

Article

Lake Volume Data Analyses: A Deep Look into the Shrinking and Expansion Patterns of Lakes Azuei and Enriquillo, Hispaniola

Mahrokh Moknatian ¹ and Michael Piasecki ^{2,*}

¹ Institute for Sustainable Cities, Hunter College, New York, NY 10065, USA; mmoknat00@citymail.cuny.edu

² Civil Engineering Department, City College of New York, New York, NY 10031, USA

* Correspondence: michael.piasecki@gmail.com or mpiasecki@ccny.cuny.edu;

Tel.: +1-610-564-8184

Received: 5 November 2019; Accepted: 20 December 2019; Published: 24 December 2019

Abstract: This paper presents the development of an evenly spaced volume time series for Lakes Azuei and Enriquillo both located on the Caribbean island of Hispaniola. The time series is derived from an unevenly spaced Landsat imagery data set which is then exposed to several imputation methods to construct the gap filled uniformly-spaced time series so it can be subjected to statistical analyses methods. The volume time series features both gradual and sudden changes the latter of which is attributed to North Atlantic cyclone activity. Relevant cyclone activity is defined as an event passing within 80 km and having regional monthly rainfall averages higher than a threshold value of 87 mm causing discontinuities in the lake responses. Discontinuities are accounted for in the imputation algorithm by dividing the time series into two sub-sections: Before/after the event. Using leave-p-out cross-validation and computing the NRMSE index the Stineman interpolation proves to be the best algorithm among 15 different imputation alternatives that were tested. The final time series features 16-day intervals which is subsequently resampled into one with monthly time steps. Data analyses of the monthly volume change time series show Lake Enriquillo's seasonal periodicity in its behavior and also its sensitivity due to the occurrence of storm events. Response times feature a growth pattern lasting for one to two years after an extreme event, followed by a shrinking pattern lasting 5–7 years returning the lake to its original state. While both lakes show a remarkable long term increase in size starting in 2005, Lake Azuei is different in that it is much less sensitive to storm events and instead shows a stronger response to just changing seasonal rainfall patterns.

Keywords: lake dynamics; Hispaniola; time series analysis; imputation; trend test; Mann-Kendall test; linear regression model; change point; Pettitt test; wavelet transform

1. Introduction

Lakes Azuei (LA) and Enriquillo (LE) located adjacent to each other in Haiti, and the Dominican Republic (DR), respectively, experienced an unprecedented expansion starting in 2005 and lasting until 2014 (Figure 1). This 9-year growth in lake expanse has had dramatic impacts on the surrounding areas [1]. Due to the very shallow sections at the West and East sides of both lakes, large swaths of arable land were inundated rendering them useless due to the high salt content, especially around LE. In addition, the small village of Boca de Cachon, DR, located at the western end of LE, had to be abandoned and resettled to higher grounds in 2014. Also, the key border crossing in Jimani, along the main trade highway between the two countries, was threatened to be completely

inundated; only extensive work to raise the road along the eastern stretch of LA kept the highway open for trade activity [2]. As a consequence, besides responding with immediate short term activities to address the most pressing issues, both countries have expressed interest in understanding the causes for these surface level raises, so they could devise medium and long term response plans to help alleviate the impacts for the local population.



Figure 1. Study Area: Lake Azuei and Lake Enriquillo extent in 2003 and 2014.

There are several approaches one could take to address the task to better understand the causes of the lake changes and also to assess future impacts and consequences. One alternative is to develop detailed numerical representations of the lakes' watersheds thus creating tools not only suitable to construct responses but also to explore what-if scenarios [3]. Another alternative would be the "paleo-approach" in which one goes back in time and analyzes long term time series of hydroclimatic variables vis-à-vis lake characteristics such as surface area and volume [4]. Yet, another alternative would be to try to use purely anecdotal evidence, i.e., a history of stories and observations passed down through generations, or to use surrogate indicators such as elements of fauna and flora to extract data that would illuminate the causes for the growth [5]. These approaches while each having its merits, are not practical in this instance however (even though they do on occasions yield supplemental information), because they require either long term hydro climatological and lake data, an abundance of detailed local data on soil, land cover and land use, subsurface hydrological characteristics, and/or ecological observations data, all of which do not exist in sufficient quantity or quality to be useful.

The application of statistical techniques and machine learning methods, on the other hand, have been very well established in many hydro-climatic studies, whether it is for the purpose of evaluating hydrologic system characteristics or modeling hydro-climate systems. The identification of key system characteristics, such as stationarity, non-stationarity, linearity, nonlinearity, periodicity, non-periodicity, complexity, correlation, and trend, provides vital information for adequately modeling geophysical systems [6]. In 2017, Medwedeff and Roe [7] studied the complete global dataset of glacier mass-balance records to analyze their characteristics. In this study, they de-trended the time series from the natural, interannual variability using least-squares regression, subsequently evaluated the normality and the presence of persistence in the time series, and then examined the existence of correlation between winter and summer records. In another study, continuous wavelet filtering, multi-resolution decomposition based on the maximal overlap discrete wavelet transform, auto-regressive-based decomposition, singular spectrum analysis, and empirical mode decomposition were all applied to the Baltic sea level time series to investigate the existence of long-term seasonal cycle changes [8]. Khelifa et al., [9] evaluated the global sea level anomaly time series using the singular spectrum analysis and the wavelet multiresolution analysis to look for signs of seasonality and trend. Moon and Lall (1995) [10] showed evidence of quasi-periodic interannual and interdecadal variability in the Great Salt Lake (GSL) volume fluctuations by performing Singular Spectrum Analysis on the lake's monthly volume changes. Results suggested variations in coherence

of GSL to the regional atmospheric variation and also Northern Hemisphere sea level pressure. In 2007 Moon et al., [11] built on the aforementioned results and developed a multivariate, non-parametric model using three atmospheric circulation indices of the Southern oscillation index (SOI), the pacific/North America (PNA) climatic index, and the central North Pacific (CNP) climatic index for short-term forecasting of the GSL monthly volume.

The main underlying data source for the work presented here are the lake Volume Time Series (VTS) of the two lakes that are derived from Lake Surface area (obtained from Landsat imagery) and bathymetric data [12] to analyze the lakes' patterns (for more on the creation of the volume data sets see [12]). This dataset reaches sufficiently far back into the past (1972) to yield a reasonable "long" (47 years) time series, which in turn can be used for both watershed budget calculations and time series analysis. The analyses considered for this study include the detection of sudden shift variations (Change Point Detection), cyclical or periodic variations (Wavelet Transforms), and steady increase/decrease (Trend Analysis), which are all meaningful aspects of the non-stationary characteristics of a data series as they help to identify internal or external stimuli [6]. None of these characteristics, however, are easily computed from remote sensing data which are not consistent and feature numerous gaps due to overpass scheduling, sensor failure, or other issues such as cloud and cloud shadow obstructions. Hence, the first task is to develop an evenly spaced VTS with meaningful intervals for the analyses scope, before carrying out the time series analyses on the Storage/Volume Change Time Series (VCTS) that helps to better understand the lakes' behavior and responses to climate and/or human-induced forcing. The work presented here is necessarily tied to the individual setting of these two lakes, an observation that is common when researching lakes, because lakes tend to be very unique in their characteristics. Yet, the authors hope that the presented sequence of steps is sufficiently generic in nature to aid other scientists carrying out similar work especially when only very limited data sources are available.

2. Materials and Methods

2.1. Data Acquisition and Observational Volume Time Series

Landsat images are available about every 16 days and have been produced since 1972 (Landsat-1) using an evolution of different satellites and instrument loads throughout the years until this day (currently the Landsat-8 mission, with Landsat-9 slated to be launched in 2020). Our study area falls into the Worldwide Reference System (WRS) of path 8 and row 47 of Landsat scenes yielding a set of more than 400 available images from all Landsat sensors. While the majority of time intervals between images are 16 days, there are several gaps of various sizes the largest of which is about 4 years (1974 to 1978 and 1992 to 1996). In very few cases, the time spacing is actually shorter, i.e., 8 days, due to the overlapping schedule of two different satellites operating simultaneously between 2000 and 2001. Of all archived images 252 and 297 images for LE and LA, respectively [13], proved suitable for constructing the lakes' surface area time series. Subsequent derivation of the lakes' observational VTS was achieved through combining the bathymetry of LE and LA and the lake surface area values; for more on this the reader is referred to [12–16]. Figure 2 presents the observational VTS' for both lakes for the period of 1972–2017.

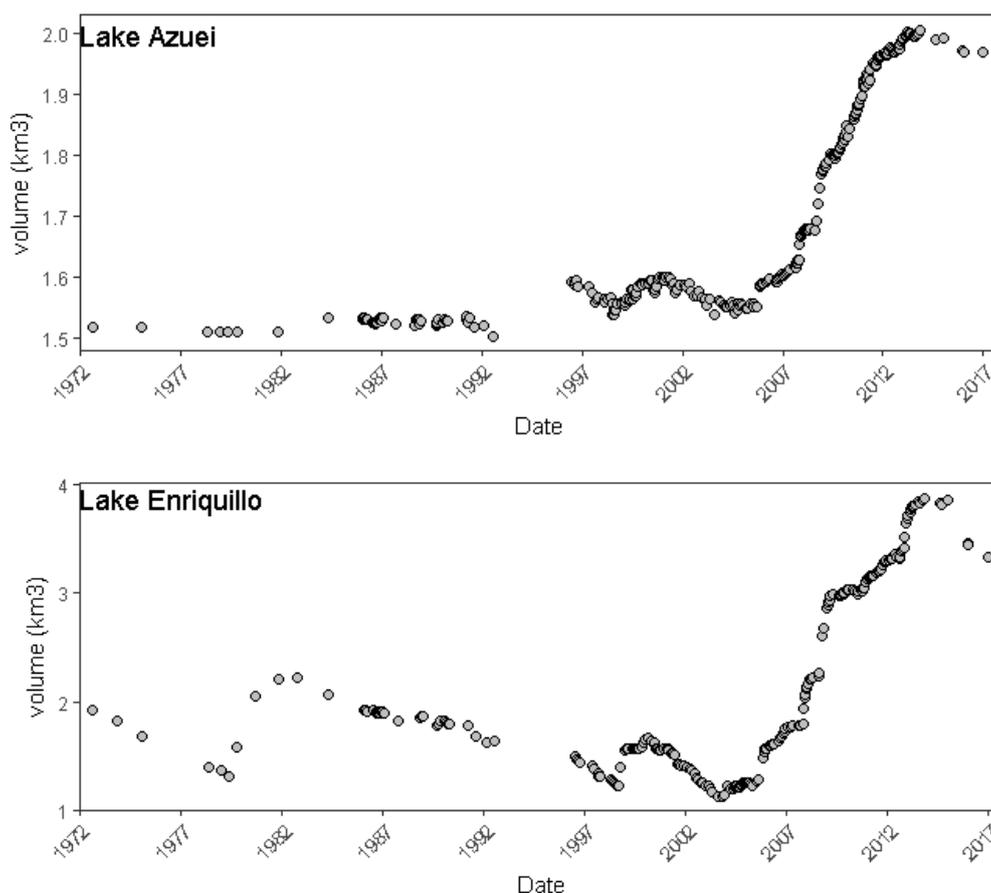


Figure 2. Lakes Azuei's and Enriquillo's volume time series.

