DAHITI - An Innovative Approach for Estimating Water Level Time Series over Inland Water using Multi-Mission Satellite Altimetry

Christian Schwatke, Denise Dettmering, Wolfgang Bosch, and Florian Seitz

Deutsches Geodätisches Forschungsinstitut der Technischen Universität München (DGFI-TUM),
Arcisstr. 21, 80333 München, Germany

Correspondence to: Christian Schwatke
(christian.schwatke@tum.de)

Abstract. Satellite altimetry has been designed for sea level monitoring over open ocean areas. However, for some years, this technology has also been used to retrieve water levels from reservoirs, wetlands and in general any inland water body, although the radar altimetry technique has been especially applied to rivers and lakes. In this paper, a new approach for the estimation of inland water level time series is described. It is used for the computation of time series of rivers and lakes available through the web service 'Database for Hydrological Time Series over Inland Water' (DAHITI). The new method is based on an extended outlier rejection and a Kalman filter approach incorporating cross-calibrated multi-mission altimeter data from Envisat, ERS-2, Jason-1, Jason-2, Topex/Poseidon, and SARAL/AltiKa, including their uncertainties. The paper presents water level time series for a variety of lakes and rivers in North and South America featuring different characteristics such as shape, lake extent, river width, and data coverage. A comprehensive validation is performed by comparisons with in situ gauge data and results from external inland altimeter databases. The new approach yields RMS differences with respect to in situ data between 4 cm and 36 cm for lakes and 8 cm and 114 cm for rivers, respectively. For most study cases, more accurate height information than from other available altimeter data bases can be achieved.

Keywords. Satellite Altimetry; Inland Water; Kalman Filter; DAHITI; Envisat; ERS-2; Topex/Poseidon; Jason-1; Jason-2; SARAL/AltiKa

1 Introduction

Since the 1990s, monitoring and modelling the water cycle of the system Earth have become a very important task (Stakhiv and Stewart, 2010). In particular, the knowledge of regional changes of water storage in rivers and lakes is fundamental for the risk assessment of natural disasters such as the droughts and floods which have been increasing over the last few decades (Guha-Sapir and Vos, 2011). Despite the growing importance of measurements, the number of in situ stations monitoring river discharge is globally declining. The number of river discharge time series provided by the Global Runoff Data Center (GRDC) decreased from about 7,300 to 1,000 stations between 1978 and 2013 (Global Runoff Data Center, 2013). In order to make a statement about the development of water level gauging stations an equivalent database such as the GRDC is required. In general, in situ water level data are managed by federal institutions which make data access very difficult. However, many remote sensing satellites have been launched in the last few years measuring parameters relevant for the investigation of the water cycle, e.g. precipitation, water level, and gravity.

Among these remote sensing techniques is satellite altimetry. Besides its main design goal of measuring water levels in the ocean, satellite altimetry can also be used for deriving water levels of inland water bodies, i.e. lakes, reservoirs, rivers, and wetlands (e.g. Birkett 1995, Crétaux and Birkett 2006 and Crétaux et al. 2011). The advantage of satellite altimetry is its global availability, which allows for estimation of water level time series even in remote areas without local infrastructure. Satellite altimetry can provide water level time series longer than two decades.

However, because its measurement geometry provides observations along specific ground tracks touching water bodies is by chance. Thereby, big water bodies have a higher probability to be passed than smaller ones. In addition, because of a repeat orbit configuration the temporal resolution is limited to 35 (ERS-2, Envisat, SARAL/AltiKa) or 10

(Topex/Poseidon, Jason-1, Jason-2) days when only single altimeter missions are used. Thus, the combination of different altimeter systems plays a key role in increasing the temporal and spatial resolution as well as the length of the time series. Satellite altimetry has to cope with different problems over inland water which are mainly caused by the large pulse-limited footprint of radar altimeters. For altimeter missions using Ku-band such as Envisat, the resulting footprint varies between 2 km over the ocean and up to 16 km over the land (Chelton et al., 2001). Even for SARAL/AltiKa, measuring in Ka-band, the footprint size is still about 8 km (Schwatke et al., 2015). The majority of problems in the field of inland satellite altimetry are the result of land contamination. This effect is two-fold: on the one hand the contamination of the radar echo leads to degraded range quality or even to unusable data sets, on the other hand so-called 'hooking' or 'off-nadir' effects occur. The second effect arises from off-nadir radar returns when the satellite is still/already over land but receives the main reflection from the off-nadir water areas. This leads to longer ranges visible in a parabolic shape of the resulting height sequence. This effect can be corrected by fitting curves to the resulting water levels (da Silva et al. 2010, Maillard et al. 2015). For each land-water transition a parabola can be fitted to the measurements that can be used to correct the off-nadir effect. In this paper, the off-nadir data are discarded since for all targets enough reliable nadir-measurements are available. The first effect is more challenging. The contamination of radar measurements by land causes a degeneration of ocean-like waveform shapes (Brown, 1977). The affected waveforms do not have typical brown-like shapes and cannot be retracked by using ocean waveform retrackers (MLE (Challenor and Srokosz, 1989), NASA β (Martin et al., 1983), etc). Therefore, additional retracking can be applied with retracking algorithms such as OCOG (Wingham et al., 1986), Improved Threshold (Hwang et al., 2006), etc. which are more robust with respect to the geometry of the waveforms and can achieve reliable heights. The choice of retracker depends on the quality of existing altimeter measurements which varies between investigated inland water bodies because of their extent, shape or ambient topography.

