



Article

Estimating Chlorophyll-*a* of Inland Water Bodies in Greece Based on Landsat Data

Vassiliki Markogianni ^{1,*}, Dionissios Kalivas ², George P. Petropoulos ³  and Elias Dimitriou ¹ 

¹ Institute of Marine Biological Resources and Inland Waters, Hellenic Centre for Marine Research, 46.7 km of Athens-Sounio Avenue, 19013 Attica, Greece; elias@hcmr.gr

² Department of Natural Resources Management and Agricultural Engineering, Agricultural University of Athens, 75 Iera Odos, 11855 Athens, Greece; kalivas@aua.gr

³ Department of Geography, Harokopio University of Athens, El. Venizelou 70, 17671 Kallithea, Athens, Greece; gpetropoulos@hua.gr

* Correspondence: vmarkogianni@hcmr.gr; Tel.: +30-22910-76349

Received: 27 May 2020; Accepted: 24 June 2020; Published: 29 June 2020



Abstract: Assessing chlorophyll-*a* (Chl-*a*) pigments in complex inland water systems is of key importance as this parameter constitutes a major ecosystem integrity indicator. In this study, a methodological framework is proposed for quantifying Chl-*a* pigments using Earth observation (EO) data from Landsat 7 Enhanced Thematic Mapper Plus (ETM+) and 8 Operational Land Imager (OLI) sensors. The first step of the methodology involves the implementation of stepwise multiple regression (MLR) analysis of the available Chl-*a* dataset. Then, principal component analysis (PCA) is performed to explore Greek lakes' potential interrelationships based on their Chl-*a* values in conjunction with certain criteria: their characteristics (artificial/natural), typology, and climatic type. Additionally, parameters such as seasonal water sampling and the date difference between sampling and satellite overpass are taken into consideration. Next, is implemented a stepwise multiple regression analysis among different groups of cases, formed by the criteria indicated from the PCA itself. This effort aimed at exploring different remote sensing-derived Chl-*a* algorithms for various types of lakes. The practical use of the proposed approach was evaluated in a total of 50 lake water bodies (natural and artificial) from 2013–2018, constituting the National Lake Network Monitoring of Greece in the context of the Water Framework Directive (WFD). All in all, the results evidenced the suitability of Landsat data when used with the proposed technique to estimate log-transformed Chl-*a*. The proposed scheme resulted in the development of models separately for natural ($R = 0.78$; $RMSE = 1.3 \mu\text{g/L}$) and artificial lakes ($R = 0.76$; $RMSE = 1.29 \mu\text{g/L}$), while the model developed without criteria proved weaker ($R = 0.65$; $RMSE = 1.85 \mu\text{g/L}$) in comparison to the other ones examined. The methodological framework proposed herein can be used as a useful resource toward a continuous monitoring and assessment of lake water quality, supporting sustainable water resources management.

Keywords: Landsat; chlorophyll-*a*; PCA; MLR analysis; lakes; WFD

1. Introduction

Accumulating passive exploitation of natural resources, improper land-use practices, and irregular development activities in lake basins undermine various significant functions of water resources [1]. Surface water provides exceptional financial benefits, regarding water supply (quantity and quality), fisheries, agriculture, wildlife resources, and recreation and tourism opportunities (Ramsar Information Paper no. 1 2007 [2]). The need for sustainable management of water bodies highlights the fact that water resources are not inexhaustible and have limited resistance under anthropogenic pressures such as ongoing drainage, conversion, and pollution.

Hence, one of the most significant aspects for the sustainable management of water bodies—lakes in particular—is the constant monitoring of their quality, as well as of their watersheds [3]. Water quality parameters comprising physical, chemical, and biological properties are conventionally measured by collecting samples from the field and then analyzing those samples in the laboratory. Although in situ monitoring provides high accuracy, it is a time-consuming procedure, and it cannot ensure a simultaneous water quality dataset on a regional or greater scale [3–5]. Furthermore, traditional point sampling methods are not capable of detecting the spatial or temporal variations in water quality, as required in extensive assessment and management of water bodies. On the other hand, geoinformation and especially geospatial technologies provide a promising direction in that respect. In particular, the combined use of Earth observation (EO) and geographical information systems (GIS) allows monitoring in an efficient and robust way lake parameters over variable spatial scales, including lakes that are otherwise inaccessible [6,7].

Various EO instruments mounted on either airborne or satellite platforms acquire spectral information and measure the energy from the water's surface at different wavelengths [3]. The most commonly used approach to inland water remote sensing involves fitting a standard linear regression between spectral band/band ratio values and temporally coincident in situ water quality measurements [8]. Visible, near-infrared (NIR), and shortwave infrared (SWIR) bands of the solar spectrum are usually used by many researchers to acquire powerful correlations—through empirical approaches—among water column reflectance values and constituents in different water bodies [3,9–14]. EO data from several satellite and airborne sensors such as Satellite Pour l'Observation de la Terre (SPOT), Moderate Resolution Imaging Spectroradiometer (MODIS), Earth Observing 1-Hyperion, and Medium Resolution Imaging Spectrometer (MERIS) were used for chlorophyll-*a* (Chl-*a*) estimation [15–19]. Nonetheless, it was revealed that Landsat seems to be more appropriate and widely used for Chl-*a* assessment due to its temporal coverage, spatial resolution, and easy accessibility [3].

Chl-*a* is the major indicator of trophic state and considered as one of the top water pollution indices related to public health, eutrophication, and deterioration of ecosystem habitat. In Case 2 waters (i.e., inland and coastal waters), the optical properties are measured based on a compound of dissolved organic matter, dead organic and inorganic particulate matter, and phytoplankton (Chl-*a*). Therefore, determination of Chl-*a* concentration is much more complex and less accurate [3,20–22]. Oligotrophic to mesotrophic waterbodies with low biomass present a Chl-*a* spectrum characterized by a sun-induced fluorescence peak centered at 680 nm [8,23], while eutrophic waterbodies (high biomass) present a fluorescence signal which is masked by absorption features and backscatter peaks around 665 nm and 710 nm, respectively [8,24]. The ratio between these two wavelengths is widely used to quantify Chl-*a* concentrations with satisfactory accuracy [8,25].

Many studies have so far focused on monitoring eutrophication or trophic state through Chl-*a* concentration retrieval in Greek lakes. For example, Peppas et al. [26] applied Chl-*a* detection algorithms in Lake Pamvotis using Sentinel-2 data, Markogianni et al. [27] developed empirical Chl-*a* quantitative models based on Landsat 5 images in the brackish urban shallow Koumoundourou lake, while Markogianni et al. [14] investigated the suitability of the OLI instrument on board the Landsat 8 satellite platform in accurately estimating Chl-*a* in the largest freshwater body of Greece (Trichonis Lake). In another study, Kontopoulou et al. [28] tried to exploit the Water Framework Directive (WFD) dataset and measure Chl-*a* concentrations by using Landsat 8 data in 11 (six natural; five artificial) of the 50 lakes comprising the national sampling network.

In this study, we discuss the utility of remotely sensed techniques in the qualitative assessment of 50 lake water bodies, particularly of Chl-*a* concentrations, derived from the WFD (2000/60/EC) monitoring network for lakes in Greece. WFD provides a scheme for the conservation and improvement of inland, ground, and coastal waters' ecological status and aims to harmonize European legislation on water. Thus, pan-European hydromorphological, physicochemical, and biological datasets are used to

determine ecological status of surface waters (Article 8) in order to assure and further improve future water quality and quantity [29].

In purview of the above, this study proposes a methodological framework that aims to provide Chl-*a* in Case 2 complex inland waters of Greece by generating accurate quantitative models with EO data from the Landsat 7 ETM+ and 8 OLI satellite series. The methodology applied initially includes the implementation of stepwise multiple regression (MLR) analysis of the whole available Chl-*a* dataset with the basic aim of exploring its potential to establish robust Chl-*a* quantitative algorithms, regardless of lake characteristics. Then, principal component analysis (PCA) is performed to highlight the most significant parameters (artificial/natural, WFD typology, water sampling season, and climatic type) affecting Chl-*a* values. This procedure was considered valuable for the next step, involving the execution of multiple stepwise MLR analyses—based on PCA results—among different groups of cases. This effort aimed at exploring different remote sensing-derived Chl-*a* algorithms for various types of lakes according to the most significant lake characteristics. The practical use of the proposed approach was evaluated in a total of 50 lake water bodies (natural and artificial) from 2013–2018, constituting the National Lake Network Monitoring of Greece in the context of the Water Framework Directive (WFD).

2. Materials and Methods

2.1. Study Area

The National Monitoring Network of Waters in Greek lakes, according to the Joint Ministerial Decision 140384/2011, is implemented by the Goulandris Natural History Museum, Greek Biotope/Wetland Centre (EKBY). The network consists of 50 lake water bodies, natural and artificial.

At the majority of the lakes, only one sampling station is detected, except for trans-boundary lakes (Megali Prespa, Mikri Prespa, and Doirani), where two sampling stations are located (Table 1; Figure 1). From the total of 53 sampling sites, there are 27 surveillance and 26 operational ones. Surveillance stations operate in water bodies of good status, for a certain period of time (one year), while operational stations are monitored on a monthly or seasonal basis, in water bodies which fail to achieve good status.

Table 1. General characteristics of the lakes comprising the National Lake Network Monitoring in Greece (Water Framework Directive (WFD) [30]).

