Article

Estimating Water pH Using Cloud-Based Landsat Images for a New Classification of the Nhecolândia Lakes (Brazilian Pantanal)

Osvaldo J. R. Pereira [1,*], **Eder R. Merino** [1] [iD], **Célia R. Montes** [2], **Laurent Barbiero** [3] [iD], **Ary T. Rezende-Filho** [4] [iD], **Yves Lucas** [5] **and Adolpho J. Melfi** [1]

[1] IEE, NUPEGEL, Universidade de São Paulo, São Paulo 05508-010, Brazil; edermerino@usp.br (E.R.M.); ajmelfi@usp.br (A.J.M.)
[2] CENA, NUPEGEL, Universidade de São Paulo, Piracicaba 13400-970, Brazil; crmlauar@usp.br
[3] GET, IRD, CNRS, UPS, OMP Toulouse 31400, France; laurent.barbiero@get.omp.eu
[4] FAENG, Universidade Federal do Mato Grosso do Sul, Campo Grande 79079-900, Brazil; ary.rezende@gmail.com
[5] Université de Toulon, Aix Marseille Université, CNRS, IM2NP, 83041 Toulon CEDEX 9, France; yves.lucas@univ-tln.fr
*   Correspondence: osvaldo.pereira@tnc.org; Tel.: +55-(61)-34219133

check for updates

Abstract: The Nhecolândia region, located in the southern portion of the Pantanal wetland area, is a unique lacustrine system where tens of thousands of saline-alkaline and freshwater lakes and ponds coexist in close proximity. These lakes are suspected to be a strong source of greenhouse gases (GHGs) to the atmosphere, the water pH being one of the key factors in controlling the biogeochemical functioning and, consequently, production and emission of GHGs in these lakes. Here, we present a new field-validated classification of the Nhecolândia lakes using water pH values estimated based on a cloud-based Landsat (5 TM, 7 ETM+, and 8 OLI) 2002–2017 time-series in the Google Earth Engine platform. Calibrated top-of-atmosphere (TOA) reflectance collections with the Fmask method were used to ensure the usage of only cloud-free pixels, resulting in a dataset of 2081 scenes. The pH values were predicted by applying linear multiple regression and symbolic regression based on genetic programming (GP). The regression model presented an R^2 value of 0.81 and pH values ranging from 4.69 to 11.64. A lake mask was used to extract the predicted pH band that was then classified into three lake classes according to their pH values: Freshwater (pH < 8), oligosaline (pH 8–8.9), and saline (≥9). Nearly 12,150 lakes were mapped with those with saline waters accounting for 7.25%. Finally, a trend surface map was created using the ALOS PRISM Digital Surface Model (DSM) to analyze the correlation between landscape features (topography, connection with the regional drainage system, size, and shape of lakes) and types of lakes. The analysis was in consonance with previous studies that pointed out that saline lakes tend to occur in lower positions compared to freshwater lakes. The results open a relevant perspective for the transfer of locally acquired experimental data to the regional balances of the Nhecolândia lakes.

Keywords: pH; Google Earth Engine; time-series; genetic programming; Landsat; lakes

1. Introduction

Lakes are transitory landscape features, sometimes created by catastrophes such as volcanic eruptions, floods, earthquakes, or human interventions, and they sometimes evolve slowly over a long period of time [1]. The morphology and distribution of lakes are related to physical, chemical, and biological events within the basins coupled to climatological constraints that play a major role

in the control of a lake's metabolism and dynamics [2]. These patterns govern the distribution of dissolved gases, nutrients, and organisms and are influenced by the geomorphology of the basin and how it has evolved [3]. Lakes are integral features of the global hydrological system interacting directly with atmospheric water, surface water, and groundwater, being greatly influenced both by their physiographic and climatic settings [4].

An outstanding fluvio-lacustrine system with ponds and lakes ranging from a few meters to dozens of kilometers in radius is seen in all regions of the Pantanal wetland [5–7]. The Pantanal is one of the world's largest freshwater wetlands and comprises an area of ~150,000 km^2 (Figure 1) [8–10]. The Nhecolândia region in the southern Pantanal is a unique lacustrine system with tens of thousands of small lakes and ponds, where fresh and saline-alkaline waters co-exist in close proximity [8]. The lakes of Nhecolândia have been classified into distinct categories, namely saline lakes, regionally referred to as "salinas," freshwater lakes known as "baías," and oligosaline or hard water lakes known as "salobras" [7,11–14]. This particular landscape has long been studied and different hypotheses have considered its origins such as the reworking of fluvial sediments by aeolian processes [15,16]; confinement of floodplain areas resulting from the cross-cutting and overlapping of marginal levees [17]; or depressions in the alluvial plains developed over karstic terrains [18].
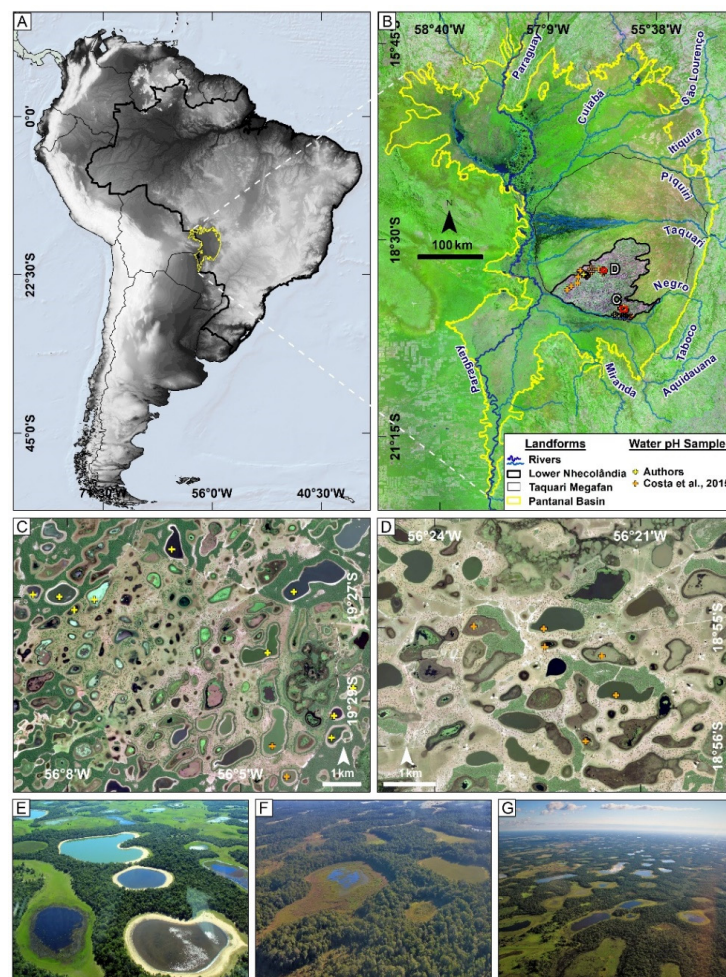

Figure 1. Pantanal wetland and Nhecolândia Region. (**A**) Location of the Pantanal wetland; (**B**) location of the Lower Nhecolândia region within the Pantanal (with water-sampled lakes indicated); (**C,D**) details of the field-surveyed areas; (**E**) picture showing the differences among salinas with their sandy beaches and surrounded by forest (cordilheiras) and baías that are normally connected to the regional drainage system (vazantes) (picture taken by Lucas Leuzinger in December 2012); (**F**) lakes being covered by aquatic macrophytes; and (**G**) general view of the Nhecolândia lakes.

Monitoring and understanding the characteristics of global inland waters are of fundamental importance to scientists and policymakers. While conventional approaches tend to be limited in terms of spatial coverage and temporal frequency, remote sensing provides invaluable sources of data from local to global scales [19]. The chemical composition of the waters that supply the Pantanal is directly influenced by the lithology and land use of the surrounding areas [20,21], but both the fresh and saline lake waters of the Nhecolândia lakes belong to the same chemical family and suggest that most of the chemical compositional changes in the system are related to chemical sedimentation mainly involving Ca, Mg, K, Si, Al, and Fe [22–25]. The freshwater lakes present sodic, calcic, carbonate, chloride, and sometimes potassic waters, whereas the saline lakes contain sodic-carbonate water [23]. These patterns result in distinct conditions among the lakes that acquire different colors, making image classification difficult. Thus, classification using pH values is a suitable alternative for the classification of the lakes.

Previous studies have performed manifold classifications of the Nhecolândia lakes using passive and active sensors at short-term periods or single images (see [11,12,14,26]). Nowadays, Google Earth Engine (GEE) offers a new suitable cloud-based platform for environmental data analysis from local to planetary scales, with rapid access and processing of multiple orbital data from different missions [27]. The data catalogue contains a variety of standard Earth Science raster datasets consisting of imagery, geophysical, climate and weather, and demographic data collections with widely used geospatial datasets, such as the entire Landsat archive. GEE has exponentially increased the feasibility and reliability of remote sensing analysis, processing large volumes of data of long-term environmental analysis. Datasets that would previously have been analyzed as single, neighboring, or temporally sequential scenes can now be processed faster and mass bulked [28].

Taking into account the capabilities of GEE, we hereby propose an estimative of water pH values for Nhecolândia lakes using long-term time-series analyses of satellite images. These lakes are suspected to be a strong source of greenhouse gases (GHGs) to the atmosphere. Recent studies have highlighted that water pH is one of the key factors in controlling the biogeochemical functioning and, consequently, the methane, carbon dioxide, and nitrous oxide production and emission of water bodies [29,30]. In this context, the pH of surface water has been shown to be one of the key tools in controlling GHG emissions. Here, we create a new pathway to analyze the pH of the Nhecolândia lakes using a cloud-based time series (2002–2017) of Landsat TM/ETM+ and OLI images, with threefold objectives: (I) To estimate pH values, (II) to analyze spectral signature variations of lakes with changes in pH, and (III) to correlate the different types of lakes with relief landforms and drainage networks.

2. Regional Settings

The Pantanal is located in the Upper Paraguay River Basin, at the center of the South-American continent (Figure 1A). The floodplain is surrounded by plateaus that consist of calcareous formations in the north (Serra das Araras) and in the south (Serra da Bodoquena), basalts of the Serra Geral Formation mainly in the south-east, sandstone formations in the east, and crystalline rocks on the wetland basement [21,31]. Despite the flat relief with a low topographic gradient of about 0.3‰ (80 to 200 m), the Pantanal consists of several sub-regions with their own specificities regarding the date and duration of flooding [8], the sedimentation rates and geologic/geomorphologic characteristics [10,32], sediment mineralogy, and water chemical composition [21], among others. The floodplain is covered by quaternary sediments and is composed of several fluvial megafans, including that of the Taquari River, considered the largest active megafan in the world [18,33,34].