The lake volume data naturally exhibit the same interval and gap patterns as the original Landsat imagery time series, i.e., between 8 days and 4 years. The volume of both lakes shows smooth oscillations throughout the temporal domain. For example, LE exhibits a decreasing trend for the years between 1984 to 1998, followed by a slight increase up to 2000; a slight decrease until 2003, before encountering a dramatic increase up to 2014 after which it seemed to have been stopped and at the time of this writing shows a decrease again. Note that both lakes, in general, show similar trends but on smaller time scales exhibit both synchronous and asynchronous behavior. Both lakes also show patterns of sudden changes. For example, in September 1998, LE's volume suddenly leaped in a matter of just a few days as a result of Hurricane activity (which was then recorded a few later by the Landsat satellite).

These types of events can be observed several times throughout the period between 1972–2017, suggesting the existence of external forcing triggering these responses. In normal situations (meaning no extreme events present such as hurricanes moving through), the volume fluctuation of the lakes stay within smaller ranges and the increase/decrease pattern of any of the two lakes spreads out over longer time intervals. These changes in parameter characteristics thus needed to be identified first before any modification was applied to the raw dataset making sure that the evenly spaced generated time series possessed the same parameter characteristics as the unevenly spaced (raw) observations.

2.2. Missing Data Imputation

Missing data issues for time series abound in many fields of earth science and many studies have addressed developing techniques to fill the data gaps [17–24]. Since there were no other attributes available for estimation of missing values in the lakes' observational VTS, focus was placed on Landsat-derived lake extent data having time intervals of about 16 days. These were characterized as

“missing at random” (MAR) [25] prompting the use of univariate methods to generate complete VTS. Univariate techniques include methods such as deletion and substitution methods as well as interpolation, smoothing, and seasonal decompositions methods [26]. The consensus is that simple univariate imputations algorithms (such as deletion or substitutions) yield inferior results while more sophisticated approaches (such as interpolations and smoothing), using Kalman smoothing interpolations, and seasonal decompositions are supposed to yield better results [26]. Due to the existence of a general trend in our target time series, deletion and replacement methods were thus considered unsuitable. Also, inspection of the time series showed that in general the changes in volume for both lakes were fairly smooth (monthly time scales), based upon which the decision was made to apply interpolation, smoothing, and seasonal decompositions approaches to construct a corresponding monthly VTS which is well matched by typical time increments (daily, weekly or monthly) available in atmospheric data sets. Note, that extreme events such as the occurrence of Hurricanes at times yielded a very fast response with time scales of just a few days prompting the need to address these rapid change periods separately. Toward this end, the lakes’ VTS had to be first constructed featuring an equal and small-time step (16 days) requiring a strategy to fill in missing values after which the time series was resampled to feature a monthly time step.

Univariate methods use time series’ characteristics to fill in the missing values [27] and thus need to be first computed. This mainly concerns the creation of the lakes’ storage/Volume Change Data Set (VCDS) which are representations of their water balance rate of change. A second step then examines the suitability of candidate imputation methods and also how the set of monthly VTS is created. Note that the imputation procedure is applied to the observational VTS to produce the monthly imputed VTS.

2.2.1. Alternative Observational ΔV Datasets and the Characteristics of Sudden Changes

The rate-of-change time series is derived by simply using the difference of two consecutive variable values. While this is readily done for an evenly spaced time series with $\Delta T = \text{constant}$, irregular time stepping requires the use of two consecutive points throughout the temporal domain. The latter is much more challenging because of the uneven spacing or missing data over prolonged periods of time. Short of interpolating in between observed data points so that the equidistant data points can be calculated, any choice of a constant ΔT therefore runs the risk of omitting crucial information because data points are disqualified for not being “on the mark”. To preserve most of the data points in the lakes’ observational VTS, it was decided to construct alternative datasets with ΔT equal to multiples of 16-day intervals, the highest frequency available. The resulting set of different ΔT -datasets provided insight into the lakes’ characteristics, some of which might have been present in one dataset but missing in the other. Examination of the alternative datasets (16-day, 32-day, 48-day, 64-day, 80-day, 96-day, 112-day) and their associated changes show which one of them would capture the most observational changes, i.e., their ability to represent all the variable outliers. Preliminary statistical analysis of the datasets showed that their distributions follow a bell shape with only slight skewness. LE values seemed to skew to the right, indicating the presence of positive outliers in the data (Figure 3). In the case of LA, the outliers were distributed on both sides however with more positive values than negative ones suggesting the importance of the positive outlier presence in the datasets.

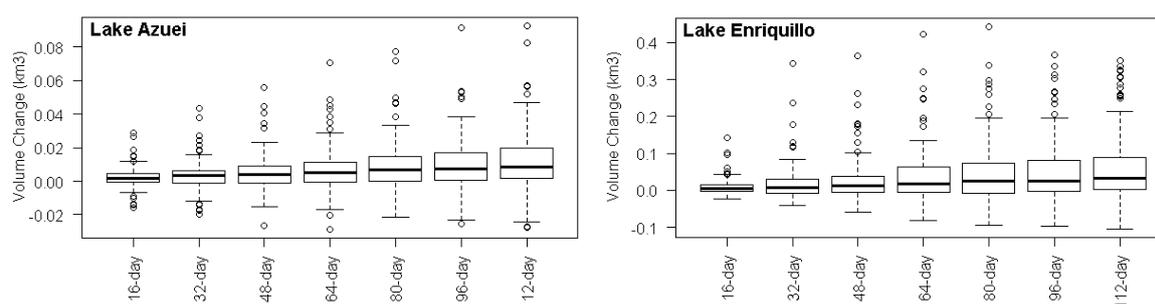


Figure 3. Datasets of volume change boxplots.

Identifying outliers and associating them with corresponding dates in the temporal domain for all datasets showed that 1979, 1998, 2005, 2007, 2008, 2009, 2012, and 2013 were the years in which the positive outliers of LE had occurred. For LA, the positive outliers were related to the years of 1988, 1999, 2000, 2005, 2007, 2008, 2009, 2010 and 2011, while the negative ones happened in 1991, 1997, 1998, 1999, 2000, 2001, 2003, and 2010. Positive outliers of the LE dataset showed that extreme anomalies always caused lake growth and that no phenomenon ever caused the lake to shrink beyond its normal fluctuations. Conversely, LA was much more affected by both gaining and losing water beyond its normal variations.

In order to differentiate the outliers to see how they changed the lake regime, focus was placed on the analysis of volume change values right after the outlier. A regime shift was observed when a positive outlier was followed by another positive outlier. Conversely, when the targeted outlier was followed by a value in the normal range, no regime shift was observed in the lakes' behavior. In the case of negative outliers no such pattern was observed, the following volume change values were always in the normal range (the normal range is between the high and low end bars in the box plot).

All outliers of LE were followed by positive changes that were higher than the normal range, except for the years of 2009 and 2013. For LA, only the outliers occurring in 2007 and 2008 were followed by other positive outliers. The need to find additional underlying causes of outlier occurrences, prompted further examination of the watersheds' physics to look for a trigger. Outliers can be the sign of errors in measurements or they could also be the response signal to an actual event [28–31]. Precipitation is typically the main contributor to closed-basin lakes both as direct deposition and run-off collection from the watershed. They occur at different time scales however, i.e., direct rainfall can cause a sudden change to the lake storage while runoff coming from the watershed would be built up gradually over time as it pours into the lake in the aftermath of a storm event. Hence, severe storm events have the ability to cause rapid lake responses, followed by slower runoff volumes being added. The only balancing process is lake surface evaporation which, however, takes place at even larger time scales. Hence, the combination of these processes, sudden strong rainfall, moderately fast runoff, and then slow evaporation, will yield sudden increases in water level which takes months to several years to return to its original equilibrium state [32–34].

The main source of extreme rainfall in our study area is the occurrence of either tropical storms or hurricanes. Hispaniola is located within the typical corridor of North Atlantic cyclones, many of which impact the Caribbean islands during the months of late summer and early fall. The impact varies however and only cyclones which pass directly over or in close proximity of the lakes' watershed tend to register a significant amount of precipitation which was identified to be within a 50-mile distance of the lakes' watersheds.

The anomalies for both lakes suggested a strong correlation between cyclone activities in the years of 1979 (Tropical Storm Claudette and Hurricane David), 1998 (Hurricane George), 2005 (Tropical Storm Alpha), 2007 (Tropical Storm Noel), 2008 (Tropical Storms Fay and Gustav), and 2012 (Tropical Storm Isaac); with each cyclone contributing to monthly rainfall rates higher than 87.6 mm over the lakes and their watersheds. Comparison with the observational VTS indicated that significant changes in lake volume happened within a window of fewer than two weeks following each of the individual cyclones thus confirming their impact and the corresponding occurrences of outliers. These "cyclone singularities" in the VTS were used to split the time series into before- and after-sections so the imputation algorithms would not be "distracted" by the singularity. Figure 4 shows an example of how to integrate a cyclone's effect in the interpolation process. In this figure, the sub-series is split into two smaller parts: One before and one after hurricane George (1998), so the general characteristics of the lake stay disconnected from this one-time extreme event and the behavioral characteristic of the lake after the storm does not affect its behavior before. The same procedure was considered for all other influential cyclones.

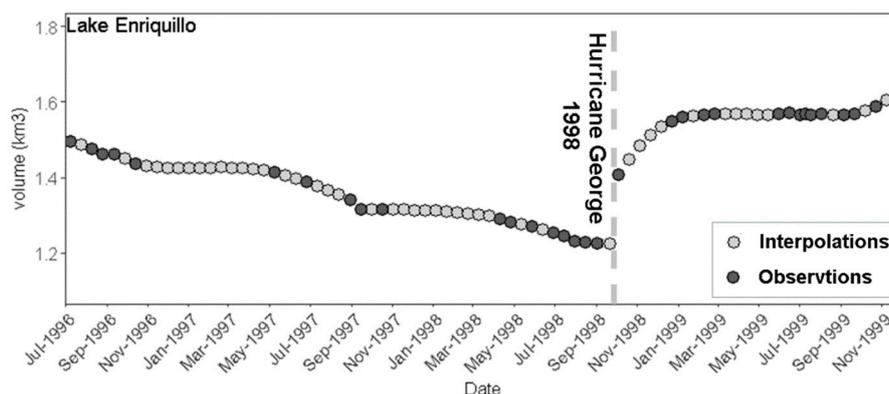


Figure 4. Lake Enriqueillo 1996–1999 observed and interpolated values as an example for incorporating a storm event in the interpolation process.

2.2.2. Evenly Spaced Time Series Construction

Due to the two 4-year gaps in lake volume data (1974 to 1978 and 1992 to 1996), the observational VTS was split into two separate time regions, i.e., before and after 1996. Since the 16-day interval featured the most data points, it was used as the reference interval. The level of missingness for both lakes was around 90% (time span 1972 to 1996), and 47% and 56% (1996 to 2014), for LA and LE, respectively. Due to the high level of missingness between 1972–1996 and also for 2014–2017 (only a few data points emerged for this time span) the focus centered on the available data between 1996–2014.