No	National Name Station	Natural (N)/ Artificial (A)	Typology	Koppen Climate Classification	Mean Depth (m)	No	National Name Station	Natural (N)/ Artificial (A)	Typology	Koppen Climate Classification	Mean Depth (m)
1	Lake Ladona	A	L-M8	Csa	-	28	Lake Petron	N	GR-VSNL	Cfa	3.1
2	Lake Pineiou	A	L-M8	Csa	15.1	29	Lake Zazari	N	GR-SNL	Cfa	3.95
3	Lake Stymfalia	N	GR-VSNL	Csa	1.3 ¹	30	Lake Cheimaditida	N	GR-VSNL	Cfa	1.01
4	Lake Feneou	A	L- M5/7W	Csa	10.5	31	Lake Kastorias	N	GR-SNL	Cfa	3.7
5	Lake Kremaston	A	L-M8	Csa	47.2	32	Lake Sfikias	A	L- M5/7W	BSk	23.2
6	Lake Kastrakiou	A	L- M5/7W	Csa	33.2	33	Lake Asomaton	A	L- M5/7W	BSk	20.8
7	Lake Stratou	A	GR-SR	Csa	9.6	34	Lake Polyfytou	A	L- M5/7W	Cfa	22.4
8	Lake Tavropou	A	L- M5/7W	Csa	15.0	35	Lake Mikri Prespa A	N	GR-SNL	Csa	3.95
9	Lake Lysimacheia	N	GR-SNL	Csa	3.5	36	Lake Mikri Prespa B	N	GR-DNL	Csa	-
10	Lake Ozeros	N	GR-SNL	Csa	3.8	37	Lake Megali Prespa A	N	GR-DNL	Csa	17
11	Lake Trichonida	N	GR-DNL	Csa	29.6	38	Lake Megali Prespa B	N	GR-DNL	Csa	-
12	Lake Amvrakia	N	GR-DNL	Csa	23.4	39	Lake Doirani 1	N	GR-SNL	Dfc	4.6
13	Lake Voulkaria	N	GR-VSNL	Csa	0.96	40	Lake Doirani 2	N	GR-SNL	Dfc	-
14	Lake Saltini	N	GR-SP1 (lagoon)	Csa	-	41	Lake Pikrolimni	N	GR-SP2 (special category)	Cfb	1.2
15	Lake Mornou	A	L- M5/7W	Csa	38.5	42	Lake Koroneia	N	GR-VSNL	Csa	3.8
16	Lake Evinou	A	L- M5/7W	Csa	31.5	43	Lake Volvi	N	GR-DNL	Csa	12.3
17	Lake Pigon Aoou	A	L- M5/7W	Csa	20.8	44	Lake Kerkini	A	GR-SR	Dfc	2.19
18	Lake Pournariou	A	L- M5/7W	Csa	29.8	45	Lake Leukogeion	A	GR-SR	Dfc	4.05
19	Lake Pamvotida	N	GR-SNL	Csa	5.3	46	Lake Ismarida	N	GR-VSNL	Csb	0.9
20	Lake Pournariou II	A	GR-SR	Csa	11.7	47	Lake Platanovrysis	A	L- M5/7W	Dfc	26.4
21	Lake Marathona	A	L-M8	Csa	15.8	48	Lake Thisavrou	A	L- M5/7W	Dfc	38.4
22	Lake Dystos	N	GR-VSNL	Csa	-	49	Lake Gratinis	A	L- M5/7W	Csb	14.2
23	Lake Yliki	N	GR-DNL	Csa	20.1	50	Lake N. Adrianis	A	GR-SR	Csb	-
24	Lake Paralimni	N	GR-SNL	Csa	2.99	51	Lake Kourna	N	GR-DNL	Csa	15
25	Lake Karlas	A	GR-SR	BSk	0.9	52	Lake Bramianou	A	L-M8	Csa	10.1
26	Lake Smokovou	A	L-M8	Csa	-	53	Lake Faneromenis	A	L-M8	Csa	9.98
27	Lake Vegoritida	N	GR-DNL	Dfa	26.52	¹ [31]					

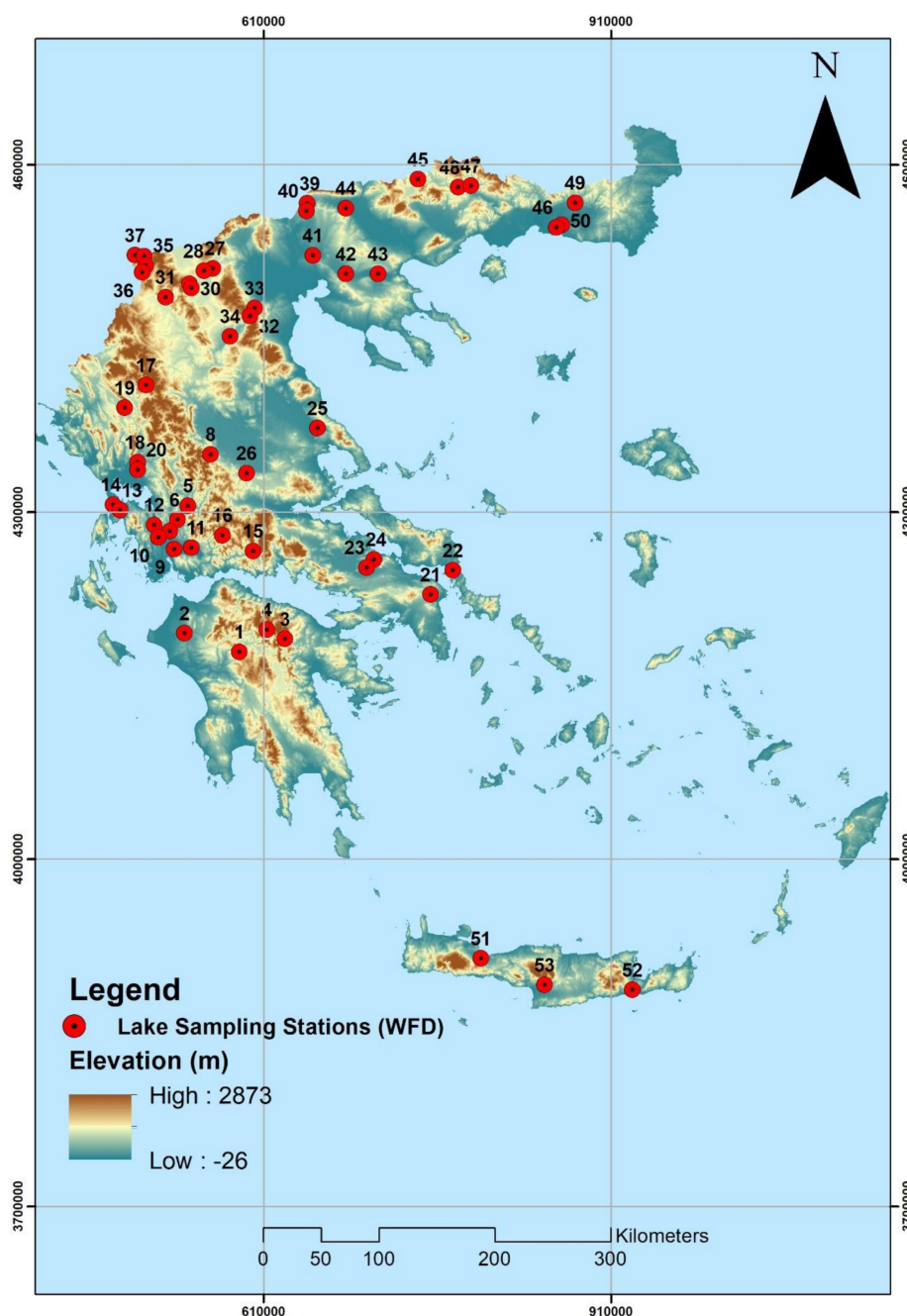


Figure 1. Monitoring network for lakes in Greece (GGRS_1987 coordinate system; Transverse Mercator projection; numbers of sampling stations coincide with the numbers presented in Table 1).

2.2. Data Acquisition

2.2.1. In Situ Data

In this study, Chl-*a* concentrations, measured from 2013 to 2018 (summer, autumn, and spring) throughout the studied lake stations, were acquired, while the Chl-*a* concentrations (in $\mu\text{g/L}$) were determined spectrophotometrically (Method 10200 H [32]). Those data were retrieved from the EKBY site (Goulandris Natural History Museum, Greek Biotope/Wetland Centre; <http://biodiversity-info.gr/index.php/el/lakes-data#!IMGP4731>), where more details about the sampling periods, stations, and variables measured can be found. Apart from the Chl-*a* measurements, some basic characteristics of the 50 studied lakes were also considered, such as whether they are natural or artificial, their

typology according to the Water Framework Directive (Table 2), and the climatic type according to the Köppen–Geiger climate classification (Table 3) [33]. The determination of the Mediterranean lake types is based on the 2013/480/EU decision (Table 2), while the typology of each lake was retrieved from the respective reports, acquired from the Environment and Energy Ministry’s website (<http://wfdver.ypeka.gr/>).

Table 2. WFD national lake types.

Type	Characteristics	Elevation (m)	Precipitation (mm) and Temperature (°C) (Mean Annual Values)	Surface (km ²)	Mean Depth (m)	Catchment (km ²)	Thermal Stratification
Artificial							
L- M5/7W	Deep, large reservoirs, silicate, wet areas	<1000	>800 and/or <15	>0.5	>15	<20000	
L-M8	Deep, large reservoirs, limestones	<1000	-	>0.5	>15	<20000	
GR-SR	Shallow reservoirs	<1000	-	>0.5	<15	-	
Natural							
GR-DNL	Deep lakes	0–1000		>0.5	>9		Thermal monomictic
GR-SNL	Shallow lakes	0–1000		>0.5	3–9		Polymictic
GR-VSNL	Very shallow lakes	0–1000		>0.5	<3		Polymictic

Table 3. Köppen–Geiger’s classification of climatic regions of Greece.

Climate Type	Description
Csa	Mediterranean hot summer climates
Csb	Mediterranean warm/cool summer climates
Cfa	Humid subtropical climates
Cfb	Oceanic climate
Dfa	Hot summer continental climates
Dfc	Subarctic or boreal climates
BSk	Cold semi-arid climate

2.2.2. EO Data

Landsat 8 satellite was launched in 2013 and bears two sensors, the OLI and the Thermal Infrared Sensor (TIR). It also includes a narrower near-infrared band and a 12-bit radiometric resolution compared to the 8 bits of previous Landsat satellites [19,34]. The main aim was to match as many records from the available Chl-*a* dataset as possible with Landsat 8 images. In cases where no Landsat 8 images were available or were cloud-covered, Landsat 7 ETM+ images were used. Landsat 7, launched in 1999, introduced the ETM+, whereby its analysis was similar to Thematic Mapper (TM) except for a 60-m thermal and a new 15-m panchromatic band [35]. Since 2003, Landsat 7 has a sensor deficiency where the scan line corrector (SCL) is off and, even though those images are characterized by black line gaps [36], its radiometric and geometric analyses remain undisturbed [37].

In order to cover the 50 lakes throughout Greece and the different sampling dates, 2013–2016 and 2018 time series of 296 Landsat imagerys (102 L7 ETM+, 194 L8 OLI) were downloaded from the USGS (United States Geological Survey) data center (<https://earthexplorer.usgs.gov/>). Images from both sensors reside in Landsat Collection 1 Level-1 category data products. Furthermore, the mean time window between the satellite overpass and the in situ measurements was about ± 15 days. More information about the bandwidths and spatial resolution of the aforementioned sensors can be found at NASA’s official website (<https://landsat.gsfc.nasa.gov/>).

2.3. Satellite Data Pre-Processing

Pre-processing was carried out for the Landsat images using the semi-automatic classification plugin (SCP) of the free and open-source cross-platform desktop geographic information system, Q-GIS v. 3.6.3-Noosa [38]. Images of both Landsat 7 ETM+ and Landsat 8 sensors were subject to the following pre-processing steps:

1. Conversion of images from digital numbers (DN) to the physical measure of top-of-atmosphere reflectance (TOA).
2. Atmospheric correction using the DOS1 method (Dark Object Subtraction 1; image-based technique), which was applied to all bands except for thermal ones.
3. Creation of a band stack set for each image. The band stack set of L7 ETM+ includes the bands B1 (blue), B2 (green), B3 (red), B4 (NIR), B5 (SWIR1), and B7 (SWIR2), while L8 set incorporates bands B2 (blue), B3 (green), B4 (red), B5 (NIR), B6 (SWIR1), and B7 (SWIR2).

