

Article

Integrating Satellite Imagery and Ground-Based Measurements with a Machine Learning Model for Monitoring Lake Dynamics over a Semi-Arid Region

Kenneth Ekpeteri ¹, Mohamed Abdelkader ^{2,*}, Sunday Ishaya ³, Edith Makwe ³ and Peter Ekpeteri ³¹ Department of Geography and Atmospheric Science, University of Kansas, Lawrence, KS 66045, USA² Department of Civil, Environmental, and Ocean Engineering (CEOE), Stevens Institute of Technology, Hoboken, NJ 07030, USA³ Department of Geography and Environmental Management, University of Abuja, Abuja PMB 117, Nigeria

* Correspondence: mabdelka@stevens.edu

Abstract: The long-term variability of lacustrine dynamics is influenced by hydro-climatological factors that affect the depth and spatial extent of water bodies. The primary objective of this study is to delineate lake area extent, utilizing a machine learning approach, and to examine the impact of these hydro-climatological factors on lake dynamics. In situ and remote sensing observations were employed to identify the predominant explanatory pathways for assessing the fluctuations in lake area. The Great Salt Lake (GSL) and Lake Chad (LC) were chosen as study sites due to their semi-arid regional settings, enabling the testing of the proposed approach. The random forest (RF) supervised classification algorithm was applied to estimate the lake area extent using Landsat imagery that was acquired between 1999 and 2021. The long-term lake dynamics were evaluated using remotely sensed evapotranspiration data that were derived from MODIS, precipitation data that were sourced from CHIRPS, and in situ water level measurements. The findings revealed a marked decline in the GSL area extent, exceeding 50% between 1999 and 2021, whereas LC exhibited greater fluctuations with a comparatively lower decrease in its area extent, which was approximately 30% during the same period. The framework that is presented in this study demonstrates the reliability of remote sensing data and machine learning methodologies for monitoring lacustrine dynamics. Furthermore, it provides valuable insights for decision makers and water resource managers in assessing the temporal variability of lake dynamics.

Keywords: lake dynamics; MODIS; CHIRPS; Landsat; machine learning

Citation: Ekpeteri, K.; Abdelkader, M.; Ishaya, S.; Makwe, E.; Ekpeteri, P. Integrating Satellite Imagery and Ground-Based Measurements with a Machine Learning Model for Monitoring Lake Dynamics over a Semi-Arid Region. *Hydrology* **2023**, *10*, 78. <https://doi.org/10.3390/hydrology10040078>

Academic Editors: Mavromatis Theodoros, Ioannis Pytharoulis, Kleoniki Demertzi and Vassilis Aschonitis

Received: 26 February 2023

Revised: 26 March 2023

Accepted: 28 March 2023

Published: 31 March 2023



Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Water is an essential component of hydrological and ecological cycles and is a vital resource for environmental sustainability [1,2]. Approximately 5.6 million km² of Earth's surface is covered by inland surface water bodies, including rivers and lakes, representing 1.1% of the Earth's surface area [3]. Although these inland surface water bodies hold only 0.013 percent of the Earth's total water, they play an important role in global and regional water cycles [4]. Thus, mapping inland water bodies is essential for understanding the hydrological cycle [5,6]. In view of the intensification of the hydrological cycle, it is imperative to provide better estimates of evaporation, lake water levels, and river stages. By using these estimates, it is possible to better assess flood and drought damage, develop a state-of-the-art flood resilience planning tool, and implement more efficient reservoir management and operation strategies.

On a regional and global scale, inland water bodies play an essential role in supporting human life. Consequently, monitoring their spatio-temporal variability is imperative for improving water management practices. The lake plays a vital role as a source of water for domestic and industrial uses [7], as well as an integral part of the riparian zone [8],

which, together, supports the ecosystem around it. A number of factors have contributed to the large variations in the surface areas of water bodies, including climate change and variability, as well as anthropogenic impacts. As a result, inland surface water mapping can provide valuable information that can contribute to the understanding of lake dynamics on a variety of spatial and temporal scales. Due to the vast extent of inland water bodies, remote-sensing-based observations could be an effective means for mapping this extent. Previous studies have shown the efficiency of remote sensing data for water-resources-related studies. These include assessments of inland water bodies, soil moisture, and groundwater on the watershed, continental, and global scales [9–11].

The surface water extent within lake basins plays a crucial role in shaping the physical and ecological processes in these areas. To achieve the proper management of the water resources in a lake basin, it is crucial to estimate and predict the lake dynamics, based on hydro-meteorological variations and anthropogenic disturbances. This task is particularly challenging in arid and semi-arid regions, where water scarcity poses a significant threat to human life. The extent of the lake and its water levels are mainly influenced by the complex interactions between the components of the hydrological cycle, such as precipitation, evaporation, and groundwater fluxes [12–15]. In contrast, the short-term variations in water levels are significantly impacted by meteorological factors, such as wind speed and pressure fluctuations over the lake surface [16,17].

Hydro-climatological factors play a crucial role in determining water volume and predictability metrics. For instance, Lake Mead, which was formed upstream of the Hoover Dam, shows a net evaporation and precipitation variability that correlates with lake water volume change [15]. Similarly, Lake Hulun in China is facing significant threats to its longevity, due to both human and natural factors [13]. Haramaya Lake in Ethiopia has experienced notable changes in its surface area, primarily due to the expansion of agriculture, unregulated human settlements, and vegetation removal [12]. Meanwhile, the Jiangnan Plain and the Dongting Lake regions in China have undergone significant size and number changes over thousands of years, influenced by both natural causes and human activities [18]. The study revealed that the decline in these lake areas appears to coincide with periods of rapid land reclamation in the middle reaches of the Yangtze River. Furthermore, these uncontrolled land reclamation activities negatively impacted the sediment depositions in the lakes, leading to further reductions in their size [19]. A recent study investigated the evaporation volume of 1.42 million global lakes between 1985 and 2018, on a global scale, using Landsat imagery [20]. The study estimated the surface areas of both natural lakes and artificial reservoirs. Meteorological data were employed to calculate the evaporation rates and volumes, demonstrating the importance of evaporation volume for evaluating the impacts of climate change on a lake's ecosystem.

These factors that affect surface water resources are expected to exacerbate water challenges. Thus, this makes it increasingly difficult to meet water demands, resulting in an insufficient water supply for people and the environment. Therefore, evaluating the temporal evolution of the changes in a lake's surface area is imperative. To understand the sustainability of lakes for supply and demand, as well as to preserve one of the most fundamental features of the hydrological landscape, it is vital to quantify the historical changes in lake dynamics. A lack of efficient lake measurement techniques has created a gap in quantifying these lake changes over time. Thus, this hampers the United Nations Sustainable Development Goals (UNSDG) on water scarcity and environmental sustainability [21].

Lake dynamics refers to the changes that occur within a lake over time, including physical, chemical, and biological features, as well as their interactions with one another. Several factors can influence these lake dynamics, such as climatic conditions, the presence of nutrients and other substances, water inputs and outputs, and the activities of living organisms. Understanding lake dynamics is crucial for several purposes, including the management of water resources, the conservation of aquatic ecosystems, and the prediction of the future conditions of a lake. The assessment of the lake area can be regarded as a critical aspect of lake dynamics. Alterations in the lake area can result from various factors,

such as changes in the temperature and precipitation, the water table's fluctuations, and the presence of natural or artificial structures that modify the water flow into or out of the lake. Several studies have found that lakes have experienced significant changes in area and volume over the past few decades, largely due to changes in precipitation, temperature, and landscape [22–25].

A succinct overview of lake area delineation methods includes visual interpretations, thresholding, supervised and unsupervised classifications, object-based image analyses, edge detection and active contours, water indices, fusions of multi-source data, and in situ measurements. These approaches encompass manual interpretation, machine learning techniques, and field surveys, with their applicability being dependent on factors such as data availability, spatial and temporal resolution requirements, and lake and landscape characteristics. The selection of an appropriate method is crucial for accurately characterizing and monitoring the lacustrine dynamics in hydrological studies.

Previous related studies have shown the efficiency of supervised and unsupervised classification methods, combined with satellite imagers, in providing an accurate delineation of a lake area when high-quality training data are available. In a study on the surface water resources in Nepal, six machine learning algorithms were evaluated for extracting this surface water information [26]. Except for naive Bayes, recursive partitioning, and regression trees, all the algorithms (neural networks, support vector machines, random forests, and gradient boosted machines) demonstrated a good performance, with the random forest showing the highest accuracy. The authors recommended separating the study areas by elevation and snow cover before applying the machine learning algorithms to the Landsat data, or the data that were supplemented with slope or NDWI information for optimal results. Dirscherl et al. (2020), suggested an automated machine-learning-based approach for the mapping of Antarctic supraglacial lakes [27]. The study aimed at understanding the impact of supraglacial lakes on the ice sheet mass balance and global sea level rise, and a machine learning algorithm was employed using Sentinel-2 and TanDEM-X topographic data for the automated mapping of these Antarctic supraglacial lakes. A random forest classifier was trained on 14 training regions and tested on 8 spatially independent regions across Antarctica. The accuracy assessment demonstrated an average kappa coefficient of 0.86. The study highlighted the potential of the random forest classifier for the automated mapping of supraglacial lakes, using Sentinel-2 data across all the Antarctic regions. Various studies have shown promising results in estimating Earth's surface features, including its water area [28–30]. However, there are still several research gaps that need to be addressed to enhance the accuracy, reliability, and generalizability of ML-based methods for the monitoring of lake dynamics. Thus, the integration of various data sources, such as remote sensing observations and ground data, can improve the lake area estimation. However, fusing multi-source data is a challenge that requires further investigation.

This study addresses whether the integration of climate variables from in situ and remote sensing sources will facilitate the assessment of long-term lake area fluctuations, using different cases from two continents with a semi-arid, dominant climate pattern. An approach that is based on machine learning is presented in the study, in which the fluctuations of lake areas are quantified through remote sensing imagery. This enables us to detect the changes in lake areas. In addition, the study assesses the long-term variability of this lake area change by incorporating hydro-climatological components, namely precipitation and evapotranspiration.