The Nhecolândia Region lies over the southern half of the Taquari megafan, comprising an area of 26,900 km^2 from the Taquari River to the Negro River in the south. Two distinct morphological zones are recognized: (I) The Upper Nhecolândia that is composed of pleistocenic morphologies such as abandoned channels/meander belts and paleolakes, nowadays gradually covered by sediments deposited by the Taquari River [35], and (II) the Lower Nhecolândia, the focus of our study (Figure 1), where thousands of lakes and ponds are spread in an area of ~10,000 km^2 [8]. The area is subject to seasonal flooding, with peaks usually occurring from February to April, and a dry period from August

to October [36]. Landscape units in the Nhecolândia Region comprise: (i) Forest woodlands, (ii) open wood savanna, (iii) open grass savanna, (iv) swampy grasslands, and (v) lakes [26]. The region presents a complex hydrographic network with water usually flowing along the hundred-meter-wide shallow waterways locally known as vazantes.

The climate is Tropical-Savanna (Aw, with "A" = Tropical and "w" = dry winter, [37]) according to the Köppen classification, with wet summers and dry winters. The mean annual temperature ranges from 20 °C to 27 °C [38] and the precipitation varies from 1200 to 1350 mm in the north-northeast, and from 710 to 1200 mm in the south-southwest [39–41]. The wetland is seasonally flooded by the inundation pulse [42], associated with low-pressure zones of the South American Summer Moonson [43]. The inundation of floodplains basically occurs in three ways [44]: (I) Overflow of natural river levees and flooding of lower-lying floodplains and waterways, draining fields locally referred to as vazantes, (II) a backwater effect of the Paraguay River, and (III) pooling of local rainfall due to low relief gradient.

The salinity observed in some lakes results from an ongoing process of accumulation controlled by the occurrence of low-permeability horizons in the soil cover surrounding the lakes [7] under the current climate conditions. In addition, recent studies pointed out that the salinity is the results of regional climate changes and biogeochemical transformations during the late Holocene [45,46]. The Nhecolândia waters evolved in an alkaline pathway under the influence of evaporation. In this context, the water pH increases with concentration, reaching values close to 10 [30]. Nevertheless, these variations in water pH are still an ongoing process of accumulation and evaporation under relatively humid climatic conditions and poor drainage conditions [7,23].

Baías and salobras are water bodies with pH < 9 and electrical conductivity (EC) below 750 μS cm^{-1}, connected to shallow waterways (vazantes) during the wet season and shrinking in the dry season, leaving the floor of some lakes totally exposed and occupied by herbaceous vegetation [26,47]. The salobras are usually deeper, less connected to the major drainage systems, and have higher pH values and total dissolved solids (TDS) concentrations than most baías [47]. Salinas are normally perennial, with circular to elliptical shapes in depressions of approximately 500–1000 m in diameter, and are 0.5–3.0 m lower in elevation than neighboring freshwater lakes [48]. This type of lake always presents pH values above 9 and TDS values between 1000 and 10,000 mgL^{-1}, and is usually free from annual floods once separated from the regional drainage system by forested sandy barriers (known as cordilheiras) that are 2–3 m higher than adjacent plains [47].

3. Dataset and Methods

3.1. Landsat Surface Reflectance Dataset

We used a 2002–2017 time-series of Landsat images (5 TM, 7 ETM+, and 8 OLI) in the GEE online platform to acquire the pH of the Nhecolândia lakes (Figure 2). To ensure the use of the real pixel values, we used calibrated top-of-atmosphere (TOA) reflectance images that are included in the "USGS Landsat (5 TM, 7, and 8) Collection 1 Tier 1 TOA Reflectance" on GEE. These collections are atmospherically corrected by the Landsat Ecosystem Disturbance Adaptive Processing System (LEDAPS) [49] that includes cloud, shadow, water, and snow masks produced using per-pixel saturation analysis and a CFMASK filter [50]. The orthorectified and atmospherically corrected images have Level-1 Terrain Precision (L1TP) processed data, with well-characterized radiometry and inter-calibration across the different Landsat sensors [50]. All the spectral bands were used in the model: 4 visible and near-infrared (VNIR), and 2 short-wave infrared (SWIR) bands of 30 m spatial resolution. Thermal bands and Landsat 8 coastal/aerosol bands were excluded. Landsat 8 OLI satellite band names were renamed in GEE to match the nomenclature of Landsat 5 and 7 (Landsat 8 OLI bands 2, 3, 4, 5, 6, and 7 renamed to 1, 2, 3, 4, 5, and 7, respectively).
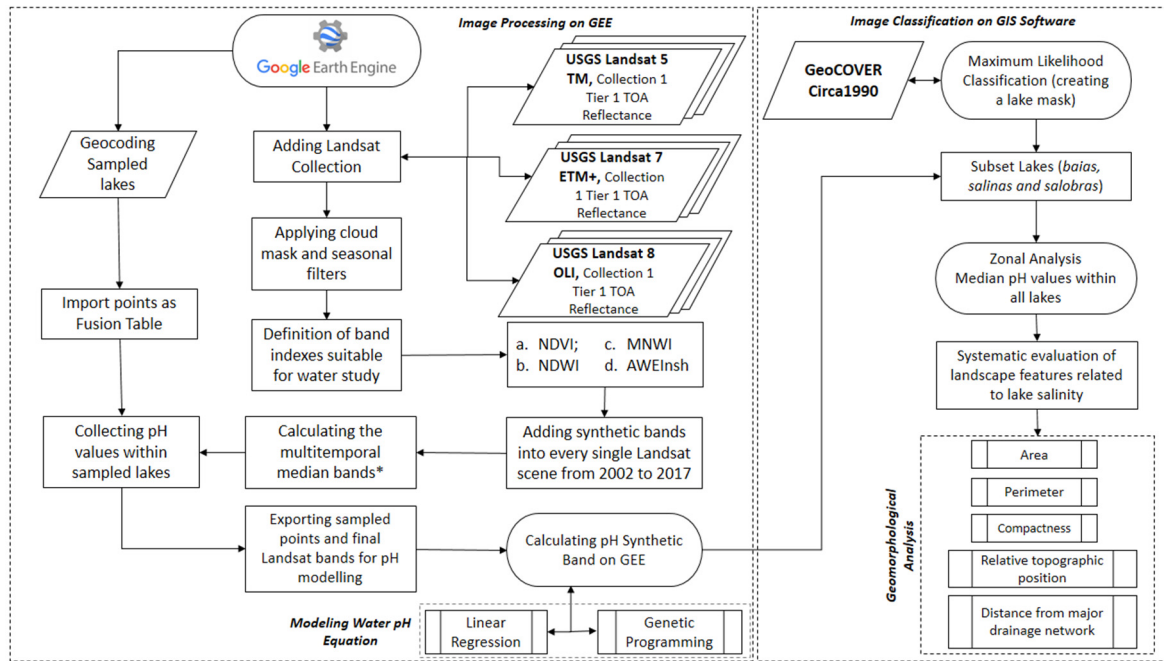
Figure 2. General workflow of the proposed method. * Every single composition represents typical median values of each band from 2002 to 2017.

Cloud coverage was masked using Quality Assessment (QA) bands contained in each Landsat sensor, using the Fmask method [51] to ensure the usage of only cloud-free pixels over the sampled lakes. Absent values of a given period were estimated through temporal interpolation of neighboring values (the mean of pixel values of scenes acquired before and after the acquisition of the scene containing the null pixel value) using the moving average smoothing method (adapted from Ivits et al. [52]) After the cloud masking, we obtained a dataset with 2081 surface reflectance scenes. The final selection of the image collections was then merged into a single collection to acquire the median values of each pixel for the 2002–2017 time-series.

3.2. Water and Vegetation Indexes

As the water pH is affected by natural variables such soil minerals, vegetation, precipitation, and temperature [53,54], we also included synthetic bands to the obtained Landsat collection (Figure 2): (i) The Normalized Difference Vegetation Index (NDVI) [55], (ii) Automated Water Extraction Index with no shadow features (AWEInsh) [56], (iii) Normalized Difference Water Index (NDWI) [57] and (iv) Modified Normalized Difference Water Index (MNDWI) [58].

NDVI is a numerical indicator that assesses the occurrence of green vegetation. It is calculated as the ratio of the measured reflectance in the red and near infrared (NIR) bands [55], computed as follows:

$$NDVI = \frac{\rho_{NIR} - \rho_{red}}{\rho_{NIR} + \rho_{red}}, \tag{1}$$

where ρ_{NIR} is the reflectance in the NIR spectral range extracted from Landsat NIR bands (~0.76 to 0.90 μm) and ρ_{red} is the reflectance in the red spectral range extracted from Landsat red bands (~0.63–0.69 μm).

The AWEI is divided into two indexes depending on the occurrence of shadow in the investigated area [56]. Here, we applied AWEInsh that is suitable for areas with no shadows, as described in Equation (2).

$$AWEInsh = 4 \times (\rho_{green} - \rho_{swir1}) - (0.25 \times (\rho_{nir} + 2.75 \times \rho_{swir2}), \tag{2}$$

where ρ_{green}, ρ_{nir}, ρ_{swir1}, and ρ_{swir2} represent reflectance in the Landsat green bands (~0.52 to 0.60 μm), NIR, SWIR 1 (~1.55 to 1.75 μm), and SWIR 2 (2.08–2.35 μm), respectively.

The NDWI is designed to maximize the reflectance of water bodies in the green band while minimizing the reflectance of those in the NIR band. It is obtained by Equation (3).

$$NDWI = \frac{\rho_{green} - \rho_{nir}}{\rho_{green} + \rho_{nir}}, \tag{3}$$

where ρ_{green} is the reflectance in the green spectral range extracted from the Landsat green band, and ρ_{nir} is the reflectance in the red spectral range. The major limitation of NDWI is that it is sensitive to signal noises coming from land cover features in built-up and disturbed areas. Xu [58] noted that water bodies have stronger absorbance in the SWIR band than that observed in the NIR band, and highlighted that the MNDWI might be more efficient in evaluating water bodies in areas with a higher influence of land cover features not related to water. The index is described according to Equation (4):

$$MNDWI = \frac{\rho_{green} - \rho_{swir}}{\rho_{green} + \rho_{swir}}, \tag{4}$$

where ρ_{green} is the reflectance in the green spectral range and ρ_{swir} is the reflectance in the red spectral range.

All the indexes were added to the merged Landsat 2002–2017 time-series images, resulting in a final raster of 10 bands (6 spectral and 4 synthetics). The new Landsat collection was then reduced to median values to avoid the influence of outliers caused by clouds or burning. The median values were calculated according to different time lags (seasonal filters), considering the following given periods: 1) Full collection from 2002 to 2017, 2) only images collected in the dry season from 2002 to 2017, 3) only images collected in the rainy season from 2002 to 2017, 4) images acquired within the period of the field campaigns in 2008 (high water season), and 5) images acquired within the period of the field campaigns from 2014 to 2017 (high water season).

3.3. Field Sample Data

Ground-truth data comprised a subset of 130 georeferenced surface water samples collected at a depth of approximately 20 to 30 cm (see location on Figure 1C) used for validation of the model. The pH values were acquired using a pH-meter (Hanna Instruments HI98140), with stabilization between 10 s and 1 min. Each sample contains pH information collected at the time of data acquisition measured directly on the surveyed lake. Some lakes were sampled more than once but in different years or seasons. The measured pH values within the 130 water samples followed a bimodal distribution (see Figure S1 in the Supplementary Materials), with values centered between 5.7 and 8.2. We also observed the occurrence of hypersaline lakes with pH values ranging from 9 to 10. The selected samples were well-distributed within the Nhecolândia region and were collected by different groups in different field campaigns. Thus, the field sample data represent the average pH condition of the Nhecolândia lakes.