The potential of satellite altimetry for the estimation of water level time series and for understanding the terrestrial water cycle was shown by Birkett (1995), Crétaux and Birkett (2006) and Crétaux et al. (2011). In most studies, only single satellite tracks were used for the computation of water level time series. The most popular study areas were the Great Lakes (Ponchaut and Cazenave (1998) used Topex/Poseidon) and the Amazon basin. For the latter, investigations were based on different missions: e.g. Topex/Poseidon (de Oliveira Campos et al., 2001; Zakharova et al., 2006), Topex/Jason-1/Jason-2 (Seyler et al., 2013) and ERS-2/Envisat (da Silva et al., 2010). In addition to these individual investigations four global databases have been developed that provide the international community water level time series over inland. The different processing strategies of these four databases are described as follows.

The *Hydroweb database*[1] was developed by the Laboratoire d'Etudes en Géophysique et Océanographie Spatiales (LEGOS). For the estimation of water level time series over lakes and rivers, a multi-mission approach using satellite altimeter data of Topex/Poseidon, ERS-1, ERS-2, Envisat, Jason-1, and GFO is applied. The physical heights are estimated in a track-wise manner and are corrected by the slope of the geoid or mean lake level and by range biases with respect to Topex/Poseidon. The final time series are computed by merging the altimeter data on a monthly basis. The applied approaches are published in Crétaux et al. (2011) and da Silva et al. (2010).

The *River & Lakes database*[2] was developed by the European Space Agency and the De Montfort University (ESA-DMU). It provides track-wise time series derived from Jason-2 and Envisat for a variety of inland waters. For each track crossing the water body of interest a single time series is processed. The methodology for the estimation uses an expert system which is based on neural networks (Berry et al., 1997).

The *Global Reservoir and Lake Monitor (GRLM)*[3] is maintained by the Foreign Agricultural Service of the United States Department of Agriculture (USDA). Water level time series of lakes and reservoirs are estimated by using a segment of one single altimeter track over the investigated target. The time series are composed of data from consecutive altimeter missions measured along the same ground track. A combination of contemporaneous missions is not performed. The method for the estimation of water level time series is described in Birkett et al. (2011).

The *Database for Hydrological Time Series over Inland Water (DAHITI)*[4] was launched by the Deutsches Geodätisches Forschungsinstitut (DGFI, now DGFI-TUM) in 2013. Currently, DAHITI provides about 250 time series of rivers, lakes, reservoirs, and wetlands. The methodology for the estimation of water level time series in DAHITI is based on an extended outlier rejection and a Kalman filter approach described in detail in the article at hand.

In contrast to the methods already published in the literature, our approach is based on a rigorous combination of a variety of altimeter missions. In addition, extended outlier detection is applied and optional waveform retracking is implemented. Moreover, the processing contains a full error propagation and provides accuracies for each height measurement. Furthermore, correlations between altimeter measurements are considered in order to achieve more reliable errors for each water level height. The current paper provides detailed information on the estimation of water level time

---

[1]http://www.legos.obs-mip.fr/soa/hydrologie/hydroweb/
[2]http://tethys.eaprs.cse.dmu.ac.uk/RiverLake/shared/main
[3]http://www.pecad.fas.usda.gov/cropexplorer/global_reservoir/
[4]http://dahiti.dgfi.tum.de

series and performs a comprehensive validation by comparing the results with in situ gauging data and time series from other databases (Hydroweb, River & Lakes, and GRLM).

The article is structured as follows: In Section 2 the altimeter data that serve as input for the processing are described. In Section 3 the methodology for the estimation of water level time series from satellite altimeter data using a Kalman filter approach is explained. Section 4 starts with the introduction of the validation areas and data before the resulting water level time series and validation results are presented. The paper concludes with a summary of the results and outlook.

2   Altimeter Data and Height Estimation

In this paper, altimeter measurements from Topex, Jason-1, Jason-2, ERS-2, Envisat, and SARAL/AltiKa are used depending on the data coverage for the inland water bodies under investigation. In principle, data from Geosat, ERS-1, HY-2A, IceSAT, and Cryosat-2 can be used. However, these missions are neglected in the current investigations for a number of reasons, i.e. lack of data over land, non/long-repeat cycle, bad data quality, or missing waveform information. The applied missions can be separated into two groups according to their orbit characteristics. Topex/Poseidon was launched in 1992 into an orbit with a repeat cycle of 9.9156 days and a track separation at the equator of about 300 km. The mission was followed by its successors, Jason-1 and Jason-2. These three altimeter satellites can be used for estimating continuous time series over more than two decades. The second group starts with ERS-2 (launched in 1995), followed by Envisat and SARAL/AltiKa. The orbit of these missions is defined by a repeat cycle of 35 days and a track separation of about 80 km at the equator. The data are available for almost two decades with a data gap between 10/2010 (end of Envisat core mission) and 03/2013 (launch of SARAL/AltiKa). The data for Envisat on its drifting orbit (10/2010-04/2012) are not used. ERS-1 is not yet ready for use in DAHITI but will be integrated in the near future. This will enable extensions of the time series back to 1991.

For the estimation of water levels, Sensor Geophysical Data Records (SGDR) altimeter products are used which provide high-frequent ranges as well as altimeter waveforms. The altimeter waveforms allows individual retracking in order to achieve more reliable altimeter ranges, especially for smaller inland water bodies. Table 1 shows a list of the altimeter missions used and provides information about the product, cycle length, frequency, along-track distance between altimeter measurements on the ground, time period, and mean range bias with respect to Topex.