Since the time shift of the Landsat-5 TM and Landsat-7 ETM data products was only 8 days, any start date for a 16-days interval based time series would necessarily negate the common use of either all of the Landsat-5 TM or the Landsat-7 ETM data points (data points would alternate on a 8-day offset). To address this issue, one could either build an 8-day time series, which causes an increase in the number of missingness or split the time series into smaller parts, analyze them separately, and then merge the results. The second option was deemed more appropriate because it permitted the retention of all data from both Landsat satellites. For this purpose, the time series was split into three sections which permitted the inclusion of the 8-day shift. The first part of the time series (June 10, 1996, to December 26, 1999) was comprised of LT5 data with a 16-day interval, the second part (January 2000, to January 2001) was a mix of Landsat-5 TM and Landsat-7 ETM data with an 8-day interval, and the third part (February 22, 2001, to August 21, 2014) corresponded to the values derived from Landsat-7 ETM, again using a 16-day interval. Note that the observational VTS was later divided into more subsections based on the date of influential cyclones. Since the response times of the lakes to forcing was slow (monthly time scales) the imputed time series (8 and 16-days intervals) were resampled to yield a VTS with monthly intervals.

2.3. Time Series Analysis

While the previous section dealt with the creation of evenly spaced VTS with monthly time steps, this section addresses the actual time series analyses, including the investigations of periodicity, abrupt changes, and monotonic increasing or decreasing pattern. One of the widely used tools to investigate the periodic pattern of time series is the wavelet transform analysis [35]. In this method, called continuous wavelet transform (CWT), the time series is decomposed in the time and frequency domains identifying the dominant frequency as well as its temporal variation [35–37] and it has been frequently used to shed some light on the complex characteristics of hydro-climatic variables [8,38–43]. Another representation of a wavelet spectrum, called the Global Wavelet Power Spectrum (GWPS), can also be used to determine the dominating time scales within a time series, whereby the coefficients of a power spectrum for one scale are averaged over the length of the entire time series [36]. Both CWT and GWPS were considered suitable and thus chosen for the purpose of this study. Note that the influential outliers were first removed using Cook's Distance method [44–46] because the preliminary analysis showed their effect on the periodicity results derived from CWT. The GWPS,

on the other hand, seemed to remain unchanged and resistant to the presence of outliers. The statistical significance for CWT was estimated using Monte Carlo methods and the results of GWPS were compared to the null red-noise which is an adequate procedure for identifying significance when performing frequency analysis [36,47].

In addition to detecting periodicity of a time series which could be correlated to regional or global climate variability [48], detecting in-homogeneities and changes in time series are critically important, as they can reveal the role of any external or internal stimulus which has triggered a shift in a phenomenon [49]. A change point (CP) is defined as a probable point with the most/significant likelihood in time from where onward the statistical characteristics of the time series change. There are a number of CP detection methods based on parametric or non-parametric statistical tools, developed to detect such abrupt changes [50–53]. Among these methods, the Pettitt test is a non-parametric test that has been widely used in hydrological and climatological studies to detect a single CP in continuous time series. Its applications in hydrology encompass studies investigating the changes in groundwater, surface runoff, and river discharges due to the climate change or human activities [54–60]. Using this test, the Pettitt's statistics of $U_{t,n}$ and K_t [52] were calculated in addition to the associated probability to determine whether a CP existed in the time series.

As a last step a trend test, which was performed on the Lakes' VCTS, was executed. The detected trend in data exhibits a steadily gradual increase or decrease of the trend over time. Depending on the characteristics of the data, such as the existence of missing values, outliers, serial correlation, non-normality, censored data, CP, and periodicity, etc., the detection of a trend is challenging, and the analysis test may result in false detection or ignorance of the trend values [61]. Identifying these characters helps with the choice of the test to be used. There are two test type classifications: parametric methods and non-parametric methods. An example for a parametric test is linear regression which requires the data to be independent and normally distributed [62]. Some of the most commonly used non-parametric methods in hydrological, climatological, and meteorological fields are Mann-Kendall, Spearman's rho and Sen's Slope method [56,58–61,63–68], among which the Mann-Kendall method has been widely used to detect the significance of the trend in a time series. The Mann-Kendall trend test, which uses the rank of observations, is known to be less sensitive to the outliers and the distribution of the data, thus is suitable for hydrological data which usually features outliers and other less-desirable characteristics of time series [67]. Along with the Mann-Kendall test, a linear regression analysis was performed for comparison purposes and also to examine the significance of CP and periodicity along with the trend which also featured the introduction of Shift and Seasonality factors to the general form of linear regression. The influential outliers and the serial correlation of the VCTS were removed using Cook's distance and pre-whitening methods (TFPW) [69,70], prior to performing linear regression tests and the results were examined to see if they follow a Gaussian distribution. Since the Mann-Kendall test is resistant to outliers, the only factor contributing to this test was the presence of CPs. Therefore, the test was adopted to the two sub-periods of before and after CP separately. Note that all analyses were carried out on the rate-of-change time series (Volume Change Time Series, VCTS) rather than the monthly imputed VTS because a focal point is also to investigate the variation of the VCTS.

3. Results

3.1. Monthly Imputed Volume and Volume Change Time Series

Imputation was carried out using 15 different algorithms (as listed in Table 1) in order to identify the algorithm producing the best results. Best results here mean to introduce the least amount of bias [71], preserving the original characteristics, and achieving a high degree of precision [72]. Using the three temporal sections (1996–2001, 2000–2001, and 2001–2014) and applying them to LE and LA, most of the algorithms, except the random value sample method, seasonally decomposition, and seasonally split methods using random value sample, showed a high degree of correlation, i.e., 98%.

Defining the performance accuracy of imputation methods is challenging because there is a dearth of comparative datasets. Performance of an applied imputation can be validated, however, in

terms of outputs instead of using reference values [19]. Methods that use this approach include leave-one-out cross-validation, leave-p-out cross-validation, and k-fold cross-validation [73]. Using the leave-p-out cross-validation approach and defining P as the number of missingness in the original time series, the Normalized Root Mean Squared Error (NRMSE) was used to test the performance of the various imputation methods. The procedure involved the random removal of different sets of the observed data points resulting in a collection of about 1000 different data sets yielding a corresponding set of NRMSE values for each imputation method used. The performance could then be evaluated by computing the average NRMSE for each imputation method considered.

The results are summarized in Table 1. As can be seen, all NRMSE values are close except those that apply random values, which is not unexpected. Among all methods, the Stineman interpolation [74] yielded the smallest NRMSE values for all six time series sections making it the method of choice for further evaluation. This is being supported as the method has previously been shown to perform well in the presence of abrupt changes [74] and also when the density of data points in the temporal domain varies significantly [75]. Clearly, the Random Value Sample, Seasonally Decomposition by Random, and Seasonally Split by Random methods did not perform well because of the nature of the methods which assigns random values to the missing points. In this procedure the assigned value was sometimes very high or very low which introduced additional disruptions in the lakes' volume time series between abrupt change episodes. Note though, that some of the methods produced similar results, i.e., there is no one perfect standout method while all others performed badly.

Table 1. The Normalized Root Mean Squared Error (NRMSE) of imputation results.

Imputation method	NRMSE					
	Lake Azuei			Lake Enriquillo		
	1996–2000	2000–2001	2001–2014	1996–2000	2000–2001	2001–2014
Linear Interpolation	0.272	0.380	0.020	0.108	0.056	0.018
spline Interpolation	0.412	0.694	0.027	0.178	0.120	0.029
Stineman Interpolation	0.269	0.375	0.019	0.094	0.048	0.016
Kalman Smoothing using Structural Model	0.282	5.865	0.068	0.105	0.206	0.089
Kalman Smoothing using ARIMA State Space Representation	0.295	6.135	0.070	0.110	0.215	0.093
Simple Moving Average	0.327	0.475	0.031	0.191	0.137	0.034
Linear Weighted Moving Average	0.306	0.434	0.027	0.171	0.121	0.029
Exponential Weighted Moving Average	0.295	0.438	0.026	0.158	0.121	0.027
Random Value Sample	1.478	1.809	1.442	1.723	1.735	1.536
Seasonally Decomposition by Linear Interpolation	0.271	0.381	0.020	0.111	0.055	0.018
Seasonally Decomposition by Random	1.599	1.633	1.405	1.470	1.818	1.579
Seasonally Decomposition by Weighted Moving Average	0.297	0.429	0.026	0.156	0.122	0.027
Seasonally Split by Linear Interpolation	0.272	0.379	0.020	0.110	0.056	0.018
Seasonally Split by Random	1.739	1.698	1.426	1.659	1.733	1.399
Seasonally Split by Weighted Moving Average	0.293	0.433	0.025	0.158	0.127	0.027

Integrating the storm information, and applying the Stineman interpolation method, the monthly values were computed with the resulting VTS and the original observations as shown in Figure 5.