2. Materials and Methods

2.1. Study Area

The study area that is shown in Figure 1 includes (1) the Great Salt Lake (GSL), which lies in the northern part of the United States, in the state of Utah, with geographic bounds covering (40° to 42° N) and (111° to 113° W); and (2) Lake Chad (LC) in Africa, which is located in the Sahelian zone of west-central Africa, around (12° to 14° N) and (13° to 15° E). The GSL is the largest saltwater lake in the Western Hemisphere and the eighth-largest

terminal lake, with significant influences from turbidity, snow, and evaporation [31]. Although both the GSL and LC are endorheic lakes, the GSL is a remnant of Lake Bonneville, a prehistoric body of water that covered much of western Utah, millions of years ago [32]. GSL's fluctuations over the years have been separately linked to the West Desert Pumping Project (WDPP), which was established in the past to mitigate flooding by pumping water from the GSL into the nearby desert. The nature of this water extraction from the GSL has caused major concern, due to its implications for the local climate; hence, the new move to restore and maintain the lake by the government is in place. Understanding the lake area's fluctuations, the climate variables, and the lake's physical parameters is crucial to making future predictions for lake areas.

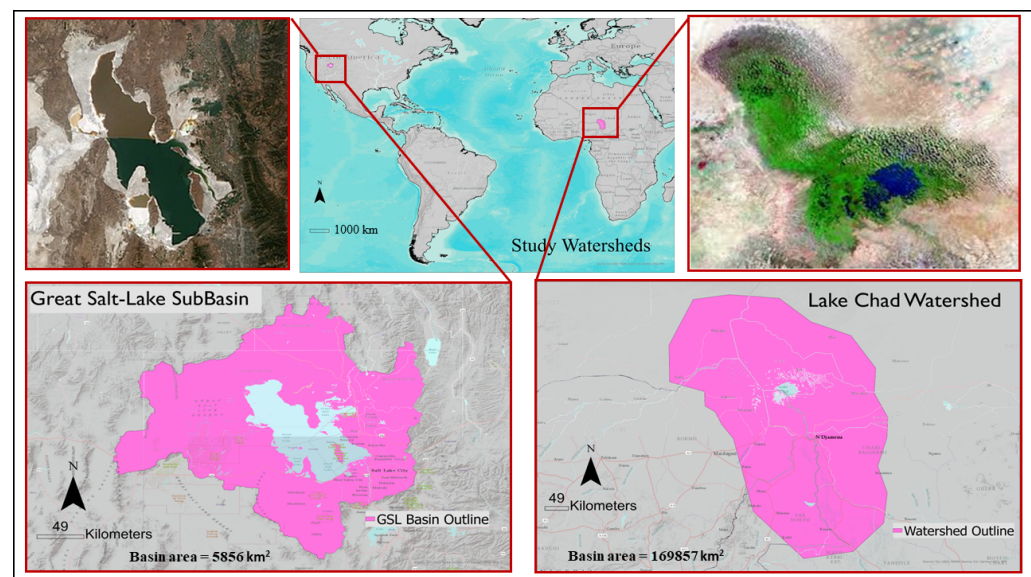


Figure 1. Geographical location of the study areas. The left insert represents the Great Salt Lake Sub-basin and the satellite image of GSL at its current state. The right insert map shows the digitized boundary of the Lake Chad watershed and the satellite image of Lake Chad in its current state.

On the other hand, Lake Chad spans Cameroun, Chad, Nigeria, Niger, and the Central African Republic, serving as a source of natural resources for over 35 million people in the region [33]. Lake Chad is situated in an interior basin that was formerly occupied by a much larger ancient sea, which was known as Mega-Chad [34]. Historically, Lake Chad is ranked among the largest lakes in Africa, regardless of its seasonal and annual surface area fluctuations [34,35]. Lake Chad is an important regional asset that is known for its archaeological discoveries and its role in trans-Saharan trade. Lake Chad is estimated to have lost more than 90% of its water mass, shrinking from an estimated 30,000 km² to 2500 km² between 1963 and 2013 [33,36].

2.2. Datasets

2.2.1. Landsat Imagery

In total, three Landsat products, the Thematic Mapper (TM), the Enhanced Thematic Mapper (ETM+), and the Operational Land Imager (OLI), were harmonized in this study to cover the study period of 1999 to 2021. This is because the TM was operational between 1984 to 2013 and the ETM+ between 1999 and 2022, which possessed scan line issues that were resolved in OLI (2013 to present). The harmonized image collection was used for the training and classification of features for LC and the GSL. Landsat-5 Thematic Mapper Collection 1 Tier 1 calibrated the top-of-atmosphere (TOA) reflectance that covered the period from 1984 to 2013; however, there were missing Landsat scenes from 1984 to 1998, lacking a complete annual coverage of the Lake Chad Basin. This Landsat data gap over the Sahelian African region was the major reason that the study period started in 1999. The

second reason was to focus on recent lake area dynamics. Landsat 7 Collection 1 Tier 1 and Real-Time data calibrated the TOA reflectance that was used in this study and covered the period from 1999 to date; however, the products were characterized by a scan line corrector (SLC) failure between 2003 and 2011 [37]. We avoided the SLC characteristic periods and used the SLC year of 2012 to bridge Landsat-5 and Landsat-8, which started in 2013 and covered the study period up to 2021. The Landsat 8 Collection 2 Tier 1 calibrated the TOA and was used to ensure consistency with the earlier Landsat products that were used in the study.

For the annual image processing, we collected all the images that intersected our study area and created an image from a combination of temporal composite and spatial mosaic from the image footprint. The annual composited images were derived for the classification algorithm to proceed. These filter methods in the Google Earth Engine (GEE) include filtering the image collection by the study watersheds, by date range, and setting a percentage cloud cover of 50% to ensure that only cloud-free images up to 50% were used for the analyses. A total of 4955 images over the GSL were used after the filter operation was applied, while Lake Chad recorded over 4211 images that covered the period of 1 January 1999 to 31 December 2021. The average number of images that were used for the yearly composite was 216 for the GSL and 184 for LC. From these numbers, the study used over 9000 images for the two lakes and covered 1999 to 2021. The filtering procedure and the creation of the image composite are summarized in Figure 2.

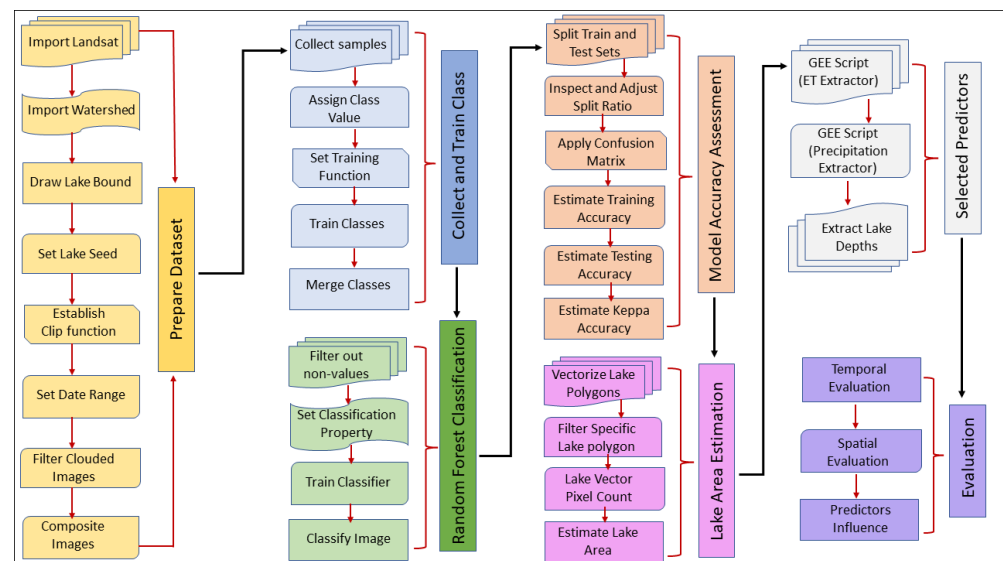


Figure 2. Flowchart of steps for collecting and processing data, and adopting machine learning techniques for estimating lake extent fluctuations. Non-values in this case are the missing values.

2.2.2. Precipitation Data

The precipitation datasets for this study were obtained from the Climate Hazards Group InfraRed Precipitation with Station data (CHIRPS), covering over 30 years of temporal scope and starting in 1981. CHIRPS incorporates $0.05^\circ \times 0.05^\circ$ resolution satellite imagery with in situ station data to create a gridded rainfall time series for trend analyses and seasonal drought monitoring [38]. The CHIRPS station processing stream incorporates data from five public data streams and several private archives. More details about these public data streams and private archives are presented in [38]. The CHIRPS precipitation dataset that was provided by the Google Earth Engine can be extracted monthly, quarterly, or yearly. This study used the yearly intervals of the CHIRPS datasets for its analysis. The annual estimated precipitation over the watershed from the CHIRPS dataset was used to find the relationship between the lake area dynamics and the precipitation for the years under consideration. A preliminary evaluation of the CHIRPS dataset was performed

using the NLDAS hourly precipitation from the Rapid Forcing Retrieval (RFR) web tool and the Global Precipitation Measurement (GPM-IMERG), which has a 0.1° resolution and global coverage [39].

2.2.3. Evapotranspiration Data

The evapotranspiration (ET) data source that was used in this study was obtained from the Terra Moderate Resolution Imaging Spectroradiometer (MODIS) MOD16A2 version 6. The MOD16A2 version 6 evapotranspiration/latent heat flux product is an 8-day composite product that was produced at 500 m pixel resolution. The algorithm that was used for the MOD16 data product collection was based on the concept of the Penman–Monteith equation, which takes inputs of daily meteorological reanalysis data along with MODIS remotely sensed data products, such as land cover, albedo, and vegetation property dynamics. The MODIS ET dataset provides global coverage, making it an acceptable option for driving global and continental research such as this current study. The pixel values for the ET layer that was used in this study were collected every 8 days within the composite period. The last 8-day period of each year was a 5- or 6-day composite period, depending on the year [40]. The annual means of the ET over the entire study watersheds were estimated for the period between 1 January 1999 to 12 December 2021, and were used for determining the impact of ET on the lake area fluctuations, and for using those relationships to make future predictions. ET datasets were also provided by the Rapid Forcing Retrieval (RFR) web tool [39].

