanquihue is an impressive natural feature and the product of the dynamic geological and climatic forces that have shaped the landscape of southern Chile. Its deep, crystal-clear waters, surrounded by volcanoes and green forests, make it a popular destination for outdoor activities such as hiking, fishing, and kayaking [38].
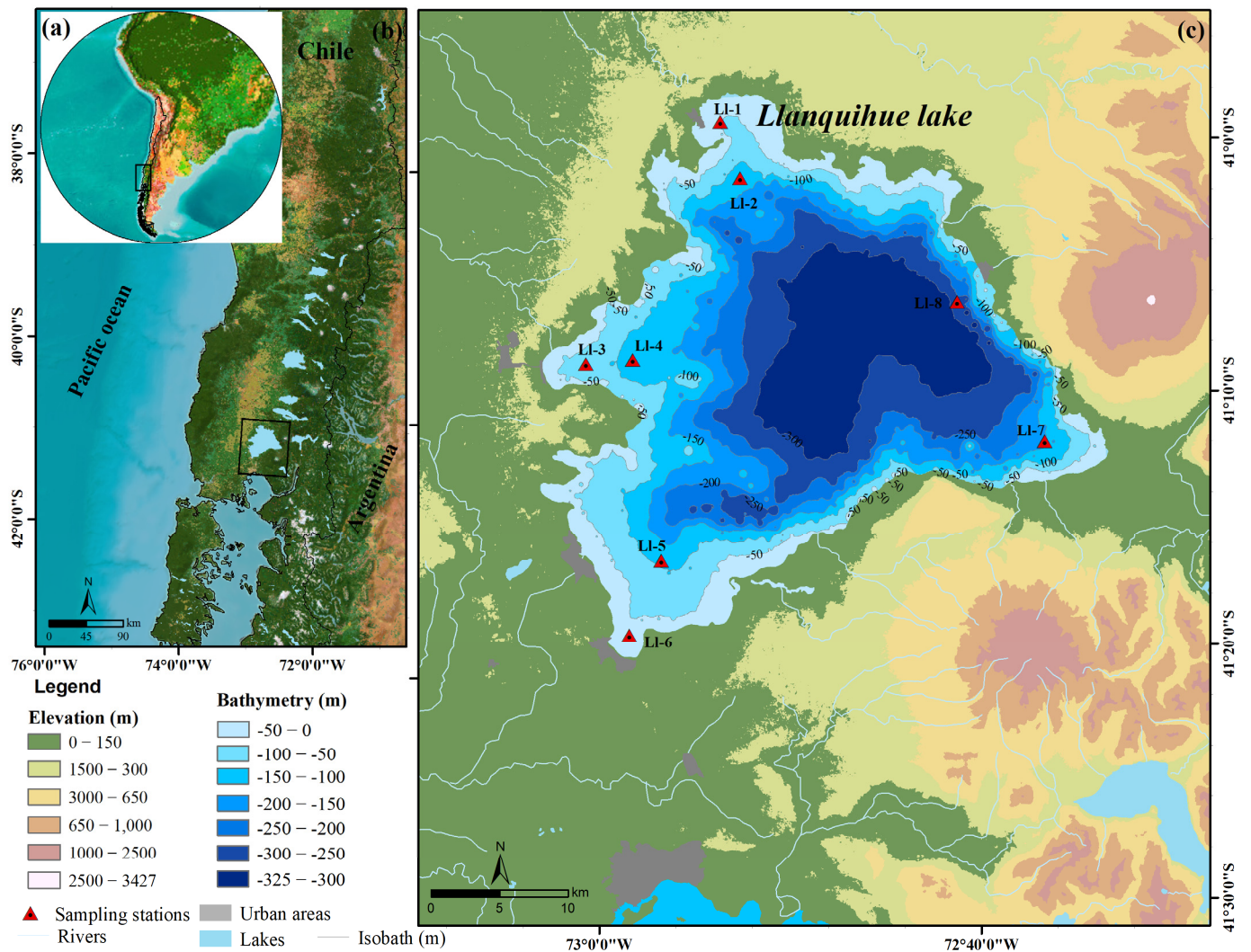

Figure 1. Location of the study area and sampling stations: (**a**) Chile in South America context, (**b**) location of Llanquihue lake in Chile, and (**c**) Llanquihue lake bathymetry and sampling stations (represented by red triangles).

2.2. Sample Collection

The Dirección General de Aguas de Chile (DGA for its acronym in Spanish) has been monitoring a group of lakes in Chile since 1986. Lake Llanquihue is within the selected group because it is the second-largest lake in the Chilean territory and because of its economic-social-cultural importance. The monitoring campaigns carried out in all seasons of the year consisted of sampling and in situ measurements of parameters at eight stations located in the lake (Ll1-Ll8, see Figure 1).

The collection of field data on physicochemical and biological parameters, as well as water samples for algae and pigment identification, is an essential part of water quality monitoring. The data collected will provide valuable information for understanding the

health of this water body and identifying potential environmental threats, such as the possible occurrence of algal blooms. The in-situ parameters selected for this study were Secchi disk depth (SD), chlorophyll-a (Chl-a) (Standard Methods N°10,200 H DGALGOCL1/2009), temperature (°C), total nitrogen (Nt) (Standard Methods N°4500-N C) and total phosphorus (Pt) (Standard Methods N°4500-P E). By following a standardized protocol and keeping detailed records, researchers and water managers can make informed decisions regarding water body management.

2.3. Preprocessing of Landsat 8 Satellite Images

Landsat-8 (L-8 OLI) images were used with a low percentage of clouds (less than 11%) covering the Llanquihue Lake (path/row: 233/89). L-8 is an Earth observation satellite of the Landsat project operated by the National Aeronautics and Spatial Administration (NASA) and the United States Geological Survey (USGS) [39,40]. It has two sensors, the OLI (Operational Land Imager) which provides nine bands in the visible, near-infrared, and shortwave spectra and covers from 0.433 µm to 1.390 µm, and the TIRS sensor (Thermal Infrared Sensor), which covers from 10.30 µm to 12.50 µm [41]. The 14 multispectral images used have a 30-m spatial resolution and were obtained from the USGS Earth Explorer (https://earthexplorer.usgs.gov/, accessed on 7 January 2023). The orthorectified and corrected images of the terrain of Collection 2 Level 1 were selected considering, the closeness to the sampling date and availability (see Table S1).

Considering a previous visual inspection through the Quality Assessment band (QA) and the Region of Interest (ROI), the images were atmospherically corrected in the ACO-LITE software (version 20211124.0) from https://github.com/acolite and accessed on 10 February 2023. ACOLITE is a generic processor that was developed specifically for marine, coastal, and inland waters, and brings together the atmospheric correction algorithms and software developed at RBINS for processing of images satellites applied to aquatic remote sensing [18,42,43]. ACOLITE uses a default atmospheric correction based on Dark Spectrum Fitting (DSF) [44–46] and Exponential Extrapolation (EXP) [43,47,48] algorithms.

From the resulting bands representing the surface-level reflectance (ρs) for L-8, the values were extracted in a matrix of 3×3 pixels per sampling point, according to [49]. Pixel values were extracted using ArcGIS software (ESRI's v. 10.8.2). Only data from cloud-free areas were used to have high-quality data and to avoid affecting the accuracy of the chlorophyll concentration estimate. These values were obtained from five multispectral bands: blue (B), green (G), red (R), near-infrared (NIR), and shortwave infrared (SWIR). In addition, the values of the Normalized Difference Vegetation Index (NDVI) and Floating Algal Index (FAI) were used. Both indexes are algorithms included in ACOLITE as part of the recovery of parameters derived from reflectance [50,51]. The limits of the lake were acquired from the DGA, and only the water body was considered for the analysis (DGA, 2023) [52].

Single bands are widely used to correlate with in situ data and estimate water quality parameters, such as chlorophyll [30,53], total suspended solids [54], turbidity [55], and temperature [56]. Surface reflectance values have shown good performance in these estimations and are even being used in artificial intelligence [57,58], including in algal bloom detection [24,32,55,59]. NDVI and FAI indices have been used in research related to chlorophyll and algal bloom estimation with good precision [18,60,61]. The NDVI is a commonly used indicator of vegetation photosynthetic activity and has been widely used in algae and chlorophyll extraction studies [62,63]. On the other hand, FAI is defined as a linear spread of reflectivity in the near-infrared, red, and shortwave infrared regions and can be applied to monitor algal blooms. The observation results of this algorithm provide a high accuracy [61,63,64].

2.4. Prediction Using Statistical and Deep Learning Models

2.4.1. SARIMAX

The SARIMAX (Seasonal Autoregressive Integrated Moving Average with Exogenous Variables) model is a type of time series model. It is an extension of the ARIMA (Autore-

gressive Integrated Moving Average) model that incorporates both seasonal and exogenous components [65]. The mathematical model is defined by Equation (1):

$$y_t = \beta_t x_t + u_t$$
$$\varphi_p(L)\tilde{\phi}_p(L^s)\Delta^d\Delta_s^D u_t = A(t) + \theta_q(L)\tilde{\theta}_Q(L^s)\zeta \tag{1}$$

where β in the first part of the formula represents external variables. The model is similar to the SARIMA model, with the following hyperparameters [66]:

p represents the order for the Autoregressive part (AR)
q represents the order for the moving average part (MA)
I represents the differencing order
P represents the seasonal AR order
Q represents the seasonal MA order
D represents the seasonal differencing
s represents the seasonal coefficients

The complete data pipeline used is shown in Figure 2. Raw data were obtained for each station (Figure 1). Subsequently, data were resampled at monthly intervals. Autocorrelation (ACF) and Partial Autocorrelation (PACF) functions were computed. Additionally, the Dickey Fuller test [67] was employed to examine whether the chlorophyll-a time series exhibited the characteristics of white noise. Subsequently, a seasonal order was applied, and, if necessary, a 1-step differentiation was performed. The optimal hyperparameters were determined through calibration using the pmdarima library (https://github.com/alkaline-ml/pmdarima, (accessed on 20 February 2023)), and the range of values for calibration was established based on the ACF and PACF results. Finally, the most suitable model was selected based on Akaike Information Criteria (AIC) and Bayesian Information Criteria (BIC).
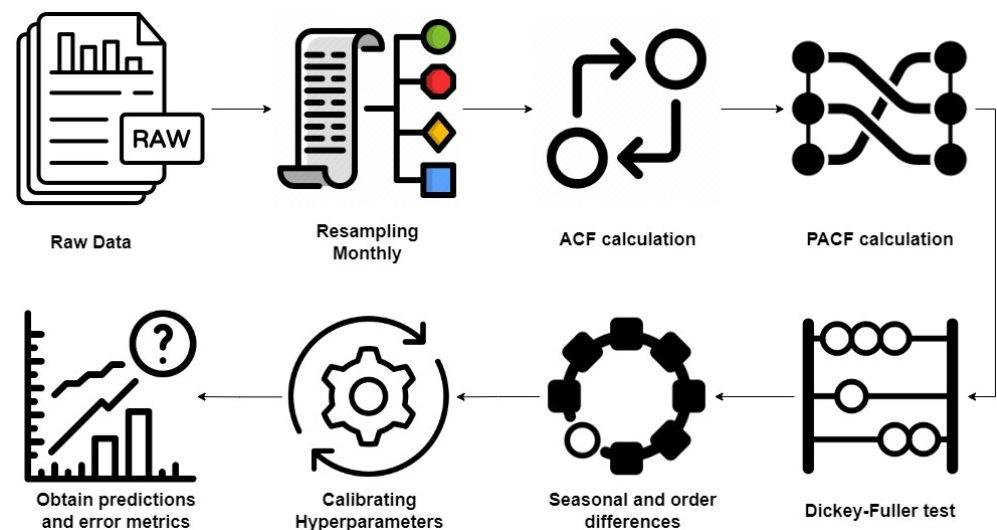

Figure 2. Flowchart of the integrated autoregressive moving average ARIMA model.