All lake categories (freshwater, saline, and oligosaline) were surveyed. The final database presented samples collected by different groups, as follows: (i) 65 samples collected by the authors; 4 in 2009, 39 in 2014, 12 in 2015, and 10 in 2016; and (ii) 66 water samples measured by Costa et al. [12] in 2008. The water pH values of Costa et al. [12] were also determined in situ by using a Hydrolab Quanta G multimeter.

Samples not related to water bodies, as well as those located in lakes smaller than the minimum possible area mapped by the Landsat data (900 m^2), or in areas sampled in swampy regions adjacent to the lakes (vazantes) were excluded to avoid outliers. The resulting databases were combined and subsequently divided into independent validation and training datasets to predict the

regional occurrence of saltwater and freshwater lakes according to the methodology described in the following sections.

3.4. Prediction of pH Values

The prediction of pH values to separate saline lakes from freshwater lakes at a broad scale was achieved by considering linear and non-linear regression methods. The filtered composition with a higher correlation between the auxiliary image bands and field-measured pH values among the five seasonal filters was selected using stepwise multiple linear regression (SMLR). The spectral values of the lakes used for pH prediction were collected for the pixel (30 × 30 m) that overlaps the georeferred water pH sample collected in the field. Field-measured pH values were set as dependent variables, while median Landsat pixels from the 2002 to 2017 time-series were set as explanatory variables. The pH values were predicted by applying linear multiple regression and symbolic regression based on genetic programming (GP), a powerful machine-learning modeling technique introduced by Koza [59]. The objective was to select the best global linear model (full collection from 2002 to 2017 or collections with specific seasonal filters) to predict the dependent variables. Unlike linear methods, the symbolic regression searches both the parameters and the form of the equations, allowing the automatic generation of GP functions [60]. The resulting equations of the regression between in situ pH and spectral bands (Landsat bands 1, 2, 3, 4, 5, 7, and synthetic bands NDVI, AWEInsh, NDWI, and MNDWI), were used to simulate pH bands along the Nhecolandia lakes.

We randomly classified the 130 samples into subsets, comprising training (90 samples) and independent validation datasets (40 samples). The training dataset was used to predict values of the pH and to derive the regression models to be used for generating the pH image band according to median Landsat time-series collection. The independent validation dataset was used to evaluate the quality of the regression models. We used the coefficient of determination (R^2), root-mean-square error (RMSE), and Akaike's Information Criterion (AIC) to access the accuracy of the predicted pH values against the measured validation values. The RMSE and AIC are defined as

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(\rho_i - \hat{\rho}_i)^2}, \tag{5}$$

$$AIC = Nln\left[\frac{1}{n}\sum_{i=1}^{n}(\rho_i\hat{\rho}_i)^2\right] + 2P, \tag{6}$$

where $\hat{\rho}_i$ and ρ_i are the observed and predicted pH values, respectively, i represents the lake sample, P is the number of parameters used, n is the total number of observations, and Nln is the natural logarithm. The best model is the one with RMSE values closer to 0 and smaller AIC values.

The GP uses a set of arithmetic and complex operators to model the relationship between pH and auxiliary Landsat pixels, by testing different operators until the best prediction model is achieved. The module chosen for the genetic programming takes numerous solutions to find the best fit for the dependent variable (pH) related to the independent variables. The equation found by the GP is then performed in the GEE environment applied to the bands or indexes with higher correlation, with the field-measured pH values that will generate the predicted pH band. Finally, we conducted principal component analysis (PCA) [61] to evaluate how different band groups and measured pH values explain distinct lake characteristics.

3.5. Analyzing the Spatial Distribution of Saline and Freshwater Lakes

We performed a supervised classification using the "Maximum Likelihood Classification" method to create a lake mask. However, for the analyzed period, most lakes of the Nhecolândia southwest portion were colonized by shrub vegetation or covered by macrophytes, resulting in the misclassification of a large number of lakes that were being classified as vazantes. Conversely, in the Geocover circa 1990,

the majority of lakes were devoid of vegetation (see Supplementary Materials), fitting our purpose perfectly. This product is a global set of regional images mosaicked from the Landsat collection imagery collected from 1989 to 1993 [62]. The Geocover multispectral data, though systematically constrained to an 8-bit dynamic range, display a wide range of reflectance intensities within a single one-degree area [62], and resulted in better classification of the area. The final lake mask was used to extract the predicted pH band that was then classified into three lake classes according to their pH values: 1) Freshwater (pH < 8), 2) oligosaline (pH 8–8.9), and 3) saline (pH ≥ 9). These ranges was based on the studies of Rezende-Filho et al. (2012) and Furian et al. (2013).

The major goal of this analysis was to evaluate, at broad scale, the correlation between landscape features (topography, connection with regional drainage system, size, and shape of lakes) and variations in the lakes' pH estimated from Landsat bands, as well as to verify the relative topographic difference between saline and freshwater lakes.

The PRISM Digital Surface Model (DSM) [63], with a 30 m spatial resolution, was the altimetric data considered in the analysis. The original DSM presents values of surface altitude in meters, projected according to the vertical reference of the EGM 96 geoid. A trend surface generated from a third-order global polynomial interpolation was subtracted from the original model, creating a residual model of altitude based on the method proposed by Zani et al. [64]. The residual model, in our case, represents the local relative difference between lake levels and the surrounding environment.

In addition to estimation of the relative altitude of lakes, we also used the DSM to estimate the preferential regions of drainage flux to estimate the mean Euclidean distance between the centroid of classified lakes and the surrounding drainage network estimated from hydrological modeling of the DSM. We also correlated other metrics such as area, perimeter, and compactness of lake classes with the regional drainage system, as previously discussed by other researchers [7,26,65]. The compactness index was estimated by considering the Polsby–Popper index that varies from 0 to 1; the closer the value is to 1, the more rounded the observed feature (lake) is.

4. Results and Discussion

4.1. Spatial Modeling of Nhecolândia Lakes

By using GEE, we were able to analyze over 2081 scenes of median values of spectral signatures with cloud-free pixels for the 2002–2017 time-series, including the TM, ETM, OLI bands, and synthetic indexes. The use of five seasonal filters was necessary because the Pantanal is seasonally flooded. The correlation values between explanatory bands and dependent variables considering five different scenarios, as well as the results of SMLR analysis, are summarized in Table 1.

Table 1. Stepwise multiple linear regression (SMLR) of pH values related to spectral responses of median bands and median synthetic bands generated in the Google Earth Engine (GEE) at different periods and with different seasonal filters.

SMLR GEE Collections	Adjusted R^2	RMSE	AIC	Considered Bands
Full Collection (2002 to 2017)	0.68	0.82	−12.91	B1/B2/B3/B7/MNDWI/NDWI
High Water Season Jan to May (2002 to 2017)	0.72	0.90	−37.18	B1/B4/B5/MNDWI/NDWI
Low Water Season Jul to Out (2002 to 2017)	0.65	0.94	−8.15	B1/B2/B3/B4/MNDWI
High Water (2014 to 2017 field samples)	0.64	0.96	−3.03	B1/B3/B4/B7/NDVI/NDWI
High Water (2008 field samples)	0.67	0.92	−13.2	B1/B2/B4/B5/NDVI/NDWI

The highest correlation was observed by considering the high-water season Landsat collection from 2002 to 2017, with an R^2 of 0.72 and RMSE of 0.90. It was followed by the collection for the same period, but with no seasonal filter. We also verified the correlation of pH values with spectral data at sampled locations for the same sample data acquisition period, considering separately the period of the two field campaigns (high water season in 2008 and the period from 2014 to 2017). We observed an R^2 of 0.67 for the median values for the 2008 samples (high water season from January to May 2008), while

the 2014 to 2017 samples resulted in a correlation with an R^2 of 0.64 between the pH and spectral values. Therefore, the longer the period considered for the median Landsat data, the better the correlation that exists between the pH and spectral values (Table 1), explaining why the full collection (2002 to 2017) within the wet season in Table 1 resulted in the best SMLR model. Thus, the high-water collection from 2002 to 2017 was used as the reference collection for modeling the pH of the Nhecolândia lakes.

The PCA of the pH and median spectral values was applied to highlight the original and synthetic Landsat bands that better explain the spatial variability of the water pH within the sampled lakes. The two first factorial axes, F1 (first principal component) and F2 (second principal component) represented 79.64% and 10.77%, respectively, of the explained total variance (Figure 3a). Therefore, the first factorial plan explains up to 90% of the total variance in the sampling values (Figure 3b). It is also interesting to note that the third factorial, F3, explained less than 7% of the variance followed by F4 that explained 3% of the variance, making it difficult to interpret the real significance of these factors within variability in the spectral signatures and pH of the lakes (Figure 3b).
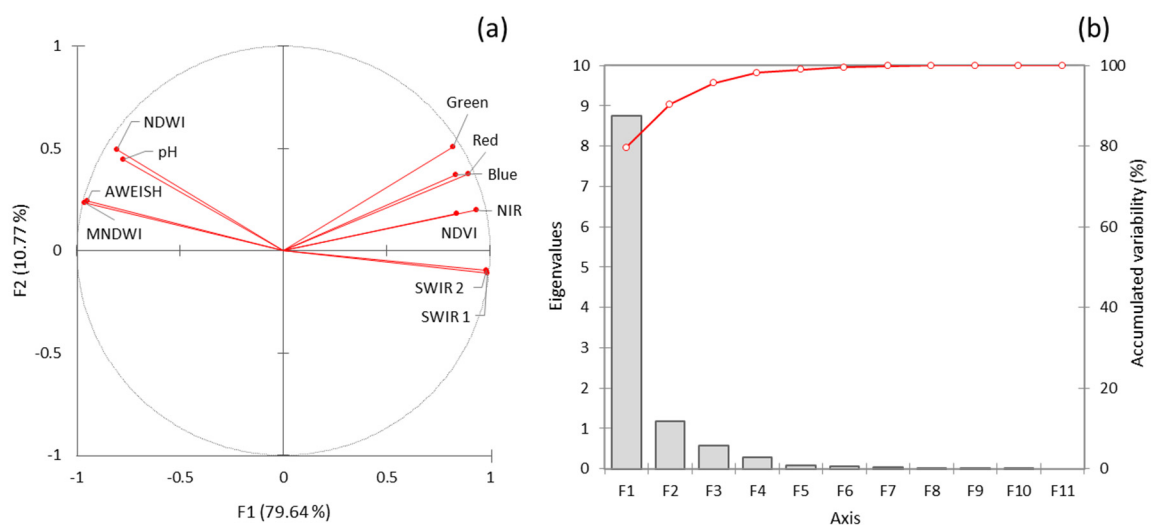

Figure 3. Principal component transformation of the independent (spectral bands) and dependent (pH) variables: (**a**) Distribution of the variables in the first axis plain considering the first two factors of the principal component transformation; (**b**) eigenvalues and cumulative values (%) for the 11 factors resulting from the principal component transformation.