2.2.4. Lake Depth Data

The lake water depth data that were measured from the reference height that was used in this study were provided by the European Theia Project (ETP). These ETP data use Jason-1 and Jason-2 altimetry datasets and are made available with global coverage. These datasets have been available at 10-day intervals since 1993 [41]. The dataset is available at <https://hydroweb.theia-land.fr/> (accessed on 7 November 2022) and requires an account sign-up for access. We compared the ETP data using complementary data from the Global Reservoir and Lake Monitor (GRLM), which were provided by the joint project of the U.S. Department of Agriculture (USDA), Foreign Agricultural Service (FAS), National Aeronautics and Space Administration (NASA) Goddard Space Flight Center (GSFC), and the University of Maryland (UMD). The GRLM records variations in surface water height for approximately 70 lakes and reservoirs worldwide, using a combination of satellite radar altimetry datasets. However, the data were only made available starting in 2003, providing a temporal limitation for studies of earlier dates [42]. The ETP was, however, validated using GRLM to boost the research confidence. For clarity and comparison, Table 1 summarizes the properties of the datasets that were used in this study.

Table 1. Summary of the remote sensing products used in the study. TOA stands for top of atmospheric reflectance, SR—surface reflectance, TM—Thematic Mapper, ETM+—Enhanced Thematic Mapper, OLI—Operational Land Imager, and NA—not applicable.

Properties	Landsat	CHIRPS	MODIS	ETP
Product	Landsat-5 (TM), 7 (ETM+), and 8 (OLI)	CHIRPS Daily	MOD16A2	Jason-1 and Jason-2 altimetry
Spectral resolution	1 to 9 bands	1 band	36 bands	2 bands
Pixel size	30 m	5566 m	250, 500 or 1000 m	NA
Scene width	185 km	0.05°	2330 km	1324 km
Temporal resolution	16 days	Day, month, pentad, and year	Twice daily	10 days
Reflectance	TOA	Satellite + station data	SR	Satellite + stations
Time span	1984 to present	1981 to present	1999 to present	1993 to present
Target	Earth features (e.g., water bodies)	Precipitation	Evapotranspiration and temperature, etc.	Lake depth, area, and volume

2.3. Classification Algorithm and Evaluation Metrics

2.3.1. Supervised Classification

The machine learning algorithm that was used in this study was the supervised classification technique (SCT) type. SCT identifies a new category of observations based on training datasets. The algorithm learns from the dataset and then classifies these new observations into several classes that are defined by a Boolean (true or false, yes or no, and negative or positive). These new classes can then be used as the targets/labels or categories [43]. In summary, SCT works by recognizing specific entities within the dataset and trying to conclude how those entities should be defined. Some SCT algorithms include support vector machines (SVM), linear classifiers, decision trees, k-nearest neighbors, and random forests [43–50].

To reduce the variance and create more accurate data predictions, this study adopted the random forest, a non-linear supervised machine learning algorithm, which applies to both classification and regression purposes. RF builds decision trees on different samples and proceeds by taking their majority votes for classification and regression [45]. RF earns its reputation by having the capability to handle datasets with continuous variables (e.g., for regression), and performs best when dealing with categorical variables (e.g., for classification). RF is most useful because it takes less training time, predicts outputs with a higher accuracy, and maintains this accuracy when a large proportion of data is missing [46]. In addition to the benefits of RF, two assumptions must be adhered to: the availability of the actual values in the feature variable, for the dataset to avoid guessing, and the predictions from the individual tree needing to have low correlations [42]. The steps for using the RF algorithm in this study included selecting a random K data point from the training set, building the decision trees that were associated with the selected data points, choosing the study number of N for the decision trees, repeating the process, and finally, for each data point, finding the predictions of each decision tree and assigning the new data points to the group that won the most votes [42].

A preliminary assessment has shown that decision trees run the risk of overfitting, as they try to fit all the samples within the training data. However, since the RF consists of multiple decision trees, its classifier will not overfit the model, because the averaging of the uncorrelated trees lowers the overall variance and prediction error. The study's analyses were conducted using the Google Earth Engine (GEE), which has standard API that uses only four classifier packages for handling traditional ML algorithms. At the moment, the GEE classifier ML algorithms include CART, RF, naïve Bayes, and SVM. An initial assessment showed that the RF algorithm is currently the best amongst the four

available GEE classifiers, as it has the capability to handle larger datasets (the collection of Landsat datasets that cover the study period).

The hyperparameters of the RF algorithm that was built in the GEE consisted of the number of trees, the variables per split, the minimum leaf per population, the bag fraction, the maximum nodes, and the seed, which were all built internally and accessed by using the GEE API. The number of trees variable determined the number of decision trees that had to be created. A total of 350 trees were used within the GEE processing capability. The initial hyperparameter tuning indicated that the more trees the better, so we tried different trees, keeping in mind the processing capability and time-out within the GEE. The 350 trees that were used were the optimal number that was obtained from hypertuning the parameters. The variables per split used the square root of the number of variables. The minimum leaf population variable was set to default as 1; this created nodes whose training set contained at least these many points. The bag fraction was the fraction of the input to the bag per tree, which was set to default as 0.5. The maximum nodes, which was the maximum number of the leaf nodes in each tree, was left as default (null), allowing for the algorithm to derive its nodes from the set of input features. The randomization seed was set to default as 0. The full script for implementing the GEE RF algorithm for our target watershed is presented in the data section, while the workflow is presented in Figure 2.

2.3.2. Model Performance Evaluation

In total, four accuracy indices and one statistical metric were used to evaluate the performance of the machine learning (ML) algorithm for this study. The accuracy indices included the overall accuracy, kappa coefficient, producer accuracy, and user accuracy, while the statistical metric that was used was the correlation coefficient metric. The producer accuracy, which measured the error of omission, showed the probability that a reference sample was correctly classified. The user accuracy measured the error of commission by representing the probability that a randomly selected sample from the satellite imagery was consistent with the observations from the reference data. The overall accuracy (OA) provided the probability that a randomly selected sample from the satellite imagery was classified accurately. The OA can be expressed mathematically as:

$$OA = \frac{NTP + NTN}{GTP} \times 100 \quad (1)$$

where NTP is the number of true positives from the pixel samples, NTN is the number of true negatives, and GTP is the ground truth pixels.

The GEE RF algorithm provides the tools for collecting features (land cover identifiable features), which are served as inputs and used for both training and validation. The reference image here is the samples that were collected for the seven classes of features that were used in the study (water, urban, forest, grassland, bare earth, cropland, and shadow). In total, 5313 samples were collected between 1999 and 2021 on the GSL watershed (for example), and from which, 70% served as reference images for the validation, while the target image was the classified image.

The kappa coefficient (k) measures the percentage of the agreement between the classified water body pixels and the ground truth. k is expressed mathematically as:

$$k = \frac{P_O - P_e}{1 - P_e} \quad (2)$$

where P_O is the overall accuracy of the model, and P_e is the measure of the agreement between the model predictions and the actual class values, as if happening by chance [11].

The study adopted the Pearson's correlation coefficient (cc) metric to investigate the degree of the shared external variabilities and relationships between the hydro-climatological variables (precipitation, ET, and lake height from reference point) and the estimated lake area. The internal variability between the lake area dynamics over the years was also measured. The mathematical expression in Equation (3) was used where the correlation

coefficient between the lake depth and lake area was estimated for the study periods (t). The cc was unitless and ranged from -1 to 1 , with a higher estimate nearing 1 suggesting a positive strong correlation, while values nearing -1 suggested a strong negative correlation.

$$cc = \frac{\sum_{t \in T} (P_t - \bar{P})(A_t - \bar{A})}{\sqrt{\sum_{t \in T} (P_t - \bar{P})^2} \sqrt{\sum_{t \in T} (A_t - \bar{A})^2}} \quad (3)$$

where P are the study's adopted parameters (precipitation, evapotranspiration, or in situ lake depth), A is the estimated lake area, and t is the i -th year category.

2.4. Data Processing Tools and Workflow

Over the 21 years of the Lake Utah (LU) surface dynamics that were studied, 4995 images were used in total, while 4211 images were used to analyze the Lake Chad (LC) surface area change for the same period. The analyses of these large datasets were performed remotely on the cloud using the Google Earth Engine (GEE), and would have proven computationally intensive and expensive if performed locally. However, the GEE cloud-based platform takes away the burden, by providing an intrinsically parallel way with a high performance and consistency of its multi-petabyte remote sensed data, which are preprocessed to be ready-to-use and efficiently accessed [46–48].

A summary of the study's methodology and workflow is shown in Figure 2, and includes the dataset preparation, training of the class, random forest classification, model accuracy assessment, lake area estimation, extraction of predictors, model, and results from the evaluations. The dataset preparation stage included importing the Landsat datasets into the GEE environment and filtering them by date, basin area geometry, and cloud cover percentage. Further clouded pixels were masked by using the GEE's simple composite algorithms. The class training stage included the sample collection, assigning values to each class (e.g., water = 0, bare earth = 1, forest = 2, grassland = 3, cropland = 4, urban = 5, and shadow = 6), setting the training function, training the class, and merging class. The next stage was the RF classification, which included the filtering out of missing values, setting the classification property, training the classifier, and finally, classifying an image. At the model assessment stage, the study split the trained samples into training and testing sets (70% and 30% for a start), inspected the results, and adjusted them to different split ratios, applying the RF confusion matrix and evaluating the accuracies (training, testing, and kappa accuracies). The final stage of the lake area estimation included vectorizing the lake polygons and filtering the specific lake polygons using the lake seed, lake pixel count, and pixel area conversion for the lake area estimation. Additional scripts for estimating the precipitation, ET, and lake depth using altimetry were developed.