Depending on the investigated inland water body the original ocean ranges in the SGDR are very often corrupted. Especially over small lakes and rivers the altimeter waveforms do not exhibit the typical ocean-like shape because of land contamination. Land-contaminated altimeter waveforms are usually more peaky and noisy which leads to flat-patched, quasi-specular and complex waveforms (Berry et al., 2005). The quality of the ranges can be improved by retracking these waveforms. In this study, the 'Improved Threshold Retracker' (Hwang et al., 2006) with a threshold of 10% is applied if additional retracking is necessary. In general, all altimeter measurements of smaller lakes and rivers are retracked if the ocean product does not lead to reliable time series because of land contamination. Testing different thresholds for the retracking of altimeter measurements showed that a threshold of 10 % gives slightly better results for smaller lakes and rivers. In our implementation of the 'Improved Threshold Retracker' the first sub-waveform is always chosen. Nor do we use a reference height for choosing the sub-waveform such as the last range over ocean as described in (Hwang et al., 2006) since this is difficult in the case of small lakes and rivers. This algorithm is very robust and delivers ranges for all surface types which are more reliable than the original ranges over small inland waters. However, over open water (i.e. larger lakes) the resulting ranges are less precise than ranges derived from retracking algorithms for ocean applications. It is known that switching retracking algorithms along a single satellite track leads to height offsets (Crétaux et al., 2009). To avoid those offsets, all altimeter measurements of an investigated inland water body are retracked with the same algorithm.

In order to convert the range measurements (original or retracked) to water levels serving as input for our Kalman filter approach numerous preprocessing steps are necessary. Eq. 1 summarizes the height computation from altimeter products (orbit height h_{sat} and (retracked) altimeter range r_{alt}). These processing steps have to be performed for each individual altimeter measurement. The derived normal heights h_{normal} serve as input for the DAHITI approach described in Section 3.

$$
\begin{aligned}
h_{normal} \quad = \quad & h_{sat} - r_{alt} - \\
& \Delta h_{wet} - \Delta h_{dry} - \Delta h_{iono} - \\
& \Delta h_{etide} - \Delta h_{ptide} - \\
& \Delta h_{rad} - N
\end{aligned} \tag{1}
$$

First, the range has to be corrected for geophysical effects. For this purpose, the models and corrections given in Table 2 are applied. It is important to apply identical geophysical corrections for all missions and over the whole time period in order to avoid inconsistencies in the resulting multi-mission time series. To correct the wet (Δh_{wet}) and dry (Δh_{dry}) tropospheric delay, products of ECMWF for Vienna Mapping Function 1 (VMF1) (Boehm et al., 2009) are used. The ionospheric delay Δh_{iono} is corrected by using the NOAA Ionospheric Climatology 2009 (NIC09) (Scharroo and Smith, 2010) model. The solid Earth tide and pole tide corrections (Δh_{etide}, Δh_{ptide}) are applied according to the IERS Conventions 2003 (McCarthy and Petit, 2004). Finally, each single altimeter measurement is corrected for its

Table 1: List of all altimeter missions used in this study together with their main characteristics.

Mission	Product	Cycle length	Data rate	Along-track distance	Time period	Mean range bias
Envisat[1]	SGDR (v2.1)	35 d	18 Hz	~374 m	2002 - 2010	450.8 ± 7.9 mm
Envisat (EM)[1]	SGDR (v2.1)	35 d	18 Hz	~374 m	2010 - 2011	441.2 ± 2.6 mm
ERS-2[1]	SGDR (REAPER)	35 d	18 Hz	~374 m	1995 - 2007	71.2 ± 6.9 mm
Jason-1[23]	SGDR-C	9.9156 d	20 Hz	~294 m	2002 - 2009	97.3 ± 1.3 mm
Jason-1 (EM)[23]	SGDR-C	9.9156 d	20 Hz	~294 m	2009 - 2012	97.2 ± 2.6 mm
Jason-1 (GM)[23]	SGDR-C	9.9156 d	20 Hz	~294 m	2012 - 2013	103.1 ± 1.7 mm
Jason-2[234]	SGDR-D	9.9156 d	20 Hz	~294 m	2008 - active	-4.7 ± 1.0 mm
Poseidon[23]	ALT SDR (L1B)	9.9156 d	10 Hz	~620 m	1992 - 2002	-1.1 ± 7.2 mm
Topex[23]	ALT SDR (L1B)	9.9156 d	10 Hz	~620 m	1992 - 2002	-0.2 ± 1.2 mm
Topex (EM)[23]	ALT SDR (L1B)	9.9156 d	10 Hz	~620 m	2002 - 2005	-0.0 ± 2.5 mm
SARAL/AltiKa[35]	SGDR-T (patch 2)	35 d	40 Hz	~173 m	2013 - active	-67.5 ± 1.7 mm

operated by: [1]ESA, [2]NASA, [3]CNES, [4]EUMETSAT, [5]ISRO

radial error Δh_{rad} in order to account for inter-mission biases. Radial errors are derived from a global multi-mission crossover analysis as described by Bosch et al. (2014). They are computed with the ocean products. Radial errors were interpolated over land to provide range bias corrections for each altimeter measurement over land. This approach works quite well as long as the ocean product is used for the computation of inland water levels. However, as soon as retracking is involved additional retracker offsets will occur. In order to minimize the relative offsets between different altimeter tracks, we use the same retracker for all measurements over one target. That minimizes the inter-mission biases which are shown later for selected results in Section 4.3 and allow us to use different altimeter missions as a single virtual altimeter system. The average values of the applied range errors are given in Table 1 for each altimeter mission. All data used in this study (the altimeter data as well as all corrections) are extracted from OpenADB[5], the open altimeter database of DGFI-TUM. More information on OpenADB is given in Section 3.1. The quality of extracted geophysical corrections is checked and altimeter measurements are rejected if they do not comply with the valid ranges given in the mission handbooks.

For the computation of water level time series within the Kalman filter approach normal heights h_{normal} are used as input data whereas altimetry provides ellipsoidal heights. However, ellipsoidal heights are purely geometrical and do not allow us to predict where the water will flow. We compute normal heights by subtracting a (quasi-)geoid model (N) from the ellipsoidal heights. For this purpose, the EIGEN-6c3stat (Förste et al., 2012) model is used which supplements the EGM2008 geoid model with additional GOCE gravity data.