More information about each correction including the theoretical background can be found in the SCP Documentation Release 6.2.0.1 [39]. To ensure the use of only cloud-free pixels over the sampled lakes, the Cloud Masking QGIS plugin (<https://smbyc.github.io/CloudMasking>) was used. By using this tool, clouds, cloud shadow, cirrus, aerosols, and ice/snow were masked for all Landsat images using the combination of the Fmask and Blue Band processes.

The pre-processing procedure of L7 images also included the retrieval of data that coincided with the aforementioned black diagonal stripes. By visual checking, sampling sites covered by those stripes were recognized and, by employing GIS operations and focal statistics, the mean value within a circle of seven cells around it for each input cell location was calculated. This radius was the most adequate among several trials. Then, by applying the Con and IsNull functions, only the cells that had no values were replaced. In cases where parts of sampled lakes were cloud-covered, the SetNull and IsNull functions were combined with the cloud mask calculated in earlier stages to remove the cloud-biased pixels.

2.4. Statistical Methods of Analysis

2.4.1. Basic Statistics and PCA

Basic statistical analysis among the Chl-*a* datasets of 2013, 2014, 2015, 2016, and 2018 was carried out including the calculation of mean, median, standard deviation, and min/max. Based on the available lake characteristics (Table 1), a factor analysis was used with the Varimax-rotated principal component analysis (PCA) as a factor extraction method to interpret the major patterns of Chl-*a* variation within the whole dataset. Furthermore, PCA's basic aim was to explore and indicate the presence of inter-correlations among Chl-*a* concentrations, lake characteristics, climatic type, typology, and seasonal sampling and further indicate which were the most significant parameters affecting Chl-*a* concentrations in the studied lakes. The seasons were defined as follows: summer (June, July, and August), autumn (September, October, and November), and spring (March, April, and May). The Chl-*a* concentration distribution was grouped and categorized by the most significant variables/criteria that PCA indicated, using box-plot diagrams. PCA was performed in SPSS Statistical Package (v. 24.0).

2.4.2. Developing Relationships between Landsat and Chlorophyll-*a* Data

There are several band combinations and transformations proposed in the relevant literature for establishing relationships between the Landsat reflectance data (independent) and Chl-*a*, log (Chl-*a*), and ln (Chl-*a*) data [14]. Matthews [40] reported that the wavelength areas where water reflects and scatters the majority of the entering solar radiation are those that are mostly used for the monitoring of water quality constituents. These wavelengths incorporate the water-leaving reflectance in visible and NIR wavelengths of the electromagnetic spectrum. Thus, to compare Chl-*a* concentrations and EO data, visible (blue, green, and red), NIR, and SWIR bands were used, while

aerosol, cirrus, panchromatic, and TIRS thermal spectral bands were excluded. In addition, Landsat band ratios, additions, subtractions, and log and ln transformations were added to the analysis to establish accurate and reliable estimation algorithms. Based on the respective literature, more than 75 available band transformations/combinations were developed, including spectral indices such as enhanced vegetation index (EVI) [41], normalized vegetation index (NRVI) [42], normalized difference water index (NDWI) [43], modified normalized difference water index (MNDWI) [44], green normalized difference vegetation index (GNDVI) [45], normalized difference vegetation index (NDVI) [46], and surface algal bloom index (SABI) [47]. As a first step, the main objective was to distinguish the highest important predictors among the aforementioned band transformations/combinations by conducting a correlation analysis among them and the Chl-*a*, log (Chl-*a*), and ln (Chl-*a*) concentrations. Taking into account only the correlations significant at the 0.01 level, a threshold value of Spearman r was set to ± 0.4 (moderate relationship [48]). Variables that presented an r value equal or greater than the aforementioned value were selected. Then, those variables/predictors were further inserted—combined in various ways—in numerous stepwise MLRs while examining for statistical performance and residuals. Further criteria consisting of multi-collinearity, tolerance factor, and variance inflation factor (VIF) were applied and checked to a subset of optimal models in order to further compare them and select the simplest models with higher accuracy (higher R^2). To develop a robust Chl-*a* algorithm, stepwise MLR analyses were evaluated in the context of two scenarios, as also illustrated in Figure 2.

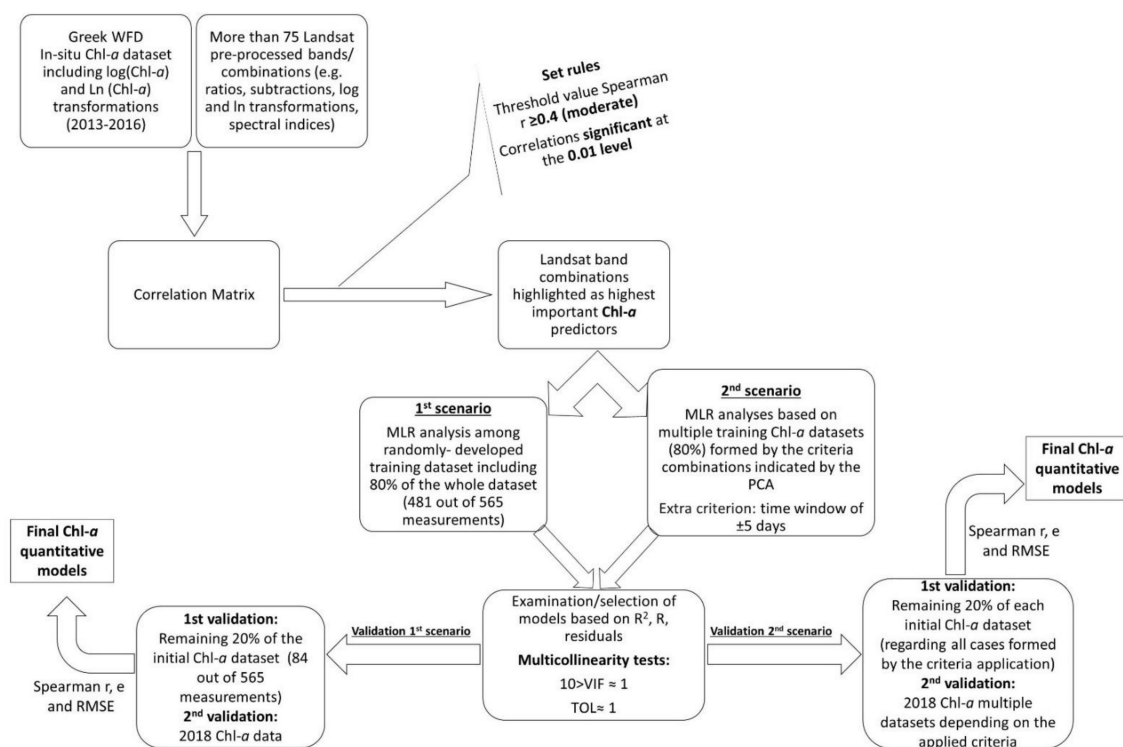


Figure 2. Flowchart summarizing the methodology followed in this study.

First scenario: The basic aim of this scenario was the exploration of the potential of stepwise MLR analysis to establish robust Chl-*a* quantitative algorithms. The training dataset was randomly established and included 80% of the whole available dataset (481 out of 565 Chl-*a* measurements, regardless of lake characteristics; Figure 2). This methodology is widely used in remote sensing of inland water quality and attempted to develop an algorithm that can be applied in multiple types of lakes.

Second scenario: This scenario aimed at exploring different remote sensing-derived Chl-*a* algorithms for various types of lakes according to the most significant lake characteristics resulting

from the PCA of Chl-*a* values, lake climatic type, typology, and seasonal sampling. PCA-indicated parameters were combined in all possible ways, forming the various modeling cases examined to highlight the most optimal model. In this scenario, the various training datasets were subsets of the whole dataset of in situ Chl-*a* values, defined by the respective combination of the PCA-derived criteria. Each training dataset was then used in a separate MLR analysis (Figure 2). As an additional criterion, a confined time window of ± 5 days was used, defined as the difference between field measurements and satellite overpass. Kloiber et al. [49] observed that in situ measurements within one day of the satellite image date resulted in the best calibrations; however, a great number of ground observations with a longer time difference balanced some of the loss of accuracy. Hence, since the available dataset of in situ Chl-*a* measurements is quite large, the ± 5 -day window was determined presuming that, during this period, the water quality usually does not exhibit large and rapid water quality fluctuations.

During the analyses which included both scenarios mentioned above, it was ensured that the separate training and validation datasets contained representative samples of the data (80% and 20%, respectively) from the initial dataset. Training datasets of both scenarios comprised the in situ Chl-*a* measurements of 2013–2016.

2.4.3. Validation Approach

The qualitative capability of the derived algorithms was evaluated by regression analysis in two ways. Algorithms derived from the MLR based on the first scenario (random dataset) were applied in Landsat images associated with the respective validation dataset (20%) and in images of the year 2018 (Figure 2). The algorithms resulting from the second scenario's training datasets (criteria) were applied in available Landsat imageries concerning the Chl-*a* of each respective validation dataset (20%) and the in situ Chl-*a* values of 2018. After applying several models on the available images, the optimal ones were verified based on Spearman's *r* correlation coefficient, mean residual value (*e*), and Root-Mean-Square Error (RMSE) statistical indices (Figure 2).

3. Results

3.1. Statistical Analyses

Total measurements of in situ Chl-*a* concentrations of years 2013–2016 and 2018 were 702, including all sampling campaigns in all lakes, while the most measurements were presented in 2015 (Table 4). Sampling campaigns of 2014 indicated the highest range of Chl-*a* values (1026.7 $\mu\text{g/L}$), while, in the remaining years, the annual ranges were significantly lower. The minimum Chl-*a* value was 0.22 $\mu\text{g/L}$ for years 2013 and 2014, whereas this value increased over the years. Despite this fact, mean Chl-*a* value remained similar between different years, indicating no special deterioration trend in Greek lake water quality. Furthermore, datasets of all years were skewed to the right with low values, while the kurtosis of all years was described as leptokurtic (fat tails).

Table 4. Descriptive statistics summary table of in situ chlorophyll-*a* (Chl-*a*) of years 2013–2016 and 2018.