3. Results and Discussion

3.1. Assessment of the Machine Learning Method Accuracy

The study used 4955 images over two decades, between 1999 to 2021, to perform the image classification task for the GSL, and 4211 images for LC during the same period (Tables 2 and 3). Satellite imagery was available in greater numbers over the GSL watershed. The study did account for the number of images per year for the study period in each watershed. For instance, over the GSL, the year 2009 had the highest number of images (351 filtered images), while 2021 recorded the least (139 filtered images) (Table 2). Unlike the GSL, LC recorded fewer numbers of images, with the highest in 2016 (313 filtered images), and the least in 2021 (10 images) (Table 2). Data scarcity was the main factor that influenced the number of images that were used in the study, particularly over the LC watershed. The discrepancies in the number of images that were used for the classification tended to impact the overall feature classification, as more images produced a better mosaic representation of reality and a realistic classification scheme [15].

Table 2. Random forest classification accuracy over the Great Salt Lake Basin. Where OA is overall accuracy, KOA is the kappa overall accuracy, TA is the training accuracy, KTA is kappa training accuracy, VA is the validation accuracy, and KVA is kappa validation accuracy.

Year	OA	KOA	TA	KTA	VA	KVA	Images	Total Samples	Train	Test
1999	0.991	0.989	0.945	0.935	0.97	0.964	159	231	70	30
2000	0.987	0.984	0.95	0.942	0.956	0.948	147	231	70	30
2001	0.982	0.979	0.921	0.908	0.923	0.909	203	231	70	30
2002	0.995	0.994	0.97	0.965	1	1	187	231	70	30
2003	0.991	0.989	0.961	0.955	0.972	0.968	256	231	70	30
2004	0.991	0.989	0.94	0.929	0.953	0.944	229	231	70	30
2005	0.995	0.994	0.974	0.97	0.958	0.949	188	231	70	30
2006	0.978	0.974	0.94	0.93	0.92	0.905	165	231	70	30
2007	0.987	0.984	0.943	0.933	0.986	0.983	338	231	70	30
2008	1	1	0.941	0.931	0.855	0.83	305	231	70	30
2009	0.991	0.989	0.948	0.94	0.909	0.891	351	231	70	30
2010	0.987	0.984	0.941	0.93	0.95	0.941	229	231	70	30
2011	0.991	0.989	0.942	0.932	0.946	0.937	210	231	70	30
2012	0.995	0.994	0.947	0.938	0.901	0.883	207	231	70	30
2013	0.987	0.984	0.961	0.955	0.945	0.936	224	231	70	30
2014	0.991	0.989	0.96	0.953	0.937	0.926	167	231	70	30
2015	0.995	0.994	0.963	0.957	0.97	0.964	174	231	70	30
2016	0.995	0.994	0.993	0.992	0.973	0.968	244	231	70	30
2017	0.991	0.989	0.963	0.957	0.97	0.964	269	231	70	30
2018	0.991	0.989	0.94	0.93	0.953	0.944	144	231	70	30
2019	0.991	0.989	0.945	0.935	0.97	0.964	231	231	70	30
2020	0.978	0.974	0.963	0.957	0.953	0.945	189	231	70	30
2021	0.995	0.994	0.954	0.946	0.982	0.978	139	231	70	30
	0.990	0.988	0.952	0.944	0.95	0.941	4955	5313	3719	1594

The study collected 5313 samples over the GSL watershed and 4876 samples over LC for the image classification (Tables 2 and 3). These feature samples included water, urban, forest, grassland, bare earth, cropland, and shadow. While the water was the target feature, it was essential to correctly classify the remaining features, in order to accurately distinguish the water bodies, especially lakes, from the non-water features, particularly shadows, which could cause a waterbody miss-classification. These many samples added confidence to the classification algorithms by ensuring that we had many reserved testing samples. In these cases, we reserved over 1600 and 1500 test samples for the GSL and LC, respectively. The training and testing columns in Tables 2 and 3, respectively, are the percentage of the total collected samples that were reserved for the training and validation in the RF algorithm. Including the class “shadow” helped to eliminate the chances of miss classifying a shadow as water. An equal number of samples (759 samples) were collected for the individual classes for the period of 1999 to 2021. The more samples that were in the sets (split into training and validation) for each year, the better the outcome. There was no fixed number of samples that was generally agreed upon, and we generally collected as many as possible where the feature was well-defined to avoid mixed classification, such as classifying a swamp as grassland or a water body.

Table 3. Random forest classification accuracy over Lake Chad watershed. Where OA is overall accuracy, KOA is the kappa overall accuracy, TA is the test sample accuracy, KTA is kappa test accuracy, VA is the validation sample accuracy, and KVA is kappa validation accuracy.

Year	OA	KOA	TA	KTA	VA (%)	KVA	Images	Total Samples	Train	Test
1999	0.97	0.96	0.903	0.886	0.789	0.753	10	212	70	30
2000	0.981	0.977	0.9	0.882	0.951	0.943	88	212	70	30
2001	0.981	0.977	0.952	0.944	0.939	0.929	83	212	70	30
2002	0.976	0.972	0.925	0.912	0.942	0.931	102	212	70	30
2003	0.985	0.983	0.931	0.919	0.91	0.894	93	212	70	30
2004	0.971	0.966	0.916	0.901	0.877	0.854	159	212	70	30
2005	0.962	0.955	0.91	0.895	0.924	0.91	134	212	70	30
2006	0.981	0.977	0.893	0.875	0.915	0.9	159	212	70	30
2007	0.99	0.988	0.895	0.877	0.898	0.88	143	212	70	30
2008	0.976	0.972	0.931	0.919	0.903	0.886	149	212	70	30
2009	0.966	0.961	0.915	0.9	0.949	0.939	126	212	70	30
2010	0.966	0.961	0.925	0.913	0.89	0.868	110	212	70	30
2011	0.99	0.988	0.904	0.888	0.927	0.914	90	212	70	30
2012	0.976	0.972	0.894	0.876	0.901	0.884	129	212	70	30
2013	0.99	0.988	0.937	0.927	0.97	0.965	196	212	70	30
2014	0.99	0.988	0.936	0.925	0.943	0.933	292	212	70	30
2015	1	1	0.949	0.94	0.932	0.92	305	212	70	30
2016	0.99	0.988	0.902	0.886	0.896	0.878	313	212	70	30
2017	0.981	0.977	0.92	0.907	0.885	0.864	312	212	70	30
2018	0.981	0.977	0.941	0.931	0.948	0.938	308	212	70	30
2019	0.985	0.983	0.963	0.957	0.959	0.952	301	212	70	30
2020	1	1	0.947	0.938	0.932	0.92	300	212	70	30
2021	0.971	0.966	0.933	0.922	0.852	0.826	309	212	70	30
	0.981	0.977	0.923	0.910	0.914	0.899	4211	4876	3413	1463

The average accuracy assessment over the GSL watershed shows that both the training accuracy and kappa training accuracy were 0.952 and 0.944, the validation accuracy and kappa validation accuracy were 0.950 and 0.941, respectively, while the overall accuracy and kappa overall accuracy were 0.99 and 0.988, respectively (Table 2). Over the LC watershed, the training accuracy and kappa training accuracy were 0.923 and 0.910, respectively, the validation accuracy and kappa validation accuracy were 0.914 and 0.899, and the overall accuracy and kappa overall accuracy were 0.981 and 0.977, respectively (Table 3). Although the overall accuracy suggests that the machine learning technique that was employed in the study performed well, the measurement of the GSL was better than that from the LC. The slight variation in the accuracy between the two lakes could be due to many factors; one of these factors happens to be the differences in the number of sampled images that were used in the classification scheme. This study, therefore, showed that the random forest classification algorithm could be a vital tool for supervised image classification and valuable for lake water extent delineation over time.

3.2. Spatial-Temporal Evolution of Lake Area

Figures 3 and 4 show the mean annual lake area changes over time. The main lake of interest (in red) represents this study's focus and shows significant fluctuations over the years. Figure 5 shows the numerical fluctuations in the lake area over the two decades, which Figures 3 and 4 attempt to depict. The study shows that the GSL lake extent declined

between 1999 and 2021, with a few years of fluctuations. The years of these intermittent fluctuations included 2005, 2011, 2014, 2017, and 2020, and will be discussed. The study found the GSL to have declined from almost 7000 km² in 1999 to less than 3500 km² in 2021, losing nearly half of its size within two decades (Figure 5).