2.4.2. Long Short-Term Memory (LSTM)

Subsequently, Long short-term memory (LSTM) was used, which is a variant of a Recurrent neural network (RNN) proposed by Hochreiter and Schmidhuber in 1997 [68]. This algorithm solves the long-term dependency problem in RNNs by introducing memory (C) and an appropriate gate structure.

The LSTM cell (Figure 3) has four gates: input (*i*), forget (*f*), control (*c*) and output gates (*o*). The input gate determines the information that can be inserted and transferred to the cell:

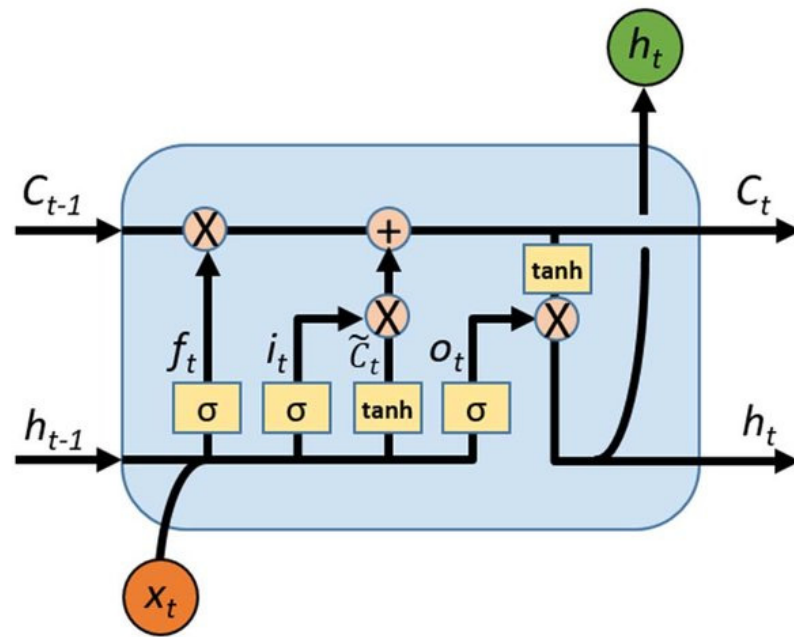$$i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i) \tag{2}$$

Figure 3. Structure of a Long Short-Term Memory (LSTM) cell. Adapted from [69].

The forget gate decides which information from the input is important from previous memory with:

$$f_t = \sigma\left(W_f \cdot [h_{t-1}, x_t] + b_f\right) \tag{3}$$

The control gate stabilizes the update in the cell state from C_{t-1} to C_t using Equations (4) and (5):

$$\widetilde{C_t} = \tanh(W_C \cdot [h_{t-1}, x_t] + b_c) \tag{4}$$

$$C_t = f_t \times C_{t-1} + \widetilde{C_t} \tag{5}$$

The output gate generates the output updating the hidden vector h_{t-1} with Equations (6) and (7):

$$o_t = \sigma_t \times \tanh(C_t) \tag{6}$$

$$h_t = o_t \times \tanh(C_t) \tag{7}$$

where σ is the activation function, W corresponds to the weight matrices calibrated during the training process, tanh is used to scale values in the range of -1 to 1, and b represents the bias in each step. During the training process, a lag of 9 is constructed from the input variables. An LSTM layer with a variable number of cells ranging from 30 to 50 was used based on the size and complexity of the dataset. Furthermore, a Dense layer is employed as the output layer to facilitate accurate predictions. The latter topology is considered to be a common configuration in the LSTM algorithm [69]. The complete training structure is described in Figure 4.

2.4.3. Recurrent Neural Networks (RNN)

We incorporated the architecture of Recurrent Neural Networks (RNN) to assess the performance of the LSTM models relative to their predecessor. RNN are a class of neural networks that are suitable for sequential data such as time series [68]. This algorithm is just a feed-forward neural network that unfolds over time (Figure 5). At each time step, the network produces an intermediate output o_t and maintains an internal state s_t; therefore, x_t creates the sequential input that is given to the network following Equations (8)–(11):

$$a_t = b + Ws_{t-1} + Ux_t \tag{8}$$

$$s_t = \tanh(a_t) \tag{9}$$

$$o_t = c + Vs_t \tag{10}$$

$$y_t = softmax(o_t) \tag{11}$$

where *U, V* and *W* represent matrices with the parameters of the model learned by standard propagation, b represents the bias, and y_t represents the final output and linear represents the activation function which in this case is the identity transformation defined in Equation (12) as previously described by Rumelhart et al. (1986) [70].

$$linear(o_t) = o_t$$

Figure 4. Schematic representation of LSTM structure used for the training in the LI-7 station. k represents the number of variables (including independent and dependent outcome) depending on the case it could be 9 (A),12 (B) or 14 (C). L represent the number of LSTM units for the training with values between 7 and 50 depending on the data complexity.

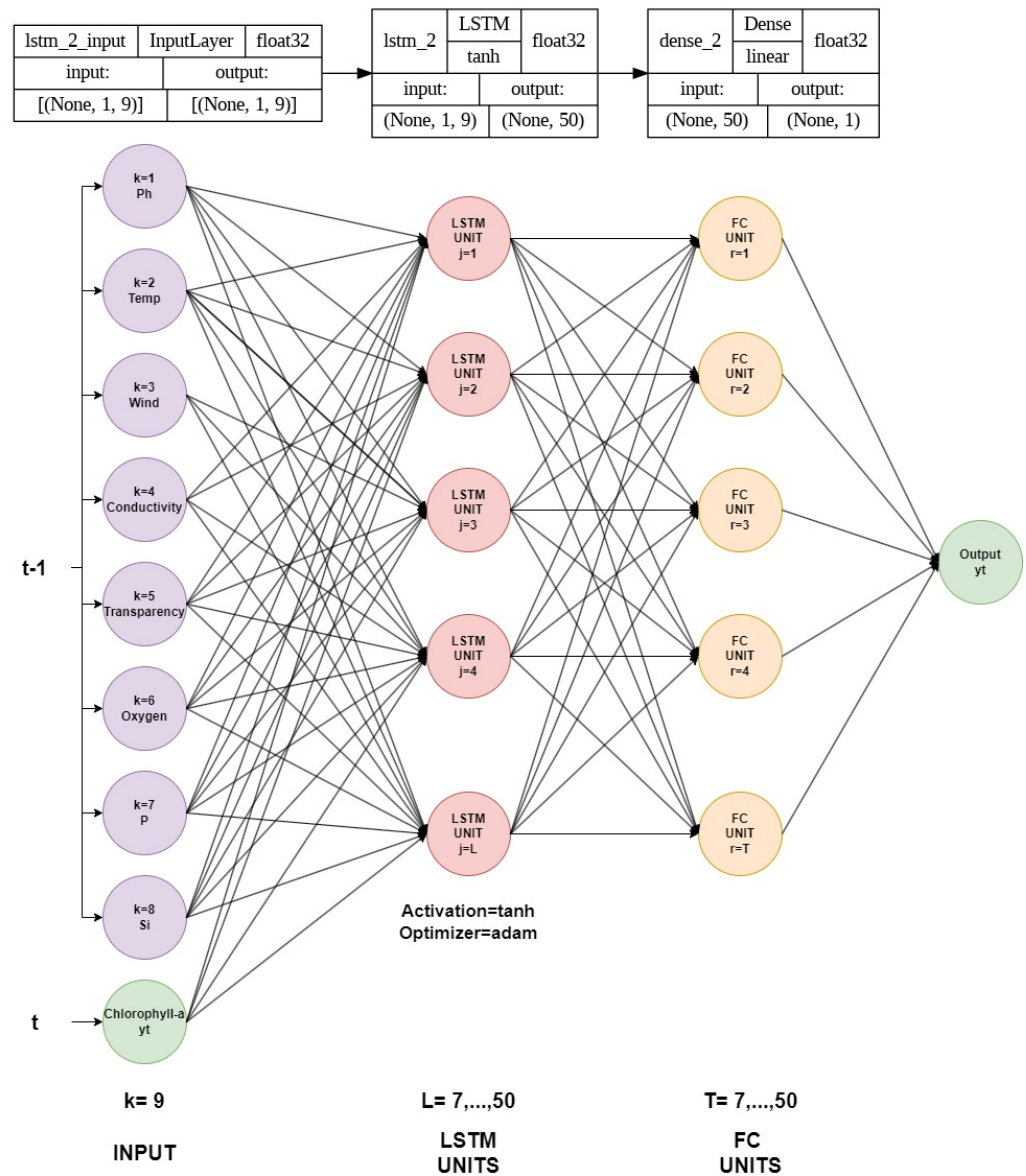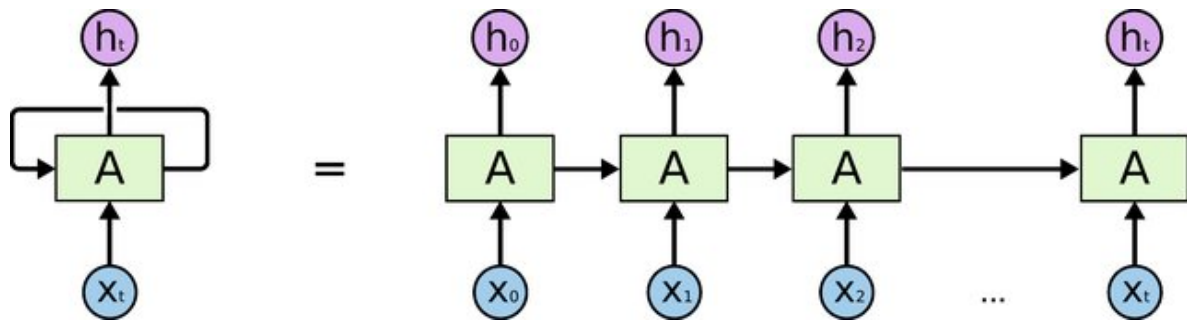**Figure 5.** RNN network architecture with a loop (adapted from [71]).
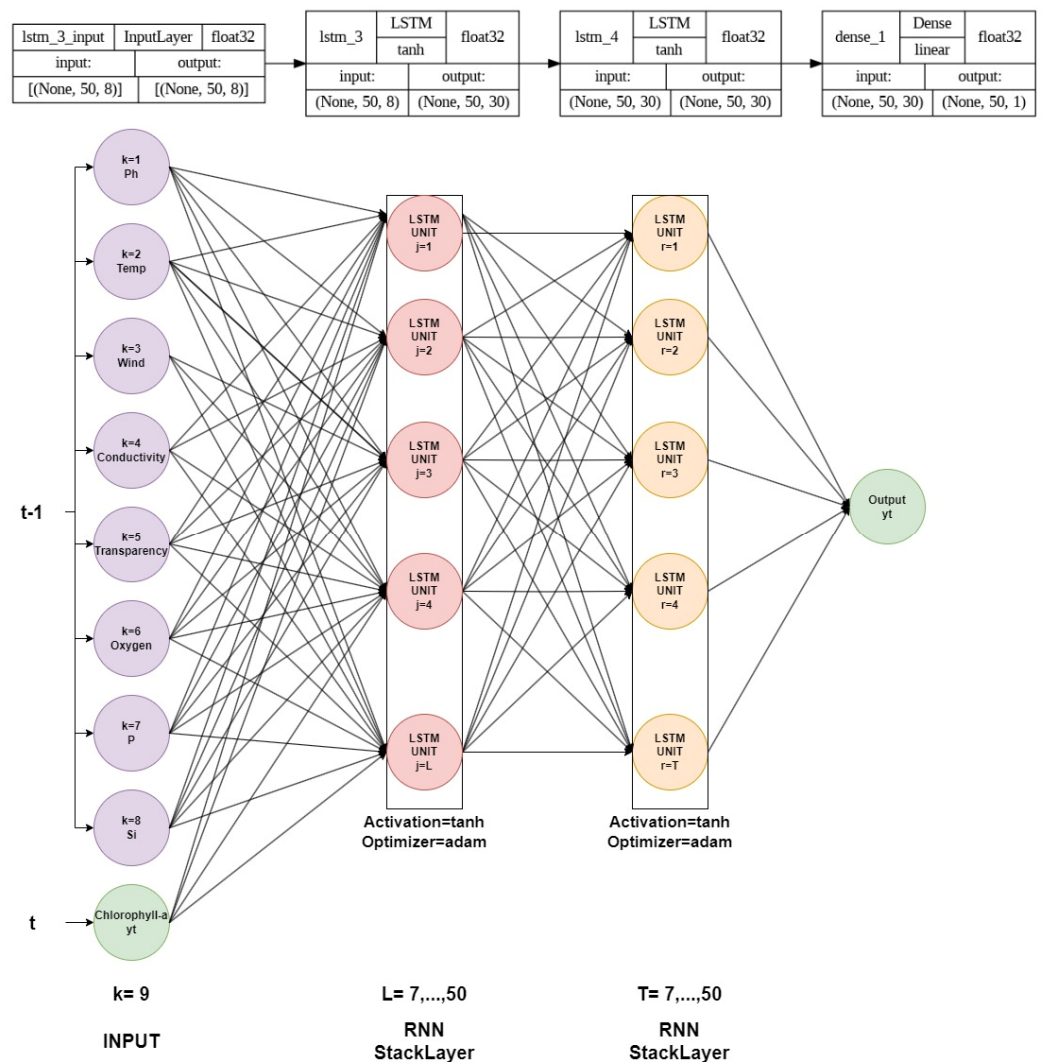
Figure 6. Schematic representation of RNN structure used for the training in the LI-7 station. k represents the number of variables (including independent and dependent outcome) depending on the case it could be 9 (A),12 (B) or 14 (C). L and T represent the number of LSTM units for the training with values between 10 and 30 depending on the data complexity.