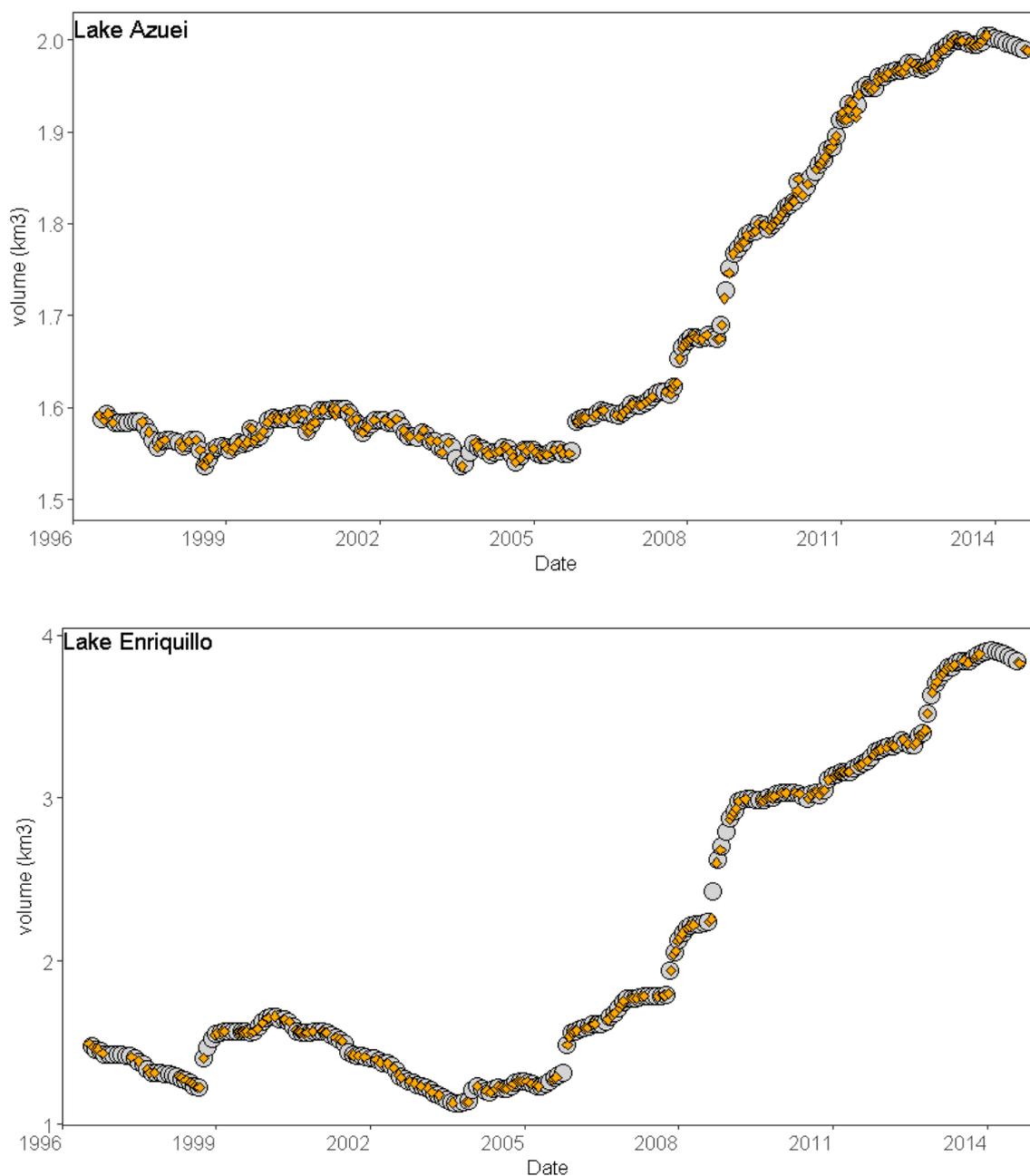


Figure 5. Lake Azuei’s and Lake Enriquillo’s monthly volume time series (grey points) and observational time series (orange diamonds).

In order to validate the monthly imputed VTS, its VCTS were constructed and compared with the observational 32-day VCDS. Note that the datasets produced from imputation corresponded to the period of 1996 to 2014, while the observations were distributed between 1972 and 2017. Although the time spans were not the same, the values of the datasets were related to the same lake characteristics, hence it was assumed that they possess the same statistical characteristics. Figure 6 shows the distribution of each VCDS, along with their boxplot and outlier position. Values of minimum, maximum, mean, median, and other statistical parameters for the datasets were calculated and shown in Table 2. The graphs for observations and imputed data values were reasonably similar as were the statistical comparison values. As a next step, it was important to establish if both data sets had the same distribution for which one could compute the significance level of their differences via a Bootstrap test.

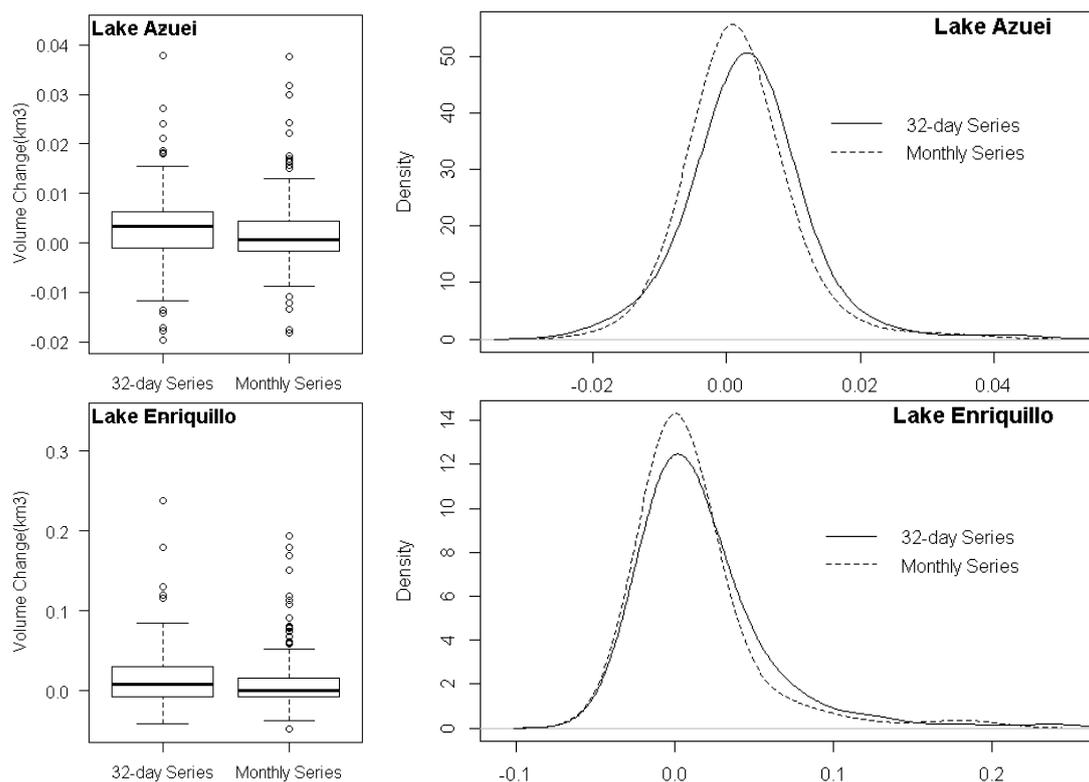


Figure 6. Comparison of the 32-day and monthly volume change datasets boxplot and distribution.

Table 2. Statistical variables of the lakes volume change datasets.

Lake	Series	No. Points	Min	1st Qu	Median	Mean	3rd Qu	Max	SD
Enriqueillo	32-day dataset	112	-0.0415	-0.0072	0.0084	0.0221	0.0320	0.3437	0.0509
	Monthly imputed	216	-0.0483	-0.0073	0.0009	0.0109	0.0165	0.1943	0.0362
Azuei	32-day dataset	165	-0.0197	-0.0014	0.0032	0.0029	0.0061	0.0435	0.0083
	Monthly imputed	216	-0.0183	-0.0017	0.0007	0.0019	0.0044	0.0377	0.0070

Statistical parameters of choice for comparison include mean, median, maximum, minimum, 25th quantile, 75th quantile, variance, and standard deviation as well as skewness and kurtosis as measures of asymmetry and “tailedness” of the probability distribution. Results of the Bootstrap test showed an agreement among all statistical characteristics of both series (except LA’s median values) yielding a 1% significance level. Hence, it was concluded that the statistical variable of the observational and imputed datasets share the same characteristics with a confidence level of 99%. The two monthly imputed VCTS for LA and LE are shown in Figure 7.

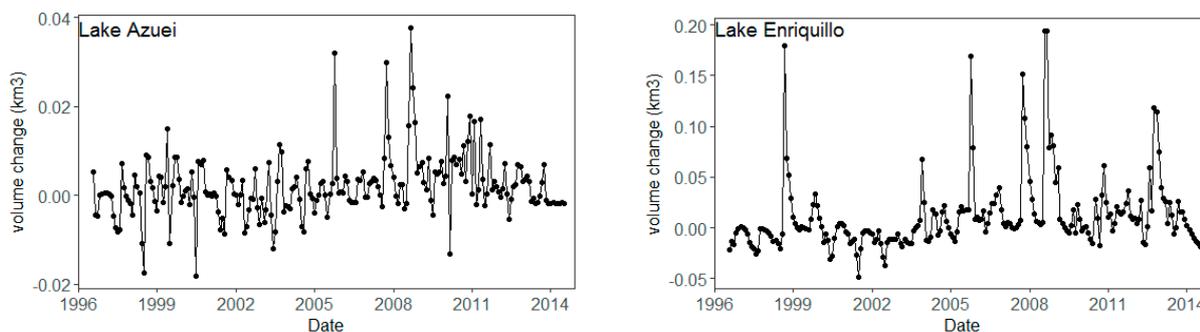


Figure 7. Monthly imputed volume change time series of both Lakes (km³).

3.2. Periodicity Detection: Wavelet Transform

The result of GWPS and the CWT analyses on the monthly VCTS are plotted in Figure 8 for both LA and LE spanning the years 1997 to 2014. Both the global and the power spectrum showed the 1-year cycle to be the most dominant scale of variation in the monthly VCTS of LE. The higher power (darker colors) showed mostly for periods in the 8–16 months bracket (annual scale) throughout the entire temporal domain, with some patches formed for sub-annual scales. The annual scale was significant for the years of 1997–2000 and 2005–2009, while it became non-significant for the rest of the time domain. Note the presence of sub-annual clusters for some years which were, however, not present for the entire time domain. None of the sub-annual clusters emerged as significant in the global wavelet power spectrum.

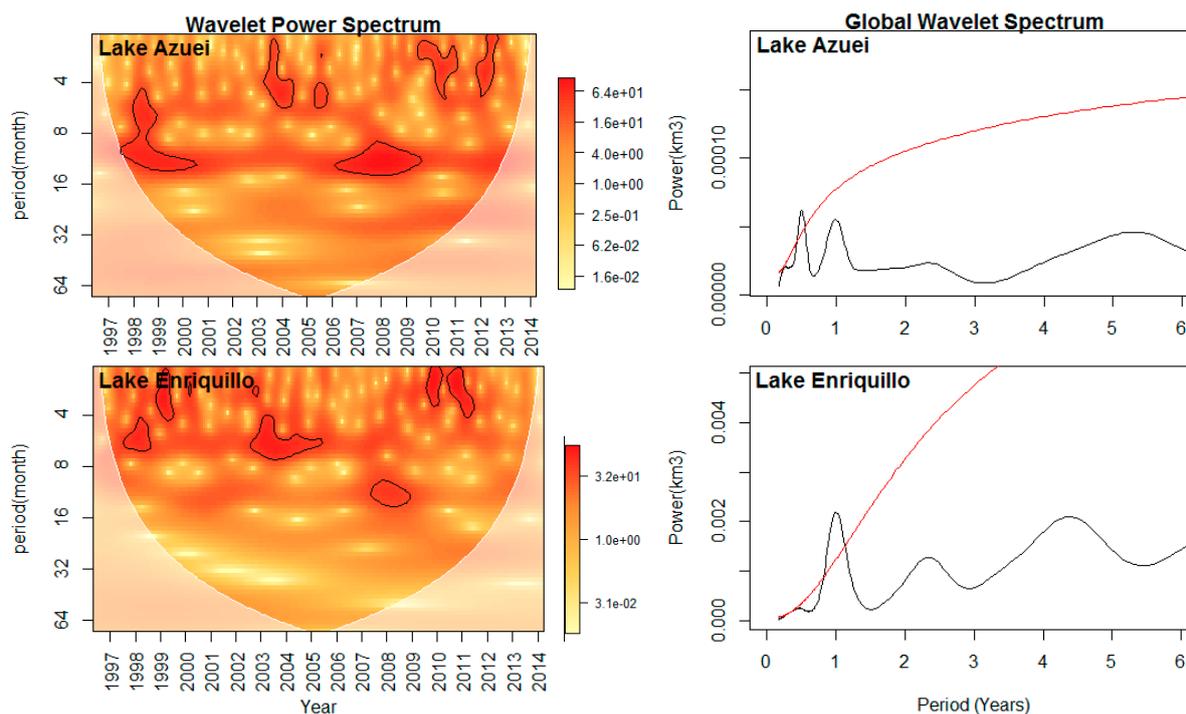


Figure 8. Wavelet power spectrum, CWT, (left), and global wavelet, GWPS, (right) for Lake Enriqueillo and Lake Azuei. [(1) Plots on the left: shades of colors are wavelet power (The darker the color, the stronger the power of the scale), lighter shaded area: the Cone of Influence (COI) where edge effect is important, black contour: the 5% significance level; (2) Plots on the right: red line is the 95% confidence level of a red-noise process].