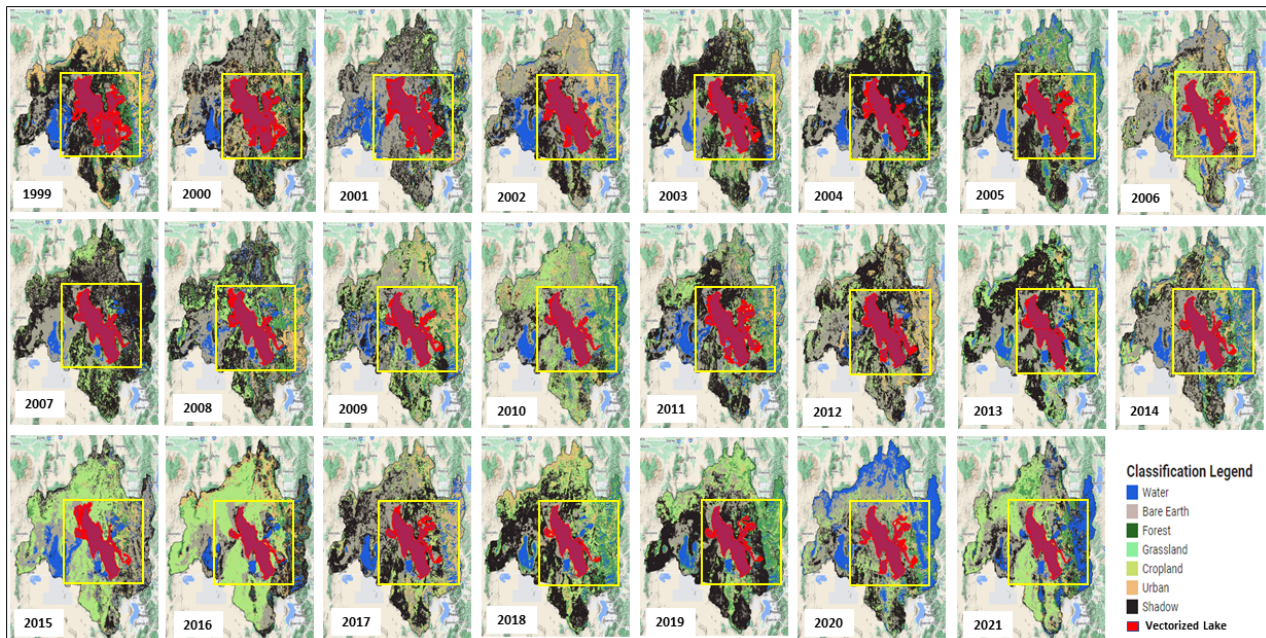


Figure 3. The Great Salt Lake surface water extent change through time. Vectorized lake of interest is in red, while the rest of unconnected pockets of water bodies are in blue. The rest of the features are represented with different colors in the image.

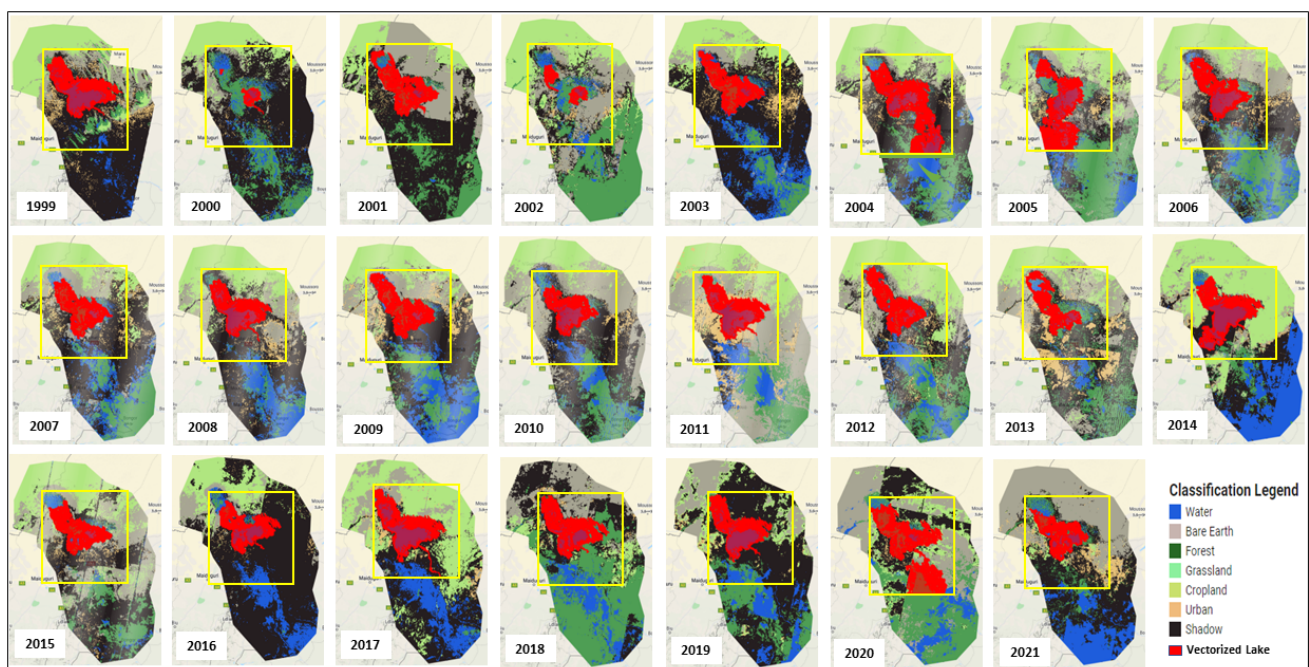


Figure 4. The Lake Chad surface water extent change through time. Vectorized lake of interest is in red, while the rest of unconnected pockets of water bodies are in blue. The rest of the features are represented with different colors in the image.



Figure 5. A comparative plot of lake surface area changes of GSL (**top**) and LC (**bottom**) between 1999 and 2021.

The study assessment of LC showed otherwise, as we mainly observed fluctuation. However, the study showed that LC had an estimated area coverage of nearly 900 km² in 1999, which declined to less than 800 km² in 2021. This LC decline in its area is not significant compared to the GSL; however, LC recorded major fluctuations in between those years of interest. Although both the GSL and LC are in arid regions, the study further explored the potential causes of these declines and fluctuations in the two lakes, which have proven to be of economic value to their local communities.

The obtained results showed that the GSL is on the decline, and that quick action is required to salvage the GSL to ensure that the lake keeps serving its purpose. Additionally, despite popular opinion that LC has been shrinking lately, our study has shown that LC is not shrinking, but somewhat fluctuating, with years of highs and lows. In the following sub-section, we try to understand what climatic predictors and physical parameters can explain the decline and fluctuations in the GSL and LC during these two decades.

3.3. Influence of Climate Variability

It is well established that changes in precipitation, inter-basin flow, evaporation, drinking water supply, and irrigation water usage may drive significant changes in lake extent, leading to ecological degradation, economic losses in the lake environment, and sometimes, irreversible changes in nature [16]. To this end, the current study investigated the temporal evolution of precipitation, evaporation, and water depth over the regions of interest, using historical records from between 1999 and 2021. Thus, one of the main

objectives of this study was to reveal whether hydro-climatological factors could explain the changes in lake extent over time.

To examine the impact of precipitation variations on the lake extent changes in the study areas, the inter-annual changes in the annual precipitation are analyzed. The precipitation records for the GSL and LC are shown in Figure 6. It is noteworthy that, although there are significant fluctuations in these annual precipitation changes, the extent of the lake does not follow the same patterns as the variations in precipitation. For example, high precipitation records were observed for the GSL in 2005, 2010, 2016, and 2019. These precipitation records did not correspond to an increase in the lake's extent during these years. Nevertheless, it should be noted that the lake's surface area increased in the years that followed an increase in the annual precipitation. As seen from the graphs below, this was primarily observed for the years of 2006, 2011, 2017, and 2020, when the lakes maintained their surface areas or showed a significant increase in their surface area. The results that were obtained are also valid for the LC case, where the area of the lake increased significantly a year after the significant increase in precipitation. For instance, a peak in the annual precipitation occurred in 2019, followed by the highest annual lake extent in 2020. Therefore, the obtained results show that annual precipitation can be employed as a predictor for estimating the lake extent for the following year. It is also imperative to note that precipitation change should not be considered as the primary factor that is responsible for an increase or decrease in the lake surface area.

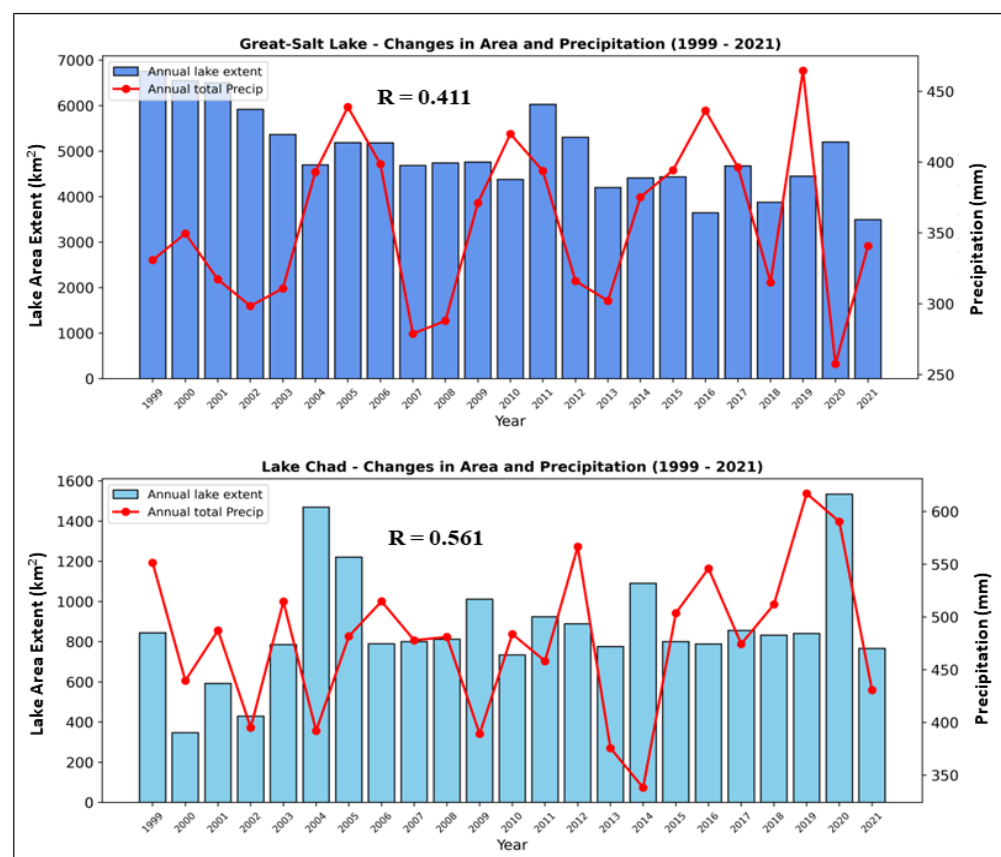


Figure 6. Temporal variation of the annual lake area and precipitation in GSL (top) and LC (bottom).