To train this network, we use a sliding window technique to generate training data for a time-series prediction model using a window of size lookback between 15 and 30 from the input time series. The corresponding target value (chlorophyll-a) was set as the next time step after the input sequence. In this way, the model can learn from the input-output pairs and make predictions based on the observed patterns in the training data.

Therefore, the data enter the RNN topology passing through a number of units between 16 and 32, depending on the amount of data and complexity, and finally, a dense layer produces the predictions. The optimizer and loss metric used were Adam and mean squared error, respectively. This configuration has been widely used in previous studies on time-series predictions [68]. The complete training structure is described in Figure 6.

To test the three models described above, cases A, B, and C are defined as follows:

Case A (Measurement Data): In the first case, we included the real variables measured in the monitoring campaigns for the four seasons of the year and in the eight stations of the lake.

Case B (Measurement and Meteorological Data): In addition to the actual variables, we included meteorological data as conditioning variables that can influence the autochthonous processes of the lake.

Case C (Measurement, Meteorological Data and Satellite Data): In this case, we include bands and indices from the L-8 satellite image processing.

2.5. Statistical Validation

To analyze the performance of the models defined in Section 2.4, several metrics were used, such as Mean Squared Error (MSE) as described in [72], Root Mean Squared Error (RMSE) described in [30], Mean Absolute Error (MAE), Maximum Error describe in [73] and R^2 described in [74] following a similar approach as the one applied in [35]. Thus, it helps in understanding the accuracy, precision, and potential limitations of estimating chlorophyll-a. Number of samples for train and test are described in Table 1. Sequential splitting with a 70/30% rule was used to calculate the different error metrics. In this method the time series are separated as follows, the earlier (later) is used to train (validate) each model across the different stations (Figure 1) and have been demonstrated good performance to assess time-series performance in deep learning models [75,76].

Table 1. Number of samples used for (train/test) in the train and validation phases for each station and case analyzed.

Case	LI-1	LI-2	LI-3	LI-4	LI-5	LI-6	LI-7	LI-8
Case A	238/59	77/19	238/59	92/23	238/59	80/35	332/83	67/29
Case B	-	36/8	-	36/10	-	36/10	36/12	36/8
Case C	-	36/8	-	36/10	-	36/10	36/12	36/8

On the other hand, we used a method called Garson's weighting to assign importance or weights to the input variables in neural networks. This method provides a measure of the relative contribution of each predictor variable explaining the variation in the dependent variable [77]. To obtain weights the method uses the magnitude and direction of the coefficients obtained from the model. Variables with larger absolute coefficients are considered to have higher importance. In this way, the most important variables in the model were obtained for each case of study.

3. Results

3.1. Limnological Behavior and Meteorological Data

The water quality and trophic level of a lake primarily depend on nutrient inputs (Nitrogen and Phosphorus) from the watershed. This is why they are selected in conjunction with the transparency, the temperature, and the study variable chlorophyll a. All these parameters influence the spatio-temporal distribution of algae. Figure 7 shows the behavior of limnological variables associated with algal blooms. The boxplot shows the distribution of the numerical dataset of the variables indicated by five key statistics: minimum value, first quartile (Q1), median (Q2), third quartile (Q3), and maximum value (Q4).
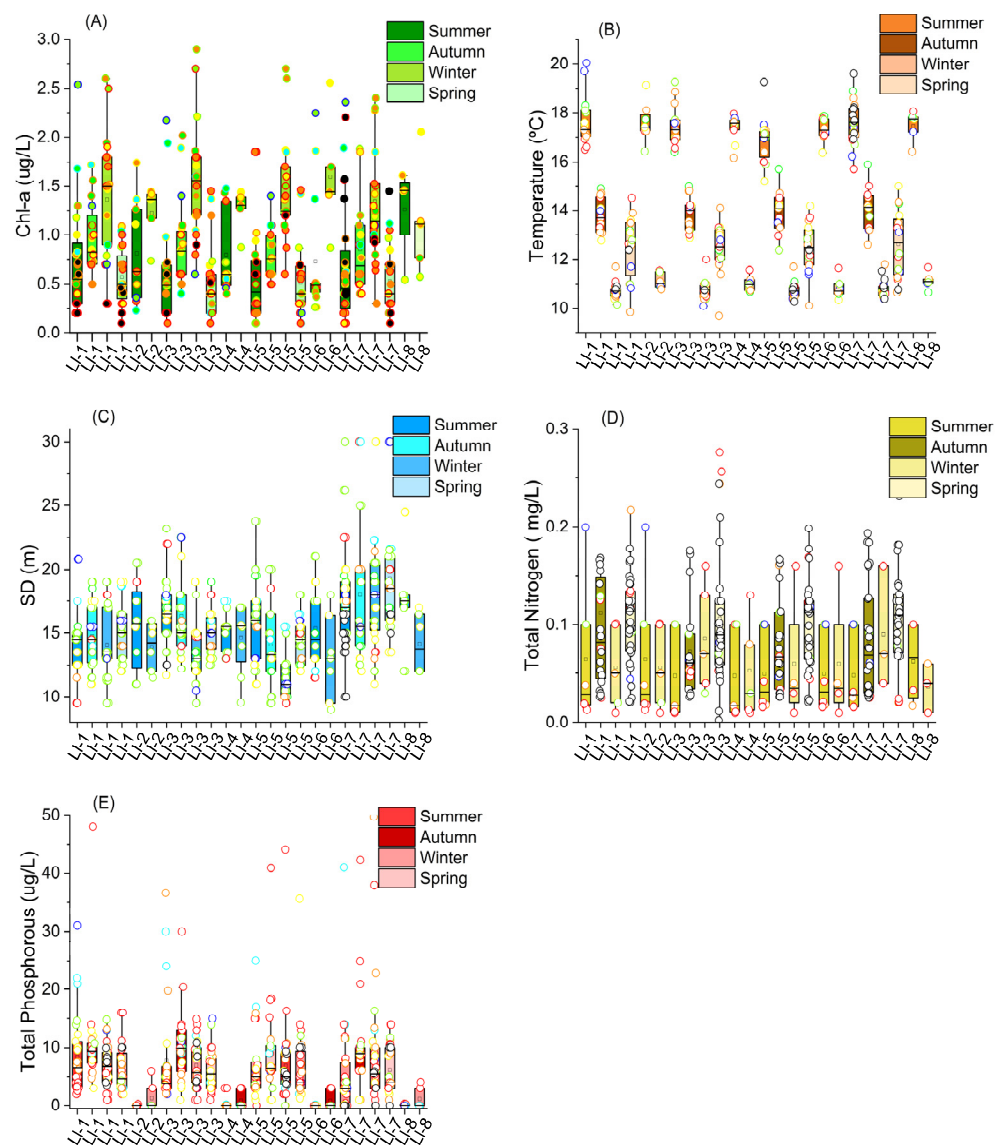
Figure 7. Boxplot of limnological parameters in Llanquihue Lake over 1986–2021: (**A**) chlorophyll-a (μg/L), (**B**) temperature (°C), (**C**) SD (m), (**D**) total nitrogen (mg/L), (**E**) total phosphorous (μg/L).

Chlorophyll-a values vary widely depending on location, season, and environmental conditions (see Table 2). Chl-a in Lake Llanquihue ranged from a minimum of Q1 = 0.50 μg/L to a maximum of Q4 = 2.90 μg/L, for all other statistics see Table S2. Water temperature varied according to seasonality, an expected result with a winter minimum of 1.6 °C and a summer maximum of 20 °C, the temperature minima and maxima coinciding with the Chl-a minima and maxima. Nitrogen and total phosphorus were analyzed in the lake system. The recorded values of nitrogen were low between 0.003 and 0.6 mg/L, while phosphorous was found in higher concentrations between 1.0 and 56.0 μg/L. Transparency is generally high in Llanquihue Lake during most of the year and for all lake seasons with a maximum of 30 m and a minimum of 6 m in winter, which may be attributed to turbidity related to precipitation events and strong winds during this time of year in the Southern Hemisphere.

Table 2. Meteorological variable for Llanquihue Lake.

