Automated Generation of Lakes and Reservoirs Water Elevation Changes From Satellite Radar Altimetry

Modurodoluwa Adeyinka Okeowo, Hyongki Lee, Faisal Hossain, and Augusto Getirana

***Abstract*—Limited access to *in-situ* water level data for lakes and reservoirs have been a major setback for regional and global studies of reservoirs, surface water storage changes, and monitoring the hydrologic cycle. Processing satellite radar altimetry data over inland water bodies on a large scale has been a cumbersome task primarily due to the removal of contaminated measurements as a result of surrounding land. In this study, we proposed a new algorithm to automatically generate time series from raw satellite radar altimetry data without user intervention. With this method, users with a little knowledge on the field can now independently process radar altimetry for diverse applications. The method is based on K-means clustering, interquartile range, and statistical analysis of the dataset for outlier detection. Jason-2 and Envisat radar altimetry data were used to demonstrate the capability of this algorithm. A total of 37 satellite crossings over 30 lakes and reservoirs located in the U.S., Brazil, and Nigeria were used based on the availability of *in-situ* data. We compared the results against *in-situ* data and root-mean-square error values ranged from 0.09 to 1.20 m. We also confirmed the potential of this algorithm over rivers and wetlands using the southern Congo River and Everglades wetlands in Florida, respectively. Finally, the different retracking algorithms in Envisat; Ice-1, Ice-2, Ocean, and Sea-Ice were compared using the proposed algorithm. Ice-1 performed best for generating water level time series for in-land water bodies and the result is consistent with previous studies.**

***Index Terms*—Lakes, outlier detection, reservoirs, satellite altimetry, water level.**

I. Introduction

Several studies have demonstrated the capability of satellite altimetry in monitoring water level changes over rivers [3], [4], lakes and reservoirs [5], [6], and floodplains and wetlands [7]–[9]. Monitoring lake level variation is an indicator of global climate change and ecological issues [10]. Further studies on this indicator were performed on the Qinghai-Tibetan Plateau by Lee *et al.* [11] on the nexus between lake level variation and its corresponding effect on climate change [12]. Tarpanelli *et al.* [13] integrated satellite altimetry and moderate resolution imaging spectroradiometer to estimate discharge in rivers. In addition, recent studies by Hossain *et al.* [14] and Biancamaria *et al.* [4] have demonstrated the capability of satellite altimetry in transboundary flood forecasting downstream for adoption by the stakeholder agencies in strategic water resource management. Such scientific applications can mitigate the loss of lives and properties of vulnerable local residents downstream [14].

Nonetheless, Gao *et al*. [15] highlighted that access to water level data has been a major challenge in the global studies of reservoirs. Alsdorf *et al.* [16] also stated the limited access to *in-situ* data for hydrologic studies due to the impracticable cost of installation over all major water bodies. Hence, satellite altimetry is commonly used as a surrogate for *in-situ* gauge as a remote sensing technique to generate water level time series.

Currently, there have been several altimetry satellites launched into orbit to observe water level for environmental studies. These include but are not limited to: ERS-1/2, Envisat, TOPEX/Poseidon, Jason-1, 2, and 3, SARAL/AltiKa, and Sentinel-3. The surface water ocean topography (SWOT) mission, scheduled to be launched in 2021 (https://swot.cnes.fr/en/SWOT/index.htm), is a Ka-band swath mapping interferometer that will provide simultaneous measurements of water elevation and inundated area for inland water bodies [17], [18]. From these observations, surface water storage changes over water bodies whose area exceeds 250 m $\times$ 250 m (lakes, reservoirs, and wetlands) can be readily calculated [19].

Despite the recent advances in satellite altimetry and its diverse applications, there has not been sufficient research on automated data processing to harness the opportunities created by the massive amount of streaming data from multiple altimetry satellites. Such research on automation of accurate height extraction can pave the way for engaging a broader community of scientists and stakeholders that need this fundamentally elusive water information from space for a wide variety of scientific and environmental applications. However, satellite altimetry observations have their accuracy reduced by the presence of outliers due to the contamination of nonwater features within the altimetry footprint [20]. According to Birkett and Beckley [21], a manual approach of outlier removal has to be adopted for quality control in addition to land mask flags. A manual removal of outliers is time consuming, and limits a global generation of reservoir elevation profiles.

Recently, there have been attempts to automate the outlier removal in satellite altimetry data. For example, Huang *et al.*

Manuscript received September 3, 2016; revised January 6, 2017 and February 16, 2017; accepted March 5, 2017. This work was supported by NASA's Applied Sciences Program (NNX13AQ89G) and SERVIR Program (NNX16AN35G). *(Corresponding author: Modurodoluwa Adeyinka Okeowo.)*

M. A. Okeowo and H. Lee are with the Department of Civil and Environmental Engineering and National Center for Airborne Laser Mapping, University of Houston, Houston, TX 77204 USA (e-mail: doluokeowo@yahoo.com; hlee@uh.edu).

F. Hossain is with the Department of Civil and Environmental Engineering, University of Washington, Seattle, WA 98195 USA (e-mail: fhossain@uw.edu).

A. Getirana is with the Hydrological Sciences Laboratory, NASA Goddard Space Flight Center, Greenbelt, MD 20771 USA (e-mail: augusto.getirana@nasa.gov).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/JSTARS.2017.2684081

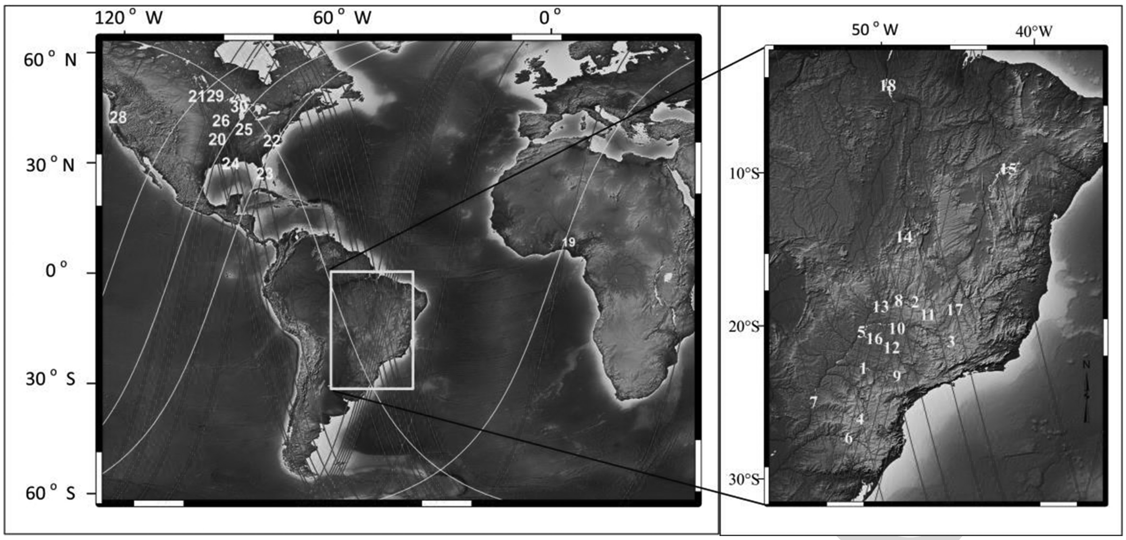


Fig. 1. Map of the study area. Table I shows the names of each reservoir and lakes in the map. The red and cyan lines show Envisat and Jason-2 tracks, respectively, over study areas.

[22] removed outliers from ICESat data using threshold values of reflectivity in the ICESat product and digital elevation model (DEM) from shuttle radar topography mission (SRTM) to detect outliers. Thus, this method relies on ancillary dataset, i.e., DEM, to effectively remove outliers. Another method of outlier detection in satellite altimetry was developed by Schwatke *et al.* [23] which involves extensive pre- and postprocessing including a Kalman filter approach. This method is robust but not significantly better than the less complex and yet effective method proposed in this study as discussed later in Section II. Nielsen *et al.* [24] removed outliers from CryoSat-2 by using a combined distribution of Cauchy and Gaussian distribution to represent the observations in order to remove outliers.

In this study, we have developed and demonstrated a new approach and algorithm to automatically generate water level time series for lakes and reservoirs without user intervention or the use of ancillary data. This algorithm can also be used to generate water level time series for rivers and floodplains. It is based on a combination of K-means clustering, interquartile range (IQR), and statistical error computation. Developing algorithms, such as the one proposed in this study, is critical toward lake and reservoir monitoring at the regional and global scales.

We performed a quantitative assessment of the result using root-mean-square error (RMSE) and R^2 to validate the proposed algorithm. This algorithm can also be used to generate time series for other altimeters, such as SARAL/AltiKa, Senntinel-3, and Jason-3.

II. Data and Methodology

A. Data

In this study, we used Jason-2 and Envisat altimetry satellite data to demonstrate the capability and consistence of this algorithm using 37 satellite crossings over 30 reservoirs in Brazil, Nigeria, and the U.S. chosen based on the availability of *in-situ* data.

We obtained the gauge data for the reservoirs in U.S. from the U.S. Geological Survey website (https://waterdata.usgs.gov/nwis/), while the *in-situ* data in Brazil and Nigeria are not publicly available.

Fig. 1 shows the location of the different lakes and reservoirs with their corresponding names cited in Table I. The next sections give a summary of the altimetry satellites information used in this study. For more detailed information, readers are referred to the Jason-2 handbook [25] and Envisat handbook [26].

1) Environmental Satellite (Envisat) Data: Envisat was built by the European Space Agency and launched into orbit in March 2002. Envisat has an orbital period of 35 days thereby providing a dense satellite track over inland and open water bodies. It has an orbital inclination of 98.55° and was designed to measure the earth's atmosphere and surface. In this study, we used the 18-Hz along-track range data in the geophysical data record (GDR) product which is publicly provided by the Centre National d'Etudes Spatiales (CNES) data center (https://aviso-data-center.cnes.fr/). Further detail about the data processing is given later in this paper.

2) Jason-2 Data: The Jason-2 satellite was launched in June 2008 as a continued mission on TOPEX/Poseidon and Jason-1. Jason-2 follows the same orbital track as TOPEX/Poseidon and Jason-1 with a temporal repeat of approximately ten days. In this study, we obtained the 20-Hz along-track range data in the GDR product [25]. We downloaded the dataset from the CNES archive (ftp://avisoftp.cnes.fr/AVISO/pub/jason-2/gdr_d/). The next section gives more details on the data extraction and processing.

B. Methodology

1) Clustering: Clustering is the processing of classifying datasets into different groups based on a measure of proximity [27]. The process of grouping datasets into different clusters can be further explored to detect outliers in our measurements.