3   DAHITI Approach

In order to use altimeter measurements from different tracks and missions a consistent and reliable combination strategy is important. The irregular spaced observations from different locations must be merged into one time series per target and the optimal combination of measurements with different uncertainties must be ensured. This requirement is fulfilled by our DAHITI approach which is based on an extended outlier rejection and a Kalman filter for the estimation of water level time series.

The processing strategy for the estimation of water level time series over inland waters using the DAHITI approach is separated into three steps: preprocessing, Kalman filtering and postprocessing (cf. Figure 1). The preprocessing step includes all necessary tasks for the preparation of the input altimeter heights such as waveform retracking, applying range corrections, calculation of height errors, and rejection of outliers. In the Kalman filtering step, the computation of the water levels of the investigated water body is performed. In this paper, we apply Kalman filtering in a single location centred on the investigated water body and obtain one computed water level for each epoch. However, there is also an option for performing Kalman filtering on a grid which can be used for investigation of the surface variability of larger lakes.

In the postprocessing step, all water levels from the previous step are merged to form a single water level time series referring to one reference location if the Kalman filtering was performed on a grid. Subsequent outlier detection can be conducted if necessary. The final time series is stored in DAHITI, accessible via the website.

3.1   Preprocessing

The Open Altimeter Database (OpenADB) holds satellite altimeter data and derived high-level products. OpenADB provides satellite altimeter data, geophysical corrections, models, etc. which are also accessible via the website. All in-

---

[5]http://openadb.dgfi.tum.de

Table 2: List of applied models and geophysical corrections

| Correction	Source/Model	Reference
Wet troposphere	ECMWF (2.5° x 2.0°) for Vienna Mapping Functions 1 (VMF1)	Boehm et al. (2009)
Dry troposphere	ECMWF (2.5° x 2.0°) for Vienna Mapping Functions 1 (VMF1)	Boehm et al. (2009)
Ionosphere	NOAA Ionosphere Climatology 2009 (NIC09)	Scharroo and Smith (2010)
Solid earth tide	IERS Convention 2003	McCarthy and Petit (2004)
Pole tide	IERS Convention 2003	McCarthy and Petit (2004)
Range Bias	MMXO14	Bosch et al. (2014)
Geoid	EIGEN-6C3stat	Förste et al. (2012)
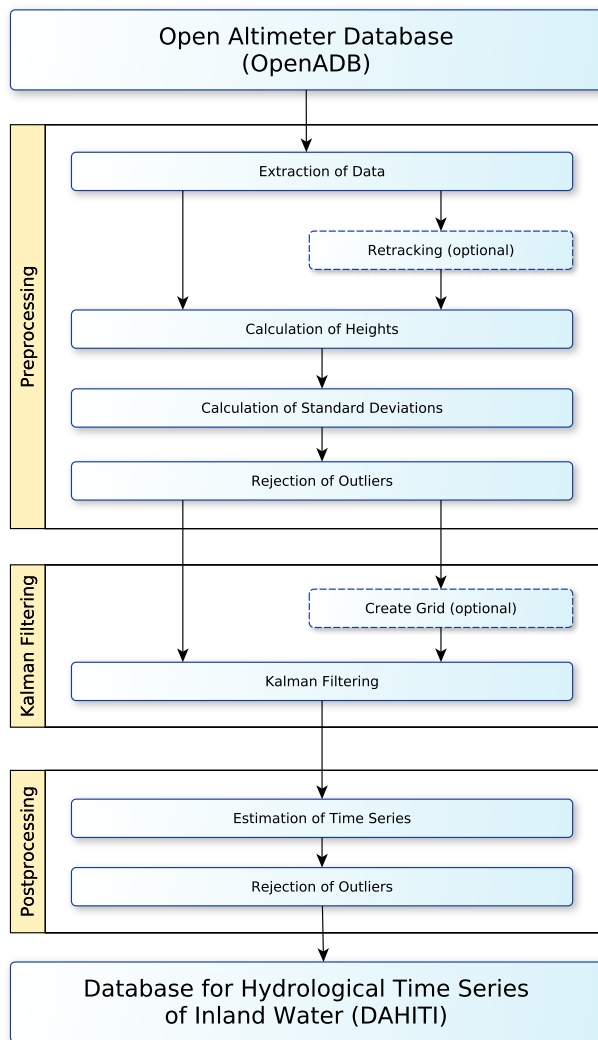
Figure 1: Processing strategy for the computation of water level time series for inland waters in DAHITI in three main steps: preprocessing, Kalman filtering, and postprocessing.

Version-Altimetry (MVA) and can be individually combined to achieve water heights. The data sets from OpenADB used for this study and the methodology used to derive individual water levels are described in Section 2.

In addition to the normal heights of the water levels the Kalman filter requires information on the quality of each measurement. This information is used for the weighting of the individual data sets as well as for the error estimation of water level products. Because of the lack of absolute accuracy, the precision of the heights is computed by analysing the along-track scatter of the measurements.

For this purpose, an 'absolute deviation around the median' (ADM) is estimated by using a sliding box along the altimeter track. The size of the sliding box varies for large lakes (±3.5km), small lakes/large rivers (±1.5km) and smaller rivers (±0.5km). The definition of the sliding box in kilometres instead of number of points allows consistent handling of missions with different data rates (10 Hz, 20 Hz, or 40 Hz) and ensures correct inter-mission weighting. The ADM is calculated by estimating a median of the water heights within the box. Then the median height is subtracted from the current water height and the absolute value of the difference is used as the 'error' of the altimeter measurement. Compared with estimated standard deviations, the ADM method is more robust against corrupted water heights and topography near shores and leads to more reliable errors as long as more than half of the altimeter measurements are over water.