We assessed the annual variations of the water surface area of the GSL and LC, and we compared these lake surface dynamics with the rainfall temporal evolution. The results indicated that the lake's water surface area changed based on the rainfall availability. A significant decrease in the area extent occurred as a succession of dry years. Additionally, it was observed that the lake surface coverage exhibited a similar trend with the precipitation. The obtained results may be explained by the significant positive effects that precipitation

had on the lake area. Because precipitation is the primary source of water for lakes, it has a positive effect on lake area. However, the large LC and GSL lake areas did not necessarily coincide with the precipitation peaks, suggesting a hysteresis in the precipitation effects. For instance, the highest annual precipitation that was recorded for the GSL (year 2019) and LC (year 2019) contributed to the significantly larger lake areas in the two regions in 2020. It was also found that precipitation played an important role in the recovery of the lake area.

The interannual variations of the ET were examined to understand the patterns and changes in the lake surface area in relation to the atmospheric pressure, humidity, temperature, wind speed, and lake environment. It is important to note that the ET was estimated from a stable land area encompassing the maximum lake extent during the study period. In this study, the watershed that encompassed the lakes of interest and other water bodies was considered for the ET calculation. While the ET observations were obtained from a stationary area, the extent of the water bodies varied during the study period (Figures 3 and 4).

Figure 7 illustrates the temporal variation of the ET over the GSL and LC watershed areas. There was negative feedback between the ET and the lake area based on the correlation between the two variables. Therefore, an increase (decrease) in the ET results in a decrease (increase) in the lake's surface area. A decrease in the annual mean ET can be observed in the years 2001, 2007, 2011, 2012, and 2020, when an increase in the GSL extent was observed. The extent of the GSL decreased between 2006 and 2019 due to increased evaporation. A similar decrease in the extent of the LC area was caused by the largest increase in the ET over the LC Basin in 2010. Conversely, a significant decrease in the ET for the year 2005 resulted in a major expansion of the lake. Overall, the temporal variations of the ET and lake area exhibit the same patterns, with an emphasis on the negative feedback between the two variables. The obtained findings are valid for both the GSL and LC.

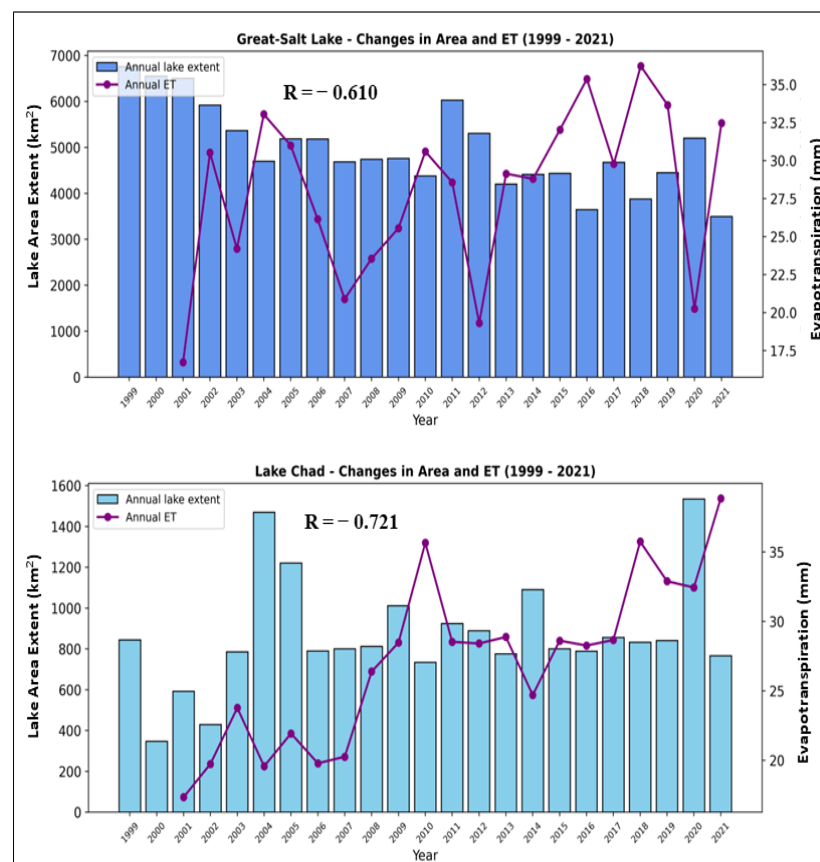


Figure 7. Temporal variation of annual lake area and evapotranspiration variation in GSL basin (top) and LC basin (bottom).

It is important to highlight that, in certain instances, a substantial increase (decrease) in the evapotranspiration did not correspond with a significant decrease (increase) in the lake area. This discrepancy can be attributed to the influence of the preceding year's precipitation on the lake's extent. For example, the Great Salt Lake exhibited an elevated mean ET in 2006, yet the lake's extent remained relatively stable. The substantial precipitation that was recorded in the region could account for this phenomenon. In the case of Lake Chad, the high precipitation levels in 2012 suggested that the lake would expand in 2013. Contrary to this expectation, there was a decrease in the LC area. This observation can be rationalized by considering the impact of the mean ET during the year 2013.

The relationship between the lake surface area and the water depth was also examined and is reported in Figure 8. In this study, a correlation between the variations in the surface area of the lake and the variations in its water depth was computed. The GSL showed a high correlation coefficient, while LC showed a low correlation coefficient. In the case of the GSL, the water depth records show variations that closely resemble the variations in the surface area. A significant relationship between the lake area and water depth did not emerge, however, it did from the variation of the lake area and the water depth time series that corresponded to LC. For instance, a significant decrease in the lake area extent was associated with the increased water depth in 2010. In contrast, LC experienced its largest surface area extent in the year 2020 for a similar depth. According to the results that were obtained, there is a site-specific relationship between the extent of lakes and their depths.

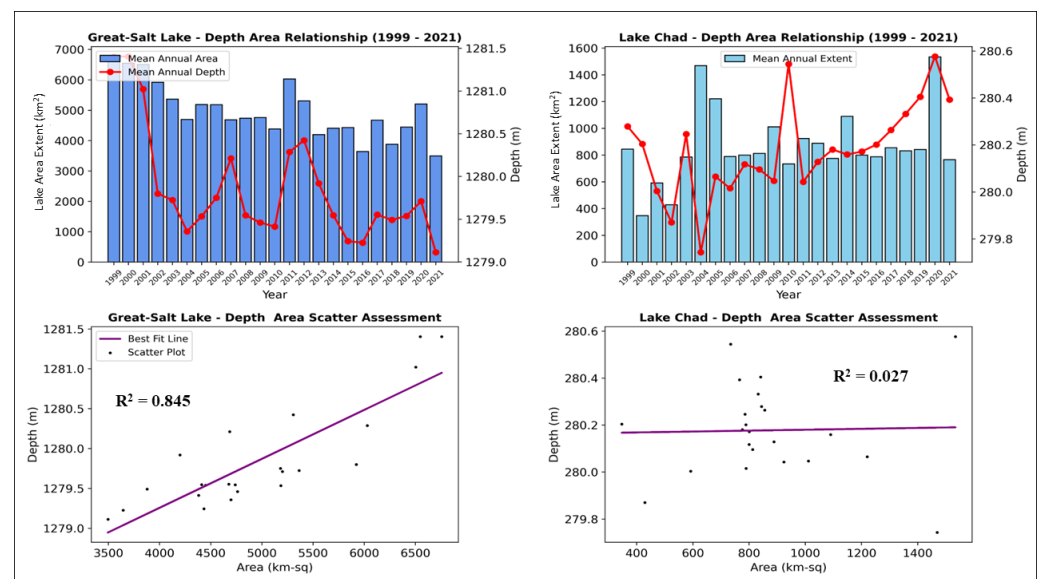


Figure 8. Temporal variation of annual lake area and water depth in GSL (**top left**) and LC (**top right**), and linear regression results between the lake area and water depth of GSL (**bottom left**) and LC (**bottom right**).

A weak correlation between the lake area and water depth, as observed in the case of LC, can be ascribed to multiple factors, encompassing morphological complexity, local geological and geomorphological influences, human interventions, seasonal and inter-annual variability, data quality and resolution, and the presence of aquatic vegetation and sedimentation. Lakes exhibit substantial variations in their shape, size, and depth profiles, which stem from their geological history, catchment characteristics, and surrounding topography. This morphological diversity generates disparate relationships between the lake area and depth, rendering the establishment of a consistent correlation challenging. Moreover, sedimentation and erosion patterns contribute to the variability in the relationship between the lake area and depth. For instance, sedimentation can diminish the water depth without significantly affecting the lake area. Furthermore, aquatic vegetation, such as floating or

submerged plants, can distort the lake area estimation, resulting in an overestimation of the true surface area.

It is also worth noting that the accuracy and resolution of the data that were employed to estimate the lake area and depth can considerably impact the observed correlations between these variables. Coarse spatial resolution or measurement errors can introduce uncertainty and mask the true relationship between the lake area and depth. This issue predominantly arises in under-observed regions, such as LC. Another factor to consider is the seasonal and inter-annual variability of regional hydrological processes, including precipitation, evaporation, and runoff, which can induce fluctuations in both the lake area and depth over time. Nevertheless, these changes might not consistently be proportional or linear, thereby weakening the correlation between the lake area and depth. These factors contribute to the tenuous relationship between the lake area and depth by introducing spatial and temporal heterogeneities, modifying the lake morphology, and affecting the data accuracy. A comprehensive understanding of these factors is indispensable for the precise assessment and monitoring of lake dynamics in hydrological studies.

According to the findings of this study, lake dynamics are influenced by a variety of factors. However, precipitation dominated the dynamics of both lakes, with a notable lag in the changes in the lake surface area. The hysteresis of the precipitation effects could also be attributed to the anthropogenic factors that contribute to lake shrinkage. These factors include agricultural irrigation, water extraction from groundwater, and flow diversion from the rivers and streams that feed the lake. According to [33], the water resources in the LC region were negatively impacted by agropastoral activities. The study indicated that deep aquifer groundwater is primarily exploited in northern Nigeria and eastern Niger. The influence of these anthropogenic factors should therefore be considered when assessing the LC area variations.