	Chl- <i>a</i> ($\mu\text{g/L}$) 2013	Chl- <i>a</i> ($\mu\text{g/L}$) 2014	Chl- <i>a</i> ($\mu\text{g/L}$) 2015	Chl- <i>a</i> ($\mu\text{g/L}$) 2016	Chl- <i>a</i> ($\mu\text{g/L}$) 2018	Chl- <i>a</i> ($\mu\text{g/L}$) all
<i>N</i>	157	155	172	82	136	702
Minimum	0.22	0.22	0.36	0.45	0.58	0.22
Maximum	263.9	1026.9	286.96	292.7	361.7	1026.9
Mean	24.44	29.13	23.06	27.99	27.03	26.06
Standard Deviation	46.94	101.82	46.27	45.34	55.92	64.31
Skewness	3.05	7.7	3.24	3.23	3.83	7.87
Kurtosis	9.7	67.8	11.7	14.1	16.9	97.5

Performance of factor analysis (using principal component analysis (PCA) as the extraction method) included five variables/criteria, e.g., sampling season, in situ Chl-*a* concentrations, lake nature (natural/artificial), typology, and climatic type. Factor analysis was implemented to obtain an indication of underlying common factors (components) that explain the interrelationships among those aforementioned variables. The analysis initially extracted five components (Table 5). Finally, only the first three components with eigenvalues greater than one were retained (as those which represent a real underlying factor) in the extraction sums of squared loadings. The percentage of the total variance explained by the three components (calculated after the implementation of the Varimax rotation method) was 35.4%, 21.9%, and 20.8%, cumulatively explaining 78% of the total variance.

Table 5. Total variance explained, initial eigenvalues, and extracted components.

Component	Initial Eigenvalues			Extraction Sums of Squared Loadings			Rotation Sums of Squared Loadings		
	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	1.78	35.5	35.6	1.78	35.6	35.6	1.8	35.4	35.4
2	1.12	22.4	57.9	1.12	22.4	57.9	1.1	21.9	57.4
3	1.01	20.2	78.1	1.01	20.2	78.1	1.04	20.8	78.12
4	0.79	15.9	93.98						
5	0.30	6.02	100						

Based on rotated component matrix results (Table 6), communalities (summation of the squared loadings in all components) of studied variables are further discussed. Thus, 53% of Chl-*a* variance was explained by the second component, which also explained 57% of the sampling season variance. The first component explained 10% of Chl-*a* variance, which also explained 81.5% of the variance of lake characteristics (natural/artificial) and 79% of the variance of lake typology. The third component explained only 5% of the Chl-*a* variance and 90% of the climatic type variance. Taking into account those results, the variables that mostly contributed to and affected the variance of Chl-*a* concentration were the lake characteristics (natural/artificial) and typology, followed by the sampling season.

Table 6. Rotated component matrix (rotation method: Varimax with Kaiser normalization; rotation converged in five iterations).

Item	Component		
	1	2	3
Sampling season	0.25	0.75	0.26
Chl- <i>a</i> (µg/L)	−0.31	0.73	−0.23
Natural/artificial	0.9	−0.03	0.06
Köppen/climate type	−0.08	0.02	0.95
Typology	0.89	0.01	−0.15

Concentrations of Chl-*a* of natural lakes were found notably higher in comparison to artificial lakes, while the highest measured values were detected in artificial lakes during the summer and autumn months (outlier values) but during autumn in natural lakes (Figure 3a). In general, the seasonality of Chl-*a* concentrations was more evident in natural rather than in artificial lakes. The highest discrepancies were detected in natural lakes (Zazari and Voulkaria lakes, July 2014) followed by autumn sampling campaigns (e.g., Zazari lake, September 2014).

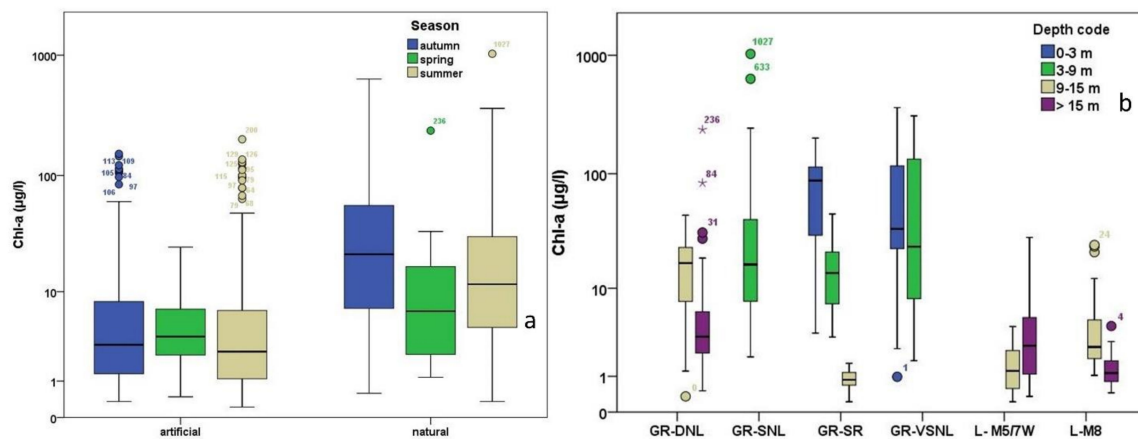


Figure 3. Boxplots presenting basic descriptive statistics (median, percentiles, min/max, outliers, and extremes) of chlorophyll-*a* concentrations of years 2013–2016 and 2018 (a) grouped by the lake characteristics and categorized by the sampling season, and (b) grouped by the lake typology and categorized by lake mean depth. Mean depth classification is based on typology characteristics as explained in Table 1.

Measured Chl-*a* concentrations in both artificial and natural lakes presented a greater range during summer, while most outliers of artificial lakes were also illustrated during summer (Kerkini August 2014, August 2016 and Karla reservoirs, July 2014) and autumn seasons (Kerkini reservoir, October 2015). Spring also presented significant differences between natural and artificial lakes, where median value of the latter was evidently decreased.

The same Chl-*a* pattern is illustrated in a different way by the second boxplot (Figure 3b). As far as the Chl-*a* distribution based on the lake typology is concerned, the highest values were detected at natural GR-VSNL and GR-SNL shallow (Zazari lake) and very shallow lakes (Voulkaria lake). These cases were followed by values measured in deep natural lakes GR-DNL (Yliki lake), while the lowest Chl-*a* values were detected in artificial shallow reservoirs (Pournari II and Stratos reservoirs).

3.2. Relationships between Landsat and Chlorophyll-*a* Data

Blue, green, red, NIR, and SWIR Landsat 8 and Landsat 7 ETM+ bands and more than 75 different band transformations and indices were used for investigating the most suitable relationship to estimate Chl-*a* concentrations throughout 50 Greek lakes.

Correlation analysis among all available variables and Chl-*a*, log (Chl-*a*), and ln (Chl-*a*) concentrations resulted in Spearman *r* values that ranged from −0.6 to +0.6. Based on the correlation matrix, the highest important predictors that met the criterion of the set threshold value of Spearman (*r*) ±0.40 in relation to log (Chl-*a*) were the following: blue/green, green/blue, blue/red, red/blue, red/green, red/SWIR1, log (blue/green), log (blue/red), log blue/log green, log blue/log red, (blue − red)/green, ln blue/ln green, ln blue/ln red, ln green/ln blue, ln red/ln blue, ln red/ln SWIR1, and ln red/ln SWIR2. Those variables/predictors were further inserted in several combinations into numerous stepwise linear regressions. As the number of the generated models was quite large, we only included only the most significant models from a statistical point of view.

Based on the first scenario, MLR analysis among the indicated variables generated several models with good performance. Based on tests on statistical significance of the bi coefficient of the independent factors (*t*-test with *p*-values less than 0.05) and on tests for multicollinearity (variance inflation factor (VIF) with values greater than one and less than 10 and tolerance higher than 0.1), the following model was finally selected as satisfactory (Equation (1); Table 7).

$$\log Chla = 3.599 - 0.63 \times \left(\frac{blue}{red} \right) - 2.183 \times \left(\frac{\ln red}{\ln swir2} \right). \quad (1)$$

Table 7. Regression analysis statistics and model summary among reflectance values and log chlorophyll-*a* concentrations (dependent variable).

Scenario/Model	<i>R</i>	<i>R</i> ²	Adjusted <i>R</i> ²	Std. Error of the Estimate	Change Statistics					Durbin–Watson
					<i>R</i> ² Change	F Change	df1	df2	Sig. F Change	
1A	0.654	0.427	0.425	0.525	0.114	94.99	1	478	0.000	1.948
Predictors: (Constant), blue/red, ln red/ln SWIR2										

Descriptive statistics results obtained for the training and validation datasets (regarding log Chl-*a*) suggested that there was no discrepancy in the mean (training: 0.8; validation: 0.79) and standard deviation (training: 0.69; validation: 0.68) values between the two groups. To ensure statistically significant results, an independent *t*-test for the mean values was also implemented. Based on these results (*t*-value 0.276, *p* = 0.783), it is assumed that there is no difference in the group means.

Random selection of Chl-*a* measurements and MLR analysis (first scenario, 1A model) yielded a model accompanied by a value of coefficient of determination equal to 0.43, including the band ratios ln red/ln SWIR2 and blue/red (Table 7). The standard error of the estimate was equal to 0.53 and the collinearity statistics were considered acceptable.

Concerning the second scenario, MLR analysis conducted among the selected predictors and the various combinations of the criteria (Table 8) indicated by the PCA generated several models. Examination of the aforementioned statistical indices highlighted two models (Table 9; Equations (2) and (3)). It should also be noted that the various criteria combinations always included the values with Chl-*a* concentrations lower than 500 µg/L in the examined dataset (based on the normal Quantile-Quantile (Q–Q) plot of Chl-*a* concentrations and outlier analysis) and mean depth greater than 5 m to surely avoid the bottom reflectance noise [50]. The criterion of the time window of ±5 days was applied in cases with statistically significant results to examine and explore any further improvement. The 13th criterion case (Table 8) aimed to establish accurate algorithms considering only the mean depth differentiation of the lakes by including the shallow reservoirs (<15 m), as well as shallow (3–9 m) and some deep (>9 m) natural lakes.

Table 8. Description of cases including all possible combinations of criteria applied. The criteria were selected according to principal component analysis (PCA) results.

No. of Case Regarding the Criteria Application	Combination of Criteria
1	Mean depth >5 m, natural lakes, all seasons
2	Mean depth >5 m, artificial lakes, all seasons
3	Mean depth >5 m, natural lakes, season spring
4	Mean depth >5 m, natural lakes, season autumn
5	Mean depth >5 m, natural lakes, season summer
6	Mean depth >5 m, artificial lakes, season spring
7	Mean depth >5 m, artificial lakes, season autumn
8	Mean depth >5 m, artificial lakes, season summer
9	Mean depth >5 m, season spring
10	Mean depth >5 m, season autumn
11	Mean depth >5 m, season summer
12	Mean depth >5 m, date difference (sampling/satellite): ±5 days
13	>3 m mean depth <15 m

Table 9. Regression analysis statistics and model summary among reflectance values and log chlorophyll-*a* concentrations (dependent variable).