Before Kalman filtering is performed various user-defined outlier rejections can be applied. Inaccurate water levels must be rejected before Kalman filtering; precise ones are used for the estimation of the resulting water levels. The following outlier criteria can be applied in the preprocessing step.

- Latitude thresholds

- Water height thresholds

- Height error (ADM) threshold

- Backscatter coefficient (Sigma0) thresholds

- Along-track outlier test (SVR)

It is important to note that the criteria for the outlier detection are very flexible and the optimal configuration strongly

formation is stored in separate parameter files named Multi-

depends on the investigated water body. As a consequence, the parameters for outlier rejection vary with the study areas. First, three outlier criteria (latitude thresholds, water height thresholds and height error threshold) are applied. The backscatter coefficients of altimeter measurements provide information about the reflectance of the surface. This information can be used to reject altimeter measurements affected by ice. Moreover, outlier detection with Support Vector Regression (SVR) (Smola and Schölkopf, 2004) is implemented. This method applies linear regression to each altimeter track to reject altimeter measurements that do not represent the flat water level of the inland water target. SVR is similar to common regression but is more flexible and robust. SVR is an advancement of the Support Vector Machine (SVM) (Boser et al., 1992) which is used as a classification algorithm for applications such as pattern recognition and machine learning. Depending on the mathematical problem, the kernel for the regression varies. One can use linear, polynomial or radial base functions (Smola and Schölkopf, 2004). In our case, SVR is applied on single altimeter tracks over an inland water body using a linear kernel and zero-slope constraint. Based on the constant representing the flat water level, an interval is defined which separates into valid and invalid data. Figure 2 shows an example of an altimeter track (Envisat, Pass 80, Cycle 007) crossing Lake Erie with an island in the middle. Blue dots indicate valid measurements, red dots indicate rejected data that exceed the ADM threshold of 5 cm (black dotted line), and green dots mean outliers detected by SVR (with rejection interval of ± 5 cm). The threshold of the SVR should be in the order of the noise of high-frequent altimeter measurements. One can see that all heights influenced by land contamination are detected as outliers and the remaining heights represent a flat surface.

3.2 Kalman Filtering

Kalman filtering is the central part of the computation of water level time series in DAHITI. It updates a model by measurement data of different accuracies and predicts the current state to the next time epoch (Kalman, 1960). In contrast to the common least-squares adjustment the Kalman filter works recursively and the number of input observations per processing step is significantly reduced because of its sequential integration. This also enables real-time applicability.

The Kalman filter performs the estimation of water level time series from the track-wise input heights by combining time-dependent input data available at irregular intervals and –in the case of larger lakes– at different locations . Different modified Kalman filter approaches have been used for geodetic applications (e.g. Yang and Gao 2006, Eicker et al. 2014 and Gruber et al. 2014) In principle, this algorithm realizes a sequential least squares adjustment by taking into account the accuracies of the input data as well as the deterministic and stochastic behaviour of the system and produces a statistically optimal estimate of the water level time series.

3.2.1 Update Interval

The Kalman filter uses input observations to update the current state of the system and predict the model of the following time epoch. This is performed in a continuous loop consisting of two steps (an update and a prediction step) running consecutively for every period of time t_k. At the beginning , an initialization is necessary in order to set the starting conditions. The work flow is illustrated in Figure 3. The time increment of the Kalman filter can be defined arbitrarily. In our case an observation-based update interval instead of a constant one is used. That means that our system is updated each time a new altimeter track is available. Thus, the update interval strongly depends on the size and the data coverage of the investigated water body. It can vary between 35 days (if only an Envisat track crosses the target area) and one day (in the case of large lakes covered by different altimeter missions). Time intervals shorter than one day are precluded by assigning the individual measurements to full days. The use of an adaptive update interval avoids smoothing effects in the case of data gaps that may occur when a fixed time increment is selected.

3.2.2 Computation Grid

All computations can be referred to one location (centre of the target) or performed on a computation grid. The latter is optional and can be applied for special investigations on surface variability of larger lakes. The standard solution -also used for all computations within this study- assumes uniform lake surfaces in balance with gravity and merges all water heights of one update step to one location. Surface differences owed to systematic height or geoid errors or hydrodynamic effects from wind and waves are neglected. In practice, our approach automatically creates a grid by means of a recursive algorithm used on an initial grid node as reference point. A land water mask provides information on the extent of the water body and the grid. The grid node separation can be chosen manually depending on the extent of the investigated inland water. Thus, normally we define only one grid node over the target. However, in cases where surface differences are expected, a smaller grid node distance can be chosen. The computations will then be performed for all grid nodes and different water levels for the whole lake surface.

3.2.3 Kalman Filter Equations

In the following, the basic equations of the Kalman filter are introduced. The algorithm consists of an observation model and a dynamic model.

The observations for each step k corresponding to epoch t_k are given in vector $\mathbf{l}_k$ and its co-variances in matrix $\boldsymbol{\Sigma}_{ll,k}$.