For LA, the only dominant scale was the six-month period as shown in the global wavelet graph. In contrast, a cluster of annual scales appeared significant in the power wavelet spectrum. The sub-annual scales also exhibited different behavior, gaining higher significant power for some years (i.e., 1998–1999 and 2002–2005), while losing their significance for other years. In general, both LE and LA showed statistically significant scales of annual and sub-annual variability, which were corresponding with the seasonal changes as most of the local atmospheric parameters such as precipitation, temperature, and relative humidity which also exhibited periodicities of either six months, 12 months, or both. Comparing the lakes' monthly variation with regional precipitation variability (Figure 9) showed that LE was more sensitive to North Atlantic cyclone occurrences, while LA was mimicking the common precipitation pattern.

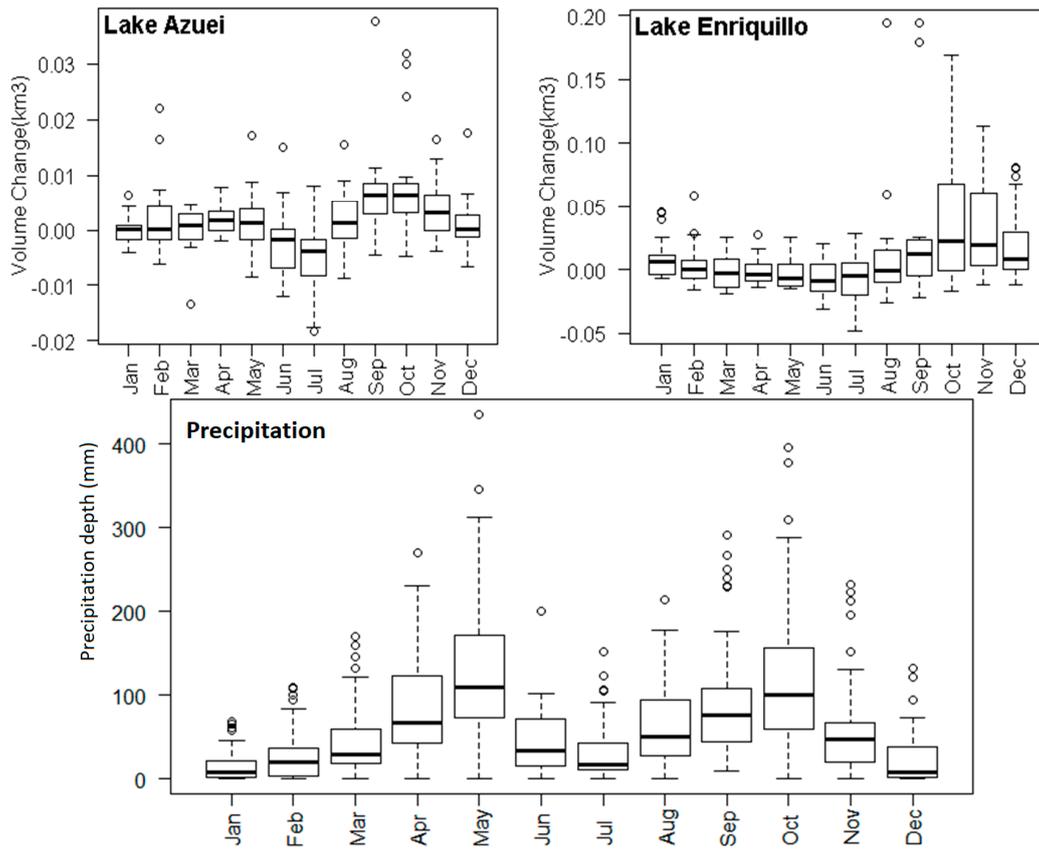


Figure 9. Top: monthly Variation of lake’s volume change for Lake Enriqueillo and Lake Azuei [the data used to graph boxplots is imputed monthly volume time series for the years of 1996 to 2014]; Bottom: monthly precipitation pattern at Jimani Station (1951–2015).

Note that the multi-annual periodicity (2 to 5 years) of the lakes’ variability did not turn out to be statistically significant which ruled out the correlation to the large-scale atmospheric teleconnections.

3.3. Change Point Detection: Pettitt Test

The Pettitt test predicted significant CPs in 2005 for both LE (April) and LA (August). The resulting monthly VCTS are plotted in Figure 10 along with $U_{t,n}$, two confidence levels of 95% and 99%, and a vertical line that denotes the estimated CP locations (red dotted line).

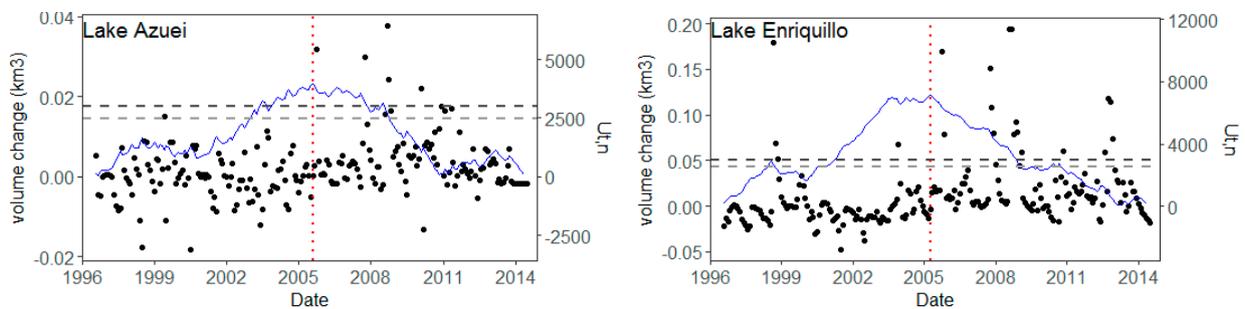


Figure 10. Application of Pettitt test to the lakes’ time series (1996–2014). [Time series values (black colored points), $U_{t,n}$ (blue line), Confidence level 99% (dark grey dashed line), Confidence level 95% (light grey dashed line), CP (red dotted line)].

In both cases, the probability value was less than 0.001, corresponding to a confidence level higher than 99.99%. Mean and variance were also used as two important parameters to detect changes in the monthly VCTS characteristics. After removing the outliers from the series, a non-parametric

hypothesis test (Bootstrap) was applied. The resulting probability for the mean showed a significant increase after the occurrence of the CP, thus explaining why both lakes had experienced constant growth after 2005. The changes in variance values, however, were not significant. Therefore, it was concluded that the positive shift in the mean value of volume change was responsible for the growth of both lakes which started in 2005 and not 2003 as previously reported in the literature. It also meant that the increase observed in both lakes volume prior to 2005 was in the normal range. Between 2005 and 2012, five North Atlantic cyclones in form of tropical storms or hurricanes had passed in the vicinity of the lakes' watersheds (less than 80 km away) with rainfall rates of more than 87 mm/month which correlated with positive shift occurrences in both lakes monthly VCTS.

3.4. Trend Test: Mann-Kendall Test and Linear Regression Model

The result of both Mann Kendall and linear regression trend tests showed statistically non-significant trends, having p-values of higher than the 10% significance level (Tables 3 and 4). The linear regression approach which was used to assess the significance of the CP alongside the trend showed that the p-values (probability) associated with the CP are 0.010 and 0.043 for LE and LA respectively thus satisfying the significant level of 5% (Table 4). This implied that after applying the shift (CP) to the series, the trend lost its significance and the CP stood out for both lakes.

Table 3. Result of Mann-Kendall trend analysis after applying the change point.

Series	Sub-Period before Shift		Sub-Period after Shift			
	Z Statistic	p-Value	Z Statistic	p-Value	N	
Lake Enriqueillo	0.013	0.848	N	-0.100	0.122	N
Lake Azuei	0.013	0.840	N	-0.082	0.211	N

Table 4. Result of linear regression analysis for trend, change point and seasonality (periodicity).

Series	Lake Enriqueillo			Lake Azuei		
	t Statistic	p-Value	N	t Statistic	p-Value	N
Trend	-0.423	0.673	N	-0.524	0.601	N
Change Point	2.596	0.010	*	0.631	0.043	*
6-month periodicity	0.413	0.680	N	3.610	0.0004	***
Annual periodicity	4.787	3.62×10^{-6}	***	2.742	0.0067	**

In the above table "N" and the star symbols mean: Significance levels of $p > 0.05$ (N), 0.05 (*), 0.01 (**), and 0.001 (***).

Both trend tests, Mann-Kendall and linear regression, produced consistent results that led to a similar conclusion of having a significant positive shift. This implies that the 2005 regime shift was able to explain the constant volume increase of both lakes since then. Both the monotonic trend and shift are illustrated in Figure 11; note though that the monotonic trend was not significant. The results for the scales of interannual and annual periodicity, however, showed that they were significant for both lakes. In the case of LE only annual periodicity was significant while for LA both 6-month and 12-month periodicity was significant. These results were consistent when compared to the wavelet spectrum analysis in which both annual and interannual scales emerged in different parts of the temporal domain.

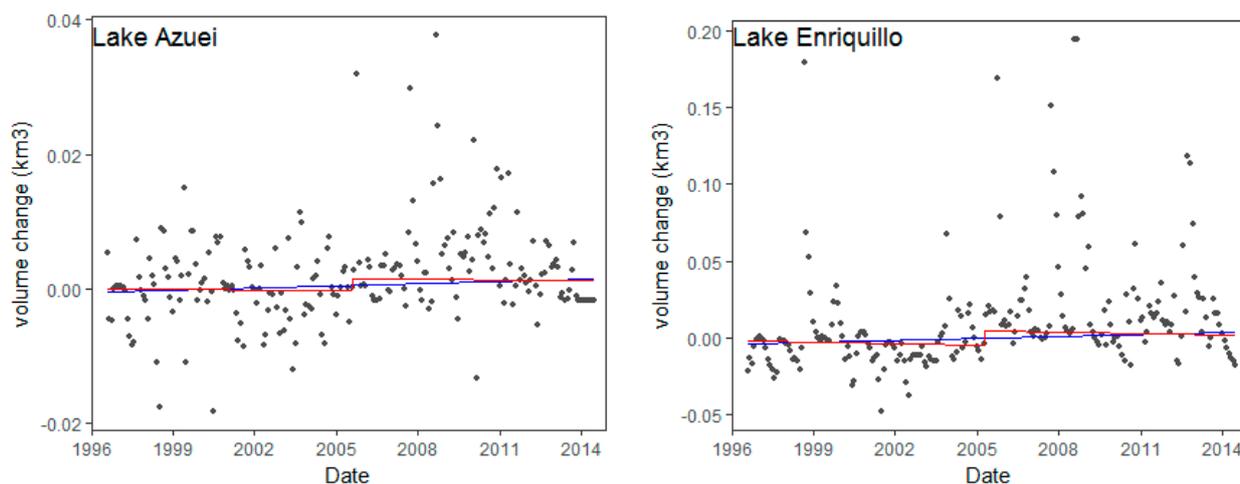


Figure 11. The demonstration of both tend and step (CP) fitted the time series for Lake Azuei and Lake Enriquillo [time series values (grey colored points), monotonic trend (blue line), trend lines and step change (CP) (red line)].