TABLE I
R^2 AND RMSEs OF 37 SATELLITE CROSSING OVER LAKES AND RESERVOIRS IN BRAZIL, NIGERIA, AND THE U.S

	Study Area	Country	R^2	RMSE (m)	Track	Satellite	Crossing Length (km)
1	Capivara Reservoir	Brazil	0.97	0.34	248	Envisat	4.86
2	Emboracacao Reservoir	Brazil	1.00	0.36	620	Envisat	18.89
3	Furnas Reservoir	Brazil	1.00	0.12	549	Envisat	9.33
4	G. B. Munhoz	Brazil	0.99	0.64	435	Envisat	0.64
5	Ilha Solteira Reservoir	Brazil	0.99	0.09	792	Envisat	22.05
6	Ita Reservoir	Brazil	0.59	1.10	248	Envisat	0.97
7	Itaipu Reservoir	Brazil	0.73	0.24	607	Envisat	27.36
8	Itumbiara Reservoir	Brazil	0.93	1.20	177	Envisat	7.56
9	Jurumirim Reservoir	Brazil	0.75	1.08	620	Envisat	5.15
10	Marimbondo Reservoir	Brazil	0.97	0.73	162	Envisat	5.95
11	Ponte Nova Reservoir	Brazil	0.95	0.57	76	Envisat	10.46
12	Promissao Reservoir	Brazil	0.84	0.34	263	Envisat	5.38
13	Sao Simao Reservoir	Brazil	0.79	0.70	263	Envisat	5.09
14	Serra da Mesa Reservoir	Brazil	1.00	0.13	706	Envisat	17.7
15	Sobradinho Reservoir	Brazil	0.99	0.32	663	Envisat	25.75
16	Tres Irmaos	Brazil	0.56	0.61	792	Envisat	7.72
17	Tres Marias Reservoir	Brazil	1.00	0.11	463	Envisat	54.56
18	Tucurui Reservoir	Brazil	1.00	0.11	420	Envisat	71.13
19a	Kainji Reservoir	Nigeria	1.00	0.23	874	Envisat	47.57
19b	Kainji Reservoir	Nigeria	0.99	0.27	135	Jason-2	28.77
20	Beaver Creek Reservoir	U.S.	0.92	0.13	76	Jason-2	2.29
21a	Devils Lake	U.S.	0.66	0.35	93	Jason-2	6.61
21b	Devils Lake	U.S.	0.68	0.19	151	Envisat	8.85
21c	Devils Lake	U.S.	0.76	0.21	196	Envisat	11.07
22	Falls Lake	U.S.	0.72	0.32	738	Envisat	2.38
23	Lake Okeechobee	U.S.	0.92	0.21	465	Envisat	53.16
24	Lake Salvador	U.S.	0.51	0.12	981	Envisat	10.78
25	Monroe Lake	U.S.	0.89	0.12	167	Jason-2	4.46
26	Rathbun Lake	U.S.	0.95	0.27	682	Envisat	5.35
27a	Sam Rayburn Reservoir	U.S.	0.83	0.33	596	Envisat	4.79
27b	Sam Rayburn Reservoir	U.S.	0.92	0.23	695	Envisat	13.72
27c	Sam Rayburn Reservoir	U.S.	0.99	0.17	41	Jason-2	12.2
28	Upper Klamath Lake	U.S.	0.39	0.46	942	Envisat	7.14
29a	Upper red Lake	U.S.	0.45	0.21	940	Envisat	23.17
29b	Upper red Lake	U.S.	0.56	0.16	895	Envisat	23.55
30a	Wheeler Lake	U.S.	0.38	0.59	710	Envisat	5.62
30b	Wheeler Lake	U.S.	0.50	0.45	723	Envisat	6.72

Several studies [28]–[30] have been done using clustering pattern for outlier detection.

Without prior knowledge of the dataset as in the case of a discriminant analysis of clusters, the method of using an unsupervised method of classification in outlier detection can, therefore, be a daunting task. Hence, it is important to understand the definition of outliers in details. Hawkins defines outlier as measurements with anomaly from the rest of the dataset [31].

2) K-Means: K-means clustering has been used in many studies to detect outliers [29], [30]. The K-means clustering is an unsupervised method of classification based on a predefined number of classes [32]. The K-means clustering is an iterative algorithm that partitions a dataset into K numbers of classes. Fig. 2 shows the schematic flowchart of K-means algorithm. Let *n* represents the number of points to be classified and K is the number of clusters. First, the user specifies the value of K (K is a positive integer) as the only input parameter, then K number of points are randomly selected as initialization centroid (mean of observations in a cluster) of the K number of clusters. Second, the remaining points (n − K) are assigned to a cluster based on their proximity (Euclidean distance to the centroid of a cluster) to the centroid of the initialization cluster until the sum of squared distance to the centroid of each cluster has been minimized

$$\min \sum_{x \in C_j} ||x - \mu_j||^2 \quad (1)$$

where C_j is the cluster j, x is the data points that belongs to cluster j, and μ_j is the centroid of cluster j.

Finally, note that during each iteration, the centroid (mean) of each cluster is recomputed and points are moved from one cluster to another. At the end of the iteration, K clusters would have been achieved. Hartigan and Wong [33] describe the process as assigning points to a cluster and minimizing the sum of squared distances within each cluster. This concept was explored due to its competitive advantage of speed [30] to partition satellite altimetry observations to generate water level time series of reservoirs and lakes.

However, some of the known limitations of the classical K-means algorithm is its sensitivity to the initialization centroid which could affect the classification of points [34] and more so, all observations are equally weighted [35]. We used the "K-means" function in MATLAB software (R2015a) which implements the K-means++ algorithm to initialize the centroid.

TABLE I
CONTINUED

	Water Level Amplitude (m)	Latitude Range Min.	Latitude Range Max.	Longitude Range Min.	Longitude Range Max.	No. of Cycles Before	No. of Cycles After	Percentage (%) of Complete Cycle Left	Percentage (%) of Complete Data Left
1	7.76	−22.8251	−22.7740	−51.0461	−51.0330	84	77	92	48
2	24.04	−18.5156	−18.3490	−47.8502	−47.8128	82	82	100	39
3	7.51	−20.8696	−20.8337	−46.1536	−46.1626	82	81	99	55
4	31.40	−26.1268	−26.1198	−51.3120	−51.3138	71	42	59	48
5	2.97	−20.3897	−20.1954	−51.1703	−51.1260	82	80	98	54
6	6.57	−27.2765	−27.2668	−52.1743	−52.1720	67	40	60	49
7	2.09	−25.2639	−25.0062	−54.4071	−54.4737	82	78	95	45
8	16.47	−18.3463	−18.2761	−48.9116	−48.9292	80	79	99	13
9	7.81	−23.2873	−23.2398	−49.0070	−48.9960	81	80	99	29
10	14.57	−20.2323	−20.1759	−48.9793	−48.9644	81	68	84	49
11	16.43	−19.2065	−19.1005	−47.2949	−47.2692	82	69	84	34
12	3.51	−21.4015	−21.3541	−49.6179	−49.6294	82	82	100	58
13	6.75	−18.7098	−18.6622	−50.2628	−50.2734	81	72	89	38
14	24.18	−14.3088	−13.9810	−48.3063	−48.2309	83	80	96	33
15	8.22	−10.0323	−9.7824	−42.2011	−42.2697	80	80	100	49
16	3.67	−20.6790	−20.6067	−51.2455	−51.2258	82	82	100	40
17	14.71	−18.6125	−18.2409	−45.2577	−45.3437	84	82	98	50
18	18.86	−4.3937	−3.7565	−49.6760	49.5356	82	80	98	28
19a	11.62	10.0174	10.5254	4.5242	4.6401	81	81	100	38
19b	11.02	10.3247	10.6316	4.4453	4.5555	258	250	97	50
20	2.52	36.0169	36.0318	−78.6829	−78.6924	256	255	100	92
21a	2.81	48.0429	48.0730	−98.8333	−98.7960	258	253	98	72
21b	1.60	48.0002	48.0648	−99.0040	−98.9805	83	81	98	80
21c	2.04	48.0376	48.0924	−99.0475	−99.0661	82	75	91	67
22	2.55	36.0106	36.0287	−78.7459	−78.7427	80	71	89	67
23	2.71	26.7036	27.0926	−80.7825	−80.8900	83	83	100	78
24	0.86	29.7090	29.7944	−90.2019	−90.2242	83	79	95	51
25	1.70	28.8124	28.8457	−81.2537	−81.2366	260	260	100	87
26	5.64	40.8401	40.8780	−93.0243	−93.0126	82	77	94	65
27a	3.78	31.3172	31.3578	−94.4668	−94.4559	85	85	100	47
27b	3.09	31.0875	31.1989	−94.1694	−94.1990	79	78	99	45
27c	6.51	31.0963	31.2107	−94.2329	−94.1797	258	258	100	68
28	2.63	42.4185	42.4638	−121.9484	−121.9328	84	78	93	50
29a	1.36	48.0679	48.2126	−94.6680	−94.6060	85	78	92	86
29b	1.05	48.0580	48.2011	−94.7474	−94.8070	82	72	88	80
30a	3.72	34.6481	34.6965	−87.0534	−87.0423	83	81	98	33
30b	2.79	34.7419	34.7917	−87.3018	−87.3147	79	71	90	55

Arthur and Vassilvitskii [36] developed the K-means++ algorithm and compared their method of choosing the initialization centroid to the random method used in classical K-means. They concluded that the K-means++ increased the stability of the classical K-means method. In this study, we did not explicitly verify this claim, however, repeated computation of RMSE over our study areas yielded a consistent result. Hence, we could infer that the limitations due to the initialization centroid has been addressed in the "K-means" function in MATLAB.

3) Data Extraction and Processing: Prior to extracting Jason-2 and Envisat data over the study areas, we overlaid the nominal altimetry tracks over Google earth image to determine an estimate of the latitude range of the overlap (see Table I, Fig. 3(b)) to truncate the dataset. It is imperative to note that some of the reservoirs used in this study (most especially in Brazil) has more than one satellite crossing. In such cases, we selected the longest track over the reservoir.

While extracting the water elevation data from the raw file, we discarded measurements with retracked range quality flags, and defaulted latitude and longitude values as recommended in [21]. This was necessary to ensure that the outliers detected were not due to the failure to remove the measurements that have been flagged either due to instrument error or data quality checks.

In this study, we used the Ice-1 retracked range measurement for Envisat, which is considered most suitable for inland water [3], and Ice retracked range for Jason-2 [37]. The equation below was used to compute the elevation above the reference datum:

$$H_{\mathrm{corr}} = Alt - [R - \Delta R + A_{\mathrm{wet}} + A_{\mathrm{dry}} + A_{\mathrm{iono}} + T_p + T_E + T_L] \quad (2)$$

where H_{corr} is the corrected elevation, *Alt* is the satellite altitude above the reference ellipsoid, *R* is the measured range to the surface of the water; $A_{\mathrm{wet}}, A_{\mathrm{dry}}, A_{\mathrm{iono}}$ are the corrections for the wet troposphere, dry troposphere, and ionosphere, respectively; T_P, T_E, T_L are corrections for pole, earth, and loading tides, respectively, and ΔR is the retracked range correction.

In case of Jason-2, we subtracted 0.7 m from H_{corr} to convert the reference from Topex ellipsoid (Jason-2) to WGS-84

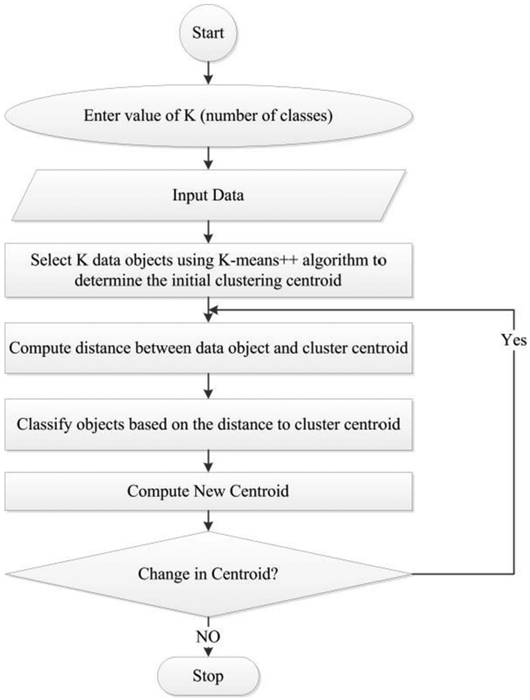


Fig. 2. K-means Flowchart modified from [1].

ellipsoid [38], [39] for consistency. Finally, this H_{corr} referenced to WGS-84 ellipsoid was subsequently passed into the algorithm. The next section gives a detailed explanation of the algorithm.