We observed two major groups of variables within the analyzed samples (group 1: MNDWI, AWEInsh, NDWI, and pH; group 2: Green, red, blue, nir, swir 1, swir 2, and NDVI bands). However, most of the variability is explained by the first factor, with the variation in the AWEInsh, NDWI, and MDWI directly related to the spatial variability in the pH (Figure 3a). The first set of data, closely related to the pH, comprises the synthetic band indexes spatially suited for studies of water bodies that refer to the group of interest in this research. The second group is composed of the reflectance of the original Landsat bands (Figure 3a). The first factorial plan, with 90% of the variance, highlights the weight of the spectral difference in the Nhecolândia lakes, with a wide range of chemical compositions from freshwater to saline lakes.

Lower spectral variability was observed in the 130 sampled lakes for the visible bands (see boxplot of band variance and basic statistics: Media, variance, standard deviation, etc., respectively available in Figure S2 and Table S2 in the Supplementary Materials). The lake with a maximum reflectance value in bands 1, 2, and 3 has a very close reflectance value compared to the lake that presented the minimum value. Band 4 (NIR) presented the higher variation (minimum of 0 and maximum of 2700) among the spectral bands. A higher variability of the reflectance values was only achieved by using the synthetic bands (water and vegetation indexes). The MNDWI and NDWI presented the higher variation of reflectance values with contrasting maximum and minimum values. The MNDWI band

presented the highest variance among the sampled lakes. This variability is essential to represent the chemical and biological diversity of the lakes, which will ultimately better represent the pH variation in the sampled lakes.

The higher variance in the NDWI and MNDWI can be explained by the seasonality of the Pantanal wetland, which is directly affected by the Summer Monsoon that causes intense rainfall and large floods [43], but also experiences annual droughts after the passage of the flood-pulse, resulting in rivers and lakes decreasing in water level, with some eventually drying out [66,67].

Regression Models for Predicting pH in Lakes

The predictive models for the pH of lakes were based on the high-water season from 2002 to 2017 that presented the highest correlation between the median pixel values and field-measured pH. We considered only models capable of explaining more than 60% of the pH spatial variation within the lakes. As the pH values are not constant, the use of a long-term time series gives us the typical value of each pixel, and by using median values, we avoid outliers.

The GP self-learning model was set to use the logarithm-squared error for automatic selection of the most suitable model. The model based on symbolic regression resulted in seven equations for solving the correlation between the pH and Landsat bands, from simple to more complex solutions, and we selected the equation (Equation (7)) that offered the higher correlation between the pH and explanatory data (Figure 4). Landsat bands 2, 3, 4, and 7 were not used in any of the seven predictive models, while MNDWI, NDWI, and AWEInsh appeared in 6, 5, and 4 of the resulting GP models, respectively. Hence, two models were selected, one derived from genetic programing based on symbolic regression (Equation (7)) and the other based on SMLR (Equation (8)). Four basic arithmetic operators (+, −, *, and /) and seven more complex operators ($\sqrt{}$, x^2, power, tang, sin, cos, and log) were considered for generating the final GP model, as shown in Equation (7).

$$pH_{(GP)} = a + b * NDWI + c * MNDWI + d * MNDWI^2 + e * sin(f + a) - g * sin(h * MNDWI), \quad (7)$$

where $pH_{(GP)}$ is the pH predicted by genetic programming, NDWI and MNDWI are the resulting Landsat bands from 2002 to 2017, while the letters represent coefficients of the equation, as follows: a: 5.502868058; b: 0.000268732; c: $5.669416074e^{-8}$; d: $4.014937768e^{-6}$; e: 0.485758883; f: 5.035647409; g: 0.375986531; h: 2.502132245.

Unlike GP, the best linear multiple regression model obtained by SMLR used bands 2, 4, AWEInsh, and NDWI:

$$pH_{(SMLR)} = a - b * \text{AWEInsh} + c * band\,2 - d * band\,4 + e * NDWI, \quad (8)$$

where $pH_{(SMLR)}$ is the pH predicted by SMLR, band 2, and 4, AWEInsh and *NDWI* are the Landsat bands from 2002 to 2017, and the letters represent coefficients of the equation, as follows: a: 1.36338; b: 0.00110; c: 0.00818; d: 0.00392; e: 0.00120.

Genetic programing based on symbolic regression (Equation (7)) resulted in a more complex solution than SMLR (Equation (8)), but the correlation results were better in explaining the spatial variability of pH in the Nhecolância lakes considering both the validation and total sample dataset (Figure 4). The model obtained by GP explained more than 85% of the spatial pH variability (Figure 4c), while the model generated by SMLR explained 74% of the variability (Figure 4d), taking into account the validation dataset. Moreover, the correlation considering both validation and training datasets (Figure 4a,b) is better in the GP regression model, with an R^2 of 0.81 against 0.73 for the SMRL model.

The GP model proved the most accurate for modeling pH in the Nhecolândia using explanatory Landsat bands (Figure 4). The GP model based on the validation dataset had R^2, RMSE, and AIC values of 0.85, 0.55, and −32.8, respectively, while the values of the R^2, RMSE, and AIC for SMLR were 0.74, 0.85, and −27.06, respectively. As a result, Equation (7) was used to generate a synthetic Landsat band in the GEE cloud platform, referred to here as the pH Band. This proposed band allows for the systematic segmentation of saline and freshwater lakes.
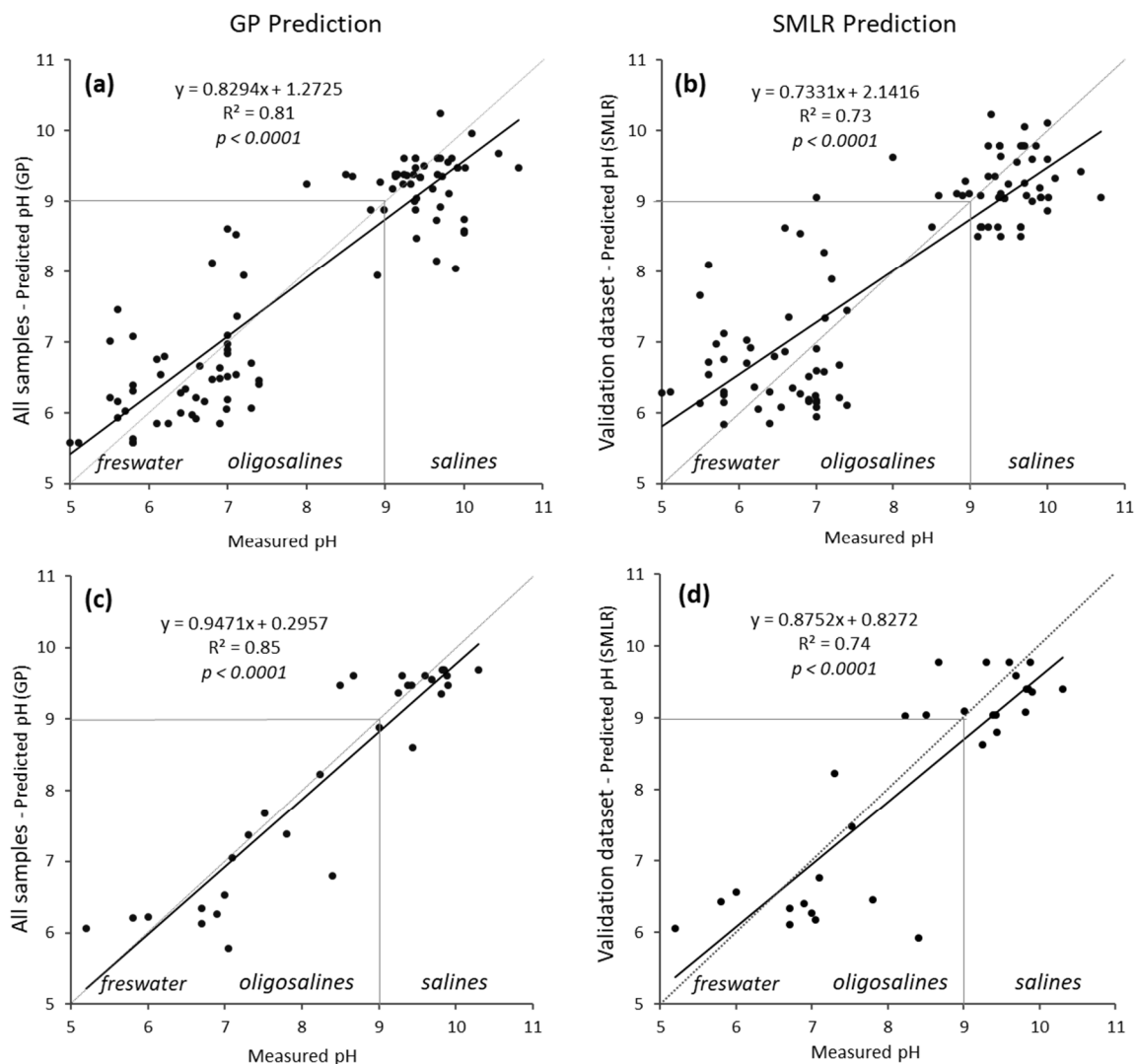
Figure 4. Correlation between the predicted and measured pH for the total samples according to (**a**) genetic programming (GP) prediction and (**b**) SMLR prediction; and for the validation dataset according to (**c**) GP prediction and (**d**) SMLR prediction.

The Nhecolândia saline lakes tend to be more perennial, normally devoid of vegetation, and with high concentrations of cyanobacteria; on the other hand, the freshwater lakes are mostly temporarily affected by the flood pulse and have large amounts of aquatic macrophytes [7,13,29,68,69]. That explains why the GP model using the NDWI and MNDWI indexes presented the best correlation with the field-measured pH values. To demonstrate that saline lakes tend to be more perennial than freshwater lakes, we performed a time-series of NDWI for all the field-surveyed lakes, showing the higher values of NDWI for saline and oligosaline lakes (Figure 5). The saline lakes may also present different colors such as green, black, or crystalline, while the freshwater lakes present crystalline water due to the bloom of cyanobacteria (see example in Figure 1E).