Months	Temperature (°C)	Wind Speed (m/s)	Relative Humidity (%)	Cloud Cover (%)	Accumulated Precipitation (mm)	Photosynthetic Active Radiation (mmol/m^2)
January	16.43	3.90	62.50	0.50	60.07	63,581.1
February	18.42	3.30	75.50	0.33	47.67	56,124.7
March	17.77	2.70	65.20	0.55	75.00	33,109.3
April	14.89	3.80	89.70	0.70	121.09	19,872.1
May	12.15	4.10	76.20	0.80	184.7	15,883.9
June	9.40	3.90	71.80	1.00	239.60	11,225.7
July	8.64	4.10	78.30	0.90	205.40	10,393.8
August	12.05	2.90	73.30	0.90	207.90	17,965.5
September	13.11	2.70	76.15	0.72	110.70	30,330.5
October	14.00	4.07	62.80	0.69	105.60	42,777.1
November	14.89	3.90	66.01	0.55	80.90	54,885.1
December	15.66	4.20	92.70	0.45	57.10	62,428.7

3.2. Results and Validation of Statistical and Deep Learning Models Cases

3.2.1. Case A (Measurement Data)

Figure 8 shows the behavior of the estimated chlorophyll-a at the eight-sampling stations during the study period. In each case, we modeled Chl-a for the three models using SARIMAX, LSTM and RNN.
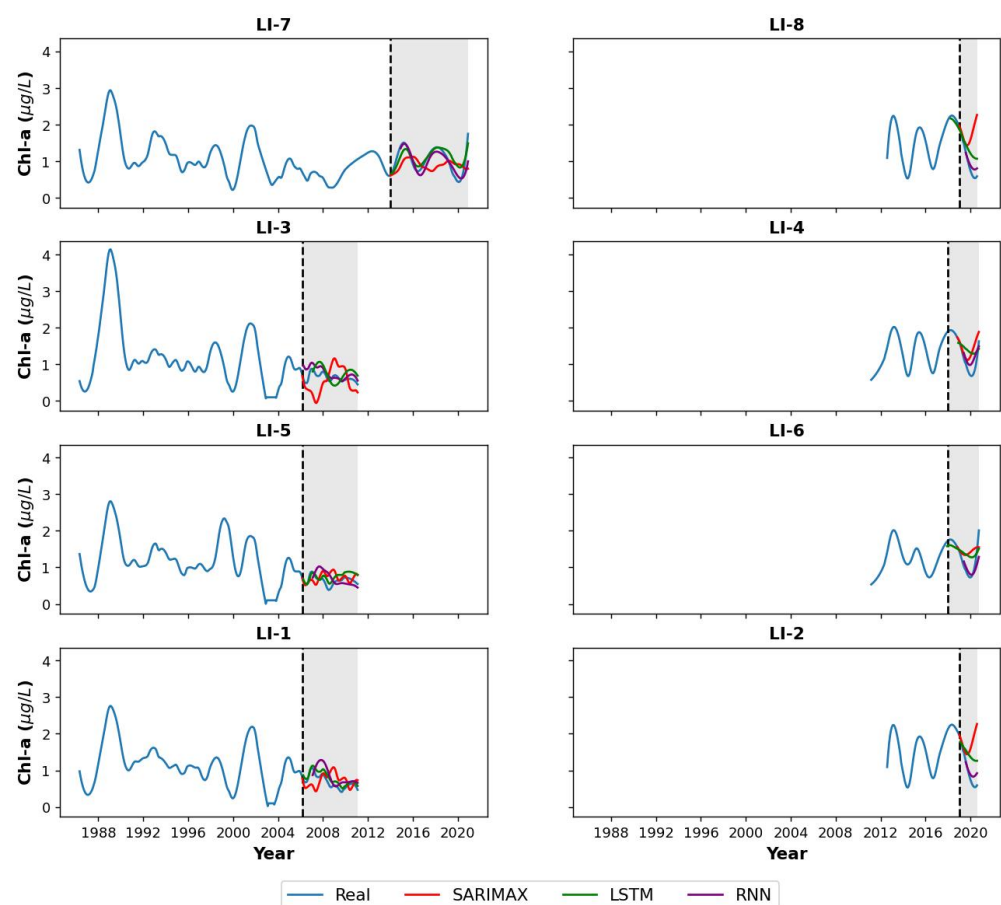

Figure 8. Chlorophyll-a estimation of case A in the eight-sampling stations of Llanquihue Lake during 1988–2020. The shaded regions represent observations for the selection of the test.

From the results, we can observe that in most of the stations, the models offer a good retrieval of the temporal variations, except for LI-3, LI-2, and LI-8 when the SARIMAX model

is used (red line). In addition, the SARIMAX model exhibits higher MSE, RMSE, MaxError, and MAE (R^2 < 0.915) values compared to the other models (RNN and LSTM) (Table 3).

Table 3. Validation metrics for all the stations and models considered in Case A (Section 2.4).

Station	Statistic	Case A							
		LI-1	LI-2	LI-3	LI-4	LI-5	LI-6	LI-7	LI-8
SARIMAX	MSE (μg/L)^2	0.083	0.787	0.160	0.173	0.043	0.190	0.139	0.787
	RMSE (μg/L)	0.288	0.887	0.400	0.415	0.206	0.436	0.372	0.887
	MaxError (μg/L)	0.505	1.676	0.783	0.726	0.459	0.700	0.955	1.676
	MAE (μg/L)	0.244	0.660	0.350	0.324	0.159	0.366	0.322	0.660
	R^2	0.892	0.724	0.857	0.864	0.915	0.685	0.793	0.795
LSTM	MSE (μg/L)^2	0.014	0.260	0.029	0.166	0.020	0.101	0.039	0.098
	RMSE (μg/L)	0.116	0.510	0.169	0.407	0.142	0.317	0.199	0.314
	MaxError (μg/L)	0.194	0.730	0.389	0.625	0.244	0.563	0.423	0.552
	MAE (μg/L)	0.106	0.442	0.136	0.348	0.121	0.263	0.152	0.247
	R^2	0.912	0.932	0.896	0.912	0.934	0.854	0.893	0.936
RNN	MSE (μg/L)^2)	0.056	0.045	0.046	0.066	0.0509	0.068	0.030	0.028
	RMSE (μg/L)	0.236	0.212	0.214	0.257	0.225	0.260	0.174	0.167
	MaxError (μg/L)	0.447	0.331	0.383	0.379	0.451	0.724	0.754	0.238
	MAE (μg/L)	0.196	0.176	0.192	0.237	0.191	0.183	0.127	0.146
	R^2	0.901	0.915	0.876	0.893	0.926	0.827	0.843	0.648

Furthermore, the LSTM model generally performs well across most stations, with lower values for MSE (<0.260 (μg/L)^2), RMSE (<0.510 ug/L), MaxError (<0.730 μg/L), and MAE (<0.442 μg/L) compared to SARIMAX, with higher values at LI-2 station. Additionally, the R^2 values for LSTM were consistently high, indicating a good fit to the data. In addition, the RNN model shows similar performance to LSTM, with relatively low MSE (<0.068 (μg/L)^2), RMSE (<0.260 ug/L), MaxError (<0.751 μg/L), and MAE (<0.283 μg/L) values, and the R^2 values were also consistently high (>0.827) (Table 3).

3.2.2. Case B (Measurement and Meteorological Data)

Figure 9 shows the results for Case B. In all cases, the estimation of the chlorophyll-a values was similar. Sampling stations 1, 3 and 5 did not have sufficient data to estimate the variable with the models.

Similarly, for Case-A, the SARIMAX model showed a moderate performance compared to the RNN and LSTM models with the MSE, RMSE, MaxError, and MAE metrics relatively lower (higher) for station LI-6 (LI-8) compared to the others. In addition, the R^2 values for SARIMAX are generally approximately equal to 0.8 (Table 4).

In contrast, the LSTM model generally performs well across all stations with relatively low MSE (<0.029 (μg/L)^2), RMSE (<0.172 ug/L), MaxError (<0.175 μg/L), and MAE (<0.172 μg/L) values. The R^2 values are lower compared to Case A but considering the smaller amount of data used for the training, there is still a good performance (>0.80). Besides the RNN model showed a similar performance to LSTM, with low MSE, RMSE, MaxError, and MAE values across all stations, and R^2 values above 0.80 (Table 4).
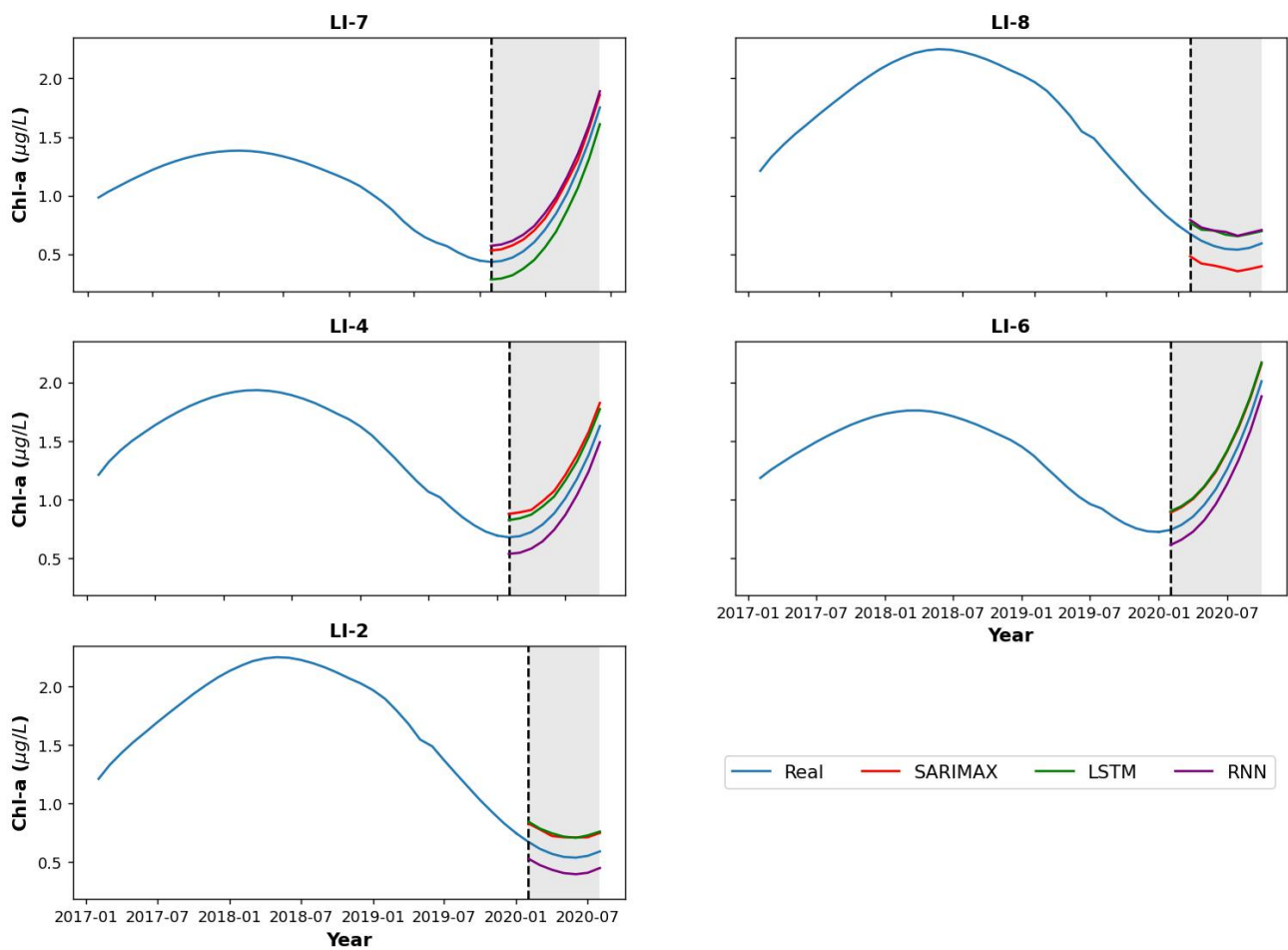
Figure 9. Chlorophyll-a estimation of Case B at the five sampling stations with data from Llanquihue Lake during 1988–2020. The shaded regions represent observations for the selection of the test.

Table 4. Validation metrics for all the stations and models considered in Case B (Section 2.4).