4) Outlier Removal: Prior to generating the time series from the cycles, we do not have *a priori* information on the range of the elevation of the reservoirs. Consequently, it becomes challenging to mask the elevation without using any ancillary data. This section explains in details the algorithm used for the automated generation of altimetry time series. Fig. 4 shows the flowchart of the outlier removal algorithm that was implemented in this study.

a) For clarity, the *complete* dataset in this context refers to all height measurements within the delineated lake or reservoir boundary for all cycles, while the *sample* dataset refers to the cycle under consideration. The first phase of the outlier removal after the data extraction was the removal of outliers from the complete dataset using the equations shown in (4) and (5) to remove extraneous outliers. *Q1* and *Q3* represent the first and third quartiles of the data, respectively, and *IQR* is computed from the difference between *Q3* and *Q1*

$$IQR = Q3 - Q1 \tag{3}$$

$$\text{Lower} = Q1 - 1.5 \cdot IQR \tag{4}$$

$$\text{Upper} = Q3 + 1.5 \cdot IQR. \tag{5}$$

For illustration, Kainji reservoir in Nigeria with Jason-2 Pass 135 (see Fig. 3(a)) dataset was used to explain this step. The lower (see (4)) and the upper limits (see (5)) of the complete dataset were 132.58 and 199.22 m, respectively. Measurements above and below the computed upper and lower limits, respectively, were removed.

b) The second phase of this algorithm was performed on the sample dataset, i.e., each cycle. In this stage, the K-means clustering was used to identify the clusters. As earlier mentioned in Section I, *a priori* knowledge of the number of clusters is necessary. Then, the critical question arises; what is the appropriate K value for this algorithm? We here set K as 2. This value was used because the measurements could either be classified as "good" or "bad." Note that the "good" or "bad" referred here does not inherently mean water and land signals, respectively. This phase was iterative till the statistical range (SR, difference between maximum and minimum height) of the heights was within a specified threshold of 5 m. Further explanations will be provided later in Section II-B 5 on the choice of this threshold value.

While the SR of the height was above the set threshold, the cluster with fewer observations was discarded. This process can be referred to as the majority vote based on the assumption that the larger cluster has a higher likelihood of being the *right* measurement. At the end of this iteration, one cluster would have been achieved.

c) Finally, the mean was computed from the cluster from the previous section and the deviation from the mean was also computed. At each iteration, the largest deviation from the mean was computed and removed until the standard deviation (std) threshold of 0.3 m was achieved. We discussed in details the choice of the threshold value of 0.3 m std and 5 m SR in Section II-B5. Finally, the average along-track height was then computed and used in the time series. These processing steps were repeated for all the cycles to generate the time-series plot.

5) SR and Standard Deviation Threshold: Using Jason-2, Pass 135 over the Kainji reservoir (see Fig. 3(b)) as an illustration, Fig. 5(a) shows that the RMSE, which was computed using *in-situ* gauge, is more sensitive to the SR of the height compared to the standard deviation used in the flowchart. Although we could obtain a lower RMSE by choosing a lower range and standard deviation, the tradeoff can be seen in Fig. 5(b) and (c), where the percentage of outliers detected increases as the RMSE reduces. Fig. 5(b) shows that a lower range and standard deviation results in fewer samples of dataset after outliers have been removed. For instance, if we used a SR of 5 m and a std of 0.3 m in the algorithm (see Fig. 4, flowchart), the RMSE obtained from Fig. 5(a) is approximately 0.2 m and the corresponding percentage data detected as outliers is approximately 30% (see Fig. 5(b) and (c)). On the other hand, if a SR of 8 m and std of 0.8 m were used in the algorithm, we obtain RMSE of about 0.4 m with an average data loss of approximately 20%.

Hence, for this paper, we decided to choose a conservative value of 0.3 and 5 m for the standard deviation and range, respectively, to process both the Envisat and Jason-2 time series. Note that the threshold values of 0.3 and 5 m are not necessarily optimal to obtain the best RMSE. Hence, users can modify these values to obtain a desired time series taking cognizance of the tradeoff in percentage of outliers and RMSE. The plots in Fig. 5 might be slightly different for different reservoirs, but the overall trend is expected to be consistent.

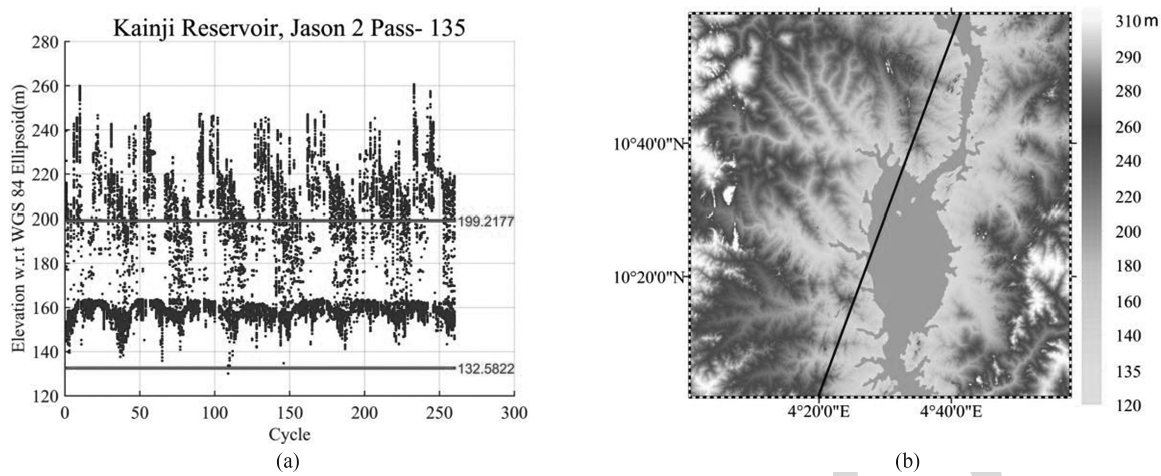


Fig. 3. (a) Upper and lower limits for outlier removal from the complete dataset of Kainji reservoir, Jason-2 Pass 135. (b) SRTM DEM over the Kainji reservoir, Nigeria. The black line represents Jason-2 Pass 135.

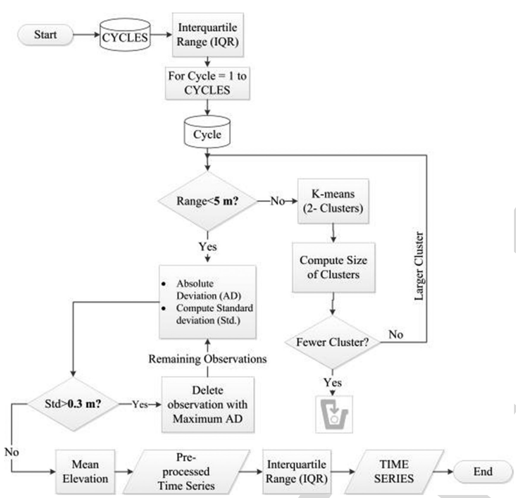


Fig. 4. Schematic diagram for the automatic generation of time series and outlier removal.

III. Result

In order to validate the water level time series generated using the proposed algorithm, we obtained the daily *in-situ* data for the lakes and reservoirs (see Section II-A).

The daily *in-situ* gauge data was downsampled to correspond to the date of the satellite passes and the difference between the resampled *in-situ* and the altimetry water level time series was computed. The mean difference was used to reduce the *in-situ* gauge to the altimetry water level for comparisons [40]. In addition, we performed a quantitative assessment of the water level time series of the algorithm using RMSE as the metrics.

We compared the altimetry water level time series to the gauge observations of reservoirs and lakes (Fig. 6 corresponds to Nigeria, Fig. 7 to Brazil, and Fig. 8 to U.S.). The error bar for the altimetry water level was not shown in the figures for clarity in illustration. However, recall from the algorithm description that the maximum standard deviation for each cycle was limited to 0.3 m.

In addition, from Figs. 6–8, we can also infer the stability of the proposed algorithm under varying climatic conditions and possible dam-controlled operations that can potentially compromise on the outlier detection capability. It is important to evaluate the stability of outlier detection algorithms based on the factors listed below as they can potentially affect the distribution of measurements which can lead to false detection as outliers

1) *First, protracted increase or decrease in time-series trend:* The continuous increase or decrease in water level can be attributed to prolonged flooding or drought, respectively. Fig. 7(1B) captures the extended drought period in Emboracacao reservoir, Brazil from 2010 with 100% of the dates (cycles) retained (see Table I). On the other hand, Fig. 8(4C) shows a protracted increase in the water level for the period of November 2014–July 2015 of approximately 21.09–24.44 m. Prior to 2014, the maximum water level from 2008 was 22.42 m. Despite the continuous increase in the water level, the proposed algorithm retained 100% (see Table I) of the dates (cycles) after outlier detection without any error due to commission.

2) *Second, momentary flood or drought:* Figs. 7(1C), 7(6C), and 8(4C) show the drought approximately in year 2007, 2003, and 2012, respectively. The percentage number of dates after the outlier detection was 99%, 98%, and 100% for Figs. 7(1C), 7(6C), and 8(4C), respectively.

3) *Third, impoundment due to reservoir operations:* Although we do not have information to substantiate the impoundment on the reservoirs, from Fig. 7 (4B, 5B, 6B), we can hypothetically state that the sudden increase which is then accompanied by a steady time series simulates the impoundment of reservoirs. A close examination of Tres Marias and Serra da Mesa reservoir in Fig. 7(6B) and (5B), respectively, shows a continuous increase in the water level from 2002 to 2006 before reaching a steady state. Despite this change in reservoir water levels, the proposed algorithm retains 98% and 96% of the available dates (cycles) over Tres Marias and Serra da Mesa reservoir, respectively (see Table I).

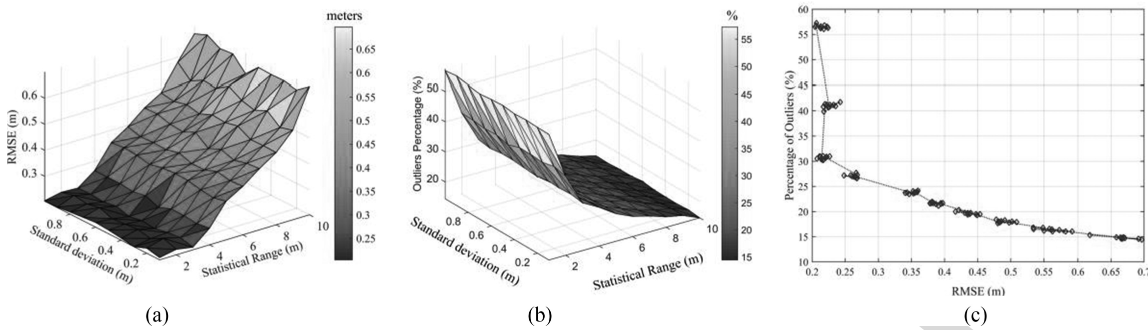


Fig. 5 Analysis of Jason-2, Pass 135 data (Kainji reservoir) in choosing a threshold value: (a) 3-D surface plot showing how RMSE, range, and standard deviation in the outlier detection algorithm varies. (b) Corresponding effect of range and standard deviation on the average percentage of data after outlier removal. (c) Percentage of outliers removed and its variation with respect to RMSE.

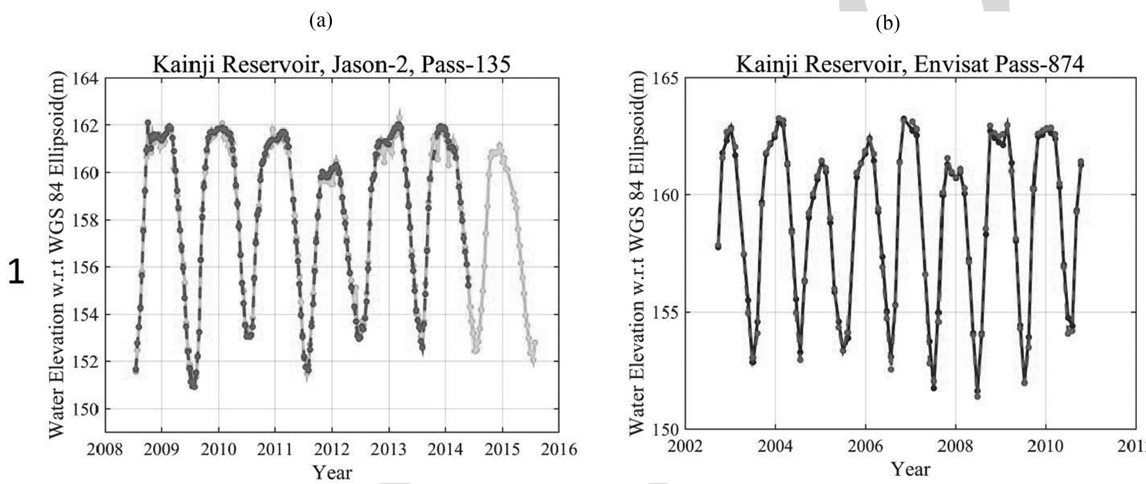


Fig. 6. Comparison of *in-situ* gauge observation (red) and altimetry-derived (Jason-2: green, Envisat: blue) water level of Kainji reservoir, Nigeria referenced to WGS 84 Ellipsoid.