The pH is an "invisible" parameter for remote sensing analysis, unlike dissolved or suspended sediment concentration (i.e., Park & Latrubesse [70]); thus, indirect factors such as the seasonality and the bloom of cyanobacteria explain the high correlation of the GP model. Due to the bloom of cyanobacteria, the saline lakes reflect even in the NIR and SWIR bands (used in the water indexes), while the baías absorb most of it. The bloom of cyanobacteria also shows an increment in the levels of chlorophyll that causes increases in the pH values [71]. Saline lakes typically have higher dissolved inorganic and organic C concentrations, and the biogeochemistry of C in saline lakes can be different

from that of freshwater lakes such as carbonate precipitation/dissolution reactions and the chemical enhancement of CO_2 exchange rates at the air/water interface, being much more prevalent in saline environments [72,73].
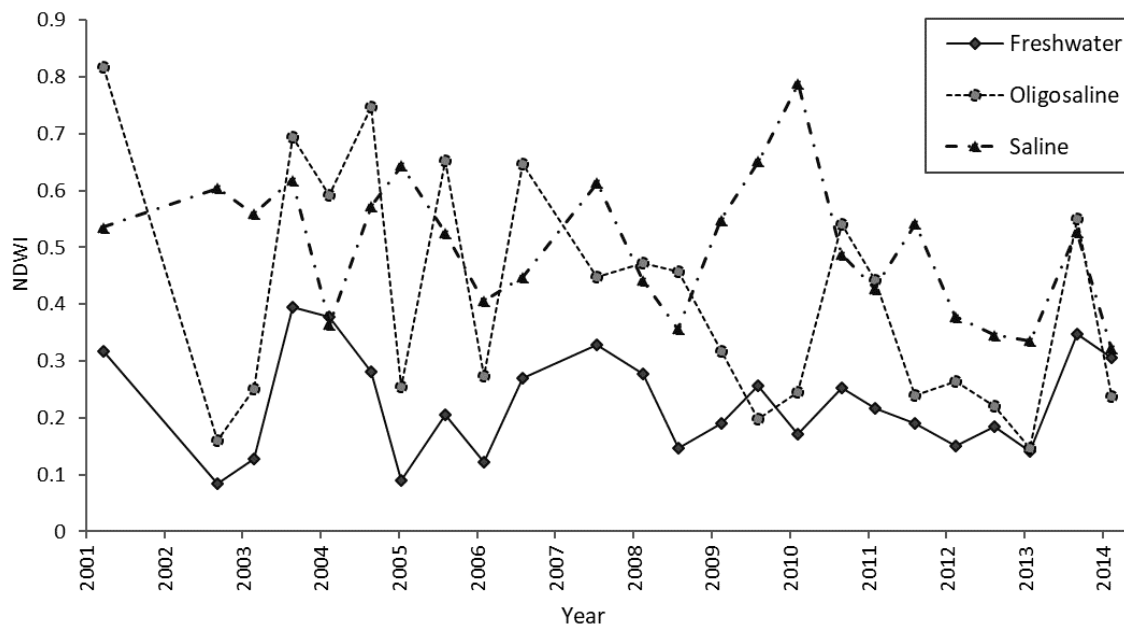

Figure 5. Normalized Difference Water Index (NDWI) of in-situ sampled lakes with a 60 m buffer. NDWI values closer to 1 represent clear water and values closer to 0 represent a greater presence of vegetation/bare soil.

In limnological studies, pH and salinity are not necessarily synonyms, and the aim of this study was not to measure the salinity of the lakes, but the pH. However, it is important to highlight that pH and electrical conductivity (EC) have a direct correlation in the Pantanal lakes. Furian et al. [7] analyzed more than 300 samples collected on surface water and found a high correlation ($R^2 = 0.87$) between pH and the logarithm of EC. Therefore, it is safe to assume that pH is a direct proxy of water salinity in the Nhecolândia lakes. In our dataset, we found an R^2 slightly smaller (0.745, see graph in the Supplementary Materials); however, one should remember that our dataset uses field samples collected in different campaigns and by distinct groups, once we also used samples from Costa et al. [12], which could explain the smaller value of the R^2. We believe the model properly represents the reality observed in the field as we could capture the variation in pH values in the Nhecolândia region with a high correlation between field samples and our pH band. However, it is important to highlight that we did not aim to propose a universal pH-predicting model for the region. Our goal was to find a way to predict pH in non-observed lakes, in order to create a systematic map of lakes' pH, which is essential to further understand the geochemistry of the region.

4.2. Landscape Features Related to the Distribution of Lakes in the Nhecolândia

Our spatial analysis mapped around 12,150 lakes for the Lower Nhecolândia region (Figure 6). Based on the predicted pH band, ~92.5% of the lakes are freshwater lakes (pH values ranging 4.69 to 7.9). Nearly 900 lakes presented pH values of oligosaline and saline waters, with values ranging from 8 to a maximum of 11.64. It is important to note that the halo of freshwater pixels (Figure 6B,C) around the saline lakes is due to the influence of seasonality.
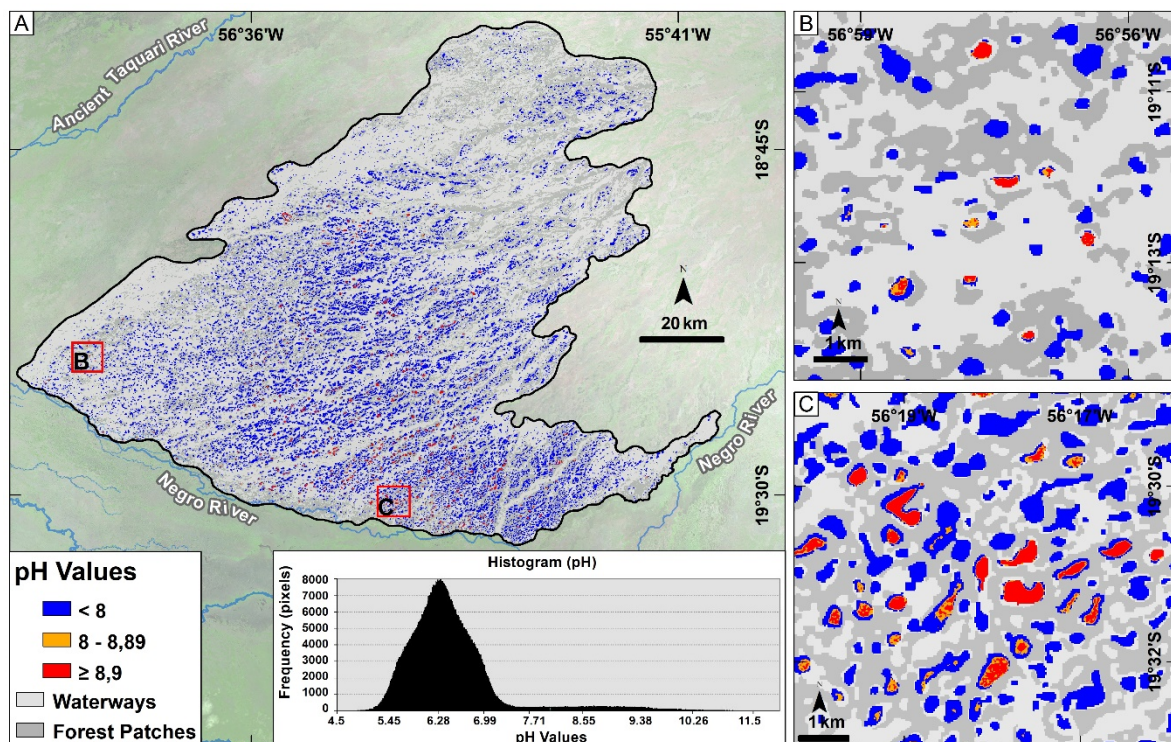
Figure 6. Spatial distribution of lakes' pH according to the pH predicted by the GP method (Equation (7)) for the entire study area. Areas B and C highlight two different regions with high and low frequencies of salt-water lakes, respectively (**B,C**). The histogram highlights the absolute frequency distribution of lakes according to the range of pH values observed in the region (**A**).

The global surface (second-order polynomial interpolator) generated from ALOS PRISM DSM had an R^2 of 0.93 compared to the original DSM PRISM that shows a high global topography component within the studied area. In our study, we found that a global third-order polynomial interpolator was useful for evaluating the relative position of the lakes (Figure 7a). The saline lakes tend to occur in lower positions compared to freshwater lakes, and there is continuous deepening from freshwater to saltwater lakes, where areas with a pH above 9 tend to occur on topographic positions lower than 4 m compared to the global DSM (Figure 7a).

We also verified two other assumptions about the Nhecolândia lakes, evolving size and typical shape. Saline lakes (Figure 7) tend to be rounded depressions ~500–1000 m, with diameters slightly larger than those of freshwater lakes, being 0.5 to 4 m lower in elevation than the freshwater areas (Figure 7a), and are cut off from floods by sandy barriers. They also tend to be larger with greater areas (Figure 7c) and perimeters (Figure 7d) and are usually deeper than freshwater lakes (Figure 7a) by an average of 0.8 to 1.5 m. The Polsby–Popper test also showed that there is a difference in shape between freshwater and saltwater lakes, where areas of saline lakes tend to be rounder (Figure 7b).

The estimation of relative distance from major waterways (Figure 8) showed no clear difference in distance from drainage networks between freshwater and oligosaline lakes (Figure 8a,b). However, the areas of saline lakes tend to be less connected than other lakes, with a higher frequency of lakes occurring at distances of 500 to 800 m from the major drainage system (Figure 8b). In addition, more than 44% of the saline lakes occurred at distances greater than 900 m from the drainage network, while 72% of freshwater lakes and 71% of oligosaline lakes occurred within less than 500 m from major drainage networks (Figure 8b). Thus, the wooded higher grounds (locally known as cordilheiras) surround these saline lakes, preserving them from the massive freshwater supply by the surface during the floods in the Pantanal [7,48]. They are exclusively supplied with subsurface flows when the freshwater level exceeds a natural soil threshold not visible within wooded areas. The dimensional