				Case B		
Station	**Statistic**	**LI-2**	**LI-4**	**LI-6**	**LI-7**	**LI-8**
SARIMAX	MSE $(ug/L)^2$	0.026	0.039	0.023	0.010	0.033
	RMSE (ug/L)	0.162	0.197	0.150	0.101	0.183
	MaxError (ug/L)	0.172	0.203	0.151	0.108	1.195
	MAE (ug/L)	0.162	0.197	0.149	0.102	0.182
	R^2	0.795	0.781	0.797	0.776	0.812
LSTM	MSE $(ug/L)^2$	0.029	0.022	0.025	0.026	0.013
	RMSE (ug/L)	0.172	0.149	0.159	0.150	0.112
	MaxError (ug/L)	0.175	0.154	0.161	0.157	0.131
	MAE (ug/L)	0.172	0.149	0.159	0.150	0.111
	R^2	0.816	0.842	0.837	0.821	0.834
RNN	MSE $(ug/L)^2$	0.019	0.020	0.016	0.019	0.016
	RMSE (ug/L)	0.141	0.141	0.129	0.139	0.125
	MaxError (ug/L)	0.144	0.144	0.131	0.144	0.146
	MAE (ug/L)	0.140	0.141	0.129	0.139	0.124
	R^2	0.805	0.840	0.830	0.824	0.806

3.2.3. Case C (Measurement, Meteorological Data and Satellite Data)

Figure 10 shows Case C, which is the most general and complete compared to the two cases presented before because it integrates all types of data available: measurement in situ data, meteorological data, and satellite data.
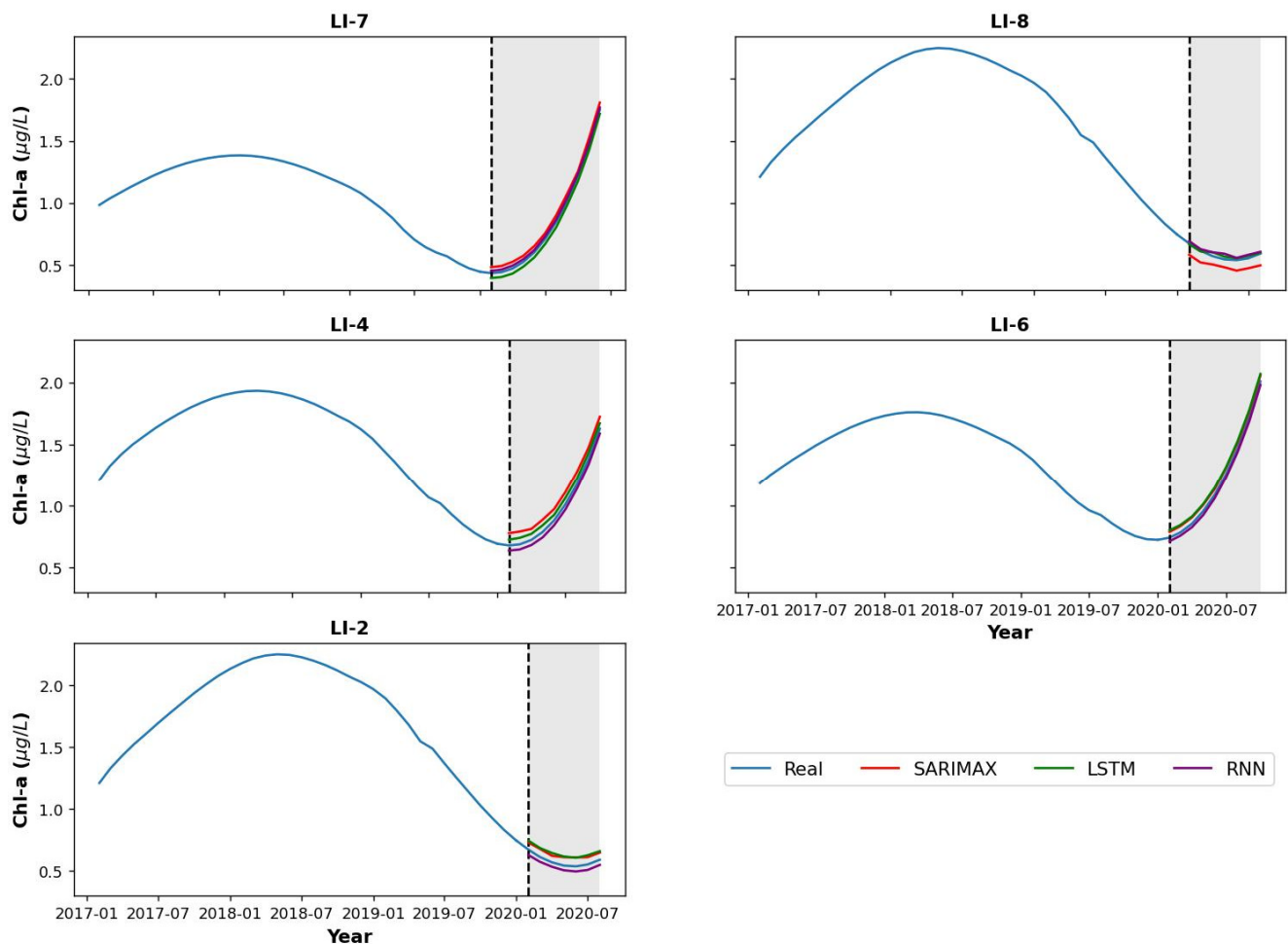

Figure 10. Chlorophyll-a estimation of case C in the five sampling stations with data from Llanquihue Lake during 1988–2020. The shaded regions represent observations for the selection of the test.

When comparing the results between Table in Case C vs. Case B, we observe that all the models' performance is better in Case C, which is evident with lower values for MSE ($<0.009 \ \mu g/L)^2$), RMSE ($<0.097 \ ug/L$), MaxError ($<0.103 \ ug/L$), and MAE ($<0.097 \ ug/L$) and higher R^2 values (<0.81); however, LSTM and RNN showed better performance against SARIMAX (Table 5). This suggests that incorporating a larger set of chlorophyll-a-related variables both directly and indirectly enhances the predictive capacity of the algorithms.

3.3. Feature Importance

Figures 11–13 also show, using Garson's weighting method [78], the relative importance or contribution of the independent variables (predictor variables) in explaining the variance of the dependent variable (outcome variable) for each case used.

Table 5. Validation metrics for all the stations and models considered in Case C (Section 2.4).

|	Case C					
SARIMAX	MSE (μg/L)^2	0.003	0.009	0.002	0.002	0.006
	RMSE (ug/L)	0.062	0.097	0.050	0.050	0.083
	MaxError (ug/L)	0.07	0.103	0.051	0.057	0.095
	MAE (ug/L)	0.062	0.097	0.049	0.050	0.082
	R^2	0.804	0.807	0.812	0.832	0.796
LSTM	MSE (μg/L)^2	0.001	0.002	0.003	0.002	0.001
	RMSE (ug/L)	0.072	0.049	0.060	0.040	0.018
	MaxError (ug/L)	0.075	0.054	0.061	0.046	0.031
	MAE (ug/L)	0.072	0.049	0.059	0.040	0.015
	R^2	0.857	0.864	0.896	0.877	0.843
RNN	MSE (μg/L)^2	0.001	0.002	0.001	0.001	0.001
	RMSE (ug/L)	0.041	0.041	0.029	0.019	0.018
	MaxError (ug/L)	0.044	0.044	0.031	0.024	0.031
	MAE (ug/L)	0.045	0.041	0.029	0.019	0.015
	R^2	0.843	0.832	0.815	0.807	0.795
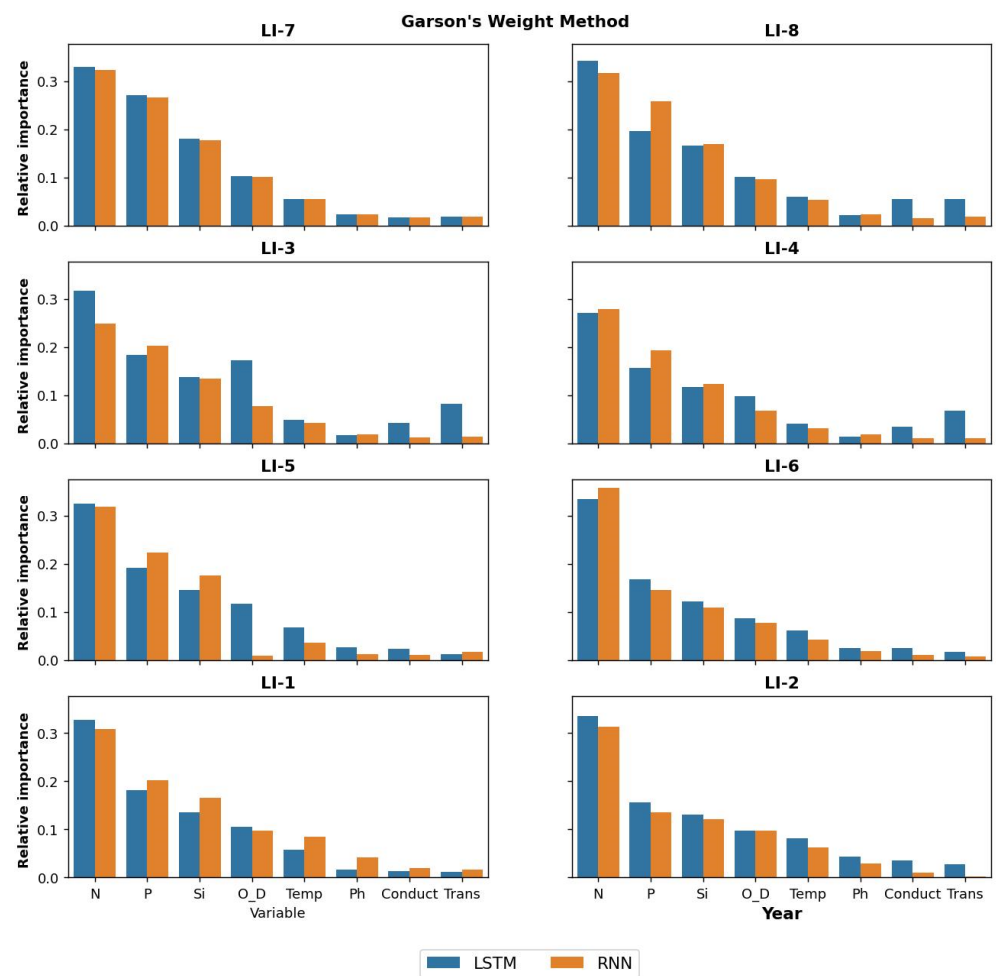
Figure 11. Relative importance using Garson's weighting method Case A for the eight-sampling stations. N is Nitrogen, P is Phosphorus, Si Silicon, O_D Dissolved Oxygen, Temp temperature, Conduct Conductivity, and Trans Transparency.
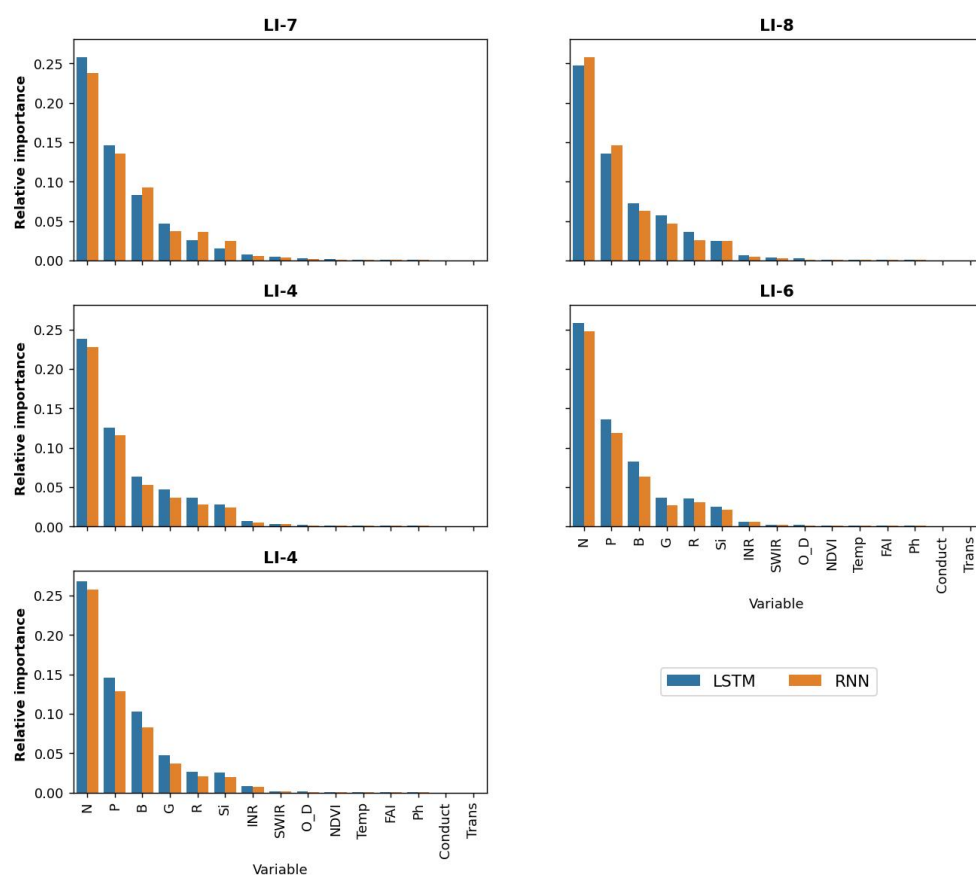
Figure 12. Relative importance through Garson's weighting method for Case B at five-sampling stations. N is Nitrogen, P is Phosphorus, Si Silicon, O_D Dissolved Oxygen, Temp temperature, Conduct Conductivity, Trans *Transparency*, B blue band, G green band, R red band, INR near infrared band, SWIR shortwave infrared band, NDVI Normalized Difference Vegetation Index, and FAI, Floating Algae Index.