4. Discussion

A key step is to take a closer look at how the lakes responded vis-à-vis the general weather patterns as well as the anomalies that are superimposed. To this end it is helpful to explore simple geophysical force inputs and their responses when compared to the observed responses. The concept of lake “response” to climate variability was first mentioned by Langbein (1961) [32], which later on was discussed and further refined by other researchers (e.g., [76,77]). In 1985, Rapley and Cooper [78] introduced the equilibrium response time of τ_{eq} as a timescale during which the lake reaches 63% of its new equilibrium as a result of a perturbation in a previously equilibrated system, using a water balance equation. It is well understood that the response of a geophysical system differs based on the system’s physical characteristics as well as characteristics of the forcing factors which all contribute to the system’s response or, as in this case, time-variant mass balance [79]. In the case of closed-basin lakes, the geometry of the lake and its surroundings, as well as precipitation and evaporation, are the variables controlling the response of the lake [34].

The value of τ_{eq} at the time of perturbation defines how a lake moves toward the changes. Having defined three simple climatic variation types of (a) step change, (b) brief duration fluctuation (a spike) and c) sinusoidal change, the lake’s path towards a new equilibrium state is depicted in Figure 12 [34].

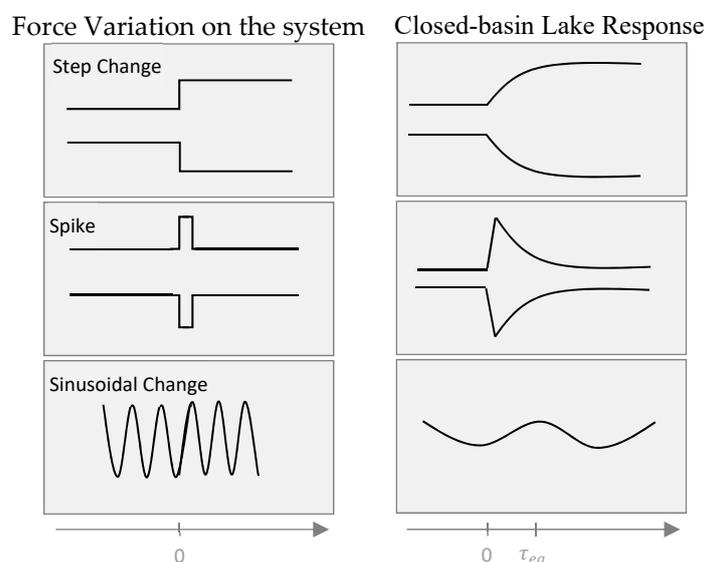


Figure 12. Illustration of closed-lake basin in equilibrium to climatic variation [34] (Reproduced with permission from [34]).

Recalling the fact that monthly and yearly precipitation does not exhibit any statistical change (trend, persistence, and/or CP), the only anomalies observed are the North Atlantic cyclones impacting the island. These extreme events, causing anomalously high daily rainfall, appear as a “spike” climatic variation over the lakes. Consequently, the lakes should have a quick response, increasing in size and then decreasing exponentially toward their previous state (Figure 12, 2nd image).

The lakes’ VTS, however, do not show such a distinct pattern. Since the climate variables over the basin are not constant and anomalies overlap (it is impossible to isolate a single spike forcing and its response), the persistent response of the lakes is therefore a combination of the response to both previous and present events and constraints. Nevertheless, a typical lake response to a “spike” can be observed between 1998 and 2004 (Figure 13) in LE’s VTS. In September 1998 Hurricane George impacted LE causing a sudden depth increase over the course of just a week with a continued albeit slowed down expansion over the next year, which does not fully qualify as a short-term spike forcing. With no other anomalies occurring in the area, the lake needed a total of 6 years (it had shrunk back by 2004) to return to its original size which is less than the 7.6 years of the equilibrium response time calculated for the lake. While it is reasonable to argue that the discrepancy is due to (a) persisting weather patterns and (b) shape of the lake that had an effect on the actual response time of the lake in comparison to the calculated τ_{eq} , the response of the lake is, in fact, a combination of responses to both a step and spike forcing, creating a bell shape with an elongated tail, as shown in Figure 13. The step forcing component is therefore due to time-lagged subsurface flow released over the course of a year (even though it is often neglected [34,80,81] and also due to a very wet year in 1999 which superimposed on the non-equilibrium state of LE.

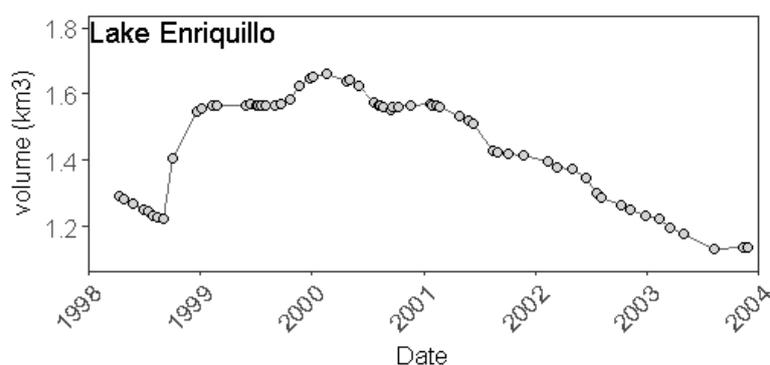


Figure 13. Lake Enriquillo’s response to the hurricane George on between the years of 1998 to 2004.

Between 2005 and 2014 (Table 5) five storms passed within 80km from the lakes’ basin, resulting in continued lake growth until 2014. Expectations would have been that the lakes continue to grow after each anomaly occurrence for up to one or two years and then shrink back to the size before the storms. Also reasonable would have been the expectation that the lakes would fall into their shrinking pattern (by 2010) after the 2007 and the 2008 storms. However, not only did the lakes continued to expand for four (rather than the expected two) more years after the 2008 storms, but also did they do so to an extent that they had never experienced before. With the arrival of tropical storm Isaac (2012) at the end of the previous 4-year growth cycle, additional growth would have been expected to continue, albeit at a slower rate because Isaac only precipitated half the amount than the previous storms. Yet, the lakes continued to grow at the same rate despite only half the precipitation amount.

Table 5. Storm and hurricane incidents in 80km of the lakes' watershed (1972–2017).

NO	Year	Month	Name	Type	Distance to Watershed (km)	Monthly Rainfall at Jimani Station
1	1979	Jul	CLAUDETTE	Tropical Storm	18.8	123.4
2	1979	Sep	DAVID	Category 1 Hurricane	51.0	170.7
3	1998	Sep	GEORGES	Category 3 Hurricane	28.2	230
4	2005	Oct	ALPHA	Tropical Storm	0.0	257.4
5	2007	Oct	NOEL	Tropical Storm	48.8	225.8
6	2008	Aug	FAYE	Tropical Storm	3.7	214.4
7	2008	Aug	GUSTAV	Tropical Storm	71.5	214.4
8	2012	Aug	ISAAC	Tropical Storm	53.1	115.6

If one accounts for the monthly rainfall introduced by the each of the storms and the low rate of annual precipitation for the years following 2011, then these added volumes do not support the extent of the lake's growth, suggesting that the characteristics of the system and/or forcing factors had changed. Three anomalies have been reported for the area of study which affected the lakes' dynamic system: (i) The years between 2003 and 2012 featured consistent above-average amount of rainfall (wet years); (ii) Change in basin land cover (continued deforestation causing less water retention and evapotranspiration); (iii) Water flow into LE due to the break of the Trujillo dike in the neighboring watershed in 2007 [82]. Cases (i) and (ii) are less probable in contributing to the lakes' dramatic growth because the lakes' basin had experienced above-average rainfall previously without responding with significant growth. Also, the LA basin had been affected by deforestation for a much longer period of time when compared to the LE basin, however, LA had never responded with a significant increase before.

Recalling that the LA's VTS has not shown the same sensitivity to the occurrence of the North Atlantic cyclones, a different reason for the start of the simultaneous growth must exist. Since one can safely assume that the same weather patterns persist for both lakes due to their proximity, alignment in W-E directions, and being bounded by the same high mountain ranges in the south and north the likelihood for not accounting for other hydro-climate components seems small. Instead, the authors believe that the growth of LA is a response to the decreased hydraulic gradient between the lakes substantially reducing the amount of water flowing from the higher lake (LA) to the lower elevation (LE). This would create a "back up" effect causing LA to expand, and at the same time increasing the hydraulic gradient again to increase the flow rate. Somewhere in this process is an equilibrium state that remains dynamic however depending on weather patterns (sequence and occurrence of wet/dry years) and the occurrences of anomalies.

While other anthropogenic influences such as the slowly continuing deforestation on both sides of the border certainly have an impact on the time scales with which water moves and how long it resides in specific stores, the authors are convinced that the Trujillo Dam failure even though only a single (anthropogenic) event that occurred in 2007, i.e., several years before the continued growth patterns, has had a long term impact on how the lakes responded and adjusted in the aftermath. Although at first glance the event seemed to affect only LE by adding a significant water volume through extra basin water transfer (from the Yaque del Sur basin) it aided significantly to the hydraulic gradient reduction between the lakes and as such indirectly to the backup effect in LA. More research work would need to be carried out in order to support and further substantiate this opinion, i.e., the land strip separating the two lakes would need to be instrumented with an array of boreholes in which to measure the actual flow rates between the lakes. Whether this is feasible in the future remains to be seen because the border between Haiti and the DR runs through the narrow land strip in an area that is notorious for its lack of safety for both equipment and people.

5. Conclusions

The paper first presented the algorithm for generating the monthly VTS of LA and LE using an imputation approach. It is based on the observational VTS that was derived using Landsat images,

hence having time intervals of about 16 days. The data series provided us with lake volume values for the period between 1972 and 2017. The level of missingness in the set, however, forced us to consider only the period between 1996–2014. Further subdivision of the observational VTS were dictated by the time frame of different Landsat satellites and the occurrence of the North Atlantic cyclones in the vicinity of the study area. The former factor influenced the choice of the suitable interval for imputation, while the latter is responsible for the occurrences of discontinuities in the VTS. Among 15 different multivariate imputation methods, the Stineman interpolation yielded the lowest NRMSE values, hence it was chosen to construct the monthly time series. To validate the resampled monthly VTS of the lakes' volume, it was compared with the 32-day VCDS. The assessment yielded non-significant differences between the two datasets with a confidence interval of 99% thus demonstrating that the constructed monthly VTS had the same statistical characteristics as the observational ones.