$$\underset{(m_k,1)}{\mathbf{l}_k} = \underset{(m_k,n)}{\mathbf{A}_k} \cdot \underset{(n,1)}{\mathbf{x}_k} - \underset{(m_k,1)}{\mathbf{v}_k} \qquad (2)$$

$$\underset{(m_k,m_k)}{\boldsymbol{\Sigma}_{ll,k}} = \underset{(m_k,m_k)}{\mathbf{I}} \cdot \underset{(m_k,1)}{\mathbf{s}_{l,k}} \qquad (3)$$
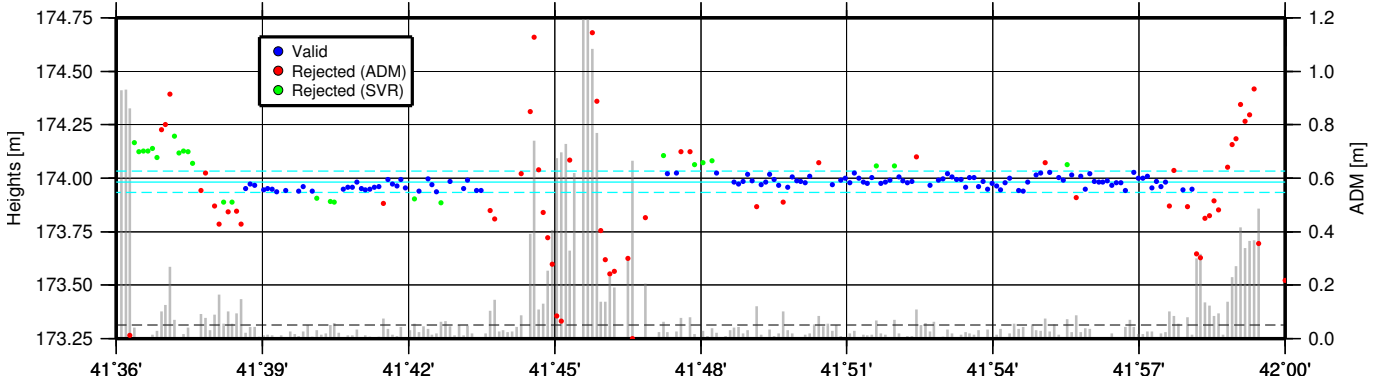
Figure 2: Example of an outlier detection using error threshold and SVR along a single satellite track over Lake Erie containing an island (between approx. 41°44' and 41°47'). The result of the regression shows valid (blue) and rejected (red, green) water heights. The height errors based on ADM are plotted as grey bars. Thresholds for height errors and SVR are marked by dashed lines (black and cyan respectively).
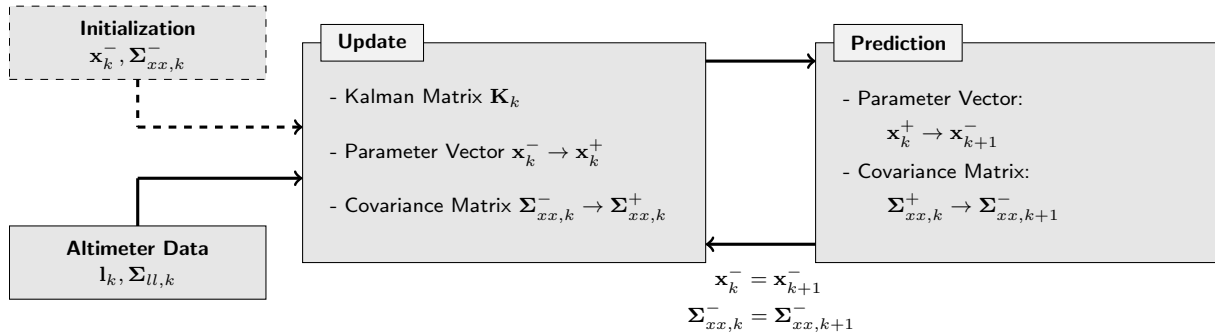

Figure 3: Procedure of Kalman filtering starting with an initialization step followed by a progressive loop containing one update and one prediction step.

The vector length of $\mathbf{l}_k$ depends on the number of water levels m_k available at each epoch t_k. The unknown grid node heights are compiled in vector $\mathbf{x}_k$. The $m_k \times n$ design matrix $\mathbf{A}_k$ is the core of the observation model and connects the water levels with the computation grid consisting of n grid points ($n = 1$ using only a single grid point). $\mathbf{A}_k$ has a dimension of $m_k \times n$ and contains ones for those grid nodes where water levels are available. Hereby, each water level height is assigned to the nearest grid node. In the case when the computation is performed on a single grid node all water level heights are assigned to them. The vector $\mathbf{v}_k$ absorbs the residuals of the observation model.

The uncertainties of the water levels are described in $\boldsymbol{\Sigma}_{ll,k}$. Since there is no information on correlation between individual water levels the matrix is defined as a diagonal matrix with variances σ_l^2 from ADM (computed in the preprocessing step) on the mean diagonal. These are collected in vector $\mathbf{s}_{l,k}$.

The dynamic model of the Kalman filter approach describes the transition of the system state from epoch t_k to t_{k+1}.

$$\underset{(n,1)}{\mathbf{x}_{k+1}^-} = \underset{(n,n)}{\boldsymbol{\Phi}_k} \cdot \underset{(n,1)}{\mathbf{x}_k^+} + \underset{(n,n)}{\boldsymbol{\Lambda}_k} \cdot \underset{(n,1)}{\mathbf{q}_k} \qquad (4)$$

$$\underset{(n,n)}{\boldsymbol{\Sigma}_{xx,k+1}^-} = \underset{(n,n)}{\boldsymbol{\Phi}_k} \cdot \underset{(n,n)}{\boldsymbol{\Sigma}_{xx,k}^+} \cdot \underset{(n,n)}{\boldsymbol{\Phi}_k^T} + \underset{(n,n)}{\boldsymbol{\Lambda}_k} \cdot \underset{(n,n)}{\mathbf{Q}_k} \cdot \underset{(n,n)}{\boldsymbol{\Lambda}_k^T} \;\; (5)$$

This includes the prediction step (cf. Figure 3) for the parameter vector $\mathbf{x}_k^+$ as well as for its covariance matrix $\boldsymbol{\Sigma}_{xx,k}^+$. The prediction of the grid node heights is done by the transition matrix $\boldsymbol{\Phi}_k$. In addition, system noise $\mathbf{q}_k$ is taken into account and mapped to the grid node heights by $\boldsymbol{\Lambda}_k$. The model uncertainties are predicted by Eq. (5) where the covariance matrix $\mathbf{Q}_k$ contains the uncertainties of the system disturbance, i.e. the system noise. Since no information on the temporal evolution of the water level is known in advance, the prediction is based purely on stochastic information. Moreover, the (deterministic) system disturbances in $\mathbf{q}_k$ are set to zero. The

system noise σ_q^2 in matrix $\mathbf{Q}_k$ is assumed to yield 5 cm² for each grid node (without correlations).