As for the GSL, human intervention began in the late 1800s with mining activities that contributed significantly to a sediment accumulation in the GSL. Furthermore, as the population of the region increased, the water feeding the lake (from rivers and streams) was diverted for use as a water supply [18,19]. Therefore, the decrease in the lake area is indicative of the impact of anthropogenic activities. Thus, both natural and anthropogenic factors have significantly affected the dynamics of the GSL and LC.

Based on the findings of this study, it was determined that the rapid expansion of lake surface area was consistent with high annual precipitation and a significant decrease in ET. Thus, the infiltration of precipitation compensates, to a certain extent, for groundwater extraction or freshwater diversion. According to the results of the study, the assessment of the water resources in the lake environment should also take into account additional factors such as recharge from nearby rivers. At this point in the analysis, the findings are in agreement with findings from previous studies regarding the control of anthropogenic factors, such as the extraction of groundwater and the diversion of river flows. In spite of this, it is less evident how the seasonal variation of regional hydrological processes, particularly precipitation, affect the lake dynamics across the domain, and should be the subject of future research.

Our study examined the inter-annual dynamics of lakes, but the intra-annual variations of lakes could also be profound in semi-arid regions, which is something that we did not consider in this study. It was evident from this study that knowledge of seasonal lake variations is critical for understanding the dynamics of lakes, especially in regions that are sensitive to climate change. It is fortunate that Sentinel-2 was launched in 2015, allowing for a greater frequency of Earth observations at a medium spatial resolution, making it possible to monitor intra-annual lake variations in upcoming studies.

4. Conclusions

This study demonstrates the capability of machine learning techniques for estimating two surface lake dynamics between 1999 and 2021. The random forest supervised classification algorithm is the ML technique that was used for the optimized feature classification,

focusing on the lakes of interest within the watersheds. The study collected seven features (water, urban, forest, grassland, bare earth, cropland, and shadow) from the mean annual Landsat Image between 1999 and 2021, using GEE tool kits over the two watersheds. The target of the classification algorithm was to delineate the dynamics of the GSL and LC between 1999 and 2021, with less focus on other non-water features. The study used over 9200 images that met the image quality characteristics for the classification purpose of the study period. A total of 10,189 remotely collected feature samples were incorporated into the GEE to perform the classification, and divided into training sets (70%) and testing sets (30%). The many feature samples ensured a high confidence in the classification algorithm.

The study's average accuracy assessment showed a high confidence in the ML feature classification algorithm that was used in the study for both watersheds. For the GSL, an overall accuracy of 0.99 and a kappa overall accuracy of 0.977 revealed a high confidence in the classification scheme. Compared to the accuracy assessment from the GSL classification algorithm, LC estimated a lower overall accuracy of 0.981 and a kappa overall accuracy of 0.977. This low accuracy for LC compared to the GSL was attributed to the difference in the number of images that were used for the classifications and the difference in the number of feature samples that were collected for the RF classification schemes. For instance, a data scarcity in the Lake Chad region led to a reduced number of available Landsat images and feature samples, with 744 fewer images and 437 fewer samples compared to the Great Salt Lake, where data scarcity was not a significant issue. This data limitation may have contributed to the lower accuracy that was observed in the LC classification results.

The assessment that was performed in this study showed that the GSL has been constantly declining in recent years, from nearly 9000 km² in 1999 to less than 3500 km² in 2021, losing almost half of its area extent. On the other hand, this paper has floored the uninvestigated assumptions that LC is shrinking at an alarming rate. The study has shown that LC is somewhat fluctuating more, with a slight decline in recent years, with an estimated 900 km² in 1999 and nearly 800 km² in 2021. The continual decline in the area extent of the GSL and the significant fluctuations in LC call for concern, as actions are required to salvage these lakes and conserve their environments, in order to meet the United Nations Sustainable Development Goal on water use efficiency and climate change mitigation.

The obtained results revealed that the dynamics of the lake systems in semi-arid regions can be attributed to an intense relationship between humans and the hydro-climatological conditions. In both study areas, the results of the study revealed significant impacts of climatological variability on the water surface area. The correlation of the lake area with the evapotranspiration and precipitation provided meaningful information regarding the predictability of the changes in the water surface area. The variation in the quantified lake extent can be explained by the temporal evolution of precipitation and evapotranspiration. According to the obtained results, remote sensing data, combined with machine learning models, can be employed to implement water regulation policies and make managerial decisions to understand temporal variability, not only for the GSL and LC, but also for other semi-arid lakes.

The findings of this study underscore the importance of accounting for morphological complexity, local geological and geomorphological influences, human interventions, seasonal and inter-annual variability, data quality and resolution, and the presence of aquatic vegetation and sedimentation when assessing lake dynamics. It is possible to gain a more comprehensive understanding of the relationship between the lake area and depth by incorporating these factors and utilizing high-quality data, in conjunction with other pertinent variables. Thus, lake dynamics can be more accurately evaluated. Moreover, it is imperative to further examine the seasonal variability of lake dynamics, keeping in mind that the temporal variation of hydrological processes, such as precipitation, evaporation, and runoff, can influence the lake depth and area.

By incorporating additional remote sensing observations from future environmental satellite missions, such as the Surface Water and Ocean Topography (SWOT) mission, this

study will serve as the foundation for future studies that incorporate additional remote sensing data, in order to estimate the changes in lake area and water volume in relation to river flow. Future research will attempt an implementation using a finer precipitation input from the Global Precipitation Measurement (GPM-IMERG).

Author Contributions: Conceptualization, research methodology, data curation, original manuscript composition, reviewing and editing and validation of data, K.E.; conceptualization, helped in research methodology, interpretation, analysis of data, original manuscript composition, reviewing and editing, M.A.; helped in research methodology, helped in editing of the manuscript, and provided suggestions S.I.; data curation, E.M.; data curation, P.E. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: Data acquisition scripts and analyses are available in: <https://code.earthengine.google.com/2a94453df3ca91cbf9047912a2ded20f> (accessed on 20 March 2023); <https://code.earthengine.google.com/3e42f50a3e1d64338977e88cd74d2078> (accessed on 20 March 2023); Area Calculator Script are available in: <https://code.earthengine.google.com/0e5188ab50a096837eae59546ce5f209> (accessed on 20 March 2023).

Acknowledgments: The authors would like to express their sincere gratitude to the FAIR Cyber Training (FACT) Fellowship for Climate and Water for providing technical support during this research. The authors would also like to acknowledge the valuable feedback and support from their FACT mentors. This research was carried out as part of the FACT Fellowship program supported by the US National Science Foundation (NSF) Grant No. 1829764. The authors would like to thank the FACT program and the NSF for their support.

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

References

1. Jones, J. *Water Sustainability: A Global Perspective*; Routledge: London, UK, 2010. [CrossRef]
2. Cornejo, P.K.; Becker, J.; Pagilla, K.; Mo, W.; Zhang, Q.; Mihelcic, J.R.; Chandran, K.; Sturm, B.; Yeh, D.; Rosso, D. Sustainability metrics for assessing water resource recovery facilities of the future. *Water Environ. Res.* **2019**, *91*, 45–53. [CrossRef] [PubMed]
3. Allen, G.H.; Pavelsky, T.M. Global extent of rivers and streams. *Science* **2018**, *361*, 585–588. [CrossRef] [PubMed]
4. Hornberger, G.M.; Wiberg, P.L.; Raffensperger, J.P.; D’Odorico, P. *Elements of Physical Hydrology*, 2nd ed.; Johns Hopkins University Press: Baltimore, MD, USA, 2014; p. 378.
5. Roberts, N.; Taieb, M.; Barker, P.; Damnati, B.; Icole, M.; Williamson, D. Timing of the Younger Dryas event in East Africa from lake-level changes. *Nature* **1993**, *366*, 146–148. [CrossRef]
6. Huang, C.; Chen, Y.; Zhang, S.; Wu, J. Detecting, Extracting, and Monitoring Surface Water From Space Using Optical Sensors: A Review. *Rev. Geophys.* **2018**, *56*, 333–360. [CrossRef]
7. Everard, M. Meeting global drinking water needs. *Nat. Sustain.* **2019**, *2*, 360–361. [CrossRef]
8. Horritt, M.; Mason, D.; Cobby, D.; Davenport, I.; Bates, P. Waterline mapping in flooded vegetation from airborne SAR imagery. *Remote Sens. Environ.* **2003**, *85*, 271–281. [CrossRef]
9. Pekel, J.-F.; Cottam, A.; Gorelick, N.; Belward, A.S. High-resolution mapping of global surface water and its long-term changes. *Nature* **2016**, *540*, 418–422. [CrossRef]
10. Wehbe, Y.; Temimi, M. A Remote Sensing-Based Assessment of Water Resources in the Arabian Peninsula. *Remote Sens.* **2021**, *13*, 247. [CrossRef]
11. Abdelkader, M.; Temimi, M.; Colliander, A.; Cosh, M.H.; Kelly, V.R.; Lakhankar, T.; Fares, A. Assessing the Spatiotemporal Variability of SMAP Soil Moisture Accuracy in a Deciduous Forest Region. *Remote Sens.* **2022**, *14*, 3329. [CrossRef]
12. Gebrehiwot, K.A.; Bedie, A.F.; Gebrewahid, M.G.; Hishe, B.K. Analysis of Surface Area Fluctuation of the Haramaya Lake using Remote Sensing Data. *Momona Ethiop. J. Sci.* **2019**, *11*, 140. [CrossRef]
13. Liu, Y.; Yue, H. Estimating the fluctuation of Lake Hulun, China, during 1975–2015 from satellite altimetry data. *Environ. Monit. Assess.* **2017**, *189*, 630. [CrossRef] [PubMed]
14. Pham-Duc, B.; Sylvestre, F.; Papa, F.; Frappart, F.; Bouchez, C.; Crétaux, J.-F. The Lake Chad hydrology under current climate change. *Sci. Rep.* **2020**, *10*, 1–10. [CrossRef]
15. Singh, A.; Seitz, F.; Eicker, A.; Güntner, A. Water Budget Analysis within the Surrounding of Prominent Lakes and Reservoirs from Multi-Sensor Earth Observation Data and Hydrological Models: Case Studies of the Aral Sea and Lake Mead. *Remote Sens.* **2016**, *8*, 953. [CrossRef]