Scenario/Model	<i>R</i>	<i>R</i> ²	Adjusted <i>R</i> ²	Std. Error of the Estimate	Change Statistics					Durbin–Watson
					<i>R</i> ² Change	<i>F</i> Change	df1	df2	Sig. <i>F</i> Change	
2A	0.776	0.602	0.587	0.432	0.024	4.818	1	81	0.031	1.666
2B	0.757	0.574	0.563	0.291	−0.006	1.666	1	120	0.199	2.034

The 2A regression model was as follows:

$$\log Chla = 4.443 - 1.421 \times \left(\frac{blue}{green} \right) - 3.454 \times \left(\frac{\ln red}{\ln swir2} \right) + 1.304 \times \left(\frac{red}{green} \right), \quad (2)$$

with 85 selected cases, based on the criteria Chl-*a* concentration <500 µg/L, mean depth >5 m, natural lakes, and all seasons.

The equation of the 2B regression model was as follows:

$$\log Chla = 2.919 - 2.011 \times \left(\frac{\ln red}{\ln swir1} \right) + 1.449 \times \left(\frac{red}{green} \right) - 1.441 \times \left(\frac{\ln red}{\ln blue} \right), \quad (3)$$

with 125 selected cases, developed based on the criteria Chl-*a* concentration <500 µg/L, mean depth >5 m, artificial lakes, all seasons, and date difference between sampling and satellite overpass ±5 days.

The optimal Chl-*a* models developed based on the second scenario incorporated the band ratios ln red/ln SWIR2, blue/green, ln red/ln blue, red/green, and ln red/ln SWIR1, while the coefficient of determination was equal to 0.57 (2B) and 0.6 (2A), and the standard error of the estimate was 0.29 (2B) and 0.4 (2A). Collinearity statistics (tolerance and VIF) of the coefficients were also considered acceptable. In general, the highest beta coefficient values accompanied the band ratios red/green, blue/red, and blue/green, while Durbin–Watson’s statistic test indicated an absence of autocorrelation, especially in the residuals of models developed in the second scenario.

3.3. Regression Model Validation

Developed regression models were validated for both scenarios, while validation datasets were different for each model. Concerning the first scenario’s analysis (1A), the model was applied in Landsat 8 and 7 ETM+ images connected to the remaining 20% of the respective validation dataset (84 out of 565 measurements; first validation) and then for the second validation by using images and all available Chl-*a* measurements of the year 2018 (*N* = 136; Table 10). Models 2A and 2B were validated based on 20% of the remaining dataset characterized by the set criteria (2A, *N* = 21 out of 106 measurements; 2B, *N* = 31 out of 156 measurements; Table 10) and, subsequently, the in situ Chl-*a* values of 2018 were used for the second validation process (2A, *N* = 28; 2B, *N* = 41). More specifically, the 2A model was validated by using the measurements concerning the natural lakes, while, for the validation of 2B model, measurements of artificial lakes with a sampling/satellite date difference of ±5 days, regardless of the sampling season, were used.

Table 10. Statistical indices used to validate the selected algorithms (correlation significant ** at the 0.01 level (two-tailed) and * at the 0.05 level (two-tailed)). RMSE—root-mean-square error.

Scenario/ Models	Spearman <i>r</i>	1st Validation (20%)				Spearman <i>r</i>	2nd Validation (2018 Data)			
		Average In Situ log Chl- <i>a</i>	Average Satellite log Chl- <i>a</i>	Average Residuals	RMSE (log Chl- <i>a</i> /Chl- <i>a</i> ; µg/L)		Average In Situ log Chl- <i>a</i>	Average Satellite log Chl- <i>a</i>	Average Residuals	RMSE(log Chl- <i>a</i> /Chl- <i>a</i> ; µg/L)
1A Training dataset <i>N</i> = 481	0.688** <i>N</i> = 84	0.79	0.76	0.03	0.29/1.95	0.758** <i>N</i> = 136	0.89	0.84	0.05	0.24/1.74
2A Training dataset <i>N</i> = 85	0.782** <i>N</i> = 21	0.85	0.83	0.02	0.095/1.25	0.697** <i>N</i> = 28	0.81	0.87	−0.05	0.13/1.35
2B Training dataset <i>N</i> = 125	0.622** <i>N</i> = 31	0.197	0.237	−0.04	0.1/1.26	0.593** <i>N</i> = 41	0.32	0.31	0.004	0.12/1.32

The first scenario yielded a model (1A) that, despite being characterized by a quite low coefficient of determination (0.43), was accompanied by two validation processes with high Spearman (*r*) values (0.69 and 0.76, significant at the 0.01 level) and quite large validation datasets (*N* = 84 and 136). Additionally, the differences among the mean in situ and satellite-derived values in both validations were not high (Figure 4a), while mean residual (Figure 5) and RMSE values were quite low for both validation processes (RMSE first validation: 1.95 µg/L; RMSE second validation: 1.74 µg/L; Table 10). Hence, this specific model could be included as one of the proposed Chl-*a* quantitative algorithms.

Concerning the models resulting from the second scenario and their validations, 2A and 2B models were those that were mostly proposed. Model 2A was derived from a regression analysis with *R*² equal to 0.6 and Spearman values equal to 0.78 (first validation) and 0.7 (second validation), respectively. Mean in situ and satellite log Chl-*a* values were almost identical (Figure 4b), and this model presented two of the lowest values of average residual (0.95 µg/L; Figure 5). RMSE values were quite low, with 1.25 µg/L (first validation) and 1.35 µg/L (second validation). Likewise, the 2B model demonstrated high enough Spearman values (0.62 and 0.59), low average residuals (Figure 5), and low RMSE values (1.26 and 1.32 µg/L), while it was also accompanied by one of the largest validation datasets (2018 dataset; *N* = 41).

Furthermore, Chl-*a* maps of selected lakes were created after application of the resulting algorithms (Figure 6). The Landsat 8 OLI satellite image of 11 August 2013 was used in order to produce the satellite-derived spatial distribution of Chl-*a* values of this day, while the respective in situ values of those lakes were sampled with −2 and +5 days of difference from the aforementioned date. Application of the 1A model resulted in Chl-*a* concentrations ranging from 0.18 to 58.9 µg/L and from 0.71 to 61.7 µg/L for artificial and natural lakes, respectively (Figure 6a). Application of the 2A model in natural lakes yielded Chl-*a* values ranging between 2.75 and 70.8 µg/L (Figure 6b), while the 2B model resulted in Chl-*a* values varying from 0.019 to 12.6 µg/L (Figure 6c), as far as the artificial lakes are concerned. The 1A model yielded quite greater Chl-*a* values in relation to the 2B model, concerning the artificial lakes, while, in natural lakes, the satellite-derived Chl-*a* values from the 1A model were slightly lower than those produced by the 2A model. Moreover, it should be noted that in situ values concerned point samples, whereas satellite-derived values related to the spatially distributed Chl-*a* concentrations along the lake surfaces.

Summarizing the information derived from the validation process, the most optimal Chl-*a* assessment model throughout the WFD Greek lakes when no information was available about their characteristics was 1A. When the studied lakes were natural, the Chl-*a* model became more complex (2A), incorporating three band ratios; when artificial lakes were the case, then model 2B was proposed with the condition that the date difference between field measurements and satellite overpass ranged from −5 to +5 days.

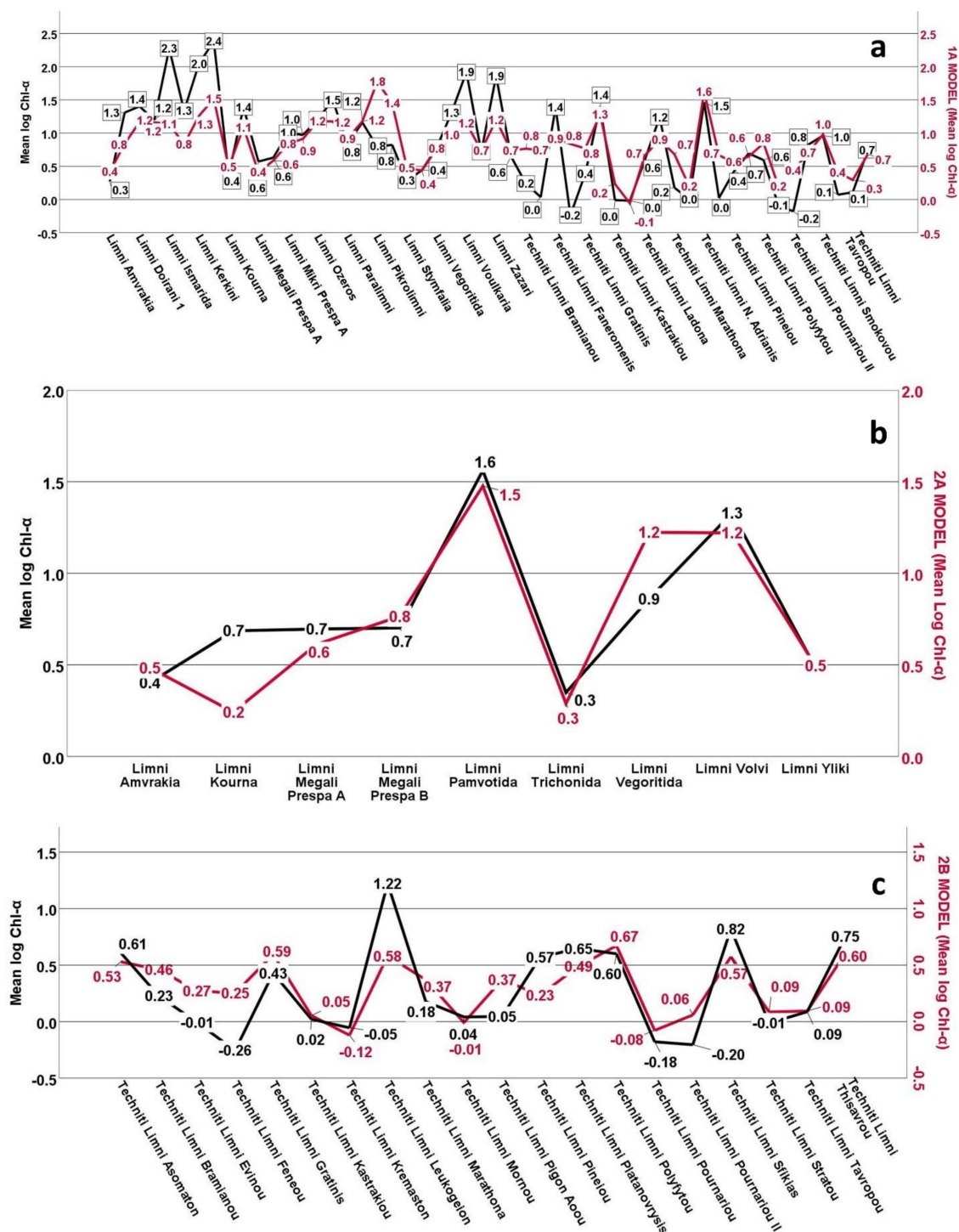


Figure 4. Spatial distribution per lake of mean in situ log Chl-*a* concentrations and mean satellite-derived log Chl-*a* concentrations based on the application of the (a) 1A model, (b) 2A model, and (c) 2B model, on both validation datasets (20% of the initial dataset, 2018 dataset).