characteristics of the lakes, and notably the relationship between the perimeter (water supply zone) and the volume (surface-depth) conditions the concentration rate of the lakes, and, therefore, their water pH.
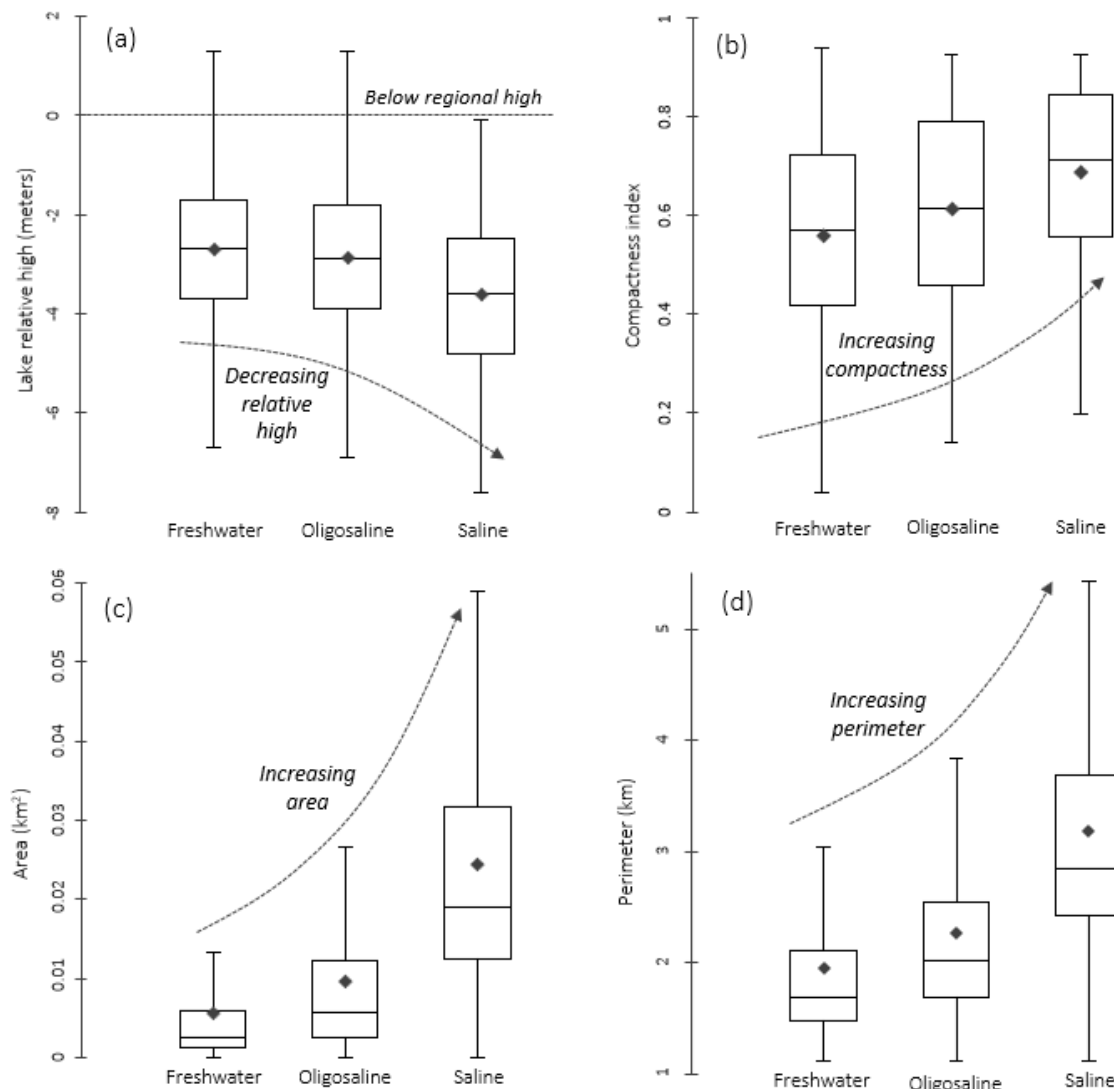

Figure 7. Major spatial features related to the Nhecolândia lakes, showing groups of freshwater (pH < 8), oligosaline (pH between 8 and 8.9), and saline lakes (pH ≥ 9), based on predicted pH within all lakes in the studied area. The charts represent values of the (**a**) relative altitude of lakes with respect to the global polynomial surface, (**b**) compactness index (Comp methods), (**c**) perimeter of lakes, and (**d**) area of lakes.

The shape of the most alkaline lakes evolves, in part, by the chemical withdrawal of elements (mainly Si, Al, Mg, and K) from the sediments and clay neoformations in the surrounding beach [22]. This evolution conditions the rounded shape of the saline lake contours and can lead to coalescence between several lakes [74].

The presence of hundreds of saline lakes among thousands of freshwater lakes is one of the unique aspects of the lower Nhecolândia region. Knowing the exact number of existing lakes in the Nhecolândia is a hard task to accomplish and different amounts have been reported. A pioneering survey [8] (Por, 1995) speculated ~10,000 lakes, with about 10% of the lakes being saline. A recent lake inventory based on field observations and orbital data [12] reported for about 8851 lakes, with 7.19% of them being salinas. The number proposed by Por [8] was an estimate, and differences between our

predictions and those by Costa et al. [12] can be attributed to two main causes: 1) Usage of different methods and instruments for lake classification; for example, they used radar while we used optical sensors; 2) toward the south-westward section of the Lower Nhecolândia, many of the lakes have been colonized by shrubs and macrophyte vegetation, which may lead to the misclassification of some areas. Morphologically, the feature is still a lake, but the vegetation causes changes in the spectral response.
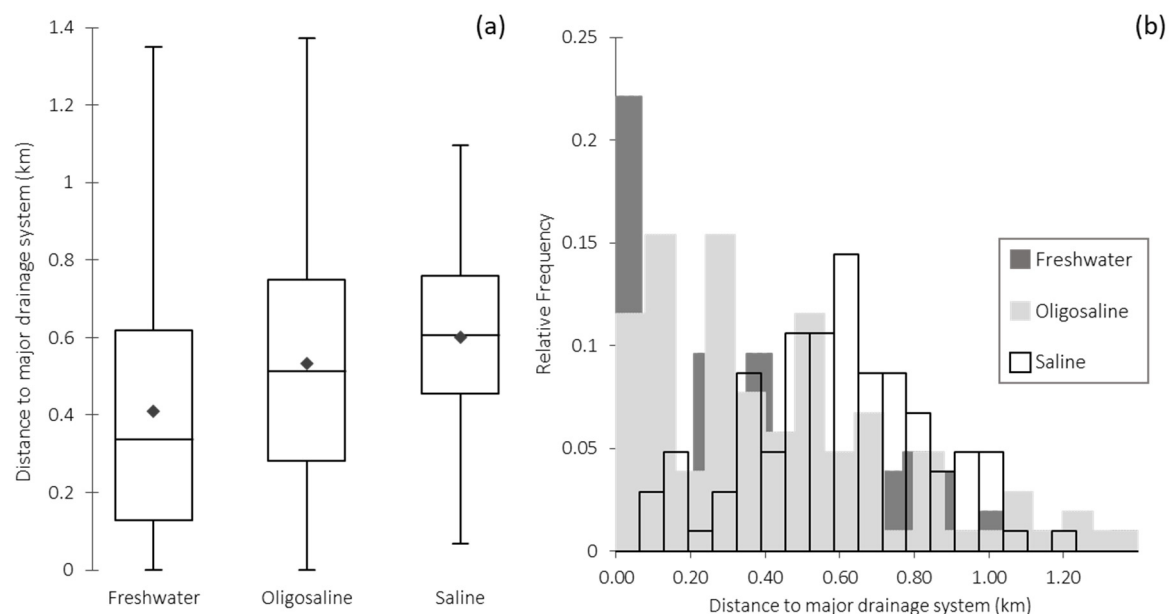

Figure 8. Distance from water flows (greater than 2nd order according to Strahler classification), considering groups of lakes organized by the estimated water pH: (**a**) Distance of the centroid of the lakes from major drainage systems, (**b**) histogram of the relative frequency distribution of the lakes, based on groups of pH values.

The co-existence of saline and freshwater lakes is due to a differential hydrological regime controlled by the morphology of the soil cover, particularly by the presence of a morphological threshold consisting of green and grey sandy loam horizons, with high sodium fractions partially cemented by silica [25]. Most of the saline lakes are concentrated on the central and southeast portion, and some on the northwest portion. Toward the southwest portion, most of the lakes are freshwater and we observed practically no salinas (Figure 6B). One of the explanations for the lack of salinas in this region may be due to the avulsion that occurred in the Taquari River. The process began crevassing in 1988 and ended with the abandonment of the former channel in 1998 [33,34], changing its mouth from the southwest to ~100 km westward (see a picture of the Taquari avulsion in the Supplementary Materials). This gradually triggered hydrological and phytophysiological changes in the pattern, intensity, and duration of the annual floods and also affected the water table dynamic, one of the factors responsible for recharging the lakes.

4.3. Remote Sensing Big Data for Inland Environments

The approach adopted in GEE for the spatial modeling and estimation of water pH was subdivided into different steps. The first and most important one, however, comprised the image tiling method implemented in GEE and was based on pixel band algebra [27]. Accordingly, processing each output tile requires retrieving only a small number of tiles for each input band that involves the concurrent processing of a limited volume of image pixels for collections with hundreds of images (Figure 9). Thus, GEE allows for the fast computation of results at any requested scale or projection that would be impractical if working offline by downloading all Landsat scenes acquired from 2002 to 2017 (Figure 9). This aspect was especially important considering the studied area as we observed an increasing

correlation between the pH values and the explanatory variables for longer periods (Figure 9). This assumption was only confirmed after considering a large number of observations, or in the case of the GEE, pixels collected at different time-periods over the same lakes. This was only possible by employing remote sensing Big-Data in GEE. The correlation (R^2) between the pH and the most correlated band (NDWI) continually increased according to the number of scenes considered in the analysis (Figure 9). This aspect is explained by the fact that we calculated the median pixel values of the lakes and a greater number of observations, probably better expressing the typical signatures of the lakes in a given location.

Figure 9. Coefficient of determination (R^2) of the pH band × different time intervals. The graph shows the R^2 of the predicted pH (using the NDWI band) versus the number of Landsat scenes within different time intervals used for acquiring the NDWI bands.

Studies based on remote sensing data have long investigated quality, characteristics, and changes in continental waters [75–77]. The measurement of water salinity through remote methods is also not novel. Thomann [78] remotely measured the coastal surface water salinity of the Mississippi and Louisiana rivers with radiometers at a wavelength of 21 cm. A pioneer in remote sensing imagery developed a regression model using Landsat multispectral scanner (MSS) images and color infrared photographs with surface measurements for salinity mapping of the San Francisco Bay Delta [79]. However, studies measuring salinity using remote sensing instruments on inland environments are, normally, created to analyze soil characteristics [80–82], with few focusing on spatio-temporal variations [83,84], and none using Big Data environments.

5. Conclusions

Here, we presented a new window into the complex classification of the Nhecolândia lakes by using orbital cloud-based imagery. We investigated the correlation between the water pH of Nhecolândia lakes and their typical spectral signatures in different multispectral satellite bands, using hundreds of Landsat observations over the same area. The use of long-term periods was more efficient in predicting pH values than using short specific time periods. The model successfully predicted pH values using Landsat and synthetic bands for 2002–2017 time-series data, achieving an R^2 correlation of over 85% for pH prediction. The use of the GEE platform and top-of-atmosphere (TOA) reflectance

images (Collection 1 Tier 1 TOA Reflectance) for Landsat 5 TM, 7, and 8 was the key tool used to achieve our major goal as we used more than 2081 Landsat scenes collected over the Nhecolândia region, which would be impractical if the images were manually downloaded. Based on our findings, we verified that ~92.5% of the lakes of the Nhecolândia region are freshwater. From these findings, we also corroborated and statistically verified previous assumptions regarding the studies that pointed out that saline water lakes tend to be larger in area, have a greater perimeter, are rounder, and are topographically lower than freshwater lakes, in addition to verifying the higher isolation of salinas. These results open a relevant perspective for the transfer of locally acquired experimental data to the regional balance of pH values and distribution of the Nhecolândia lakes.

References

1. Löffler, H. The origin of lake basins. In *The Lakes Handbook*; O'Sullivan, P.E., Reynolds, C.S., Eds.; Blackwell Science: Oxford, UK, 2004; Volume 1, pp. 8–60.

2. Hutchinson, G.E. *A Treatise on Limnology. I. Geography, Physics, and Chemistry*; John Wiley & Sons: New York, New York, USA, 1957; 1015p.

3. Wetzel, R.G. *Limnology: Lake and River Ecosystems*; Gulf Professional Publishing: San Diego, CA, USA, 2001; 1014p.

4. Winter, T.C. The Hydrology of Lakes. In *The Lakes Handbook*; O'Sullivan, P.E., Reynolds, C.S., Eds.; Blackwell Science: Oxford, UK, 2004; Volume 1, pp. 61–78.