Results showed that Nitrogen (N), Phosphorus (P), and Silica (Si) present the highest feature importance values (ranging from 0.115 to 0.336) in predicting Chlorophyll-a across all stations in case A. Subsequently, Dissolved Oxygen (O_D) and Temperature (Temp) showed relative importance ranging from 0.053 to 0.157, with O_D being more significant than Temp. Conversely, pH, Conductivity (Conduct), and Transparency (SD) exhibited relatively lower importance, all having values below 0.1.

For Case B (Figure 12), we can see that adding satellite images modifies the order of importance of some of the variables. Similarly, N and P variables showed higher relative importance, with contributions of >0.15. However, the B, G, and R bands were more important (ranging from 0.033 to 0.046) with respect to the Si variable, with contributions < 0.038.

Finally, in Case C (Figure 13), a pattern comparable to that of Case B was observed. The variables N and P remained the most crucial factors in the predictions (relative importance above 0.110), followed by the B, G, and R bands, and subsequently the Si variable. Therefore, the variables INR, SWIR, and O_D exhibit the most significant contributions, and the remaining variables have a negligible importance of <0.035.
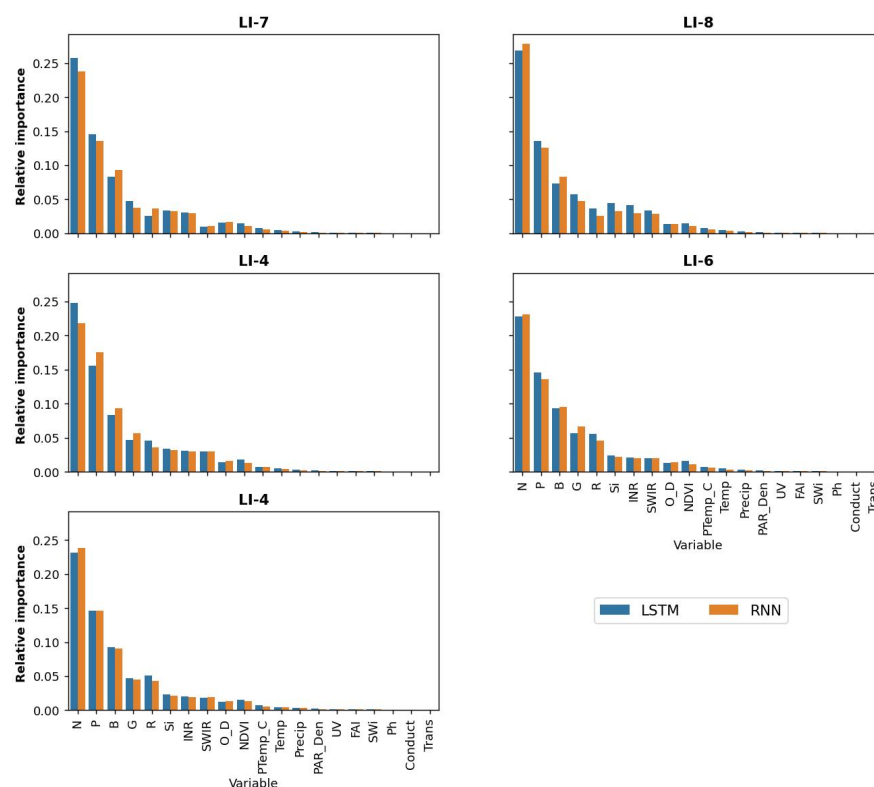
Figure 13. Relative importance using Garson's weighting method for Case C. In five-sampling station. N is Nitrogen, P is Phosphorus, Si Silicon, O_D Dissolved Oxygen, Temp temperature, Conduct Conductivity, Trans *Transparency*, B blue band, G green band, R red band, INR near infrared band, SWIR shortwave infrared band, NDVI Normalized Difference Vegetation Index, and FAI, Floating Algae Index.

4. Discussion

It is important to monitor the behavior of the lakes since they are sentinels or indicators of climate change [79]. Human activities accelerate the eutrophication processes of these southern freshwater ecosystems. In this study, using combined remote sensing and machine learning techniques, models were created to estimate water quality parameters such as chlorophyll-a in a lake in southern Chile. Llanquihue Lake is the lake body most vulnerable to contamination among the lakes of the Araucanian Lake district, owing to its slower water mass renewal time than the rest (estimated to be 74 years), in addition to the intense use of its shores [35]. It is the only lake in Chile, with four municipalities located on its shores, all of which are the main capitals: Puerto Octay, Frutillar, Llanquihue and Puerto Varas. Currently, Llanquihue Lake, despite maintaining the trophic characteristics of oligotrophic lakes, the values of key water quality parameters such as chlorophyll-a and nutrients have increased in Llanquihue Lake, and their trophic level can change in a shorter time than the natural succession process of aquatic systems. In a previous study, we calibrated and validated a set of nine artificial intelligence algorithms over a period longer than 30 years to estimate the chlorophyll-a variable at different points of the lake [35].

In the present work, we aimed to add, in addition to the historical data from the lake monitoring campaigns conducted by the General Water Directorate of Chile, data from the Meteorological Directorate of Chile and data from Landsat-8 satellite images. Excellent and accurate results were obtained for each season of the year in this lake. Model validation metrics showed that all three models effectively predicted chlorophyll as an indicator of the presence of algae in this water body. Coefficient of determination values ranging from 0.64 to 0.93 were obtained, with the LSTM model showing the best statistics in any of the cases tested, and similar results were obtained in [80] when predicting chlorophyll using the LSTM model. The LSTM model generally performs well across most stations, with lower

values for MSE (<0.260 (μg/L)^2), RMSE (<0.510 ug/L), MaxError (<0.730 μg/L), and MAE (<0.442 μg/L) compared to SARIMAX, with higher values at LI-2 station. Additionally, the R^2 values for the LSTM were consistently high, indicating a good fit to the data. In addition, the RNN model shows similar performance as LSTM, with relatively low MSE (<0.068 (μg/L)^2), RMSE (<0.260 ug/L), MaxError (<0.751 μg/L), and MAE (<0.283 μg/L) values, and the R^2 values were also consistently high (>0.827). When comparing the results between Case C (Measurement, Meteorological and Satellite Data) vs. Case B (Measurement and Meteorological Data), we observe that all the models' performance is better in case C, which is evident with lower values for MSE (<0.009 (μg/L)^2), RMSE (<0.097 ug/L), MaxError (<0.103 μg/L), and MAE (<0.097 μg/L) and higher R^2 values (<0.81); however, LSTM and RNN showed better performance against SARIMAX (Table 5). This suggests that incorporating a larger set of chlorophyll-a-related variables both directly and indirectly enhances the predictive capacity of the algorithms. Good chlorophyll predictions have been obtained in investigations that have used deep learning models and Landsat-8 in lakes, such as the case of [77–79]; as in the case of this research, these techniques can be improved by incorporating meteorological variables. The methodology of this study and other similar methodologies have applications in monitoring water quality and serve as an early warning tool for hydro-environmental management in inland water ecosystems, according to [35,81]. It is important to clarify that the precision in the models will always be greater when more input data are provided. Only available meteorological and satellite data were used in this manuscript. The more images included in the model, the better the estimate should be, and image quality can be affected by cloud density (cloud percentage) as it can alter the pixel value (band or calculated index) and decrease the precision of the model accuracy. The image quality was not a limitation in our work, but the fact of not having some images close to the monitoring due to the high cloud percentage, which prevented the use of some in situ data, was. Generally, in similar investigations, cloudiness can be a limiting factor in the availability and quality of satellite data, and therefore, affect the precision of the estimation of water quality parameters.

The "Ley de Bases del Medio Ambiente" (Ley Nº 19.300 de 1994) in Chile defines aquatic pollution in terms of the existence of standards that establish permissible limits for the presence of substances, elements, or energies, susceptible to causing environmental damage. Lake Llanquihue and the Villarrica side are the only lake systems in Chile that have a secondary water quality standard that seeks to safeguard the use of water resources, protect, and conserve the aquatic communities and ecosystems of the lake, and maximize the benefits that the ecosystem services associated with the lake provide [35]. Therefore, it is of vital importance to maintain a follow-up of these inland aquatic bodies, as Chl-a is a bioindicator parameter of algae presence commonly used in research. It is relevant to inform the authorities and the population of the current state and evolution of the lake through research such as this one. It also provides valuable base information for the management of water resources that provide us with multiple uses. In the future, we intend to use the models tested in the estimation of parameters at times of the year when the conditions of in-situ monitoring of Llanquihue Lake represent a limitation (intense wind or rainy periods). However, these estimation models are relevant in autumn and winter, when multispectral satellite images present a high percentage of cloud cover.

5. Conclusions

Combined remote sensing and machine learning techniques have proven to be valuable tools for the estimation of environmental proxies, such as chlorophyll-a. Coefficient of determination values ranging from 0.64 to 0.93 were obtained, with the LSTM model showing the best statistics in any of the cases tested. The LSTM model generally performs well across most stations, with lower values for MSE (<0.260 (μg/L)^2), RMSE (<0.510 ug/L), MaxError (<0.730 μg/L), and MAE (<0.442 μg/L). This parameter has been widely used in different aquatic ecosystems as an indicator of algal biomass and water quality. In this study, a series of in-situ data from 1989 to 2021 recorded at eight monitoring stations

spatially distributed in Llanquihue Lake was used to study the behavior of limnological variables at different points in the lake. The three estimation models employed demonstrated strong performance in estimating Chl-a, with the LSTM model yielding the most accurate results. Of the three cases applied in this study, Case C (all variables integrated), meteorological, water quality measurements, and satellite data showed the most accurate results for all stations in the lake. These models will be employed in future research focused on seasonal periods such as autumn and winter, characterized by frequent episodes of rain or "Puelches" (strong winds). Traditional monitoring methods face increased complexity during these periods. Therefore, as an alternative, recovery models like the ones presented above have emerged, taking advantage of deep learning tools to integrate real-time data with satellite observations, allowing early tools to be developed for monitoring lakes in tracking algal bloom phenomena. In addition, by combining these data sets, these models provide a more effective approach to monitoring and analyzing weather conditions during challenging periods.