4) *Finally, what if there are no errors in the observation?* The analyses in this section is incomplete without considering the tendency of false detection of dates (cycles) as outliers when the actual observations do not contain outliers, i.e., the error due to commission. Consequently, we examine the Beaver Creek reservoir (see Table I), which is deemed most appropriate for this analysis since it has the highest percentage (92%) of data after the outlier detection. A case of no outlier in the dataset implies that 100% of the data should be retained after running through the outlier algorithm. Nonetheless, we explained earlier in Section II-B5 how the choice of the threshold used in the algorithm impacts the percentage of dataset after outlier detection. Hence, if the raw dataset is within the thresholds then, 100% of the complete dataset will eventually be used without any observations detected as outliers. In conclusion, the proposed algorithm is not susceptible to error due to commission.

The different scenarios exemplified above do not intend to generalize on the performance of the proposed algorithm in those circumstances but to highlight its flexibility and effectiveness in handling such unique conditions.

Table I shows the summary of the quantitative assessment of all the reservoirs used in this study. It can be observed that most of the reservoirs have high R^2 and good RMSE value. Nonetheless, some of the lakes can be observed to have low R^2 (0.4–0.5), but good RMSE value (0.16–0.59 m). In general, the lakes with low R^2 values also had low water level fluctuation, as shown in Table I. We performed further investigation (see Section IV) to ascertain the reason for the poor RMSE of 1.2 m obtained in the Itumbiara reservoir with over 80% of the data detected as outlier.

Table II shows RMSE values computed for four different retracking algorithms applied to Envisat GDR to evaluate the performance of the proposed algorithm. Using the RMSEs, the ranks of the retracking algorithms were determined for the lakes and reservoirs. The rank was based on a scale of 1–4 which, represents the retracking algorithms; 1 being the lowest or the best RMSE value, while 4 represents the highest or worst RMSE value. It can be observed from Table III below that Ice-1 retracking algorithm outperformed the other retrackers over the study regions with an average rank of 1.3 which represents 77.8% of the 26 Envisat crossings used in this study. This is consistent with the study done by Frappart *et al*. [3] and Da Silva *et al.* [41] that Ice-1 retracked range is the best. Ocean retracking algorithm has the least performance with an average rank of

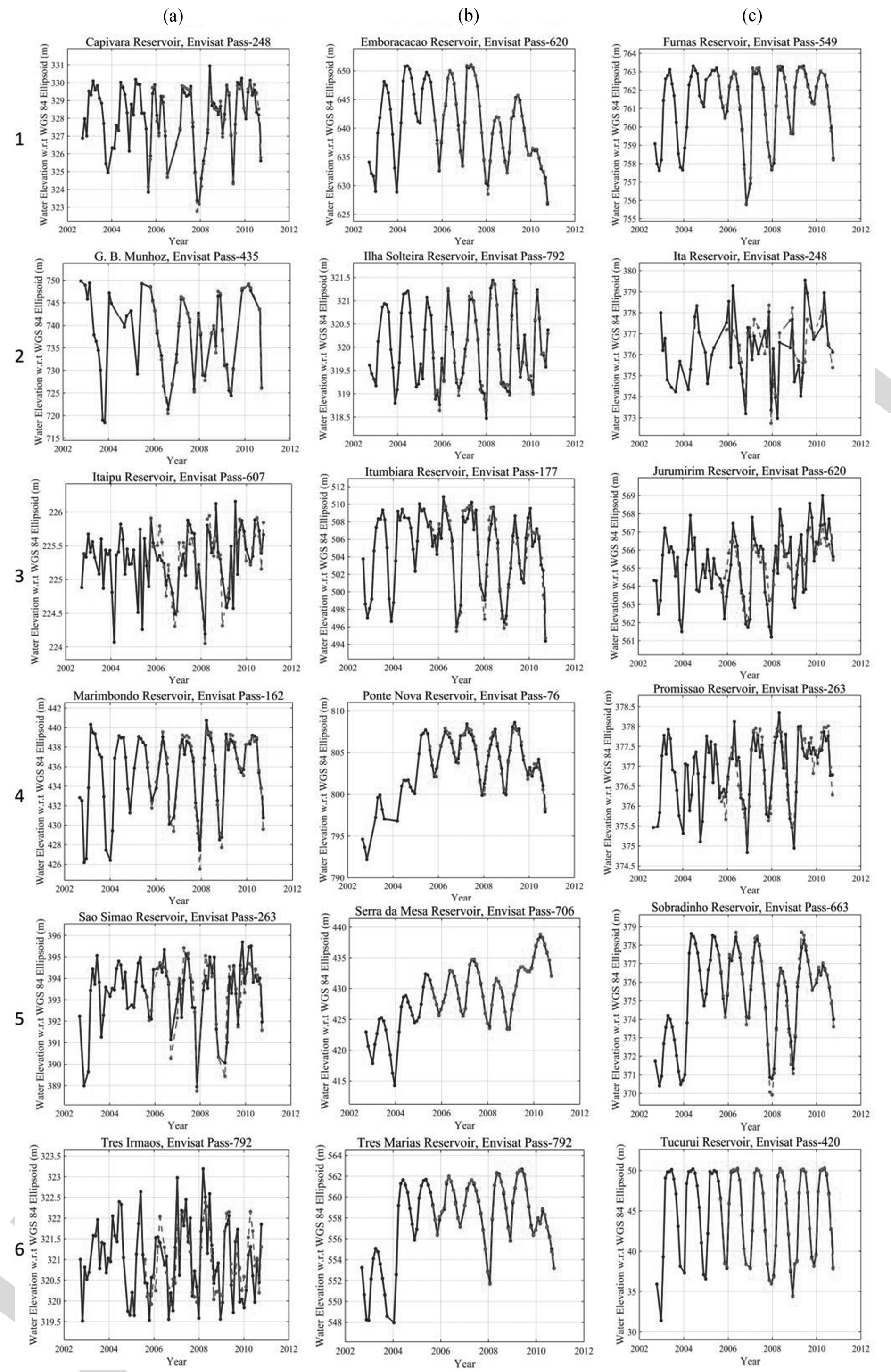


Fig. 7. Comparison of *in-situ* gauge observation (red) and altimetry-derived (Envisat) water level (blue) of 18 reservoirs and lakes in Brazil.

3.4. Hence, Ice-1 retracking algorithm shows more consistent performance in studying the in-land water bodies.

IV. Discussion

A. Time Series and Algorithm Analyses

The high RMSE values reported on some of the reservoirs could be attributed to diverse reasons; contamination of the altimetry measurements due to orbital error, retracking error [42], surrounding topography [10], etc. It is challenging to isolate these sources of errors or attribute it to the shortcoming of the algorithm.

Consequently, we performed in-depth analyses over the Itumbiara reservoir which has the worst RMSE of 1.2 m and the Kainji reservoir with a relatively good RMSE of 0.27 m to investigate the selection pattern of the proposed algorithm. For both reservoirs, we evaluated the first high and low water cycles of the time series indicated in Fig. 9.

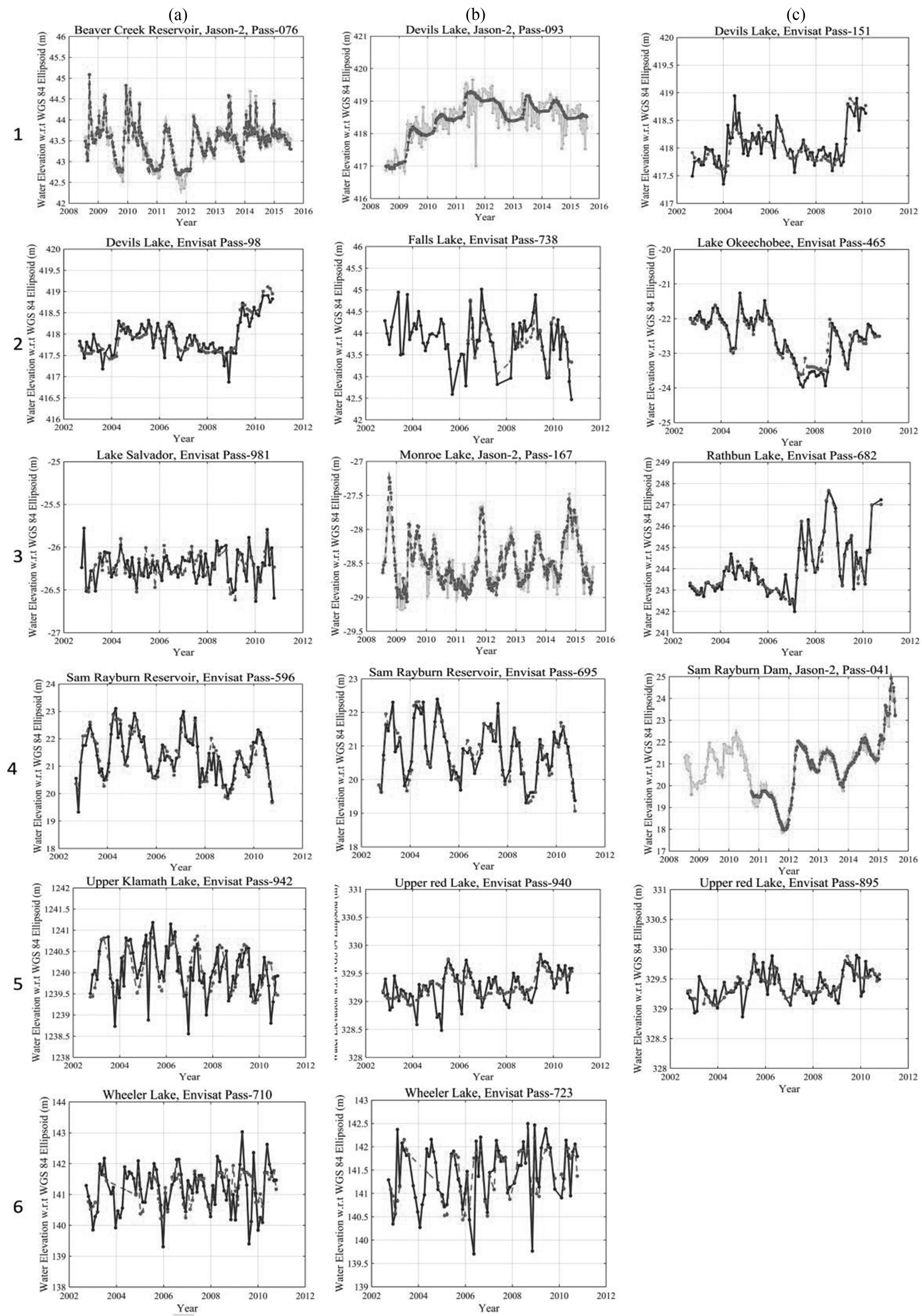


Fig. 8. Comparison of *in-situ* gauge observation (red) and altimetry-derived (Envisat: blue, Jason-2: green) water level of 11 lakes in the U.S.

Fig. 10(a) and (c) shows the dispersion of the measured height over the Itumbiara reservoir both in the first high and low water cycles, respectively. Previous studies have indicated that the complexity in shape of inland water bodies and the contamination due to surrounding topography affect the quality of radar altimetry measurements [10]. The Landsat 8 image used at the background shows the complexity of the lake boundary. Hence, it is speculated that the high RMSE of 1.2 m obtained in Itumbiara is due to the aforementioned reasons. In addition, Fig. 10(d) shows fewer observations in the low water period than Fig. 10(b) due to the data quality checks described in Section II-B3.