5. Lo, E.L.; McGlue, M.M.; Silva, A.; Bergier, I.; Yeager, K.M.; de Azevedo Macedo, H.; Swallom, M.; Assine, M.L. Fluvio-lacustrine sedimentary processes and landforms on the distal Paraguay fluvial megafan (Brazil). *Geomorphology* **2019**, *342*, 163–175. [CrossRef]

6. Stevaux, J.C.; Macedo, H.d.A.; Assine, M.L.; Silva, A. Changing fluvial styles and backwater flooding along the Upper Paraguay River plains in the Brazilian Pantanal wetland. *Geomorphology* **2020**, *350*, 106906. [CrossRef]

7. Furian, S.; Martins, E.R.C.; Parizotto, T.M.; Rezende-Filho, A.T.; Victoria, R.L.; Barbiero, L. Chemical diversity and spatial variability in myriad lakes in Nhecolândia in the Pantanal wetlands of Brazil. *Limnol. Oceanogr.* **2013**, *58*, 2249–2261. [CrossRef]

8. Por, F.D. *The Pantanal of Mato Grosso (Brazil): World's Largest Wetlands*; Springer Science & Business Media: Dordrecht, Netherlands, 1995; Volume 73.

9. Keddy, P.A.; Fraser, L.H.; Solomeshch, A.I.; Junk, W.J.; Campbell, D.R.; Arroyo, M.T.; Alho, C.J. Wet and wonderful: The world's largest wetlands are conservation priorities. *BioScience* **2009**, *59*, 39–51. [CrossRef]

10. Assine, M.L.; Merino, E.R.; Pupim, F.N.; Macedo, H.A.; Santos, M.G.M. The Quaternary alluvial systems tract of the Pantanal Basin, Brazil. *Braz. J. Geol.* **2015**, *45*, 475–489. [CrossRef]

11. Galvão, L.S.; Pereira Filho, W.; Abdon, M.M.; Novo, E.M.M.L.; Silva, J.S.V.; Ponzoni, F.J. Spectral reflectance characterization of shallow lakes from the Brazilian Pantanal wetlands with field and airborne hyperspectral data. *Int. J. Remote Sens.* **2003**, *24*, 4093–4112. [CrossRef]

12. Costa, M.; Telmer, K.H.; Evans, T.L.; Almeida, T.I.; Diakun, M.T. The lakes of the Pantanal: Inventory, distribution, geochemistry, and surrounding landscape. *Wetl. Ecol. Manag.* **2015**, *23*, 19–39. [CrossRef]

13. Almeida, T.I.R.; Calijuri, M.d.C.; Falco, P.B.; Casali, S.P.; Kupriyanova, E.; Paranhos Filho, A.C.; Sigolo, J.B.; Bertolo, R.A. Biogeochemical processes and the diversity of Nhecolândia lakes, Brazil. *Ann. Braz. Acad. Sci.* **2011**, *83*, 391–407. [CrossRef] [PubMed]

14. Almeida, T.I.R.; Sígolo, J.B.; Fernandes, E.; Queiroz-Neto, J.P.; Barbiero, L.; Sakamoto, A.Y. Proposta de classificação e gênese das lagoas da baixa Nhecolândia-MS com base em sensoriamento remoto e dados de campo. *Braz. J. Geol.* **2003**, *33* (Suppl. 2), 83–90. [CrossRef]

15. Almeida, F.F.M. *Geology of the Midwest Matogrossense*; Bulletin of the Division of Geology and Mineralogy; SERGRAF: Rio de Janeiro, Brazil, 1964; 123p.

16. Klammer, C. Die Paläovüste des Pantanal von Mato Grosso und Die Pleistozäne Klimageschichte der Brasilianischen Randtropen. *Ann. Geomorphol.* **1982**, *26*, 393–416.

17. Wilhelmy, H. Meander and embankment lakes of tropical lowland rivers. *Ann. Geomorphol.* **1958**, *2*, 27–54. [CrossRef]

18. Braun, E.W. Cone aluvial do Taquari, unidade geomórfica marcante da planície quaternária do Pantanal. *Braz. J. Geogr.* **1977**, *39*, 164–180.

19. Palmer, S.C.J.; Kutser, T.; Hunter, P.D. Remote sensing of inland waters: Challenges, progress and future directions. *Remote Sens. Environ* **2015**, *157*, 1–8. [CrossRef]

20. Rezende-Filho, A.T.; Valles, V.; Furian, S.; Oliveira, C.M.S.C.; Ouardi, J.; Barbiero, L. Impacts of Lithological and Anthropogenic Factors Affecting Water Chemistry in the Upper Paraguay River Basin. *J. Environ. Qual.* **2015**, *44*, 1832–1842. [CrossRef]

21. Rezende Filho, A.T.; Furian, S.; Victoria, R.L.; Mascré, C.; Valles, V.; Barbiero, L. Hydrochemical variability at the Upper Paraguay Basin and Pantanal wetland. *Hydrol. Earth Syst. Sci.* **2012**, *16*, 2723–2737. [CrossRef]

22. Barbiero, L.; Berger, G.; Rezende Filho, A.T.; Meunier, J.-F.; Martins-Silva, E.R.; Furian, S. Organic Control of Dioctahedral and Trioctahedral Clay Formation in an Alkaline Soil System in the Pantanal Wetland of Nhecolândia, Brazil. *PLoS ONE* **2016**, *11*, e0159972. [CrossRef]

23. Barbiero, L.; Queiroz Neto, J.P.; Ciornei, G.; Sakamoto, A.Y.; Capellari, B.; Fernandes, E.; Valles, V. Geochemistry of water and ground water in the Nhecolândia, Pantanal of Mato Grosso, Brazil: Variability and associated processes. *Wetlands* **2002**, *22*, 528–540. [CrossRef]

24. Furquim, S.A.C.; Graham, R.C.; Barbiero, L.; de Queiroz Neto, J.P.; Vallès, V. Mineralogy and genesis of smectites in an alkaline-saline environment of pantanal wetland, Brazil. *Clays Clay. Miner.* **2008**, *56*, 579–595. [CrossRef]

25. Furquim, S.A.C.; Graham, R.C.; Barbiero, L.; Queiroz Neto, J.P.; Vidal-Torrado, P. Soil mineral genesis and distribution in a saline lake landscape of the Pantanal Wetland, Brazil. *Geoderma* **2010**, *154*, 518–528. [CrossRef]

26. Evans, T.L.; Costa, M. Landcover classification of the Lower Nhecolândia subregion of the Brazilian Pantanal Wetlands using ALOS/PALSAR, RADARSAT-2 and ENVISAT/ASAR imagery. *Remote Sens. Environ.* **2014**, *128*, 118–137. [CrossRef]

27. Gorelick, N.; Hancher, M.; Dixon, M.; Ilyushchenko, S.; Thau, D.; Moore, R. Google Earth Engine: Planetary-scale geospatial analysis for everyone. *Remote Sens. Environ.* **2017**, *202*, 18–27. [CrossRef]

28. Casu, F.; Manunta, M.; Agram, P.S.; Crippen, R.E. Big Remotely Sensed Data: Tools, applications and experiences. *Remote Sens. Environ.* **2017**, *202*, 1–2. [CrossRef]

29. Andreote, A.P.D.; Dini-Andreote, F.; Rigonato, J.; Machineski, G.S.; Souza, B.C.E.; Barbiero, L.; Rezende-Filho, A.T.; Fiore, M.F. Contrasting the Genetic Patterns of Microbial Communities in Soda Lakes with and without Cyanobacterial Bloom. *Front. Microbiol.* **2018**, *9*. [CrossRef]

30. Barbiero, L.; Siqueira Neto, M.; Braz, R.R.; Carmo, J.B.; Rezende-Filho, A.T.; Mazzi, E.; Fernandes, F.A.; Damatto, S.R.; Camargo, P.B. Biogeochemical diversity, O2-supersaturation and hot moments of GHG emissions from shallow alkaline lakes in the Pantanal of Nhecolândia, Brazil. *Sci. Total Environ.* **2018**, *619–620*, 1420–1430. [CrossRef] [PubMed]

31. Assine, M.; Merino, E.; Pupim, F.; Warren, L.; Guerreiro, R.; McGlue, M. Geology and Geomorphology of the Pantanal Basin. In *The Handbook of Environmental Chemistry*; Springer: Berlin/Heidelberg, Germany, 2015; pp. 1–28. [CrossRef]

32. Bergier, I. Effects of highland land-use over lowlands of the Brazilian Pantanal. *Sci. Total Environ.* **2013**, *463–464*, 1060–1066. [CrossRef]

33. Assine, M.L. River avulsions on the Taquari megafan, Pantanal wetland, Brazil. *Geomorphology* **2005**, *70*, 357–371. [CrossRef]

34. Buehler, H.A.; Weissmann, G.S.; Scuderi, L.A.; Hartley, A.J. Spatial and temporal evolution of an avulsion on the Taquari River distributive fluvial system from satellite image analysis. *J. Sediment Res.* **2011**, *81*, 630–640. [CrossRef]

35. Assine, M.L. Sedimentation in the Pantanal Sedimentar Basin, West-Central Brazil. Ph.D. Thesis, São Paulo State University, Rio Claro, Brazil, 2003.

36. Hamilton, S.K.; de Souza, O.C.; Coutinho, M.E. Dynamics of floodplain inundation in the alluvial fan of the Taquari River (Pantanal, Brazil). *SIL Proc. 1922–2010* **1998**, *26*, 916–922. [CrossRef]

37. Alvares, C.A.; Stape, J.L.; Sentelhas, P.C.; de Moraes, G.; Leonardo, J.; Sparovek, G. Köppen's climate classification map for Brazil. *Meteorol. J.* **2013**, *22*, 711–728. [CrossRef]

38. Alho, C.J.R. The Pantanal. In *The World's Largest Wetlands—Ecology and Conservation*; Fraser, L.H., Keddy, P.A., Eds.; Cambridge University Press: New York, NY, USA, 2005; pp. 203–271.

39. Zhou, L.; Lau, K.M. Does a Monsoon Climate Exist over South America? *J. Clim.* **1998**, *11*, 1020–1040. [CrossRef]

40. Garreaud, R.D.; Vuille, M.; Compagnucci, R.; Marengo, J. Present-day South American climate. *Palaeogeogr. Palaeoclimatol. Palaeoecol.* **2009**, *281*, 180–195. [CrossRef]

41. Assine, M.; Macedo, H.; Stevaux, J.; Bergier, I.; Padovani, C.; Silva, A. Avulsive Rivers in the Hydrology of the Pantanal Wetland. In *Dynamics of the Pantanal Wetland in South America. The Handbook of Environmental Chemistry*; Bergier, I., Assine, M.L., Eds.; Springer: Berlin/Heidelberg, Germany, 2015; pp. 83–110. [CrossRef]

42. Junk, J.W.; Bayley, P.B.; Sparks, R.E. The flood pulse concept in river floodplain systems. *Can. Spec. Publ. Fish. Aquat. Sci.* **1989**, *106*, 110–127.