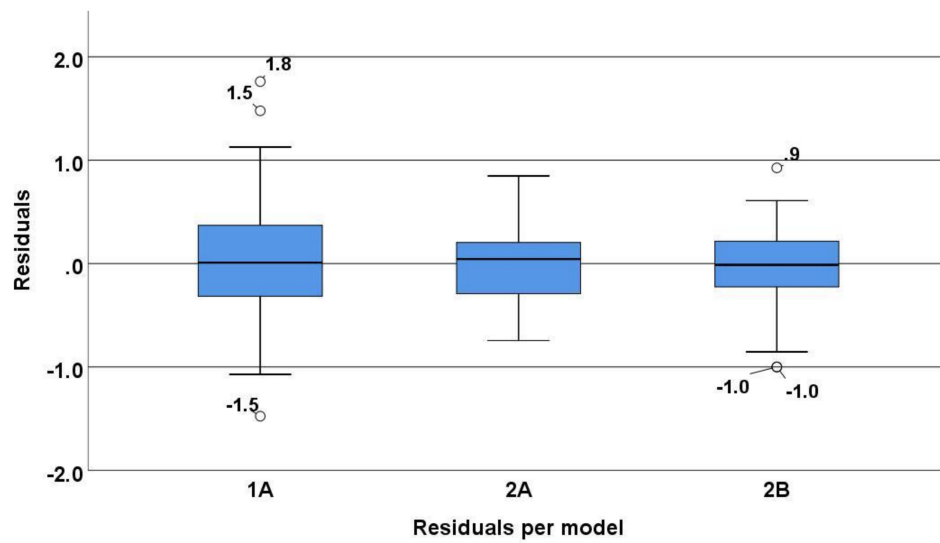


Figure 5. Boxplots presenting basic descriptive statistics of residuals per model.

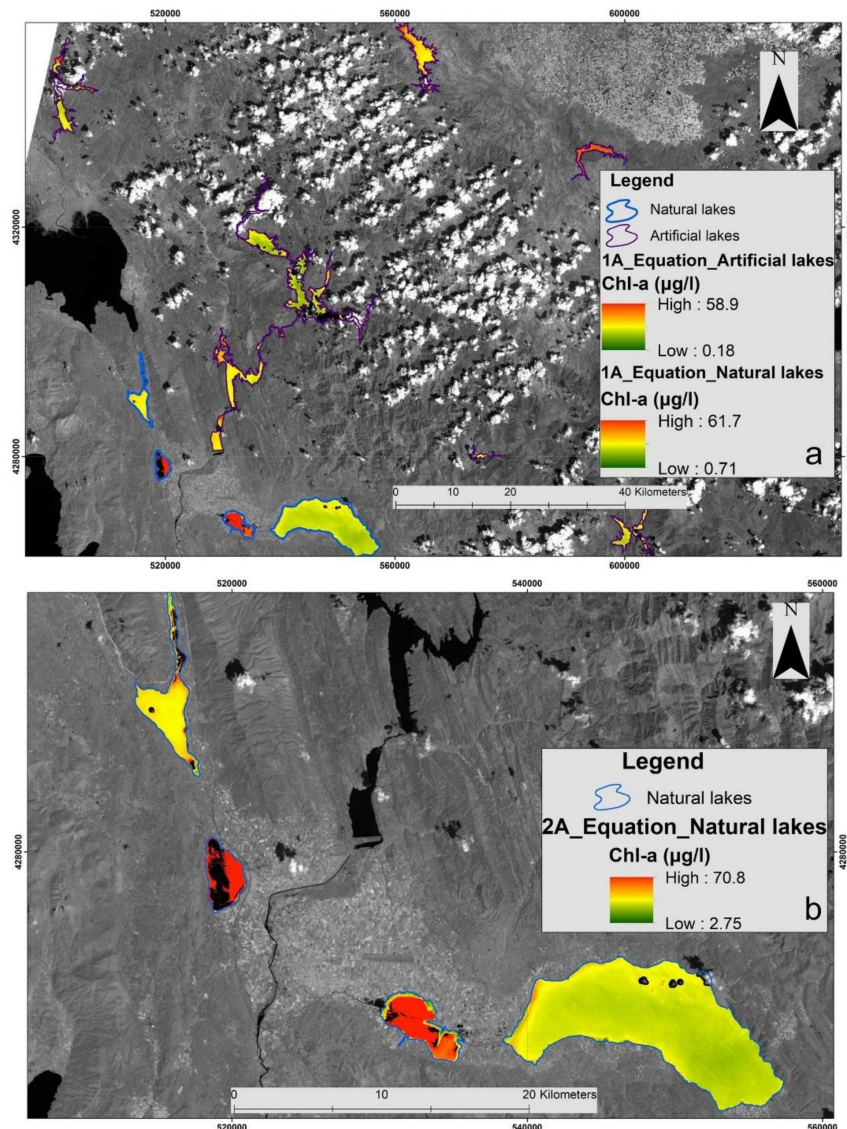


Figure 6. Cont.

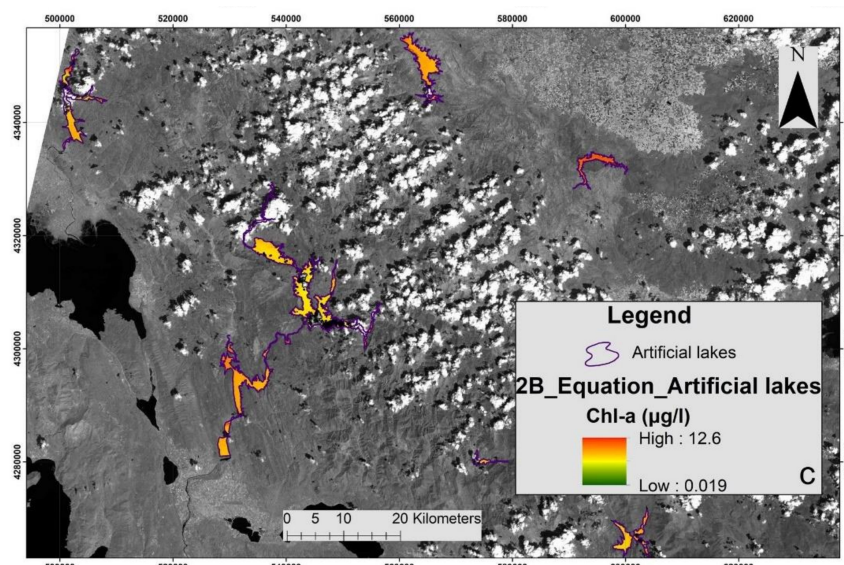


Figure 6. Satellite-derived Chl-*a* maps (on 11 August 2013) of selected lakes after the application of 1A (a), 2A (b), and 2B (c) models (WGS_1984, UTM Zone 34 N Coordinate system).

4. Discussion

WFD application in Greece concerning lake waterbodies yielded a significant Chl-*a* dataset including years 2013–2016 and 2018. Statistical analyses of Chl-*a* measurements indicated that natural lakes presented notably higher concentrations in relation to artificial. Those results indicated a degraded water quality of natural lakes compared to artificial with some lakes being characterized as vulnerable to eutrophication. Dense algal blooms, provoking high turbidity, are a frequent indicator of lake eutrophication [51]. Independent studies on natural lakes demonstrated a strong correlation between Chl-*a* and total phosphorus concentrations [52–54]. Canfield [54] predicted total phosphorus concentrations and trophic states (by measuring Chl-*a* and Secchi depths) in natural and artificial lakes through the development of empirical phosphorus models. Canfield [54] concluded that, while total phosphorus concentrations can be predicted equally well in natural and artificial lakes, prediction of lake trophic state was less reliable in artificial lakes. Relationships of Chl-*a* concentrations to total phosphorus and Secchi disc transparency were less precise in artificial lakes than natural lakes, while this difference was attributed to non-algal turbidities. Non-algal turbidities were detected in the artificial lakes and recorded as important water clarity determinants for many of this type of lake [55]. Based on this hypothesis, phosphorus concentrations in studied lakes may be examined to ascertain the drivers of decreased water clarity and levels of Chl-*a* concentrations in both types of Greek lakes.

Harmonization of Landsat 7 ETM+ and 8 OLI images yielded three Chl-*a* qualitative models including the ratios of blue to green and red, red to green and blue, and the ln-transformed bands SWIR1 and SWIR2. According to Barrett and Frazier [56], ratios between either chlorophyll absorption bands (red and blue) or chlorophyll reflectance bands (green and NIR) with either of the two SWIR bands highlight the part of the spectrum influenced by chlorophyll. Thus, modeled values are better correlated with actual in situ ones.

MLR analysis using the total amount of the available Chl-*a* dataset resulted as a quite reliable assessment model ($R = 0.65$). Next, the most significant parameters affecting the variance of Chl-*a* concentrations in studied lakes were explored, and the outcome implied the lake characteristics (natural/artificial) and typology, followed by the sampling season. The final models were separately developed for natural ($R = 0.78$) and artificial lakes ($R = 0.76$), with the latter being accompanied by a date difference between in situ and satellite data ranging ± 5 days. Those models were proven to be better—based on statistical indices concerning the validation process—in relation to the one yielded

from the total amount of the dataset. This superiority highlights the significance of the information acquisition concerning the studied areas.

Results of this study are in accordance with others similar studies published exploring properties of inland waters using either Landsat or other EO spaceborne sensors. Gholizadeh et al. [3] conducted a detailed review on water quality parameters that are widely estimated using remote sensing techniques. As the authors noted, most Chl-*a* assessment models use a wavelength near 675 nm and 700 nm. Many researchers developed empirical Chl-*a* algorithms using various but basically common image bands. Nas et al. [15] used the visible near-infrared (VNIR) and the shortwave infrared (SWIR) of Terra/ASTER for Chl-*a* mapping, presenting an R^2 value of 0.86. Zhang and Han [16] used the coastal, blue, red, and green Landsat 8 OLI bands to map Chl-*a* concentrations in Laizhou Bay, while Kim et al. [17] utilized the blue, NIR, and the ratio of blue to red Landsat 8 OLI bands to measure Chl-*a* in the Fjord of Svalbard, in the Arctic Sea, with an R^2 value of 0.6. Lim and Choi [18] also used Landsat-8/OLI in order to monitor the water quality of Nakdong River in Korea and presented high correlations among Chl-*a* and OLI bands, especially the green and NIR bands and the band ratio of NIR to green (Pearson's correlation coefficients of -0.7 , 0.71 , and -0.64 , respectively). Bonansea et al. [19] tried generating a different Chl-*a* model for different Landsat sensors (5 TM, 7 ETM+, and 8 OLI) in the largest artificial reservoir in Córdoba province (Rio Tercero, Argentina). Overall, they observed that each Landsat sensor can be used to estimate Chl-*a* in the reservoir, while the best model for the TM sensor included a combination of green, red, and NIR bands, as well as the ratio green/red ($R^2 = 0.92$), and that for the ETM+ sensor ($R^2 = 0.91$) included the green and SWIR-1 bands and the ratio red/green.