The applied Kalman filter procedure as used in the DAHITI approach is described in detail below.

Initialization

The Kalman filter approach begins with an initialization step which is necessary before starting the recursive loop. The initial state vector $\mathbf{x}_k^-$ is filled by setting all elements to the observed water level with the smallest height error in the first epoch t_k. The covariance matrix $\mathbf{\Sigma}_{xx,k}^-$ is initialized by an identity matrix of size $n \times n$.

Update

In the update step, new altimeter water levels are introduced in order to update the parameters of the current state $\mathbf{x}_k^-$ to a new state $\mathbf{x}_k^+$. The update is done by comparing the estimated observations (based on the current model, cf. Eq. 2) with the water levels. The weighting of this so-called innovation is described by matrix $\mathbf{K}_k$. It can be computed based on the design matrix and the covariance matrices of observations and parameters using

$$\underset{(n,m_k)}{\mathbf{K}_k} = \underset{(n,n)}{\mathbf{\Sigma}_{xx,k}^-} \cdot \underset{(n,m)}{\mathbf{A}_k^T} \cdot (\underset{(m_k,m_k)}{\mathbf{\Sigma}_{ll,k}^-} + \underset{(m_k,n)}{\mathbf{A}_k} \cdot \underset{(n,n)}{\mathbf{\Sigma}_{xx,k}^-} \cdot \underset{(n,m_k)}{\mathbf{A}_k^T})^{-1} \quad (6)$$

The parameter update of vector $\mathbf{x}_k^+$ describes the updated water levels for each grid node at the current epoch t_k.

$$\underset{(n,1)}{\mathbf{x}_k^+} = \underset{(n,1)}{\mathbf{x}_k^-} + \underset{(n,m_k)}{\mathbf{K}_k} \cdot (\underset{(m_k,1)}{\mathbf{l}_k} - \underset{(m_k,n)}{\mathbf{A}_k} \cdot \underset{(n,1)}{\mathbf{x}_k^-}) \quad (7)$$

$$\underset{(n,n)}{\mathbf{\Sigma}_{xx,k}^+} = (\underset{(n,n)}{\mathbf{I}} - \underset{(n,m_k)}{\mathbf{K}_k} \cdot \underset{(m_k,n)}{\mathbf{A}_k}) \cdot \underset{(n,n)}{\mathbf{\Sigma}_{xx,k}^-} \quad (8)$$

In parallel the corresponding covariance matrix $\mathbf{\Sigma}_{xx,k}^+$ of the height estimates is updated using Eq. (8). The uncertainties of new altimeter data are taken into account by applying the Kalman matrix as weighting matrix. It can easily be seen that the parameter accuracies will become smaller within the updating step.

Prediction

After the parameter vector and the covariance matrix of the current epoch t_k have been updated, the prediction of $\mathbf{x}_k^+$ and $\mathbf{\Sigma}_{xx,k}^+$ to the next epoch t_{k+1} is performed and $\mathbf{x}_{k+1}^-$ and $\mathbf{\Sigma}_{xx,k+1}^-$ are computed. The predictions are used as start parameters for the next update step, and the computation loop then continues until all water levels have been processed. In our case, no additional information about the temporal propagation of the parameter vector and the covariance matrix is introduced. Therefore, no deterministic model is applied and the transition matrices $\mathbf{\Phi}_k$ for data and $\mathbf{\Lambda}_k$ for disturbances in Eq. (4) and (5) can be identity

matrices. Furthermore, only system noise is taken into account by setting the disturbance value $\mathbf{q}_k$ equal to zero and its uncertainties $\mathbf{Q}_k$ to variances of 5 cm² for each grid node without any correlations.

3.3   Post-processing

The Kalman filter provides water heights $\mathbf{x}_k$ and their formal errors $\mathbf{\Sigma}_{xx,k}$ for each epoch t_k and grid node.

If Kalman filtering is performed on a single grid node, the final water level and error are immediately available. If it is computed on a grid, a 'mean' one-dimensional time series is computed. Instead of simply averaging all grid node heights, we select only the best water levels per epoch. Only water levels are selected that which fulfill certain error criteria. In general, the limit for the maximum height error is set to values between 5 and 10 cm. The selected limit depends on the resulting height errors. Therefore, the limit is selected manually in such a manner that only reliable heights are used for the final time series. The remaining water levels are averaged for each epoch by using the formal errors for the weighting factors. Finally, a time series of water levels and their formal errors over the entire period of time are obtained.

In a last step an outlier rejection is performed. The water level time series can still contain outliers because of bad quality of data, ice coverage, orbit manoeuvres, etc. For the detection of those outliers, SVR can be applied again - now on the full time series. Complete tracks showing significant differences with respect to the other points of the water level time series can be rejected. This time, radial base functions instead of a linear kernel are used to perform the regression since a constant water level over time cannot be assumed. The radial base function as kernel of the SVR allows us to fit the time series including seasonal variations and trends. Figure 4 shows the results of an applied SVR on a six-year subset of the time series of Lake Erie. The fitted model is plotted as a cyan line together with its manually defined confidence interval. The confidence interval is selected depending on the noise of the water level time series which varies between 7.5 cm and 100 cm. Water levels which fulfill the limit of the SVR are kept (blue) whereas outliers are rejected (red).