43. Plink-Björklund, P. Morphodynamics of rivers strongly affected by monsoon precipitation: Review of depositional style and forcing factors. *Sediment. Geol.* **2015**, *323*, 110–147. [CrossRef]

44. Hamilton, S.; Sippel, S.; Melack, J. Inundation patterns in the Pantanal wetland of South America determined from passive microwave remote sensing. *Fundam. Appl. Limnol.* **1996**, *137*, 1–23.

45. Guerreiro, R.L.; McGlue, M.M.; Stone, J.R.; Bergier, I.; Parolin, M.; da Silva Caminha, S.A.F.; Warren, L.V.; Assine, M.L. Paleoecology explains Holocene chemical changes in lakes of the Nhecolândia (Pantanal-Brazil). *Hydrobiologia* **2017**, *815*, 1–29. [CrossRef]

46. McGlue, M.M.; Guerreiro, R.L.; Bergier, I.; Silva, A.; Pupim, F.N.; Oberc, V.; Assine, M.L. Holocene stratigraphic evolution of saline lakes in Nhecolândia, southern Pantanal wetlands (Brazil). *Quat. Res.* **2017**, *88*, 472–490. [CrossRef]

47. Costa, M.P.F.; Telmer, K.H. Utilizing SAR imagery and aquatic vegetation to map fresh and brackish lakes in the Brazilian Pantanal wetland. *Remote Sens. Environ.* **2006**, *105*, 204–213. [CrossRef]

48. Barbiero, L.; Filho, A.R.; Furquim, S.A.C.; Furian, S.; Sakamoto, A.Y.; Valles, V.; Graham, R.C.; Fort, M.; Ferreira, R.P.D.; Neto, J.P.Q. Soil morphological control on saline and freshwater lake hydrogeochemistry in the Pantanal of Nhecolândia, Brazil. *Geoderma* **2008**, *148*, 91–106. [CrossRef]

49. Masek, J.; Vermote, E.F.; Saleous, N.; Wolfe, R.; Hall, F.G.; Huemmrich, F.; Lim, T.K. LEDAPS Calibration, Reflectance, Atmospheric Correction Preprocessing Code, Version 2. Oak Ridge National Laboratory Distributed Active Archive Center, 2012. Available online: https://doi.org/10.3334/ORNLDAAC/1146 (accessed on 22 August 2019).

50. U.S.G.S. USGS EarthExplorer. USGS Science for a Changing World. 2015. Available online: http://earthexplorer.usgs.gov/ (accessed on 1 July 2019).

51. Zhu, Z.; Wang, S.; Woodcock, C.E. Improvement and expansion of the Fmask algorithm: Cloud, cloud shadow, and snow detection for Landsats 4–7, 8, and Sentinel 2 images. *Remote Sens. Environ* **2015**, *159*, 269–277. [CrossRef]

52. Ivits, E.; Cherlet, M.; Sommer, S.; Mehl, W. Addressing the complexity in non-linear evolution of vegetation phenological change with time-series of remote sensing images. *Ecol. Indic.* **2013**, *26*, 49–60. [CrossRef]

53. Hem, J.D. *Study and Interpretation of the Chemical Characteristics of Natural Water*, 3rd ed.; US Geological Survey: Alexandria, VA, USA, 1985; pp. 28–30.

54. Meybeck, M.; Helmer, R. Introduction. In *Water Quality Assessments. A Guide to the Use of Biota, Sediments and Water in Environmental Monitoring*, 2nd ed.; Chapman, D., Ed.; CRC Press: London, UK, 1996.

55. Tucker, C.J. Red and photographic infrared linear combinations for monitoring vegetation. *Remote Sens. Environ* **1979**, *8*, 127–150. [CrossRef]

56. Feyisa, G.L.; Meilby, H.; Fensholt, R.; Proud, S.R. Automated Water Extraction Index: A new technique for surface water mapping using Landsat imagery. *Remote Sens. Environ.* **2014**, *140*, 23–35. [CrossRef]

57. 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]

58. Xu, H. Modification of normalised difference water index (NDWI) to enhance open water features in remotely sensed imagery. *Int. J. Remote Sens.* **2006**, *27*, 3025–3033. [CrossRef]

59. Koza, J.R. Evolving a Computer Program to Generate Random Numbers Using the Genetic Programming Paradigm. In Proceedings of the 4th International Conference on Genetic Algorithms, San Diego, CA, USA, July 1991.

60. Schmidt, M.; Lipson, H. Symbolic Regression of Implicit Equations. In *Genetic Programming Theory and Practice VII*; Riolo, R., O'Reilly, U.-M., McConaghy, T., Eds.; Springer US: Boston, MA, USA, 2010; pp. 73–85. [CrossRef]

61. Parinet, B.; Lhote, A.; Legube, B. Principal component analysis: An appropriate tool for water quality evaluation and management—application to a tropical lake system. *Ecol Modell.* **2004**, *178*, 295–311. [CrossRef]

62. MDA Federal, 2004. Landsat GeoCover 1990/TM Edition Mosaics. Tile S-21-15. ETM-EarthSat-MrSID, 1.0. USGS: Sioux Falls, SD, USA, 1990. Available online: https://landsatlook.usgs.gov/ (accessed on 1 July 2019).

63. Takaku, J.; Tadono, T.; Tsutsui, K. Generation of high resolution global DSM from ALOS Prism. *Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci.* **2014**, *XL-4*, 243–248. [CrossRef]

64. Zani, H.; Assine, M.L.; McGlue, M.M. Remote sensing analysis of depositional landforms in alluvial settings: Method development and application to the Taquari megafan, Pantanal (Brazil). *Geomorphology* **2012**, *161*, 82–92. [CrossRef]

65. Martins, E.R.C. Typology of Saline Lakes in the Pantanal of Nhecolândia (MS). Ph.D. Thesis, University of São Paulo, São Paulo, Brazil, 2012.

66. Merino, E.R.; Assine, M.L. Hidden in plain sight: How finding a lake in the Brazilian Pantanal improves understanding of wetland hydrogeomorphology. *Earth Surf Process Landf.* **2019**, *45*, 440–458. [CrossRef]

67. Pott, A.; da Silva, J.S.V. Terrestrial and aquatic vegetation diversity of the Pantanal wetland. In *Dynamics of the Pantanal Wetland in South America*; Bergier, I., Assine, M.L., Eds.; Springer International Publishing: Cham, Switzerland, 2015; pp. 111–131. [CrossRef]

68. Freitas, J.G.; Furquim, S.A.C.; Aravena, R.; Cardoso, E.L. Interaction between lakes' surface water and groundwater in the Pantanal wetland, Brazil. *Environ. Earth Sci.* **2019**, *78*, 139. [CrossRef]

69. Andreote, A.P.D.; Vaz, M.G.M.V.; Genuário, D.B.; Barbiero, L.; Rezende-Filho, A.T.; Fiore, M.F. Nonheterocytous cyanobacteria from Brazilian saline-alkaline lakes. *J. Phycol.* **2014**, *50*, 675–684. [CrossRef]

70. Park, E.; Latrubesse, E.M. Modeling suspended sediment distribution patterns of the Amazon River using MODIS data. *Remote Sens. Environ.* **2014**, *147*, 232–242. [CrossRef]

71. Zang, C.; Huang, S.; Wu, M.; Du, S.; Scholz, M.; Gao, F.; Guo, Y.; Dong, Y. 2011. Comparison of Relationships Between pH, Dissolved Oxygen and Chlorophyll a for Aquaculture and Non-aquaculture Waters. *Water Air Soil Pollut.* **2011**, *219*, 157–174. [CrossRef]

72. Mariot, M.; Dudal, Y.; Furian, S.; Sakamoto, A.; Vallès, V.; Fort, M.; Barbiero, L. Dissolved organic matter fluorescence as a water-flow tracer in the tropical wetland of Pantanal of Nhecolândia, Brazil. *Sci. Total Environ.* **2007**, *388*, 184–193. [CrossRef] [PubMed]

73. Duarte, C.M.; Prairie, Y.T.; Montes, C.; Cole, J.J.; Striegl, R.; Melack, J.; Downing, J.A. CO_2 emissions from saline lakes: A global estimate of a surprisingly large flux. *J. Geophys. Res. Biogeosci.* **2008**, *113*. [CrossRef]

74. Martins, E.R.C.; Furian, S.; Barbiero, L. Dynamic of Lake Morphology on the Taquari Alluvial Fan: Pantanal of Nhecolândia (Ms), Brazil. In Proceedings of the VII National Congress of Geomorphology, Lisboa, Portugal, 8–10 October 2015.

75. Carpenter, D.J.; Carpenter, S.M. Modeling inland water quality using Landsat data. *Remote Sens. Environ.* **1983**, *13*, 345–352. [CrossRef]

76. Gallie, E.A.; Murtha, P.A. A modification of chromaticity analysis to separate the effects of water quality variables. *Remote Sens. Environ.* **1993**, *44*, 47–65. [CrossRef]

77. Barnes, B.B.; Hu, C.; Holekamp, K.L.; Blonski, S.; Spiering, B.A.; Palandro, D.; Lapointe, B. Use of Landsat data to track historical water quality changes in Florida Keys marine environments. *Remote Sens. Environ.* **2014**, *140*, 485–496. [CrossRef]

78. Thomann, G.C. Remote measurement of salinity in an estuarine environment. *Remote Sens. Environ* **1971**, *2*, 249–259. [CrossRef]

79. Khorram, S. Remote sensing of salinity in the San Francisco Bay Delta. *Remote Sens. Environ* **1982**, *12*, 15–22. [CrossRef]

80. Scudiero, E.; Skaggs, T.H.; Corwin, D.L. Regional scale soil salinity evaluation using Landsat 7, western San Joaquin Valley, California, USA. *Geoderma Reg.* **2014**, *2–3*, 82–90. [CrossRef]

81. Muller, S.J.; van Niekerk, A. An evaluation of supervised classifiers for indirectly detecting salt-affected areas at irrigation scheme level. *Int. J. Appl. Earth Obs. Geoinf.* **2016**, *49*, 138–150. [CrossRef]

82. Gorji, T.; Sertel, E.; Tanik, A. Monitoring soil salinity via remote sensing technology under data scarce conditions: A case study from Turkey. *Ecol. Indic.* **2017**, *74*, 384–391. [CrossRef]

83. Zhang, T.-T.; Qi, J.-G.; Gao, Y.; Ouyang, Z.-T.; Zeng, S.-L.; Zhao, B. Detecting soil salinity with MODIS time series VI data. *Ecol. Indic.* **2015**, *52*, 480–489. [CrossRef]

84. El Harti, A.; Lhissou, R.; Chokmani, K.; Ouzemou, J.; Hassouna, M.; Bachaoui, E.M.; El Ghmari, A. Spatiotemporal monitoring of soil salinization in irrigated Tadla Plain (Morocco) using satellite spectral indices. *Int. J. Appl. Earth Obs. Geoinf.* **2016**, *50*, 64–73. [CrossRef]