16. Wurtsbaugh, W.; Miller, C.; Null, S.; Wilcock, P.; Hahnenberger, M.; Howe, F. Impacts of Water Development on Great Salt Lake and the Wasatch Front. 2016, p. 9. Available online: https://digitalcommons.usu.edu/wats_facpub/875 (accessed on 20 March 2023). [CrossRef]
17. Wurtsbaugh, W.A.; Leavitt, P.R.; Moser, K.A. Effects of a century of mining and industrial production on metal contamination of a model saline ecosystem, Great Salt Lake, Utah. *Environ. Pollut.* **2020**, *266*, 115072. [CrossRef] [PubMed]
18. Zhan, S.; Song, C.; Wang, J.; Sheng, Y.; Quan, J. A Global Assessment of Terrestrial Evapotranspiration Increase Due to Surface Water Area Change. *Earth's Future* **2019**, *7*, 266–282. [CrossRef] [PubMed]
19. Du, Y.; Xue, H.-P.; Wu, S.-J.; Ling, F.; Xiao, F.; Wei, X.-H. Lake area changes in the middle Yangtze region of China over the 20th century. *J. Environ. Manag.* **2011**, *92*, 1248–1255. [CrossRef]
20. Zhao, G.; Li, Y.; Zhou, L.; Gao, H. Evaporative water loss of 1.42 million global lakes. *Nat. Commun.* **2022**, *13*, 1–10. [CrossRef]
21. Maihemuti, B.; Aishan, T.; Simayi, Z.; Alifujiang, Y.; Yang, S. Temporal Scaling of Water Level Fluctuations in Shallow Lakes and Its Impacts on the Lake Eco-Environments. *Sustainability* **2020**, *12*, 3541. [CrossRef]
22. Chen, J.; Duan, Z. Monitoring Spatial-Temporal Variations of Lake Level in Western China Using ICESat-1 and CryoSat-2 Satellite Altimetry. *Remote Sens.* **2022**, *14*, 5709. [CrossRef]
23. Chen, J.; Liao, J.; Lou, Y.; Ma, S.; Shen, G.; Zhang, L. High-resolution datasets for lake level changes in the Qinghai-Tibetan Plateau from 2002 to 2021 using multi-altimeter data. *Earth Syst. Sci. Data Discuss.* **2022**, 1–18. [CrossRef]
24. Deus, D.; Gloaguen, R. Remote Sensing Analysis of Lake Dynamics in Semi-Arid Regions: Implication for Water Resource Management. Lake Manyara, East African Rift, Northern Tanzania. *Water* **2013**, *5*, 698. [CrossRef]
25. Cooley, S.W.; Smith, L.C.; Ryan, J.C.; Pitcher, L.H.; Pavelsky, T.M. Arctic-Boreal Lake Dynamics Revealed Using CubeSat Imagery. *Geophys. Res. Lett.* **2019**, *46*, 2111–2120. [CrossRef]
26. Acharya, T.D.; Subedi, A.; Lee, D.H. Evaluation of Machine Learning Algorithms for Surface Water Extraction in a Landsat 8 Scene of Nepal. *Sensors* **2019**, *19*, 2769. [CrossRef] [PubMed]
27. Dirscherl, M.; Dietz, A.J.; Kneisel, C.; Kuenzer, C. Automated Mapping of Antarctic Supraglacial Lakes Using a Machine Learning Approach. *Remote Sens.* **2020**, *12*, 1203. [CrossRef]
28. Belgiu, M.; Drăguț, L. Random forest in remote sensing: A review of applications and future directions. *ISPRS J. Photogramm. Remote Sens.* **2016**, *114*, 24–31. [CrossRef]
29. Berhane, T.M.; Lane, C.R.; Wu, Q.; Autrey, B.C.; Anenkhonov, O.A.; Chepinoga, V.V.; Liu, H. Decision-Tree, Rule-Based, and Random Forest Classification of High-Resolution Multispectral Imagery for Wetland Mapping and Inventory. *Remote Sens.* **2018**, *10*, 580. [CrossRef]
30. Son, N.-T.; Chen, C.-F.; Chen, C.-R.; Minh, V.-Q. Assessment of Sentinel-1A data for rice crop classification using random forests and support vector machines. *Geocarto Int.* **2018**, *33*, 587–601. [CrossRef]
31. YCC Team. Utah's Great Salt Lake Is Shrinking, Worsening Risk of Dust Storms. Yale Climate Connections, 8 October 2021. Available online: <http://yaleclimateconnections.org/2021/10/utahs-great-salt-lake-is-shrinking-worsening-risk-of-dust-storms/> (accessed on 3 March 2023).
32. LaVere, B.M. Utah Lake: A Few Considerations. Nov. 2017. Available online: <http://wfwqc.org/wp-content/uploads/2017/11/UL-info-Nov-2017> (accessed on 26 March 2023).
33. Buma, W.G.; Lee, S.-I.; Seo, J.Y. Recent Surface Water Extent of Lake Chad from Multispectral Sensors and GRACE. *Sensors* **2018**, *18*, 2082. [CrossRef]
34. Gritzner, J.A. Lake Chad. Encyclopedia Britannica, 19 December 2019. Available online: <https://www.britannica.com/place/Lake-Chad> (accessed on 26 March 2023).
35. Abbott, M.B.; Anderson, L. Lake-Level Fluctuations. In *Encyclopedia of Paleoclimatology and Ancient Environments*; Gornitz, V., Ed.; Springer: Dordrecht, The Netherlands, 2009; pp. 489–492. [CrossRef]
36. World Meteorological Organization (WMO); Lake Chad Basin Commission (LCBC). *Lake Chad-HYCOS, A Component of the World Hydrological Cycle Observing System (WHYCOS)*; WMO: Geneva, Switzerland, 2015.
37. Tulbure, M.G.; Broich, M.; Stehman, S.V.; Kommareddy, A. Surface water extent dynamics from three decades of seasonally continuous Landsat time series at subcontinental scale in a semi-arid region. *Remote Sens. Environ.* **2016**, *178*, 142–157. [CrossRef]
38. Funk, C.; Peterson, P.; Landsfeld, M.; Pedreros, D.; Verdin, J.; Shukla, S.; Husak, G.; Rowland, J.; Harrison, L.; Hoell, A.; et al. The climate hazards infrared precipitation with stations—A new environmental record for monitoring extremes. *Sci. Data* **2015**, *2*, 150066. [CrossRef]
39. Metadata for the Rapid Forcing Retrieval (RFR) Web Tool. 2022. Available online: <http://www.hydroshare.org/resource/ad37a792a6144c9a1d45e05621e4230> (accessed on 20 March 2023).
40. FAO. Terra Net Evapotranspiration 8-Day Global 500m (MOD16A2.006). Food and Agricultural Organization of the United Nations, April 2022. Available online: https://lpdaac.usgs.gov/documents/494/MOD16_User_Guide_V6.pdf (accessed on 26 March 2023).
41. Worqlul, A.W.; Ayana, E.K.; Dile, Y.T.; Moges, M.A.; Gitaw, M.G.; Tegegne, G.; Kibret, S. Spatiotemporal Dynamics and Environmental Controlling Factors of the Lake Tana Water Hyacinth in Ethiopia. *Remote Sens.* **2020**, *12*, 2706. [CrossRef]
42. Birkett, C.; Reynolds, C.; Beckley, B.; Doorn, B. From Research to Operations: The USDA Global Reservoir and Lake Monitor. In *Satellite Altimetry for Geodesy, Geophysics and Oceanography*; Hwang, C., Cheng, Y., Shum, C.K., Eds.; Springer: Berlin/Heidelberg, Germany, 2011. [CrossRef]

43. Khatami, R.; Mountrakis, G.; Stehman, S.V. A meta-analysis of remote sensing research on supervised pixel-based land-cover image classification processes: General guidelines for practitioners and future research. *Remote Sens. Environ.* **2016**, *177*, 89–100. [[CrossRef](#)]
44. Kratzert, F.; Klotz, D.; Shalev, G.; Klambauer, G.; Hochreiter, S.; Nearing, G. Towards learning universal, regional, and local hydrological behaviors via machine learning applied to large-sample datasets. *Hydrol. Earth Syst. Sci.* **2019**, *23*, 5089–5110. [[CrossRef](#)]
45. Sruthi, E.R. Random Forest | Introduction to Random Forest Algorithm. Analytics Vidhya, June 2021. Available online: <https://www.analyticsvidhya.com/blog/2021/06/understanding-random-forest/> (accessed on 16 July 2022).
46. 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](#)]
47. Mutanga, O.; Kumar, L. Google Earth Engine Applications. *Remote Sens.* **2019**, *11*, 591. [[CrossRef](#)]
48. Wang, Y.; Ma, J.; Xiao, X.; Wang, X.; Dai, S.; Zhao, B. Long-Term Dynamic of Poyang Lake Surface Water: A Mapping Work Based on the Google Earth Engine Cloud Platform. *Remote Sens.* **2019**, *11*, 313. [[CrossRef](#)]
49. Druce, D.; Tong, X.; Lei, X.; Guo, T.; Kittel, C.; Grogan, K.; Tottrup, C. An Optical and SAR Based Fusion Approach for Mapping Surface Water Dynamics over Mainland China. *Remote Sens.* **2021**, *13*, 1663. [[CrossRef](#)]
50. Swanson, D.K. Thermokarst and precipitation drive changes in the area of lakes and ponds in the National Parks of northwestern Alaska, 1984–2018. *Arct. Antarct. Alp. Res.* **2019**, *51*, 265–279. [[CrossRef](#)]

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