Analysis of the generated monthly VCTS using wavelet decomposition for both lakes showed seasonal periodicity in the behavior of the lakes. The Pettitt test detected a significant change point for the lakes occurring in 2005, which was correlated to consecutive cyclone incidents starting in 2005 and continuing to 2012. Moreover, the outlier study showed the importance of cyclones striking the island. Further assessment of the trend (using linear regression and Man-Kendall tests), CP, and periodicity revealed that the VCTS did not feature a trend and that the continuous growth of the lakes from 2005 to 2014 was related to changes in the inside the study area that caused a shift in the storage of the lakes starting in 2005.

Having compared the VTS of LE with the typical response shape (graph) of the closed-basin lake to climate perturbation showed how the lake behaves during storm events and during the years of recovery. As a typical response to a high-intensity event, the lake will grow for about one to two years suggesting that the water stored in the subsurface of the surrounding watershed will, over the course of 2 years, flow into the lake. After that, it takes about 5 years for the lake to return to its original state, i.e., the sequence storm event, filling (growth), and subsidence spans an average period of about 6–7 years.

Based on this result, it is clear that for the years between 2005 to 2014, which did not behave as expected, another factor/event came into play that contributed to the growth of LE besides the North Atlantic cyclone events. This event occurred in 2007 when a dike preventing water of the Yaque del Sur (which is a different watershed) from flowing into Lake Rincon (which is connected to LE through a canal) was destroyed due to tropical storms Noel and Olga passing over the island. This resulted in a prolonged uncontrolled discharge into Lake Rincon and from there onwards into LE. It is thus obvious that anthropogenic events such as the Trujillo, significantly added to the lakes' state going beyond just climatologic forcing.

The paper also examined the lake characteristics and their synchronous and asynchronous behavior. Given the fact that both lakes experience the same weather patterns, thus promulgating the notion of similar if not exact rainfall and draught rates, it clear that other mechanisms need to be at play to explain the asynchronous shrinking and growth patterns. Evidence was presented to the fact that the two lakes must be connected leading to time variant flow rates that are based on the respective lake levels encountered at any given point in time. This connectedness leads to asynchronous shrinkage/expansion patterns in the absence of extreme precipitation such as encountered during hurricanes.

Lastly, the authors would want to point out some key aspects and findings that must be seen in a larger context. Firstly, there is the realization that the explanation of the lakes' shrinkage and expansion patterns is complex. There is no ONE answer or reason that causes the behavior, rather it is a multitude of processes and reasons that contribute to it some of which are local, some of which are regional (Caribbean), others are global (North Atlantic Oscillations), some are entirely environmental (hydro climatological, land cover) while others are of anthropogenic nature (deforestation and levee construction and subsequent failure). Secondly, we are introducing a set of statistical or signal processing analyses steps that can aid in understanding what forces and events were at play at certain points in time and their impact on the lakes' response patterns. This should help or at least be of interest to researchers who are interested in conducting lake research elsewhere

in the context of changing climate and weather patterns. Thirdly, we hope that the results obtained are of interest to the general research community and also governmental research laboratories that conduct research around the lakes, which could also be biologists and ecologists. Lastly, the research effort encountered substantial adversity in terms of accessibility in both countries, in addition to a significant dearth of missing information on almost all physical aspects that are related to the lakes be it hydro geological, hydro climatological, or land surface in nature. We provided a path forward using information and data that mostly (except the bathymetric data, but this is data we ourselves collected and which was not available in either country) does not originate in either of the countries but is available as public global information (United States websites). In other words, we tried to develop defensible scientific insights from limited albeit generally accessible data sources to answer the questions posted.

Author Contributions: M.M., the first author, has been responsible for all work associated with the conceptualization, methodology, validation, formal analysis, and the investigation of the work presented. She has also been the lead on the writing-original draft preparation and all visualizations (graphics) for the manuscript. M.P. has been responsible for the writing-review and editing component, supervision, project administration, funding acquisition and in part for the data curation. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the US National Science Foundation, grant numbers 1264466 and 1513512. Additional funding was provided by the Department of Civil Engineering at the City College of New York through graduate assistantships.

Acknowledgments: We would like to acknowledge of Naresh Devineni for his valuable advice and critique with regard to the time series analysis. Appreciation also goes to all those researchers at The City College of New York, the Instituto Tecnológico de Santo Domingo and the DR state agencies of INDHRI and ONAMET who were instrumental in providing data sets for our analyses.

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

References

1. Sheller, M.; León, Y.M. Uneven socio-ecologies of Hispaniola: Asymmetric capabilities for climate adaptation in Haiti and the Dominican Republic. *Geoforum* **2016**, *73*, 32–46.
2. Kushner, J. The Relentless Rise of Two Caribbean Lakes Baffles Scientists. 2016. Available online: <https://www.nationalgeographic.com/news/2016/03/160303-haiti-dominican-republic-lakes/> (accessed on 3 March 2016).
3. Aldenberg, T.; Janse, J.H.; Kramer, P.R.G. Fitting the dynamic model PCLake to a multi-lake survey through Bayesian Statistics. *Ecol. Model.* **1995**, *78*, 83–99.
4. Kieniewicz, J.M.; Smith, J.R. Paleoenvironmental reconstruction and water balance of a mid-Pleistocene pluvial lake, Dakhleh Oasis, Egypt. *Bull. Geol. Soc. Am.* **2009**, *121*, 1154–1171.
5. Collins, G.S.; Artemieva, N.; Wünnemann, K.; Bland, P.A.; Reimold, W.U.; Koeberl, C. Evidence that Lake Cheko is not an impact crater. *Terra Nov.* **2008**, *20*, 165–168.
6. Sivakumar, B. *Chaos in Hydrology: Bridging Determinism and Stochasticity*, 1st ed.; Springer: Dordrecht, The Netherlands, 2017; pp. 29–62.
7. Medwedeff, W.G.; Roe, G.H. Trends and variability in the global dataset of glacier mass balance. *Clim. Dyn.* **2017**, *48*, 3085–3097.
8. Barbosa, S.M.; Donner, R.V. Long-term changes in the seasonality of Baltic sea level. *Tellus A Dyn. Meteorol. Oceanogr.* **2016**, *68*, 30540.
9. Khelifa, S.; Gourine, B.; Rami, A.; Taibi, H. Assessment of nonlinear trends and seasonal variations in global sea level using singular spectrum analysis and wavelet multiresolution analysis. *Arab. J. Geosci.* **2016**, *9*, 1–8.
10. Moon, Y.I.; Lall, U. Atmospheric flow indices and interannual Great Salt Lake variability. *J. Hydrol. Eng.* **1995**, *1*, 55–62.
11. Moon, Y.I.; Lall, U.; Kwon, H.H. Non-parametric short-term forecasts of the Great Salt Lake using atmospheric indices. *Int. J. Climatol.* **2007**, *4*, 1549–1555.
12. Moknatian, M.; Piasecki, M. *Observational Time Series for Lakes Azuei and Enriquillo: Surface Area, Volume, and Elevation*; CUNY Academic Works: New York, NY, USA, 2019.

13. Moknatian, M.; Piasecki, M.; Gonzalez, J. Development of geospatial and temporal characteristics for Hispaniola's Lake Azuei and Enriquillo using Landsat imagery. *Remote Sens.* **2017**, *9*, 510.
14. Piasecki, M.; Moknatian, M.; Moshary, F.; Cleto, J.; Leon, Y.; Gonzalez, J.; Comarazamy, D. *Bathymetric Survey for Lakes Azuei and Enriquillo, Hispaniola*; CUNY Academic Works: New York, NY, USA, 2016.
15. Piasecki, M.; Moknatian, M. *Bathymetry Data for Lakes Azuei and Enriquillo*; CUNY Academic Works: New York, NY, USA, 2018.
16. Moknatian, M.; Piasecki, M.; Moshary, F.; Gonzalez, J. Development of digital bathymetry maps for Lakes Azuei and Enriquillo using sonar and remote sensing techniques. *Trans. Gis* **2019**, *23*, 841–859.
17. Teegavarapu, R.S.V.; Chandramouli, V. Improved weighting methods, deterministic and stochastic data-driven models for estimation of missing precipitation records. *J. Hydrol.* **2005**, *312*, 191–206.
18. Teegavarapu, R.S.V. Statistical corrections of spatially interpolated missing precipitation data estimates. *Hydrol. Process.* **2014**, *28*, 3789–3808.
19. Junninen, H.; Niska, H.; Tuppurainen, K.; Ruuskanen, J.; Kolehmainen, M. Methods for imputation of missing values in air quality data sets. *Atmos. Environ.* **2004**, *38*, 2895–2907.
20. Jönsson, P.; Eklundh, L. TIMESAT—A program for analyzing time-series of satellite sensor data. *Comput. Geosci.* **2004**, *30*, 833–845.
21. Shen, H.; Li, X.; Cheng, Q.; Zeng, C.; Yang, G.; Li, H.; Zhang, L. Missing Information Reconstruction of Remote Sensing Data: A Technical Review. *IEEE Geosci. Remote Sens. Mag.* **2015**, *3*, 61–85.
22. Simolo, C.; Brunetti, M.; Maugeri, M.; Nanni, T. Improving estimation of missing values in daily precipitation series by a probability density function-preserving approach. *Int. J. Climatol.* **2010**, *30*, 1564–1576.
23. Pappas, C.; Papalexiou, S.M.; Koutsoyiannis, D. A quick gap-filling of missing hydrometeorological data. *J. Geophys. Res. Atmos.* **2014**, *119*, 9290–9300.
24. Elshorbagy, A.; Simonovic, S.P.; Panu, U.S. Estimation of missing streamflow data using principles of chaos theory. *J. Hydrol.* **2002**, *255*, 123–133.
25. Rubin, D.B. Inference and missing data. *Biometrika* **1976**, *63*, 581–592.
26. Moritz, S.; Bartz-Beielstein, T. imputeTS: Time series missing value imputation in R. *R J.* **2017**, *9*, 207–218.
27. Moritz, S.; Sardá, A.; Bartz-Beielstein, T.; Zaefferer, M.; Stork, J. Comparison of Different Methods for Univariate Time Series Imputation in R. *arXiv* **2015**, arXiv:1510.03924.
28. Barnett, V.; Lewis, T. *Outliers in Statistical Data*, 2nd ed.; John Wiley & Sons: Hoboken, NJ, USA, 1980; p. 2434.
29. González-Rouco, J.F.; Jiménez, J.L.; Quesada, V.; Valero, F. Quality control and homogeneity of precipitation data in the Southwest of Europe. *J. Clim.* **2001**, *14*, 964–978.
30. Chen, S.; Li, Y.; Kim, J.; Kim, S.W. Bayesian change point analysis for extreme daily precipitation. *Int. J. Climatol.* **2017**, *37*, 3123–3137.
31. Afshari, S.; Fekete, B.M.; Dingman, S.L.; Devineni, N.; Bjerklie, D.M.; Khanbilvardi, R.M. Statistical filtering of river survey and streamflow data for improving At-A-Station hydraulic geometry relations. *J. Hydrol.* **2017**, *547*, 443–454.
32. Langbein, W.B. *Salinity and Hydrology of Closed Lakes*; U.S. Geological Survey Professional Paper 412; U.S. Government Printing Office: Washington, DC, USA, 1961.
33. Street-Perrott, F.A.; Roberts, N. Fluctuations in closed-basin lakes as an indicator of past atmospheric circulation patterns. *Var. Glob. Water Budg.* **1983**, 331–345.
34. Mason, I.M.; Guzkowska, M.A.J.; Rapley, C.G.; Street-Perrott, F.A. The response of lake levels and areas to climatic change. *Clim. Chang.* **1994**, *27*, 161–197.
35. Addison, P.S. *The Illustrated Wavelet Transform. Handbook: Introductory Theory and Applications in Science, Engineering, Medicine and Finance*, 2nd ed.; CRC Press, Taylor & Francis: Boca Raton, FL, USA, 2002; pp. 6–64.
36. Torrence, C.; Compo, G.P. A practical guide to wavelet analysis. *Bull. Am. Meteorol. Soc.* **1998**, *79*, 61–78.
37. Daubechies, I. The wavelet transform, time-frequency localization and signal analysis. *IEEE Trans. Inf. Theory* **1990**, *36*, 961–1005.
38. Elsanabary, M.H.; Gan, T.Y. Wavelet analysis of seasonal rainfall variability of the Upper Blue Nile Basin, its teleconnection to global sea Surface temperature, and its forecasting by an artificial neural network. *Mon. Weather Rev.* **2014**, *142*, 1771–1791.
39. Morales-Pineda, M.; Cózar, A.; Laiz, I.; Úbeda, B.; Gálvez, J.A. Daily, biweekly, and seasonal temporal scales of pCO₂ variability in two stratified Mediterranean reservoirs. *J. Geophys. Res. Biogeosci.* **2014**, *119*, 509–520.