Fig. 11(a) and (c) over the Kainji reservoir shows clearly two different clusters from visual inspection. The data gaps in the satellite altimetry track in Fig. 11(d) are due to the land mask and quality flags. We further examined the spatial distribution

TABLE II
RMSE VALUES OF 26 ENVISAT SATELLITE CROSSINGS OVER LAKES AND RESERVOIRS IN THE U.S. AND BRAZIL USING ICE-1, ICE-2, OCEAN, AND SEA-ICE RETRACKING ALGORITHMS

		RMSE (m)			
	Country	Ice-1	Ice-2	Ocean	Sea-Ice
Ilha Solteira Reservoir	Brazil	**0.09**	0.14	0.26	0.17
Tucurui Reservoir	Brazil	**0.10**	0.11	0.45	0.13
Tres Marias Reservoir	Brazil	**0.12**	0.18	0.33	0.22
Furnas Reservoir	Brazil	**0.12**	0.27	0.28	0.30
Serra da Mesa Reservoir	Brazil	**0.13**	0.20	0.35	0.23
Itaipu Reservoir	Brazil	**0.24**	0.41	0.44	0.49
Sobradinho Reservoir	Brazil	0.32	**0.28**	0.57	0.34
Emboracacao Reservoir	Brazil	**0.33**	0.44	0.74	0.64
Promissao Reservoir	Brazil	**0.34**	0.59	0.53	0.62
Ponte Nova Reservoir	Brazil	0.56	2.81	0.68	**0.45**
Tres Irmaos	Brazil	0.61	**0.50**	0.66	0.51
G. B. Munhoz	Brazil	0.64	**0.58**	6.95	0.61
Sao Simao Reservoir	Brazil	**0.70**	0.92	0.99	1.42
Lake Salvador	US	**0.12**	0.28	0.23	0.36
Upper red Lake	US	**0.16**	0.27	0.31	0.28
Devils Lake	US	**0.19**	0.33	0.40	0.35
Devils Lake	US	**0.21**	0.34	0.40	0.49
Upper red Lake	US	**0.21**	0.31	0.37	0.33
Lake Okeechobee	US	**0.21**	0.22	0.22	0.22
Sam Rayburn Reservoir	US	**0.23**	**0.23**	0.31	0.27
Rathbun Lake	US	0.27	0.29	0.47	**0.20**
Falls Lake	US	**0.32**	0.36	0.42	0.37
Sam Rayburn	US	**0.33**	0.59	0.39	0.52
Upper Klamath Lake	US	**0.47**	0.75	0.93	0.93
Wheeler Lake	US	**0.48**	0.52	0.52	0.54
Wheeler Lake	US	**0.61**	0.76	0.69	0.75

TABLE III
MEAN RANK BASED ON THE RMS ERROR OF THE 26 ENVISAT PASSES OVER LAKES AND RESERVOIRS IN THE U.S. AND BRAZIL

	Ice-1	Ice-2	Ocean	Sea-Ice
Mean Rank	1.3	2.3	3.4	3.0
First Rank (%)	77.8	11.1	0.0	11.1

of the outliers based on the high and low water elevation using Landsat 8 images that coincides with the water level. Pass 135 shown in Fig. 11(b) and (d) is the ascending ground track and Jason-2 tracking unit may remain locked to the reservoir shore as it crosses from land to open water which resulted into error in water elevation measurements [21]. The phenomenon perhaps explains why most of the outliers are located on the lower part of the track which is further worsened by the proximity to land. In summary, during the low water period in cycle 1, see Fig. 11(d), more observations were detected as outlier compared to the high water period in Fig. 11(b).

In order to substantiate the need to remove outliers from the complete dataset, further critical analyses were performed over the Kainji Reservoir using Jason-2, Pass 135 cycles 57 and 10 (see Fig. 12) as an example. First, Fig. 12(a) shows the histogram distribution of elevation measurements in cycle 57. Without prior knowledge of the elevation, a quick glance at this cycle visually or using any clustering algorithm will detect measurements between 160 and 180 m, as the outliers represent only 14% of the dataset in the cycle. This was due to the highest frequency of the measurements being within 220–240-m range, which corresponds to a greater proportion of the sample dataset. Recall from Fig. 9(a) that the actual elevation is below 163 m. Hence, Fig. 12(a) and (b) shows a classic example of the need to perform the removal of outliers in the complete dataset before removing outliers in the sample dataset (each cycle).

Similarly, Jason-2, Pass 135 cycle 10 from Fig. 12(b) shows that none of the measurements observed was within the elevation range of 150–163 m (see Fig. 9(a)). Thus, performing outlier detection on the complete dataset will eliminate all the measurements in this cycle. The measurements would have otherwise been averaged if the outlier detection was only performed on this cycle or sample dataset without due consideration of the complete dataset as a significant indicator of the outliers.

Sulistioadi *et al*. [20] described the process of removing outliers in altimetry measurements by simply using IQR. However, this method may not effectively remove the outliers in the dataset. We examined some scenarios to further substantiate the shortcoming of using IQR exclusively in removing outliers. Fig. 13 shows the water level time series generated using the IQR. These scenarios are not uncommon when the IQR method is used depending on the noise level of the radar altimetry measurement. When compared with the result of the algorithm in Table I, Figs. 7 and 8, we observe a significant improvement and consistence in the proposed algorithm from the RMSEs.

Furthermore, we explored the potential of using the altimeter's backscattering coefficients (BC) to aid in the detection of outliers using the Kainji reservoir (Jason-2, Pass 135) by examining the distribution pattern of the altimeter BC. From Fig. 14(a), it can be observed that the pattern of the time series is clear and outliers are distinguishable. In order to analyze the dataset, we set an arbitrary threshold of 170 m which clearly discriminates the noise from the time-series pattern observed over all cycles. Fig. 14(b) shows the histogram distribution of Jason-2 BC below (red line) and above (blue color) the arbitrarily set threshold of 170 m. It can be observed that the highest frequency is within the range of 20 dB which has an overlap with the range of BC from elevations above 170 m. Therefore, if we used 20 dB as a threshold to distinguish between *true* elevations and outliers, 49.5% of the *complete* dataset will be classified as outliers. This represents 34% of the probable *true* elevation measurement, i.e., the measurements below arbitrary 170 m earlier mentioned. Moreover, Birkett and Beckley [21] stated that the BC is a highly variable quantity which indicates conditions, such as wind and ice, and this was also reinforced by Kleinherenbrink *et al*. [43]. As a result, we did not incorporate BC as a part of the outlier detection criteria in our algorithm.

Finally, using the result shown in Table I, we investigated the relationship between the RMSEs and the lengths of the satellite crossings on the reservoirs as in Fig. 15. We infer that small satellite crossings have a larger tendency to have high RMSE when compared to its longer counterpart. However, it can be seen from the scatter plot that lower RMSEs can also be obtained in case of short cross lengths.

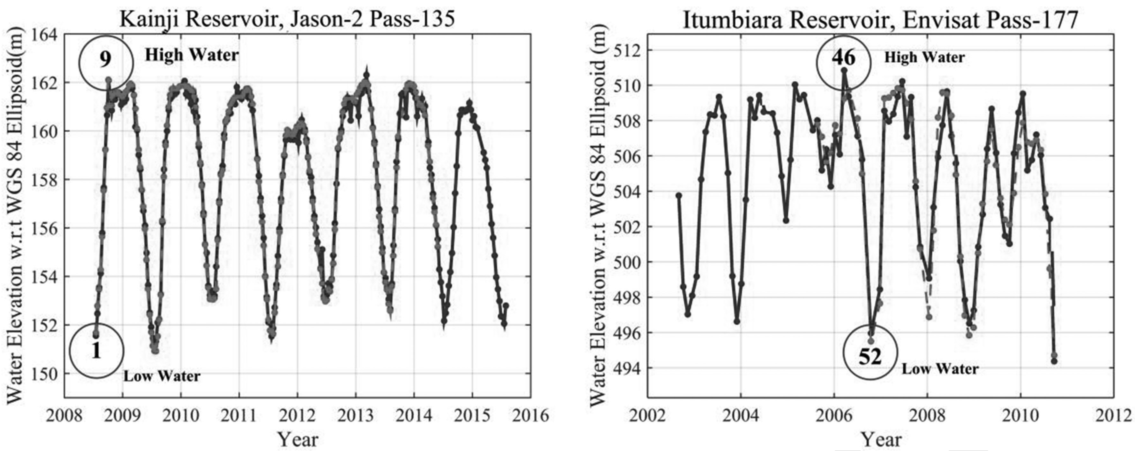


Fig. 9. (a) Cycles 1 and 9 correspond to the low and high water seasons, respectively, of the Kainji Reservoir from Jason-2 Pass 135. (b) Cycles 46 and 52 correspond to the high and low water seasons, respectively, of the Itumbiara Reservoir from Envisat Pass 177.

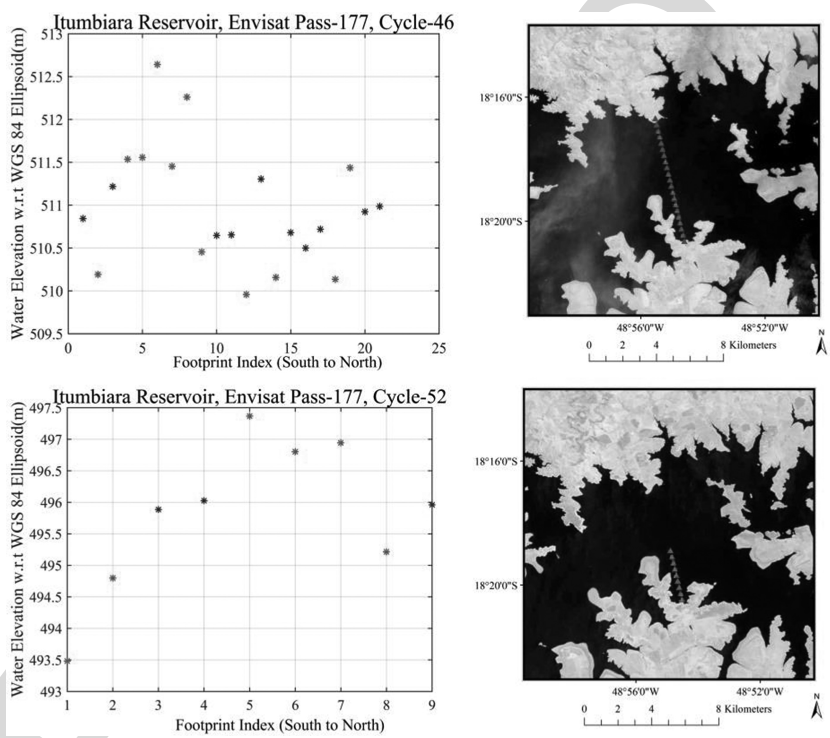


Fig. 10. Itumbiara Reservoir, Brazil: The red and blue represent the outliers detected and the averaged measurements, respectively. (a) High Water, Cycle 46 (March 20, 2006): Distribution of outliers and averaged measurements. (b) High Water, Cycle 46: Spatial distribution of outliers and averaged measurements. Background is Landsat-8 true color image acquired on March 31, 2016. (c) Low Water Cycle 52 (October 15, 2006): Distribution of outliers and averaged measurements. (d) Low Water Cycle 52: Spatial distribution of outliers and averaged measurements. Background is Landsat-8 true color image acquired on October 7, 2015.

B. Prospect of the Proposed Algorithm Over River and Wetlands

Since the algorithm is not restrained by the type of inland water body, we took a further step in testing the prospect of the algorithm over a river and wetland. Based on the availability of *in-situ* data, we used the Congo River in Africa and the Everglades wetland in Florida [2]. Yuan *et al*. [2] validated the performance of Envisat altimetry over the aforementioned locations.