While several satellite sensors can be used for Chl-*a* determination, mapping Chl-*a* in Case 2 waters is a complicated task since the optical properties are measured based on a compound of dissolved organic matter, dead organic and inorganic particulate matter, and phytoplankton (Chl-*a*). Therefore, Chl-*a* determination is characterized by less accuracy as these constituents are not statistically correlated. Taking this shortcoming into account, we tried to use spectral band ratios which decrease irradiance and atmospheric biases in the sensor signal [20] and more than one band, since the scattering and absorption of Chl-*a* are then better studied [57]. Furthermore, according to Kloiber et al. [49], all significant band combinations for chlorophyll include at least one of the shortwave infrared bands; thus, SWIR bands were incorporated into this study's analysis in order to produce optimal assessment Chl-*a* models.

A major difference in relation to aforementioned studies is that the study area used in the present study incorporated 50 different lake systems throughout Greece covering a broad geographic area and a wide range of limnological conditions, while the majority of the respective literature focused on regional scales and discrete inland water bodies. Large spatial scales require greater computational potential; thus, the release of the Google Earth Engine platform dramatically increased the scale at which earth observation research can take place [8]. An example of this extended study area is the research by Lin et al. [58] who combined in situ Chl-*a* data from 1157 lakes (2007) with Landsat data and developed a well-validated lake national model (RMSE = $34.9 \mu\text{g/L}$), by using machine learning algorithms built into the Google Earth Engine. Another example regarding Greece is the study by Kontopoulou et al. [28] who used Landsat 8 images and WFD Chl-*a* and turbidity datasets concerning 11 lakes for the years 2013–2015. They conducted regression analysis by using Matlab scripts and also examined the effect of the time difference between satellite and field data. In that study, an R^2 of 0.78 (log Chl-*a*, $n = 168$) was reported for a time window 0–15 days, while R^2 reached 0.8 ($n = 39$) for a narrower three-day time difference.

Since Chl-*a* concentrations in lakes cannot be accurately determined due to aforementioned restrictions, we consider that the proposed empirical models are reliable and should be applicable to most natural and artificial lakes within Greece. One limitation of empirical models is their restriction to confident assessments only within the range and setting of the input data. This restriction limits their application across spatiotemporal domains [8], a risk which was restrained to a large extent since training datasets of this study included the majority of Greek lakes and three sampling seasons.

However, the general applicability and potential limitations of this approach were not thoroughly addressed; hence, further improvement will be explored as soon as the latest WFD datasets are released.

Furthermore, there are clearly some factors that should be taken into consideration in this study, affecting the accuracy of Chl-*a* quantification in Greek studied lakes.

1. An uncertainty of accuracy regarding the location of sampling points. Sampling in lakes requires special attention as winds and other external factors (e.g., season, lake depth and changes in water level, ease of proximity) contribute to potential transpositions of sampling sites.
2. The implementation of the DOS1 atmospheric correction method was not validated in order to assure that atmosphere biases were completely removed. However, this method is widely used by the EO community and proved useful when no atmospheric measurements are available.
3. Optimal models were applied to lake surfaces accrued from lake shapefiles, acquired from the Environment and Energy Ministry's website. Since no classification between land and water was conducted, there is the possibility that some pixels, covering land, hindered Chl-*a* quantification with great accuracy.

All in all, it should be noted that EO is recommended to be combined with conventional in situ water sampling in order to achieve great assessment accuracy. Such a synergistic approach, in conjunction with cooperation among government and scientists, contributes to increased data retrieval, with obtained knowledge of the lake water quality, and by extension to better protection and pollution mitigation measures.

5. Conclusions

In this study, a methodological framework was proposed for quantifying Chl-*a* pigments using Earth observation (EO) data from Landsat 7 ETM+ and 8 OLI sensors. Its practical use was evaluated in a total of 50 lake water bodies (natural and artificial) from 2013–2018, constituting the National Lake Network Monitoring of Greece in the context of the Water Framework Directive (WFD).

Use of geoinformation technologies, such as EO and GIS, in combination with conventional field surveying and spatial data analysis methods, is the most efficient way forward for monitoring water quality parameters in lakes. Application of WFD in Greece resulted in a significant dataset of various water quality parameters concerning, in this case, 50 lake water bodies with different morphological characteristics and other properties. The integration of spectral information from two Landsat sensors and statistical analyses employing principle component analysis and stepwise MLR analyses yielded statistically significant results. Optimal models were developed in this study, separately for natural and artificial lakes, and they increased the feasibility of Chl-*a* assessment with high accuracy. The majority of the respective literature focused on discrete inland water bodies, reporting the most accurate and statistically significant models. Monitoring of water quality in large spatial scales, however, as in this study, may result in sustainable water resource management even though the models may be statistically weaker.

Since Chl-*a* is the major indicator of trophic state, considered as one of the top water pollution indices related to eutrophication, this study supports WFD application concerning the perpetual water quality monitoring of Greek lakes. WFD application throughout Europe aims at the monitoring of hydromorphological, physicochemical, and biological data to assess the ecological status of surface waters. Those data include optically active constituents of water that interact with light (e.g., Chl-*a*) and can be measured using remote sensing, as well as those that lack optical properties. Some examples are pH, dissolved oxygen, and nutrient concentrations. Monitoring of those properties, characterized by a low signal-to-noise ratio, by using geoinformation technologies—in particular EO and GIS—is a challenging task, and this motivated us to pursue it in the near future, exploiting the WFD monitoring results.

Author Contributions: Conceptualization, V.M.; methodology, V.M., E.D., G.P.P., and D.K.; data curation, V.M.; writing—original draft preparation V.M.; writing—review and editing V.M., G.P.P., E.D., and D.K.; supervision, G.P., D.K., and E.D. All authors read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Acknowledgments: Authors wish to thank for the freely shared data used in this study. Those data were collected in the framework of the National Water Monitoring Network for lakes, according to the Joint Ministerial Decision 140384/2011, implemented by The Goulandris Natural History Museum, Greek Biotope/Wetland Center (EKBY). The network is supervised by the Directorate for the Protection and Management of Water Resources of the Special Secretariat for Waters of the Ministry of Environment and Energy.

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

References

1. Alymkulova, B.; Abuduwaili, J.; Issanova, G.; Nahayo, L. Consideration of Water Uses for Its Sustainable Management, the Case of Issyk-Kul Lake, Kyrgyzstan. *Water* **2016**, *8*, 298. [\[CrossRef\]](#)
2. Ramsar Convention Bureau. *Wetlands Values and Functions*; Ramsar Convention Bureau: Gland, Switzerland, 2001.
3. Gholizadeh, M.; Melesse, A.; Reddi, L. A comprehensive review on water quality parameters estimation using remote sensing techniques. *Sensors* **2016**, *16*, 1298. [\[CrossRef\]](#) [\[PubMed\]](#)
4. Duan, W.; Takara, K.; He, B.; Luo, P.; Nover, D.; Yamashiki, Y. Spatial and temporal trends in estimates of nutrient and suspended sediment loads in the Ishikari river, Japan, 1985 to 2010. *Sci. Total Environ.* **2013**, *461*, 499–508. [\[CrossRef\]](#) [\[PubMed\]](#)
5. Duan, W.; He, B.; Takara, K.; Luo, P.; Nover, D.; Sahu, N.; Yamashiki, Y. Spatiotemporal evaluation of water quality incidents in Japan between 1996 and 2007. *Chemosphere* **2013**, *93*, 946–953. [\[CrossRef\]](#) [\[PubMed\]](#)
6. Whyte, A.; Feredinos, K.; Petropoulos, G.P. A new synergistic approach for monitoring wetlands using Sentinels -1 and 2 data with object-based machine learning algorithms. *Environ. Modell. Software* **2018**, *104*, 40–57. [\[CrossRef\]](#)
7. MacKay, H.; Finlayson, C.M.; Fernández-Prieto, D.; Davidson, N.; Pritchard, D.; Rebelo, L.M. The role of earth observation (EO) technologies in supporting implementation of the Ramsar convention on wetlands. *J. Environ. Manag.* **2009**, *90*, 2234–2242. [\[CrossRef\]](#)
8. Topp, S.; Pavelsky, T.; Jensen, D.; Simard, M.; Ross, M. Research trends in the use of remote sensing for inland water quality science: Moving towards multidisciplinary applications. *Water* **2020**, *12*, 169. [\[CrossRef\]](#)
9. El-Din, M.S.; Gaber, A.; Koch, M.; Ahmed, R.S.; Bahgat, I. Remote sensing application for water quality assessment in lake Timsah, Suez canal, Egypt. *J. Remote Sens. Technol.* **2013**, *1*, 61–74. [\[CrossRef\]](#)
10. Giardino, C.; Bresciani, M.; Cazzaniga, I.; Schenk, K.; Rieger, P.; Braga, F.; Matta, E.; Brando, V.E. Evaluation of multi-resolution satellite sensors for assessing water quality and bottom depth of lake Garda. *Sensors* **2014**, *14*, 24116–24131. [\[CrossRef\]](#)
11. Ritchie, J.C.; Zimba, P.V.; Everitt, J.H. Remote sensing techniques to assess water quality. *Photogramm. Eng. Remote Sens.* **2003**, *69*, 695–704. [\[CrossRef\]](#)
12. Gitelson, A.A.; Dall’Olmo, G.; Moses, W.; Rundquist, D.C.; Barrow, T.; Fisher, T.R.; Gurlin, D.; Holz, J. A simple semi-analytical model for remote estimation of chlorophyll-a in turbid waters: Validation. *Remote Sens. Environ.* **2008**, *112*, 3582–3593. [\[CrossRef\]](#)
13. Olmanson, L.G.; Bauer, M.E.; Brezonik, P.L. A 20-year Landsat water clarity census of Minnesota’s 10,000 lakes. *Remote Sens. Environ.* **2008**, *112*, 4086–4097. [\[CrossRef\]](#)
14. Markogianni, V.; Kalivas, D.; Petropoulos, G.; Dimitriou, E. an appraisal of the potential of Landsat 8 in estimating chlorophyll-a, ammonium concentrations and other water quality indicators. *Remote Sens.* **2018**, *10*, 1018. [\[CrossRef\]](#)
15. Nas, B.; Karabork, H.; Ekercin, S.; Berkay, A. Assessing water quality in the Beysehir lake (Turkey) by the application of GIS, geostatistics and remote sensing. In Proceedings of the 12th World Lake Conference, (Taal 2007), Jaipur, India, 29 October–2 November 2007; p. 646.
16. Zhang, C.; Han, M. Mapping chlorophyll-a concentration in Laizhou bay using Landsat 8 OLI data. In Proceedings of the 36th IAHR World Congress, Hague, The Netherlands, 28 June–3 July 2015.
17. Kim, S.I.; Kim, H.C.; Hyun, C.U. High resolution ocean color products estimation in Fjord of Svalbard, Arctic sea using Landsat-8 OLI. *Korean J. Remote Sens.* **2014**, *30*, 809–816. [\[CrossRef\]](#)