References

1. Sheferaw Ayele, H.; Atlabachew, M. Review of Characterization, Factors, Impacts, and Solutions of Lake Eutrophication: Lesson for Lake Tana, Ethiopia. *Environ. Sci. Pollut. Res.* **2021**, *28*, 14233–14252. [CrossRef] [PubMed]
2. El-Sheekh, M.; Abdel-Daim, M.M.; Okba, M.; Gharib, S.; Soliman, A.; El-Kassas, H. Green Technology for Bioremediation of the Eutrophication Phenomenon in Aquatic Ecosystems: A Review. *Afr. J. Aquat. Sci.* **2021**, *46*, 274–292. [CrossRef]
3. Wurtsbaugh, W.A.; Paerl, H.W.; Dodds, W.K. Nutrients, Eutrophication and Harmful Algal Blooms along the Freshwater to Marine Continuum. *Wiley Interdiscip. Rev. Water* **2019**, *6*, e1373. [CrossRef]
4. Mishra, R.K. The Effect of Eutrophication on Drinking Water. *Br. J. Multidiscip. Adv. Stud.* **2023**, *4*, 7–20. [CrossRef]
5. Hakeem, K.R.; Bhat, R.A.; Qadri, H. *Bioremediation and Biotechnology: Sustainable Approaches to Pollution Degradation*; Springer International Publishing: Berlin/Heidelberg, Germany, 2020; ISBN 9783030356910.
6. Zhang, Y.; Luo, P.; Zhao, S.; Kang, S.; Wang, P.; Zhou, M.; Lyu, J. Control and Remediation Methods for Eutrophic Lakes in the Past 30 Years. *Water Sci. Technol.* **2020**, *81*, 1099–1113. [CrossRef] [PubMed]
7. Erarto, F.; Getahun, A. Impacts of Introductions of Alien Species with Emphasis on Fishes. *Int. J. Fish. Aquat. Stud.* **2020**, *8*, 207–216.
8. Henriksson, P.J.G.; Troell, M.; Banks, L.K.; Belton, B.; Beveridge, M.C.M.; Klinger, D.H.; Pelletier, N.; Phillips, M.J.; Tran, N. Interventions for Improving the Productivity and Environmental Performance of Global Aquaculture for Future Food Security. *One Earth* **2021**, *4*, 1220–1232. [CrossRef]
9. Kibuye, F.A.; Zamyadi, A.; Wert, E.C. A Critical Review on Operation and Performance of Source Water Control Strategies for Cyanobacterial Blooms: Part II-Mechanical and Biological Control Methods. *Harmful Algae* **2021**, *109*, 102119. [CrossRef]

10. Zhan, Q.; Teurlincx, S.; van Herpen, F.; Raman, N.V.; Lürling, M.; Waajen, G.; de Senerpont Domis, L.N. Towards Climate-Robust Water Quality Management: Testing the Efficacy of Different Eutrophication Control Measures during a Heatwave in an Urban Canal. *Sci. Total Environ.* **2022**, *828*, 154421. [CrossRef]

11. Mandal, R.; Dutta, G. From Photosynthesis to Biosensing: Chlorophyll Proves to Be a Versatile Molecule. *Sens. Int.* **2020**, *1*, 100058. [CrossRef]

12. Sayyed, R.; Uarrotaaeditors, V.G. *Secondary Metabolites and Volatiles of PGPR in Plant-Growth Promotion*; Springer: Berlin/Heidelberg, Germany, 2022.

13. Gomes, P.; Valente, T.; Geraldo, D.; Ribeiro, C. Photosynthetic Pigments in Acid Mine Drainage: Seasonal Patterns and Associations with Stressful Abiotic Characteristics. *Chemosphere* **2020**, *239*, 124774. [CrossRef] [PubMed]

14. Wang, X.; Li, Y.; Wei, S.; Pan, L.; Miao, J.; Lin, Y.; Wu, J. Toxicity Evaluation of Butyl Acrylate on the Photosynthetic Pigments, Chlorophyll Fluorescence Parameters, and Oxygen Evolution Activity of *Phaeodactylum tricornutum* and *Platymonas subcordiformis*. *Environ. Sci. Pollut. Res.* **2021**, *28*, 60954–60967. [CrossRef] [PubMed]

15. Li, J.; Xiao, X.; Guo, L.; Chen, H.; Feng, M.; Yu, X. A Novel QPCR-Based Method to Quantify Seven Phyla of Common Algae in Freshwater and Its Application in Water Sources. *Sci. Total Environ.* **2022**, *823*, 153340. [CrossRef] [PubMed]

16. Wu, D.; Li, R.; Liu, J.; Khan, N. Monitoring Algal Blooms in Small Lakes Using Drones: A Case Study in Southern Illinois. *J. Contemp. Water Res. Educ.* **2023**, *177*, 83–93. [CrossRef]

17. Jankowski, K.J.; Mejia, F.H.; Blaszczak, J.R.; Holtgrieve, G.W. Aquatic Ecosystem Metabolism as a Tool in Environmental Management. *Wiley Interdiscip. Rev. Water* **2021**, *8*, e1521. [CrossRef]

18. Rodríguez-López, L.; Duran-Llacer, I.; Bravo Alvarez, L.; Lami, A.; Urrutia, R. Recovery of Water Quality and Detection of Algal Blooms in Lake Villarrica through Landsat Satellite Images and Monitoring Data. *Remote Sens.* **2023**, *15*, 1929. [CrossRef]

19. Pei, T.; Xu, J.; Liu, Y.; Huang, X.; Zhang, L.; Dong, W.; Qin, C.; Song, C.; Gong, J.; Zhou, C. GIScience and Remote Sensing in Natural Resource and Environmental Research: Status Quo and Future Perspectives. *Geogr. Sustain.* **2021**, *2*, 207–215. [CrossRef]

20. Katkani, D.; Babbar, A.; Mishra, V.K.; Trivedi, A.; Tiwari, S.; Kumawat, R.K. A Review on Applications and Utility of Remote Sensing and Geographic Information Systems in Agriculture and Natural Resource Management. *Int. J. Environ. Clim. Chang.* **2022**, *12*, 1–18. [CrossRef]

21. Khruschev, S.S.; Plyusnina, T.Y.; Antal, T.K.; Pogosyan, S.I.; Riznichenko, G.Y.; Rubin, A.B. Machine Learning Methods for Assessing Photosynthetic Activity: Environmental Monitoring Applications. *Biophys. Rev.* **2022**, *14*, 821–842. [CrossRef]

22. Stramski, D.; Joshi, I.; Reynolds, R.A. Ocean Color Algorithms to Estimate the Concentration of Particulate Organic Carbon in Surface Waters of the Global Ocean in Support of a Long-Term Data Record from Multiple Satellite Missions. *Remote Sens. Environ.* **2022**, *269*, 112776. [CrossRef]

23. Wang, W.; Shi, K.; Zhang, Y.; Li, N.; Sun, X.; Zhang, D.; Zhang, Y.; Qin, B.; Zhu, G. A Ground-Based Remote Sensing System for High-Frequency and Real-Time Monitoring of Phytoplankton Blooms. *J. Hazard. Mater.* **2022**, *439*, 129623. [CrossRef] [PubMed]

24. Xu, D.; Pu, Y.; Zhu, M.; Luan, Z.; Shi, K. Automatic Detection of Algal Blooms Using Sentinel-2 MSI and Landsat OLI Images. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2021**, *14*, 8497–8511. [CrossRef]

25. Dwivedi, Y.K.; Hughes, L.; Ismagilova, E.; Aarts, G.; Coombs, C.; Crick, T.; Duan, Y.; Dwivedi, R.; Edwards, J.; Eirug, A.; et al. Artificial Intelligence (AI): Multidisciplinary Perspectives on Emerging Challenges, Opportunities, and Agenda for Research, Practice and Policy. *Int. J. Inf. Manag.* **2021**, *57*, 101994. [CrossRef]

26. Kim, J.H.; Shin, J.K.; Lee, H.; Lee, D.H.; Kang, J.H.; Cho, K.H.; Lee, Y.G.; Chon, K.; Baek, S.S.; Park, Y. Improving the Performance of Machine Learning Models for Early Warning of Harmful Algal Blooms Using an Adaptive Synthetic Sampling Method. *Water Res.* **2021**, *207*, 117821. [CrossRef]

27. Li, H.; Qin, C.; He, W.; Sun, F.; Du, P. Improved Predictive Performance of Cyanobacterial Blooms Using a Hybrid Statistical and Deep-Learning Method. *Environ. Res. Lett.* **2021**, *16*, 124045. [CrossRef]

28. Cao, H.; Han, L.; Li, L. A Deep Learning Method for Cyanobacterial Harmful Algae Blooms Prediction in Taihu Lake, China. *Harmful Algae* **2022**, *113*, 102189. [CrossRef]

29. Guan, Q.; Feng, L.; Hou, X.; Schurgers, G.; Zheng, Y.; Tang, J. Eutrophication Changes in Fifty Large Lakes on the Yangtze Plain of China Derived from MERIS and OLCI Observations. *Remote Sens. Environ.* **2020**, *246*, 111890. [CrossRef]

30. Rodríguez-López, L.; González-Rodríguez, L.; Duran-Llacer, I.; Cardenas, R.; Urrutia, R. Spatio-Temporal Analysis of Chlorophyll in Six Araucanian Lakes of Central-South Chile from Landsat Imagery. *Ecol. Inform.* **2021**, *65*, 101431. [CrossRef]

31. Ma, J.; He, F.; Qi, T.; Sun, Z.; Shen, M.; Cao, Z.; Meng, D.; Duan, H.; Luo, J. Thirty-Four-Year Record (1987–2021) of the Spatiotemporal Dynamics of Algal Blooms in Lake Dianchi from Multi-Source Remote Sensing Insights. *Remote Sens.* **2022**, *14*, 4000. [CrossRef]

32. Rolim, S.B.A.; Veettil, B.K.; Vieiro, A.P.; Kessler, A.B.; Gonzatti, C. Remote Sensing for Mapping Algal Blooms in Freshwater Lakes: A Review. *Environ. Sci. Pollut. Res.* **2023**, *30*, 19602–19616. [CrossRef]

33. Norambuena, J.A.; Poblete-Grant, P.; Beltrán, J.F.; De Los Ríos-Escalante, P.; Farías, J.G. Evidence of the Anthropic Impact on a Crustacean Zooplankton Community in Two North Patagonian Lakes. *Sustainability* **2022**, *14*, 6052. [CrossRef]

34. McNamara, K.; Rust, A.C.; Cashman, K.V.; Castruccio, A.; Abarzúa, A.M. Comparison of Lake and Land Tephra Records from the 2015 Eruption of Calbuco Volcano, Chile. *Bull. Volcanol.* **2019**, *81*, 10. [CrossRef]

35. Rodríguez-López, L.; Bustos Usta, D.; Bravo Alvarez, L.; Duran-Llacer, I.; Lami, A.; Martínez-Retureta, R.; Urrutia, R. Machine Learning Algorithms for the Estimation of Water Quality Parameters in Lake Llanquihue in Southern Chile. *Water* **2023**, *15*, 1994. [CrossRef]

36. Segovia, M.J.; Diaz, D.; Slezak, K.; Zuñiga, F. Magnetotelluric Study in the Los Lagos Region (Chile) to Investigate Volcano-Tectonic Processes in the Southern Andes. *Earth Planets Space* **2021**, *73*, 5. [CrossRef]

37. DMC. Dirección Meteorológica de Chile. 2023. Available online: https://climatologia.meteochile.gob.cl/ (accessed on 10 March 2023).