Fig. 16 shows the result obtained using our proposed algorithm in generating water level time series. We compared the result with *in-situ* gauge and obtained RMSE and R^2 of 0.44 and 0.88, respectively, over the Congo River. On the other hand, in the Everglades wetland, we obtained RMSE and R^2 of 0.11 and 0.73, respectively. Yuan *et al*. [2] reported an RMSE of 0.35 m and R^2 of 0.95 in the Congo River and RMSE of 0.12 m and R^2 of 0.83 over the Everglades wetland, Florida. The results obtained using the proposed algorithm is

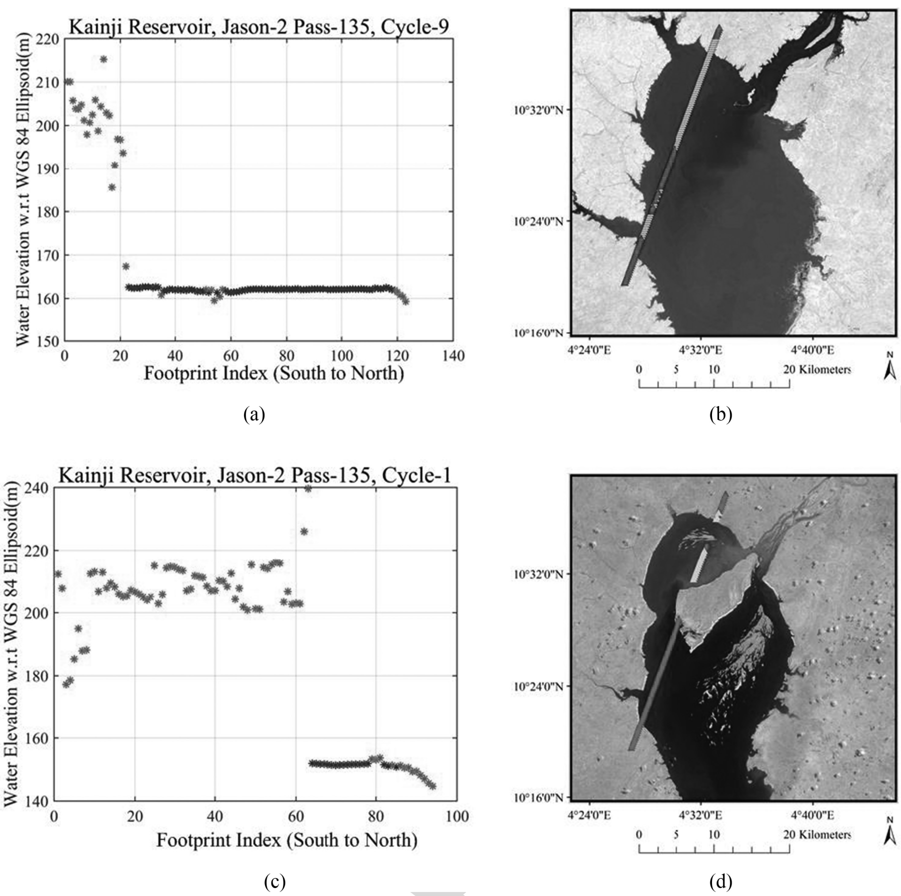


Fig. 11. Kainji Reservoir, Nigeria: The red and blue represents the outliers detected and the averaged measurements, respectively. (a) High Water, Cycle 9 (October 4, 2008): Distribution of outliers and averaged measurements. Background is Landsat-8 true color image acquired on November 22, 2015. (b) High Water, Cycle 9: Spatial distribution of outliers and averaged measurements. (c) Low Water Cycle 1 (July 17, 2008): Distribution of outliers and averaged measurements. (d) Low Water Cycle 1: Spatial distribution of outliers and averaged measurements. Background is Landsat-8 true color image acquired on July 1, 2015.

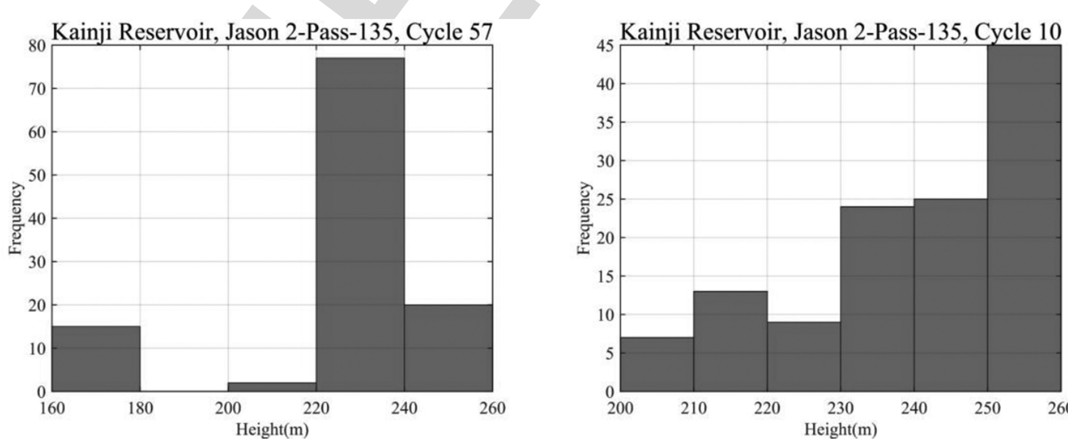


Fig. 12. (a) Histogram distribution of elevation measurements of Jason-2, Pass-135 Cycle 57. (b) Histogram distribution of elevation measurements of Jason-2 Pass-135 Cycle 10.

comparable to the manual approach with 100% of dates (cycles) retained in both cases in the Congo River, while 98% of dates (cycles) in the Everglades wetland. These results show the prospect of this algorithm over rivers and wetlands. However, an extensive study was not performed on other areas to verify this claim.

C. *Comparison of the Proposed Algorithm With Global Water Level Web Databases*

In order to show the competitiveness of the proposed algorithm, we compared the result of the algorithm with the time series publicly available from the websites of Database

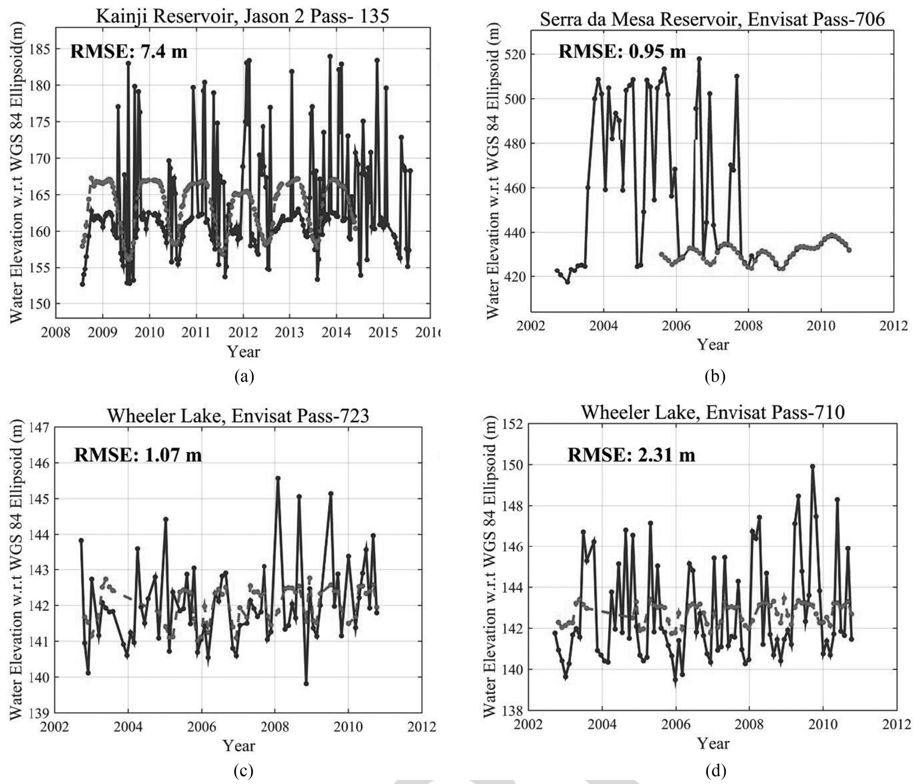


Fig. 13. Satellite altimetry water level time series derived from conventional IQR outlier editing. Red and blue colors represent *in-situ* observation and altimetry-derived (Envisat and Jason-2) water level, respectively.

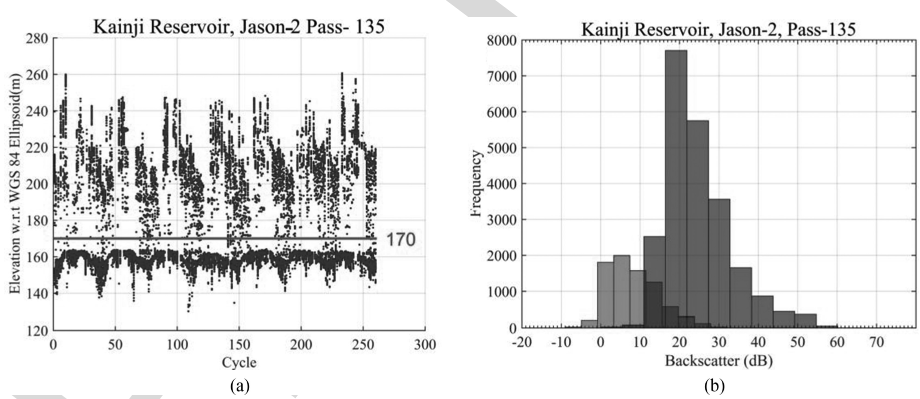


Fig. 14. Analyses of Kainji reservoir, Jason-2 Pass 135. (a) Data plots of the elevation measurements before outlier removal for all cycles over the Kainji reservoir. (b) Histogram distribution of Jason-2's BC before outlier removal of all elevation measurements from all cycles over the Kainji reservoir. The red bars show all BC above the threshold line of 170 m in (a) while the blue bars show BC below the threshold of 170 m.

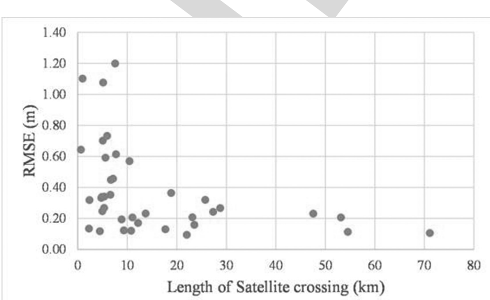
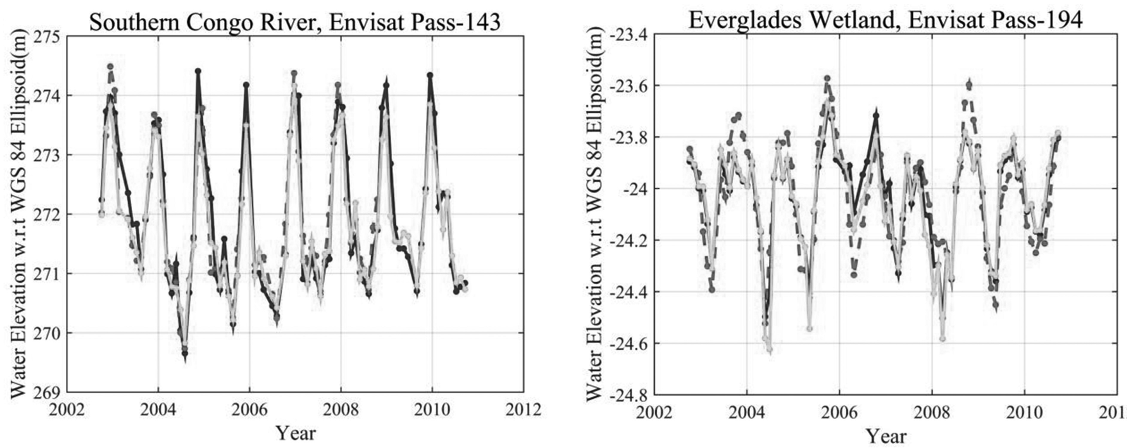


Fig. 15. Scatter plot of RMSEs and lengths of satellite crossings over reservoirs.

for Hydrological Time Series of Inland Waters (DAHITI) [23] (http://dahiti.dgfi.tum.de/en/), Hydroweb [6] (http://hydroweb.theia-land.fr/), River and Lake (http://tethys.eaprs.cse.dmu.ac.uk/RiverLake/shared/main), and U.S. Department of Agriculture (USDA, http://www.pecad.fas.usda.gov/cropexplorer/global_reservoir/) [12]. A detailed description of the data processing methods used in each of the altimetry time series web database can be found in the references. Since the aforementioned websites did not have the time series for all the lakes and reservoirs used in this study, we have limited our comparison to the common lakes and reservoirs.