18. Lim, J.; Choi, M. Assessment of water quality based on Landsat 8 operational land imager associated with human activities in Korea. *Environ. Monit. Assess.* **2015**, *187*, 1–17. [CrossRef]
19. Bonansea, M.; Rodriguez, C.; Pinotti, L. Assessing the potential of integrating Landsat sensors for estimating chlorophyll-a concentration in a reservoir. *Hydrology Res.* **2018**, *49*, 1608–1617. [CrossRef]
20. Dekker, A.; Peters, S. The use of the thematic mapper for the analysis of eutrophic lakes: A case study in the Netherlands. *Int. J. Remote Sens.* **1993**, *14*, 799–821. [CrossRef]
21. George, D. The airborne remote sensing of phytoplankton chlorophyll in the lakes and tarns of the english lake district. *Int. J. Remote Sens.* **1997**, *18*, 1961–1975. [CrossRef]
22. Ritchie, J.C.; Cooper, C.M.; Schiebe, F.R. The relationship of MSS and TM digital data with suspended sediments, chlorophyll, and temperature in Moon lake, Mississippi. *Remote Sens. Environ.* **1990**, *33*, 137–148. [CrossRef]
23. Gitelson, A.; Mayo, M.; Yacobi, Y.Z.; Parparov, A.; Berman, T. The use of high-spectral-resolution radiometer data for detection of low chlorophyll concentrations in lake Kinneret. *J. Plankton Res.* **1994**, *16*, 993–1002. [CrossRef]
24. Matthews, M.W.; Bernard, S.; Robertson, L. An algorithm for detecting trophic status (chlorophyll-a), cyanobacterial-dominance, surface scums and floating vegetation in inland and coastal waters. *Remote Sens. Environ.* **2012**, *124*, 637–652. [CrossRef]
25. Le, C.; Li, Y.; Zha, Y.; Wang, Q.; Zhang, H.; Yin, B. Remote sensing of phycocyanin pigment in highly turbid inland waters in lake Taihu, China. *Int. J. Remote Sens.* **2011**, *32*, 8253–8269. [CrossRef]
26. Peppas, M.; Vasilakos, C.; Kavroudakis, D. Eutrophication monitoring for lake Pamvotis, Greece, using sentinel-2 data. *ISPRS Int. J. Geo-Inf.* **2020**, *9*, 143. [CrossRef]
27. Markogianni, V.; Dimitriou, E.; Karaouzas, I. Water quality monitoring and assessment of an urban mediterranean lake facilitated by remote sensing applications. *Environ. Monit. Assess.* **2014**, *186*, 5009–5026. [CrossRef] [PubMed]
28. Kontopoulou, E.; Kolokoussis, P.; Karantzalos, K. Water quality estimation in greek lakes from Landsat 8 multispectral satellite data. Special Issue: 10th Word Congress on Water Resources and Environment (EWRA2017—Issue II). *Eur. Water* **2017**, *58*, 191–196.
29. Mavromati, E.; Kagalou, I.; Kemitoglou, D.; Apostolakis, A.; Tsiaoussi, D. Linkages between physicochemical status and hydromorphology in greek lakes under WFD policy. European Water Resources Association. Special Issue: 10th Word Congress on Water Resources and Environment (EWRA2017—Issue II). *Eur. Water* **2017**, *58*, 273–279.
30. Mavromati, E.; Kagalou, I.; Kemitoglou, D.; Apostolakis, A.; Seferlis, M.; Tsiaoussi, V. Relationships among land use patterns, hydromorphological features and physicochemical parameters of surface waters: WFD lake monitoring in Greece. *Environ. Process.* **2018**, *5* (Suppl. S1), 139–151. [CrossRef]
31. Zacharias, I.; Bertachas, I.; Skoulikidis, N.; Koussouris, T. Greek lakes: Limnological overview. *Lakes Reservoirs Res. Manag.* **2002**, *7*, 55–62. [CrossRef]
32. American Public Health Association (APHA). *Standard Methods for the Examination of Water and Wastewater*, 17th ed.; American Public Health Association: Washington, DC, USA, 1989.
33. Peel, M.C.; Finlayson, B.L.; McMahon, T.A. Updated world map of the Köppen-Geiger climate classification. *Hydrol. Earth Syst. Sci.* **2007**, *11*, 1633–1644. [CrossRef]
34. Olmanson, L.; Brezonik, P.; Finlay, J.; Bauer, M. Comparison of Landsat 8 and Landsat 7 for regional measurements of cdom and water clarity in lakes. *Remote Sens. Environ.* **2016**, *185*, 119–128. [CrossRef]
35. Loveland, T.; Dwyer, J. Landsat: Building a strong future. *Remote Sens. Environ.* **2012**, *122*, 22–29. [CrossRef]
36. Tebbs, E.; Remedios, J.; Harper, D. Remote sensing of chlorophyll-a as a measure of cyanobacterial biomass in lake Bogoria, a hypertrophic, saline-alkaline, flamingo lake, using Landsat ETM+. *Remote Sens. Environ.* **2013**, *135*, 92–106. [CrossRef]
37. Bonansea, M.; Bazán, R.; Ledesma, C.; Rodriguez, C.; Pinotti, L. Monitoring of regional lake water clarity using Landsat imagery. *Hydrol. Res.* **2015**, *46*, 661–670. [CrossRef]
38. QGIS Development Team. QGIS Geographic Information System. *Open Source Geospatial Foundation Project*. Available online: <http://qgis.osgeo.org/> (accessed on 1 June 2019).
39. Congedo, L.; Semi-automatic Plugin Documentation. Release 6.2.0.1. February 2019. Available online: <https://fromgistors.blogspot.com/p/user-manual.html> (accessed on 1 May 2019).

40. Matthews, M. A current review of empirical procedures of remote sensing in inland and near-coastal transitional waters. *Int. J. Remote Sens.* **2011**, *32*, 6855–6899. [[CrossRef](#)]
41. Liu, H.Q.; Huete, A.R. A feedback based modification of the NDVI to minimize canopy background and atmospheric noise. *IEEE Trans. Geosci. Remote Sens.* **1995**, *33*, 457–465. [[CrossRef](#)]
42. Baret, F.; Guyot, G. Potentials and limits of vegetation indices for LAI and PAR assessment. *Remote Sens. Environ.* **1991**, *35*, 161–173. [[CrossRef](#)]
43. 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](#)]
44. Xu, H. Modification of Normalized Difference Water Index (NDWI) to Enhance Open Water Features in Remotely Sensed Imagery. *Int. J. Remote Sens.* **2006**, *27*, 3025–3033. [[CrossRef](#)]
45. Gitelson, A.A.; Kaufman, Y.J.; Merzlyak, M.N. Use of a green channel in remote sensing of global vegetation from EOS-MODIS. *Remote Sens. Environ.* **1996**, *58*, 289–298. [[CrossRef](#)]
46. Rouse, J.W.; Haas, R.H.; Schell, J.A.; Deering, D.W. *Monitoring Vegetation Systems in the Great Plains with ERTS*; NASA SP-351; NASA: Washington, DC, USA, 1974; Volume 1, pp. 309–317.
47. Alawadi, F. Detection of surface algal blooms using the newly developed algorithm surface algal bloom index (SABI). In Proceedings of the International Society for Optics and Photonics, Toulouse, France, 12 June 2010; Volume 7825, pp. 1–14.
48. Dancey, C.P.; Reidy, J. *Statistics without Maths for Psychology*, 4th ed.; Pearson Education: Harlow, UK, 2007.
49. Kloiber, S.; Brezonik, P.; Olmanson, L.; Bauer, M. A Procedure for regional lake water clarity assessment using Landsat multispectral data. *Remote Sens. Environ.* **2002**, *82*, 38–47. [[CrossRef](#)]
50. McKinna, L.; Werdell, P. Approach for identifying optically shallow pixels when processing ocean-color imagery. *Opt. Express* **2018**, *26*, A915–A928. [[CrossRef](#)] [[PubMed](#)]
51. Schindler, W.D.; Hecky, E.R.; Findlay, L.D.; Stainton, P.M.; Parker, R.B.; Paterson, J.M.; Beaty, G.K.; Lyng, M.; Kasian, S.E.M. Eutrophication of lakes cannot be controlled by reducing nitrogen input: Results of a 37-year whole-ecosystem experiment. *Proc. Natl. Acad. Sci. USA* **2008**, *105*, 11254–11258. [[CrossRef](#)] [[PubMed](#)]
52. Sakamoto, M. Primary production by phytoplankton community in some Japanese lakes and its dependence on lake depth. *Archiv Hydrobiol.* **1966**, *62*, 1–28. [[CrossRef](#)]
53. Jones, J.R.; Bachmann, R.W. Prediction of phosphorus and chlorophyll levels in lakes. *J. Water Pollut. Control Fed.* **1976**, *48*, 2176–2182.
54. Canfield, D.E., Jr. Prediction of Total Phosphorus Concentrations and Trophic States in Natural and Artificial Lakes: The Importance of Phosphorus Sedimentation. Ph.D. Thesis, Retrospective Theses and Dissertations, Iowa State University, Ames, IA, USA, 1979.
55. Canfield, D., Jr.; Bachmann, R. Prediction of total phosphorus concentrations, chlorophyll a, and secchi depths in natural and artificial lakes. *Can. J. Fish. Aquat. Sci.* **1981**, *38*, 414–423. [[CrossRef](#)]
56. Barrett, D.; Frazier, A. Automated method for monitoring water quality using Landsat imagery. *Water* **2016**, *8*, 257. [[CrossRef](#)]
57. Dekker, A.G.; Malthus, T.J.; Seyhan, E. Quantitative modeling of inland water quality for high-resolution MSS systems. *IEEE Trans. Geosci. Remote Sens.* **1991**, *29*, 89–95. [[CrossRef](#)]
58. Lin, S.; Novitski, L.N.; Qi, J.; Stevenson, R.J. Landsat TM/ETM+ and machine-learning algorithms for limnological studies and algal bloom management of inland lakes. *J. Appl. Remote Sens.* **2018**, *12*, 1. [[CrossRef](#)]