38. Arismendi, I.; Nahuelhual, L. Non-Native Salmon and Trout Recreational Fishing in Lake Llanquihue, Southern Chile: Economic Benefits and Management Implications. *Rev. Fish. Sci.* **2007**, *15*, 311–325. [CrossRef]

39. Wulder, M.A.; Loveland, T.R.; Roy, D.P.; Crawford, C.J.; Masek, J.G.; Woodcock, C.E.; Allen, R.G.; Anderson, M.C.; Belward, A.S.; Cohen, W.B.; et al. Current Status of Landsat Program, Science, and Applications. *Remote Sens. Environ.* **2019**, *225*, 127–147. [CrossRef]

40. Chatenoux, B.; Richard, J.P.; Small, D.; Roeoesli, C.; Wingate, V.; Poussin, C.; Rodila, D.; Peduzzi, P.; Steinmeier, C.; Ginzler, C.; et al. The Swiss Data Cube, Analysis Ready Data Archive Using Earth Observations of Switzerland. *Sci. Data* **2021**, *8*, 295. [CrossRef]

41. USGS 2019. Landsat 8 Data Users Handbook, Version 5.0 LSDS-1574. 2019. Available online: https://www.Usgs.Gov/Media/Files/Landsat-8-Data-Users-Handbook (accessed on 7 January 2023).

42. Ilori, C.O.; Pahlevan, N.; Knudby, A. Analyzing Performances of Different Atmospheric Correction Techniques for Landsat 8: Application for Coastal Remote Sensing. *Remote Sens.* **2019**, *11*, 469. [CrossRef]

43. 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; pp. 9–13.

44. Vanhellemont, Q.; Ruddick, K. Atmospheric Correction of Metre-Scale Optical Satellite Data for Inland and Coastal Water Applications. *Remote Sens. Environ.* **2018**, *216*, 586–597. [CrossRef]

45. Vanhellemont, Q. Adaptation of the Dark Spectrum Fitting Atmospheric Correction for Aquatic Applications of the Landsat and Sentinel-2 Archives. *Remote Sens. Environ.* **2019**, *225*, 175–192. [CrossRef]

46. Vanhellemont, Q. Sensitivity Analysis of the Dark Spectrum Fitting Atmospheric Correction for Metre- and Decametre-Scale Satellite Imagery Using Autonomous Hyperspectral Radiometry. *Opt. Express* **2020**, *28*, 29948. [CrossRef] [PubMed]

47. Vanhellemont, Q.; Ruddick, K. Turbid Wakes Associated with Offshore Wind Turbines Observed with Landsat 8. *Remote Sens. Environ.* **2014**, *145*, 105–115. [CrossRef]

48. Vanhellemont, Q.; Ruddick, K. Advantages of High Quality SWIR Bands for Ocean Colour Processing: Examples from Landsat-8. *Remote Sens. Environ.* **2015**, *161*, 89–106. [CrossRef]

49. Rodríguez-López, L.; Duran-Llacer, I.; González-Rodríguez, L.; Cardenas, R.; Urrutia, R. Retrieving Water Turbidity in Araucanian Lakes (South-Central Chile) Based on Multispectral Landsat Imagery. *Remote Sens.* **2021**, *13*, 3133. [CrossRef]

50. Rodríguez-López, L.; González-Rodríguez, L.; Duran-Llacer, I.; García, W.; Cardenas, R.; Urrutia, R. Assessment of the Diffuse Attenuation Coefficient of Photosynthetically Active Radiation in a Chilean Lake. *Remote Sens.* **2022**, *14*, 4568. [CrossRef]

51. Parra, M.; Jimenez, J.M.; Lloret, J.; Parra, L. Description of the Processing Technique for the Monitoring of Marine Environments with a Sustainable Approach Using Remote Sensing. In *Water, Land, and Forest Susceptibility and Sustainability: Insight Towards Management, Conservation and Ecosystem Services: Volume 2: Science of Sustainable Systems*; Elsevier: Amsterdam, The Netherlands, 2023; Volume 2, pp. 165–188. ISBN 9780443158476.

52. DGA. *Ministerio de Obras Públicas Nombre Consultores: Director del Proyecto Profesionales*; DGA: Santiago, Chile, 2018.

53. Smith, B.; Pahlevan, N.; Schalles, J.; Ruberg, S.; Errera, R.; Ma, R.; Giardino, C.; Bresciani, M.; Barbosa, C.; Moore, T.; et al. A Chlorophyll-a Algorithm for Landsat-8 Based on Mixture Density Networks. *Front. Remote Sens.* **2021**, *1*, 623678. [CrossRef]

54. Lobo, F.L.; Costa, M.P.F.; Novo, E.M.L.M. Time-Series Analysis of Landsat-MSS/TM/OLI Images over Amazonian Waters Impacted by Gold Mining Activities. *Remote Sens. Environ.* **2015**, *157*, 170–184. [CrossRef]

55. Pamula, A.S.P.; Gholizadeh, H.; Krzmarzick, M.J.; Mausbach, W.E.; Lampert, D.J. A Remote Sensing Tool for near Real-Time Monitoring of Harmful Algal Blooms and Turbidity in Reservoirs. *J. Am. Water Resour. Assoc.* **2023**, *8*, 295. [CrossRef]

56. Attiah, G.; Kheyrollah Pour, H.; Scott, K.A. Lake Surface Temperature Retrieved from Landsat Satellite Series (1984 to 2021) for the North Slave Region. *Earth Syst. Sci. Data* **2023**, *15*, 1329–1355. [CrossRef]

57. Magrì, S.; Ottaviani, E.; Prampolini, E.; Federici, B.; Besio, G.; Fabiano, B. Application of Machine Learning Techniques to Derive Sea Water Turbidity from Sentinel-2 Imagery. *Remote Sens. Appl.* **2023**, *30*, 100951. [CrossRef]

58. Arias-Rodriguez, L.F.; Tüzün, U.F.; Duan, Z.; Huang, J.; Tuo, Y.; Disse, M. Global Water Quality of Inland Waters with Harmonized Landsat-8 and Sentinel-2 Using Cloud-Computed Machine Learning. *Remote Sens.* **2023**, *15*, 1390. [CrossRef]

59. Sagan, V.; Peterson, K.T.; Maimaitijiang, M.; Sidike, P.; Sloan, J.; Greeling, B.A.; Maalouf, S.; Adams, C. Monitoring Inland Water Quality Using Remote Sensing: Potential and Limitations of Spectral Indices, Bio-Optical Simulations, Machine Learning, and Cloud Computing. *Earth Sci. Rev.* **2020**, *205*, 103187. [CrossRef]

60. Yin, Z.; Li, J.; Zhang, B.; Liu, Y.; Yan, K.; Gao, M.; Xie, Y.; Zhang, F.; Wang, S. Increase in Chlorophyll-a Concentration in Lake Taihu from 1984 to 2021 Based on Landsat Observations. *Sci. Total Environ.* **2023**, *873*, 162168. [CrossRef] [PubMed]

61. Luo, J.; Ni, G.; Zhang, Y.; Wang, K.; Shen, M.; Cao, Z.; Qi, T.; Xiao, Q.; Qiu, Y.; Cai, Y.; et al. A New Technique for Quantifying Algal Bloom, Floating/Emergent and Submerged Vegetation in Eutrophic Shallow Lakes Using Landsat Imagery. *Remote Sens. Environ.* **2023**, *287*, 113480. [CrossRef]

62. Rodríguez-López, L.; Duran-Llacer, I.; González-Rodríguez, L.; Abarca-del-Rio, R.; Cárdenas, R.; Parra, O.; Martínez-Retureta, R.; Urrutia, R. Spectral Analysis Using LANDSAT Images to Monitor the Chlorophyll-a Concentration in Lake Laja in Chile. *Ecol. Inform.* **2020**, *60*, 101183. [CrossRef]

63. Ma, J.; Jin, S.; Li, J.; He, Y.; Shang, W. Spatio-Temporal Variations and Driving Forces of Harmful Algal Blooms in Chaohu Lake: A Multi-Source Remote Sensing Approach. *Remote Sens.* **2021**, *13*, 427. [CrossRef]

64. Hu, C. A Novel Ocean Color Index to Detect Floating Algae in the Global Oceans. *Remote Sens. Environ.* **2009**, *113*, 2118–2129. [CrossRef]

65. Box, G.E. *Box 2015*; John Wiley & Sons: Hoboken, NJ, USA, 2015.

66. Korstanje, J. *Advanced Forecasting with Python*; Apress: New York, NY, USA, 2021.

67. Dickey, D.A.; Fuller, W.A. Distribution of the Estimators for Autoregressive Time Series With a Unit Root. *J. Am. Stat. Assoc.* **1979**, *74*, 427–431. [CrossRef]

68. Hochreiter, S.; Urgen Schmidhuber, J. Long Short-Term Memory. *Neural Comput.* **1997**, *9*, 1735–1780. [CrossRef]

69. Yu, Y.; Si, X.; Hu, C.; Zhang, J. A Review of Recurrent Neural Networks: Lstm Cells and Network Architectures. *Neural Comput.* **2019**, *31*, 1235–1270. [CrossRef]

70. Rumelhart, D. Learning Representations by Back-Propagating Errors. *Nature* **1986**, *323*, 533–536. [CrossRef]

71. Gers, F.A.; Urgen Schmidhuber, J.; Cummins, F. Learning to Forget: Continual Prediction with LSTM. *Neural Comput.* **2020**, *12*, 2451–2471. [CrossRef]

72. Das, K.; Jiang, J.; Rao, J.N.K. Mean Squared Error of Empirical Predictor. *Ann. Statist.* **2004**, *32*, 818–840. [CrossRef]

73. Maier, A.K.; Syben, C.; Stimpel, B.; Würfl, T.; Hoffmann, M.; Schebesch, F.; Fu, W.; Mill, L.; Kling, L.; Christiansen, S. Learning with Known Operators Reduces Maximum Error Bounds. *Nat. Mach. Intell.* **2019**, *1*, 373–380. [CrossRef]

74. Everitt, B.; Howell, D.C. *Encyclopedia of Statistics in Behavioral Science*; John Wiley & Sons: Hoboken, NJ, USA, 2005; ISBN 0470860804.

75. Luetkepohl, H. *New Introduction to Multiple Time Series Analysis*; Springer: Berlin/Heidelberg, Germany, 2005.

76. Hyndman, R.J.; Athanasopoulos, G. *Forecasting: Principles and Practice*; Springer: Berlin/Heidelberg, Germany, 2018.

77. Garson, G.D. Interpreting Neural-Network Connection Weights. *AI Expert* **1991**, *6*, 46–51.

78. Luíza da Costa, N.; Dias de Lima, M.; Barbosa, R. Evaluation of Feature Selection Methods Based on Artificial Neural Network Weights. *Expert Syst. Appl.* **2021**, *168*, 114312. [CrossRef]

79. Adrian, R.; O'Reilly, C.M.; Zagarese, H.; Baines, S.B.; Hessen, D.O.; Keller, W.; Livingstone, D.M.; Sommaruga, R.; Straile, D.; Van Donk, E.; et al. Lakes as Sentinels of Climate Change. *Limnol. Oceanogr.* **2009**, *54*, 2283–2297. [CrossRef]

80. Ramaraj, M.; Sivakumar, R. Integration of Band Regression Empirical Water Quality (BREWQ) Model with Deep Learning Algorithm in Spatiotemporal Modeling and Prediction of Surface Water Quality Parameters. *Model. Earth Syst. Environ.* **2023**, *9*, 3279–3304. [CrossRef]

81. Zhang, H.; Xue, B.; Wang, G.; Zhang, X.; Zhang, Q. Deep Learning-Based Water Quality Retrieval in an Impounded Lake Using Landsat 8 Imagery: An Application in Dongping Lake. *Remote Sens.* **2022**, *14*, 4505. [CrossRef]