(a) (b)

Fig. 16. Water level time series generated using the proposed algorithm in blue, manual method used by Yuan *et al*. [2] in green, and *in-sit*u data in red. (a) Congo River, Envisat Pass 143 4.300° S 15.499° W. (b) Everglades Wetland in Florida. Envisat Pass 194, 25.515° N 80.910° W.

TABLE IV
STATISTICAL COMPARISON OF RIVER AND LAKE, USDA, DAHITI, HYDROWEB, AND THE PROPOSED ALGORITHM W.R.T *IN-SITU* GAUGE MEASUREMENTS

Number of available		Altimetry Web Database			Proposed Algorithm			
				Number of remaining				Number of remaining
Dates (Cycles)	Lake/Reservoir	RMSE	R^2	Dates (Cycles)	Source	RMSE	R^2	Dates (Cycles)
82	Emborcacao	1.32	0.94	72	USDA	0.36	1.00	82
82	Tucurui	0.32	1.00	74	USDA	0.11	1.00	80
83	Okeechobee	0.21	0.91	78	USDA	0.21	0.92	83
82	Kainji	1.22	0.90	76	USDA	0.23	1.00	81
82	Ilha Solteira	0.21	0.93	NA	DAHITI	0.09	0.99	80
83	Okeechobee	0.19	0.92	NA	DAHITI	0.21	0.92	83
81	Kainji	0.26	0.99	NA	DAHITI	0.23	1.00	81
84	Tres Marias	0.36	0.98	79	HydroWeb	0.11	1.00	82
82	Ilha Solteira	0.36	0.84	60	HydroWeb	0.09	0.99	80
81	Kainji	0.42	0.99	81	HydroWeb	0.23	1.00	81
82	Furnas	0.60	0.95	81	HydroWeb	0.12	1.00	81
82	Tucurui	0.72	0.98	74	HydroWeb	0.11	1.00	80
82	Tucurui	0.16	1.00	76	River & Lake	0.11	1.00	80
81	Kainji	0.58	0.98	78	River & Lake	0.23	1.00	81

The number of available dates in DAHITI is not applicable (NA) due to the temporal resampling of the time series.

Fig. 17 shows a comparison of DAHITI, Hydroweb, River and Lake, and USDA with the *in-situ* gauge data. From Table IV, we observe that RMSEs obtained from our proposed algorithm in most cases outperformed the existing publicly available altimetry time series based on the limited common samples used for comparison. The proposed algorithm is not aimed as a surrogate to the web time series, but a supplementary tool for stakeholders to process any inland water body of their choice without limitations due to availability of public dataset. It is also important to emphasize on the reproducibility and the simplicity of the proposed algorithm compared to the convoluted method, for example, used in DAHITI without compromising on accuracy.

In addition, we also compared the number of cycles (dates) deleted either by error of omission or commission in the process of removing outlier (see Table IV). Except in Hydroweb (Tucurui and Kainji), the proposed algorithm consistently has more number of dates without compromising on higher accuracy of the lakes (see Table IV). We are unable to compare the number of dates in the proposed algorithm with DAHITI product due to its temporal resampling.

The next section highlights some of the limitations we observed in our proposed algorithm.

D. Limitations of the Proposed Algorithm

We observed some limitations in the algorithm during the data processing. First, in some rare cases, where the two clusters from the K-means have equal number of observations, then the selection due to the majority vote becomes random. For future work, further improvement in the selection criteria should include the variance of each cluster. This is based on the premise that the outliers have a larger variance when compared to the actual observations.

Furthermore, from Fig. 8(1B), between years 2011 and 2012, we can observe an outlier at an elevation of approximately

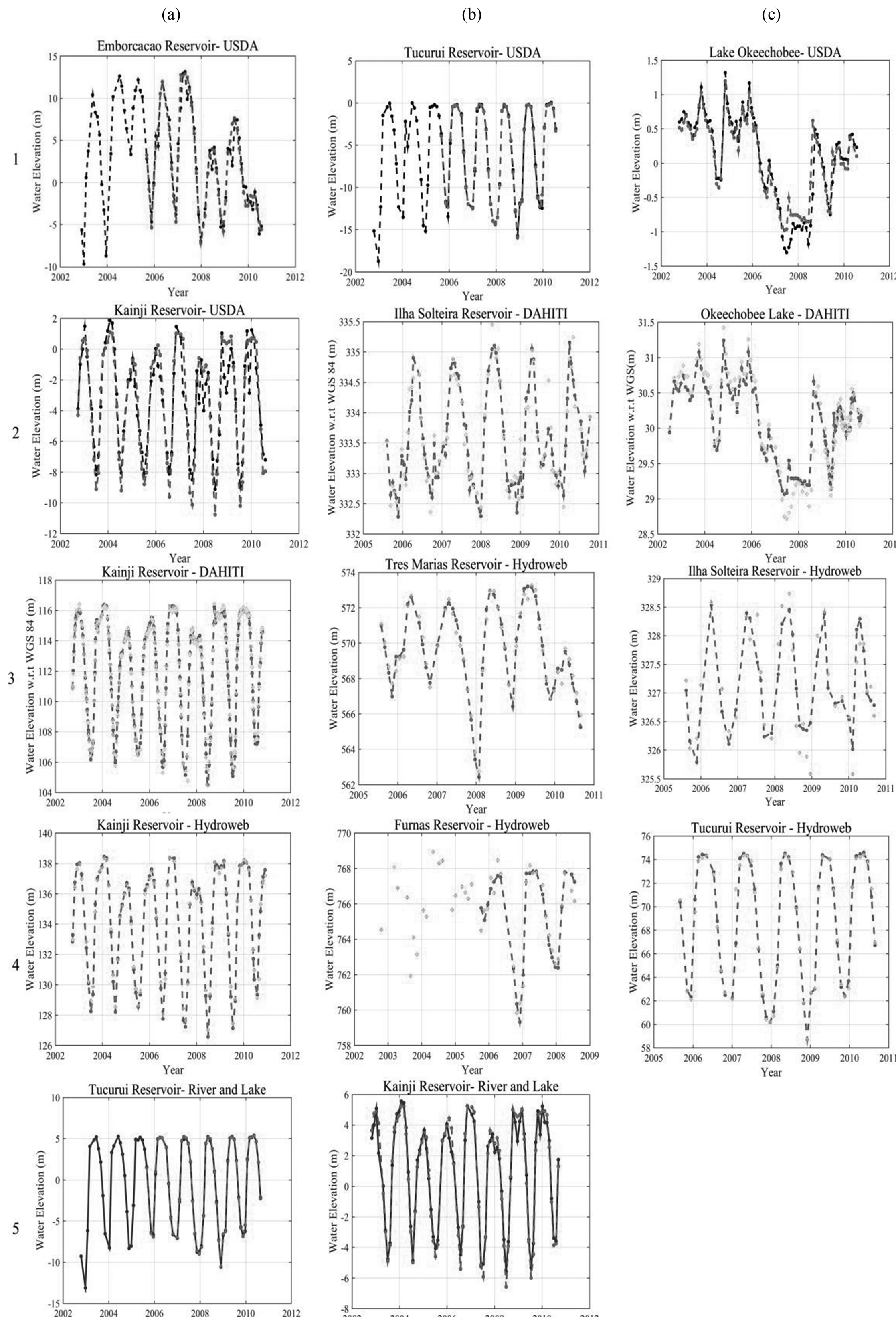


Fig. 17. Comparison of *in-situ* and other publicly available web database (dashed red lines: *in situ*, cyan circles: DAHITI, green circles: HydroWeb, black circles: USDA, and the blue line: River and Lake water level time series) over the common lakes and reservoirs.

417 m. Although this is not a common occurrence in the proposed algorithm, we suggest computing the elevation difference between successive dates (cycles) to detect such significant difference.

Finally, in a case where the percentage of the outliers is significantly more than the *actual* observations, the algorithm is more likely to fail. In such cases, a more advanced method, such as a machine learning approach, can be used for outlier detection. This limitation due to the percentage of outlier is not peculiar to this algorithm as Schwatke *et al*. [23] also suggested that more than 50% of the observations should be over water to effectively detect outliers. Nonetheless, Troitskaya *et al*. [44] stated that using adaptive retracking will significantly increase the number of valid observations, which potentially increases the accuracy of measurements.

V. Conclusion

We have been able to successfully demonstrate and validate the results of the automated altimetry time-series algorithm using 37 Jason-2 and Envisat satellite altimetry crossings, representing 30 reservoirs located in the U.S., Brazil, and Nigeria.

The result of the automated times series was compared to the *in-situ* gauge data and the RMSE computed ranges from 0.09 to 1.20 m. Interestingly, we did observe that small satellite crossing length over in-land waterbodies also have a tendency to have a lower or higher RMSE. We were able to use the algorithm to generate water level time series for a reservoir length of 0.64 km to several kilometers with low RMSE. The result of this algorithm is consistent and capable of processing water level of lakes and reservoirs at a regional and global scale with a high degree of reliability. The algorithm can also be extended to generate water level time series using SARAL/AltiKa data and recent altimeters such as Jason-3 and Sentinel-3. This study has also reinforced a previous study [3] that Ice-1 retracked range is most suitable for inland water body studies using Envisat data.

Finally, the automated time-series generation algorithm will be developed into an open-source tool for the general user community to increase the use of altimetry time-series data in studying the anthropogenic impacts on the water cycle [45], effects of climate change at a regional or global scale [12], as well as improving existing hydrologic models [46]. Hence, users will have the flexibility in processing and investigating any in-land water body of their interest that has an altimeter overpass. It should also be noted that users can try applying different thresholds of SR and std and generate the *best* time series on their discretion over water bodies of their interest where *in-situ* data may not exist to compute the RMSE and outlier percentage as done in Fig. 5.

Overall, the algorithm has increased efficiency in processing radar altimetry data and eliminated inconsistency in data processing. Finally, this algorithm is not limited to studying lakes and reservoirs in regional and global scale but can be potentially applied to rivers and wetlands.

Acknowledgment

The authors would like to thank T. Omotoso for providing the *in-situ* gauge data for Kainji reservoir, Nigeria. They would also like to thank the anonymous reviewers for their constructive comments in improving the quality of this paper.

References

[1] M. K. I. Rahmani, N. Pal, and K. Arora, “Clustering of image data using K-means and fuzzy K-means,” *Int. J. Adv. Comput. Sci. Appl.*, vol. 5, pp. 160–163, 2014.

[2] T. Yuan, H. Lee, and H. C. Jung, “Toward estimating wetland water level changes based on hydrological sensitivity analysis of PALSAR backscattering coefficients over different vegetation fields,” *Remote Sens.*, vol. 7, pp. 3153–3183, 2015.

[3] F. Frappart, S. Calmant, M. Cauhopé, F. Seyler, and A. Cazenave, “Preliminary results of ENVISAT RA-2-derived water levels validation over the Amazon basin,” *Remote Sens. Environ.*, vol. 100, pp. 252–264, 2006.

[4] S. Biancamaria, F. Hossain, and D. Lettenmaier, “Forecasting transboundary river water elevations from space,” *Geophys. Res. Lett.*, vol. 38, 2011 Q3

[5] C. Birkett, “The contribution of TOPEX/POSEIDON to the global monitoring of climatically sensitive lakes,” *J. Geophys. Res., Oceans*, vol. 100, pp. 25179–25204, 1995.

[6] J.-F. Crétaux *et al.*, “SOLS: A lake database to monitor in the Near Real Time water level and storage variations from remote sensing data,” *Adv. Space Res.*, vol. 47, pp. 1497–1507, 2011.