40. Yu, H.L.; Lin, Y.C. Analysis of space-time non-stationary patterns of rainfall-groundwater interactions by integrating empirical orthogonal function and cross wavelet transform methods. *J. Hydrol.* **2015**, *525*, 585–597.
41. Elsanabary, M.H.; Gan, T.Y.; Mwale, D. Application of wavelet empirical orthogonal function analysis to investigate the nonstationary character of Ethiopian rainfall and its teleconnection to nonstationary global sea surface temperature variations for 1900–1998. *Int. J. Climatol.* **2014**, *34*, 1798–1813.
42. Jiang, R.; Gan, T.Y.; Xie, J.; Wang, N. Spatiotemporal variability of Alberta’s seasonal precipitation, their teleconnection with large-scale climate anomalies and sea surface temperature. *Int. J. Climatol.* **2014**, *34*, 2899–2917.
43. Najibi, N.; Devineni, N.; Lu, M. Hydroclimate drivers and atmospheric teleconnections of long duration floods: An application to large reservoirs in the Missouri River Basin. *Adv. Water Resour.* **2017**, *100*, 153–167.
44. Cook, R.D. Assessment of local influence. *J. R. Stat. Soc.* **1986**, *48*, 133–169.
45. Jönsson, C.A.; Tarukoski, E. *How Does an Appointed CEO Influence the Stock Price? A Multiple Regression Approach*; KTH Royal Institute of Technology, School of Engineering Sciences: Stockholm, Sweden, 2017.
46. Lawrance, A.J. Deletion influence and masking in regression. *J. R. Stat. Soc.* **1995**, *58*, 267–288.
47. Cazelles, B.; Cazelles, K.; Chavez, M. Wavelet analysis in ecology and epidemiology: Impact of statistical tests. *J. R. Soc. Interface* **2014**, *11*, 20130585.
48. Mallakpour, I.; Villarini, G. Investigating the relationship between the frequency of flooding over the central United States and large-scale climate. *Adv. Water Resour.* **2016**, *92*, 159–171.
49. Mosqueiro, T.; Strube-Bloss, M.; Tuma, R.; Pinto, R.; Smith, B.H.; Huerta, R. Non-parametric change point detection for spike trains. In Proceedings of the 2016 Annual Conference on Information Science and Systems (CISS), Princeton, NJ, USA, 16–18 March 2016; pp. 545–550.
50. Basseville, M.; Nikiforov, I.V. *Detection of Abrupt Changes: Theory and Application*; Prentice Hall, Inc.: Englewood Cliffs, NJ, USA, 1993; pp. 25–66.
51. Matteson, D.S.; James, N.A. A nonparametric approach for multiple change point analysis of multivariate data. *J. Am. Stat. Assoc.* **2014**, *109*, 334–345.
52. Pettitt, A.N. A non-parametric approach to the change-point problem. *Appl. Stat.* **1979**, *28*, 126–135.
53. Ha, K.J.; Ha, E. Climatic change and interannual fluctuations in the long-term record of monthly precipitation for Seoul. *Int. J. Climatol.* **2006**, *26*, 607–618.
54. Ma, Z.; Kang, S.; Zhang, L.; Tong, L.; Su, X. Analysis of impacts of climate variability and human activity on streamflow for a river basin in arid region of northwest China. *J. Hydrol.* **2008**, *352*, 239–249.
55. Figura, S.; Livingstone, D.M.; Hoehn, E.; Kipfer, R. Regime shift in groundwater temperature triggered by the Arctic Oscillation. *Geophys. Res. Lett.* **2011**, *38*, L23401.
56. Li, D.; Xie, H.; Xiong, L. Temporal change analysis based on data characteristics and nonparametric test. *Water Resour. Manag.* **2014**, *28*, 227–240.
57. Li, J.; Tan, S.; Wei, Z.; Chen, F.; Feng, P. A new method of change point detection using variable fuzzy sets under environmental change. *Water Resour. Manag.* **2014**, *28*, 5125–5138.
58. Gao, P.; Zhang, X.; Mu, X.; Wang, F.; Li, R.; Zhang, X. Trend and change-point analyses of streamflow and sediment discharge in the Yellow River during 1950–2005. *Hydrol. Sci. J.* **2010**, *55*, 275–285.
59. Xuedong, L.; Yili, Z.; Zhijun, Y.; Tongliang, G.; Hong, W.; Duo, C.; Linshan, L.; Fei, Z. The trend on runoff variations in the Lhasa River Basin. *J. Geogr. Sci.* **2008**, *18*, 95–106.
60. Liu, D.; Chen, X.; Lian, Y.; Lou, Z. Impacts of climate change and human activities on surface runoff in the Dongjiang River basin of China. *Hydrol. Process.* **2010**, *24*, 1487–1495.
61. Kisi, O.; Ay, M. Comparison of Mann-Kendall and innovative trend method for water quality parameters of the Kizilirmak River, Turkey. *J. Hydrol.* **2014**, *513*, 362–375.
62. Bates, B.C.; Chandler, R.E.; Bowman, A.W. Trend estimation and change point detection in individual climatic series using flexible regression methods. *J. Geophys. Res. Atmos.* **2012**, *117*, D16106.
63. Tomozeiu, R.; Lazzeri, M.; Cacciamani, C. Precipitation fluctuations during the winter season from 1960 to 1995 over Emilia-Romagna, Italy. *Theor. Appl. Climatol.* **2002**, *72*, 221–229.
64. Ahmad, I.; Tang, D.; Wang, T.; Wang, M.; Wagan, B. Precipitation trends over time using Mann-Kendall and spearman’s Rho tests in swat river basin, Pakistan. *Adv. Meteorol.* **2015**, *2015*, 431860.
65. Gocic, M.; Trajkovic, S. Analysis of changes in meteorological variables using Mann-Kendall and Sen’s slope estimator statistical tests in Serbia. *Glob. Planet. Chang.* **2013**, *100*, 172–182.

66. Hamed, K.H.; Rao, A.R. A modified Mann-Kendall trend test for autocorrelated data. *J. Hydrol.* **1998**, *204*, 182–196.
67. Hamed, K.H. Trend detection in hydrologic data: The Mann-Kendall trend test under the scaling hypothesis. *J. Hydrol.* **2008**, *349*, 350–363.
68. Tarhule, A.; Woo, M.K. Changes in rainfall characteristics in Northern Nigeria. *Int. J. Climatol.* **1998**, *18*, 1261–1271.
69. Yue, S.; Pilon, P.; Phinney, B. Canadian streamflow trend detection: Impacts of serial and cross-correlation. *Hydrol. Sci. J.* **2003**, *48*, 51–63.
70. Yue, S.; Pilon, P.; Phinney, B.; Cavadias, G. The influence of autocorrelation on the ability to detect trend in hydrological series. *Hydrol. Process.* **2002**, *16*, 1807–1829.
71. Allison, P.D. *Missing Data*; Sage University Papers Series on Quantitative Applications in the Social Sciences: Thousand Oaks, CA, USA, 2001; Volume 07–136.
72. Little, R.J.A.; Rubin, D.B. *Statistical Analysis with Missing Data*, 2nd ed.; John Wiley & Sons, Inc.: Hoboken, NJ, USA, 2002; pp. 246–252.
73. Li, J.; Heap, A.D. *A Review of Spatial Interpolation Methods for Environmental Scientists*; Record 2008/23; Geoscience Australia: Canberra, Australia, 2008; 137p.
74. Stineman, R.W. A consistently well behaved method of interpolation. *Creat. Comput.* **1980**, *6*, 54–57.
75. Leclercq, P.W.; Oerlemans, J. Global and hemispheric temperature reconstruction from glacier length fluctuations. *Clim. Dyn.* **2012**, *38*, 1065–1079.
76. Szesztay, K. Water balance and water level fluctuations of lakes. *Hydrol. Sci. J.* **1974**, *19*, 73–84.
77. Gates, D.J.; Diesendorf, M. On the fluctuations in levels of closed lakes. *J. Hydrol.* **1977**, *33*, 267–285.
78. Rapley, C.G.; Griffiths, H.D.; Squire, V.A.; Oliver, J.G.; Birks, A.R.; Cooper, A.P.R.; Cowan, A.M.; Drewry, D.J.; Gorman, M.R.; Guzkowska, M.; et al. *Applications and Scientific Uses of ERS-1 Radar Altimeter Data*; ESA report 5684/83/NL/BI; ESA: Paris, France, 1985.
79. Roe, G.H.; O’Neal, M.A. The response of glaciers to intrinsic climate variability: Observations and models of late-Holocene variations in the Pacific Northwest. *J. Glaciol.* **2009**, *55*, 839–854.
80. Morrill, C.; Small, E.E.; Sloan, L.C. Modeling orbital forcing of lake level change: Lake Gosiute (Eocene), North America. *Glob. Planet. Chang.* **2001**, *29*, 57–76.
81. Todhunter, P.E. Mean hydroclimatic and hydrological conditions during two climatic modes in the Devils Lake Basin, North Dakota (USA). *Lakes Reserv. Res. Manag.* **2016**, *21*, 338–350.
82. Rising Water Levels at Lake Enriquillo, Dominican Republic: Advice on Potential Causes and Pathways Forward. Available online: https://iciwarm.info/wp-content/uploads/2018/01/Lake_Enriquillo_report_1-26-2012.pdf (accessed on 26 January 2012).