[7] H. Lee *et al.*, “Louisiana wetland water level monitoring using retracked TOPEX/POSEIDON altimetry,” *Marine Geodesy*, vol. 32, pp. 284–302, 2009.

[8] H. Lee, T. Yuan, H. C. Jung, and E. Beighley, “Mapping wetland water depths over the central Congo Basin using PALSAR ScanSAR, Envisat altimetry, and MODIS VCF data,” *Remote Sens. Environ.*, vol. 159, pp. 70–79, 2015.

[9] J. S. da Silva, S. Calmant, F. Seyler, O. C. Rotunno Filho, G. Cochonneau, and W. J. Mansur, “Water levels in the Amazon basin derived from the ERS 2 and ENVISAT radar altimetry missions,” *Remote Sens. Environ.*, vol. 114, pp. 2160–2181, 2010.

[10] J.-F. Crétaux and C. Birkett, “Lake studies from satellite radar altimetry,” *Comptes Rendus Geosci.*, vol. 338, pp. 1098–1112, 2006.

[11] H. Lee, C. Shum, K.-H. Tseng, J.-Y. Guo, and C.-Y. Kuo, “Present-day lake level variation from Envisat altimetry over the northeastern Qinghai-Tibetan plateau: Links with precipitation and temperature,” *Terrestrial, Atmos. Ocean. Sci.*, vol. 22, pp. 169–175, 2011.

[12] C. M. Birkett and I. M. Mason, “A new global lakes database for a remote sensing program studying climatically sensitive large lakes,” *J. Great Lakes Res.*, vol. 21, pp. 307–318, 1995.

[13] A. Tarpanelli, L. Brocca, S. Barbetta, M. Faruolo, T. Lacava, and T. Moramarco, “Coupling MODIS and radar altimetry data for discharge estimation in poorly gauged river basins,” *IEEE J. Sel. Topics Appl. Earth Obs. Remote Sens.*, vol. 8, no. 1, pp. 141–148, Jan. 2015.

[14] F. Hossain, L. C. Mazumder, S. M. ShahNewaz, S. Biancamaria, H. Lee, and C. Shum, “Proof of concept of an altimeter-based river forecasting system for transboundary flow inside Bangladesh,” *IEEE J. Sel. Topics Appl. Earth Obs. Remote Sens.*, vol. 7, no. 2, pp. 587–601, Feb. 2014.

[15] H. Gao, C. Birkett, and D. P. Lettenmaier, “Global monitoring of large reservoir storage from satellite remote sensing,” *Water Resour. Res.*, vol. 48, 2012.

[16] D. Alsdorf, C. Birkett, T. Dunne, J. Melack, and L. Hess, “Water level changes in a large Amazon lake measured with spaceborne radar interferometry and altimetry,” *Geophys. Res. Lett.*, vol. 28, pp. 2671–2674, 2001.

[17] H. Lee, M. Durand, H. C. Jung, D. Alsdorf, C. Shum, and Y. Sheng, “Characterization of surface water storage changes in Arctic lakes using simulated SWOT measurements,” *Int. J. Remote Sens.*, vol. 31, pp. 3931–3953, 2010.

[18] M. Durand, L.-L. Fu, D. P. Lettenmaier, D. E. Alsdorf, E. Rodriguez, and D. Esteban-Fernandez, “The surface water and ocean topography mission: Observing terrestrial surface water and oceanic submesoscale eddies,” *Proc. IEEE*, vol. 98, no. 5, pp. 766–779, May 2010.

[19] E. Rodríguez and P. S. Callahan, “Surface water and ocean topography mission (SWOT), science requirements document,” JPL Document, Mar. 18, 2016.

[20] Y. Sulistioadi *et al.*, “Satellite radar altimetry for monitoring small rivers and lakes in Indonesia,” *Hydrol. Earth Syst. Sci.*, vol. 19, pp. 341–359, 2015.

[21] C. M. Birkett and B. Beckley, “Investigating the performance of the Jason-2/OSTM radar altimeter over lakes and reservoirs,” *Marine Geodesy*, vol. 33, pp. 204–238, 2010.

[22] X. Huang, H. Xie, G. Zhang, and T. Liang, "A novel solution for outlier removal of ICESat altimetry data: A case study in the Yili watershed, China," *Front. Earth Sci.*, vol. 7, pp. 217–226, 2013.

[23] C. Schwatke, D. Dettmering, W. Bosch, and F. Seitz, "DAHITI—An innovative approach for estimating water level time series over inland waters using multi-mission satellite altimetry," *Hydrol. Earth Syst. Sci.*, vol. 19, pp. 4345–4364, 2015.

[24] K. Nielsen, L. Stenseng, O. B. Andersen, H. Villadsen, and P. Knudsen, "Validation of CryoSat-2 SAR mode based lake levels," *Remote Sens. Environ.*, vol. 171, pp. 162–170, 2015.

[25] J. Dumont *et al.*, "OSTM/Jason-2 products handbook," CNES: SALP-MU-M-OP-15815-CN, EUMETSAT: EUM/OPS-JAS/MAN/08/0041, JPL: OSTM-29-1237, NOAA/NESDIS: Polar Series/OSTM J, vol. 400, p. 1, 2009.

[26] J. Benveniste *et al.*, *Envisat RA-2/MWR Product Handbook*. Frascati, Italy: Eur. Space Agency, 2002.

[27] A. K. Jain, M. N. Murty, and P. J. Flynn, "Data clustering: A review," *ACM Comput. Surv.*, vol. 31, pp. 264–323, 1999.

[28] L. Duan, L. Xu, Y. Liu, and J. Lee, "Cluster-based outlier detection," *Ann. Oper. Res.*, vol. 168, pp. 151–168, 2009.

[29] R. Pamula, J. K. Deka, and S. Nandi, "An outlier detection method based on clustering," in *Proc. 2011 2nd Int. Conf. Emerging Appl. Inf. Technol.*, 2011, pp. 253–256.

[30] K. Thangavel and A. K. Mohideen, "Semi-supervised k-means clustering for outlier detection in mammogram classification," in *Proc. 2010 Trendz Inf. Sci. Comput.*, 2010, pp. 68–72.

[31] D. M. Hawkins, *Identification of Outliers*, vol. 11. New York, NY, USA: Springer, 1980.

[32] M. Kantardzic, *Data Mining: Concepts, Models, Methods, and Algorithms*. New York, NY, USA: Wiley, 2011.

[33] J. A. Hartigan and M. A. Wong, "Algorithm AS 136: A k-means clustering algorithm," *Appl. Statist.*, vol. 28, pp. 100–108, 1979.

[34] F. U. Siddiqui and N. A. M. Isa, "Enhanced moving K-means (EMKM) algorithm for image segmentation," *IEEE Trans. Consum. Electron.*, vol. 57, no. 2, pp. 833–841, May 2011.

[35] J. Z. Huang, M. K. Ng, H. Rong, and Z. Li, "Automated variable weighting in k-means type clustering," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 27, no. 5, pp. 657–668, May 2005.

[36] D. Arthur and S. Vassilvitskii, "k-means++: The advantages of careful seeding," in *Proc. 18th Annu. ACM-SIAM Symp. Discrete Algorithms*, 2007, pp. 1027–1035.

[37] H. Lee *et al.*, "Validation of Jason-2 altimeter data by waveform retracking over california coastal ocean," *Marine Geodesy*, vol. 33, pp. 304–316, 2010.

[38] H. Lee, C. K. Shum, Y. Yi, A. Braun, and C.-Y. Kuo, "Laurentia crustal motion observed using TOPEX/POSEIDON radar altimetry over land," *J. Geodyn.*, vol. 46, pp. 182–193, 2008.

[39] K. J. Bhang, F. W. Schwartz, and A. Braun, "Verification of the vertical error in C-band SRTM DEM using ICESat and Landsat-7, Otter Tail County, MN," *IEEE Trans. Geosci. Remote Sens.*, vol. 45, no. 1, pp. 36–44, Jan. 2007.

[40] C. Birkett, "Radar altimetry: A new concept in monitoring lake level changes," *Eos, Trans. Amer. Geophys. Union*, vol. 75, pp. 273–275, 1994.

[41] J. S. da Silva, S. Calmant, F. Seyler, O. C. Rotunno Filho, G. Cochonneau, and W. J. Mansur, "Water levels in the Amazon basin derived from the ERS 2 and ENVISAT radar altimetry missions," *Remote Sens. Environ.*, vol. 114, pp. 2160–2181, 2010.

[42] C. H. Davis, "Growth of the Greenland ice sheet: A performance assessment of altimeter retracking algorithms," *IEEE Trans. Geosci. Remote Sens.*, vol. 33, no. 5, pp. 1108–1116, Sep. 1995.

[43] M. Kleinherenbrink, P. Ditmar, and R. Lindenbergh, "Retracking cryosat data in the SARIn mode and robust lake level extraction," *Remote Sens. Environ.*, vol. 152, pp. 38–50, 2014.

[44] Y. Troitskaya, G. Rybushkina, I. Soustova, and S. Lebedev, "Adaptive retracking of Jason-1, 2 satellite altimetry data for the Volga River reservoirs," *IEEE J. Select. Topics Appl. Earth Obs. Remote Sens.*, vol. 7, no. 5, pp. 1603–1608, May 2014.

[45] A. Getirana *et al.*, "Forecasting water availability in data sparse and heavily managed catchments in africa and the middle east," *Gewex News*, vol. 27, pp. 8–11, 2015.

[46] A. C. Getirana, A. Boone, D. Yamazaki, and N. Mognard, "Automatic parameterization of a flow routing scheme driven by radar altimetry data: Evaluation in the Amazon basin," *Water Resour. Res.*, vol. 49, pp. 614–629, 2013.

Modurodoluwa Adeyinka Okeowo received the B.S. (First Class Hons.) degree in surveying and geoinformatics engineering from the University of Lagos, Lagos, Nigeria, in 2008, and the M.S. degree in geomatics engineering from Purdue University, West Lafayette, IN, USA, in 2012. He is currently working toward the Ph.D. degree in the Department of Geosensing Systems Engineering and Sciences, University of Houston, Houston, TX, USA.

His research focuses on the algorithm development for radar altimetry, remote sensing, and GIS.

Hyongki Lee received the B.S. and M.S. degrees in civil engineering from Yonsei University, Seoul, South Korea, in 2000 and 2002, respectively, and the Ph.D. degree in geodetic science from Ohio State University, Columbus, OH, USA, in 2008. He is currently an Assistant Professor in the Department of Civil and Environmental Engineering and the National Center for Airborne Laser Mapping, University of Houston, Houston, TX, USA. He specializes in using spaceborne and airborne geodetic instruments including satellite altimetry, SAR/InSAR, and GRACE to better understand earth system sciences. His primary research focuses on investigating terrestrial water dynamics using remote sensing data toward applications for water resources management.

Dr. Lee received the NASA New Investigator Award in Earth Science in 2014.

Faisal Hossain received the B.S. degree from Indian Institute of Technology, Varanasi, India, in 1996, the M.S. degree from The National University of Singapore, Singapore, in 1999, and the Ph.D. degree from The University of Connecticut, Storrs, CT, USA, in 2004.

He is currently an Associate Professor in the Department of Civil and Environmental Engineering, University of Washington, Seattle, WA, USA. His research interests include hydrologic remote sensing, sustainable water resources engineering, transboundary water resources management, and engineering education.

Augusto Getirana received the Ph.D. degree in hydrology and civil engineering from the University of Toulouse III, Toulouse, France, and the Federal University of Rio de Janeiro, Rio de Janeiro, Brazil, in 2009.

He is currently an Assistant Research Scientist with the Hydrological Sciences Laboratory, NASA Goddard Space Flight Center, Greenbelt, MD, USA, and the University of Maryland, College Park, MD. His research interests include the better understanding of the global spatial and temporal freshwater availability and dynamics through satellite observations and computational modeling, including quantifying the impacts of anthropogenic activities on the water cycle, and the related interactions of these activities with the atmosphere.