4   Results and Validation

In this chapter, water level time series resulting from the Kalman approach are presented and validated. Since it is not possible to show results for all inland water bodies we focus on the selected study areas introduced in Section 4.1. Three inland water targets are described in more detail. They represent different target types, i.e. large lakes, small lakes, and rivers. Moreover, results from 16 lakes and 20 river crossings
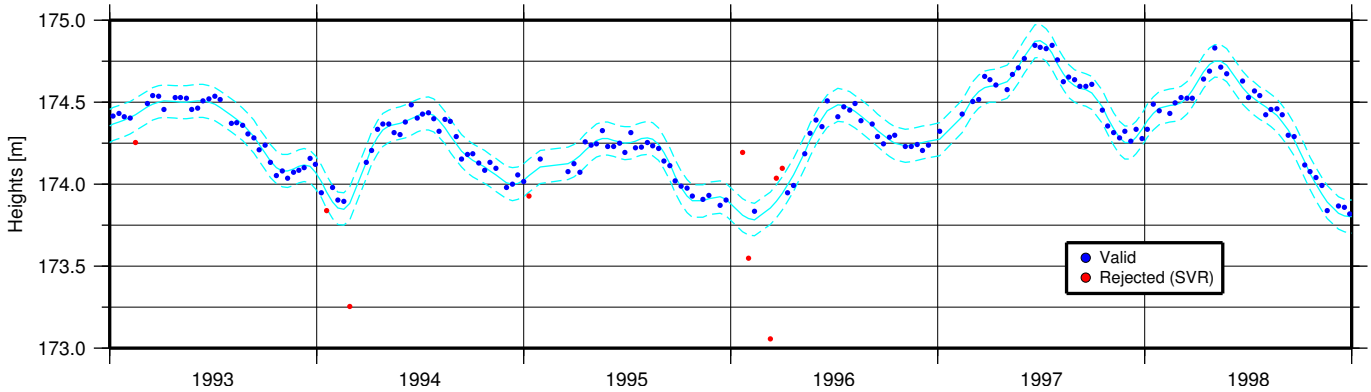
Figure 4: Example of applied SVR using radial base functions for outlier rejection on a resulting water level time series (Lake Erie) of the Kalman filtering step. The estimated regression function (cyan) and its confidence intervals (dotted cyan) are plotted. The result of the regression shows valid (blue) and rejected (red) altimeter heights. Each rejected water level height represents one complete satellite overflight.

are validated by comparison with in situ data and altimeter time series provided by other groups.

4.1 Study Areas

For altimetry-derived water level time series, in situ measurements from gauging stations are the most important validation data sets. In order to perform reliable comparisons, only those inland water bodies are selected as study areas for which in situ data are available. Since we have access to many gauging stations in North and South America we focus our study on these two continents.

Another criterion for the selection of inland water bodies is the availability of external altimetry-derived time series to demonstrate the performance of our Kalman filter method compared with other approaches. Each study case is observed by at least one other group (i.e. Hydroweb, River & Lakes or GRLM). Thus, those targets in North and South America are selected which are best represented by other inland altimetry databases for a time period as long as possible. We end up with the 16 lakes and 20 river crossings illustrated in Figure 5. For almost all investigated inland water bodies at least one in situ gauging station and one external altimetry-derived time series is available.

The first study areas are the Great Lakes of North America comprising Lake Superior (82,000km²), Lake Huron (59,000km²), Lake Michigan (58,000km²), Lake Erie (25,000km²), and Lake Ontario (19,000km²). The size of the these lakes leads to ocean-like conditions which means that the altimeter measurements are not disturbed by land. Only a few altimeter measurements near the lake shore are contaminated by land. The Great Lakes show seasonal variations of about 1 m. They are well-observed inland waters with many in situ stations provided by the 'Tides & Currents' platform of the National Oceanic and Atmospheric Adminis-

tration (NOAA)[6]. For the validation of Lake Superior, in situ stations of Duluth, Grand Marais, Marquette, Ontonagon and Point Iroqouis are used. Lake Huron has five stations for validation which are Essexville, Harbor Beach, Lakeport, Mackinaw City, and de Tour Village. The stations Calumet Harbor, Holland, Kewaunee, Ludington, Milwaukee, and Port Inland are used for Lake Michigan. Lake Erie has seven stations for validation which are Buffalo, Cleveland, Fairport, Fermi Power Plant, Marblehead, Sturgeon Point, and Toledo. For the validation of Lake Ontario, the in situ stations of Cape Vincent, Olcott, Oswego, and Rochester are used.

In addition to the Great Lakes, the Great Slave Lake (27,200 km²), Lake Winnipeg (24,000 km²), Lake Athabasca (7,800 km²), Lake Winnipegosis (5,100km²), Lake Manitoba (4,600 km²), Lake of the Woods (4,300 km²), Great Salt Lake (4,000 km²), Lake Claire (1,400 km²), and Cedar Lake (1,300 km²), which are located in Canada and the United States, are investigated. These lakes differ significantly in surface extent by up to a factor of 20. Estimation of water level time series in the Canadian lakes is made difficult by the winter conditions. Several lakes are frozen for several months which makes the water level computation challenging (Table 3). For the validation of the water level time series, in situ data provided by the government of Canada[7] and the U.S. Geological Survey (USGS)[8] are used.

In addition to the lakes in North America two lakes in the very south of South America are selected for validating our approach. Lake Argentino (1,466 km²) and Lake Buenos Aires (1,850 km²) are located in Argentina next to the Andes. The lakes are party surrounded by mountains, which can affect the altimeter measurements. The lakes have a sim-

---

[6]http://tidesandcurrents.noaa.gov/

[7]http://wateroffice.ec.gc.ca/
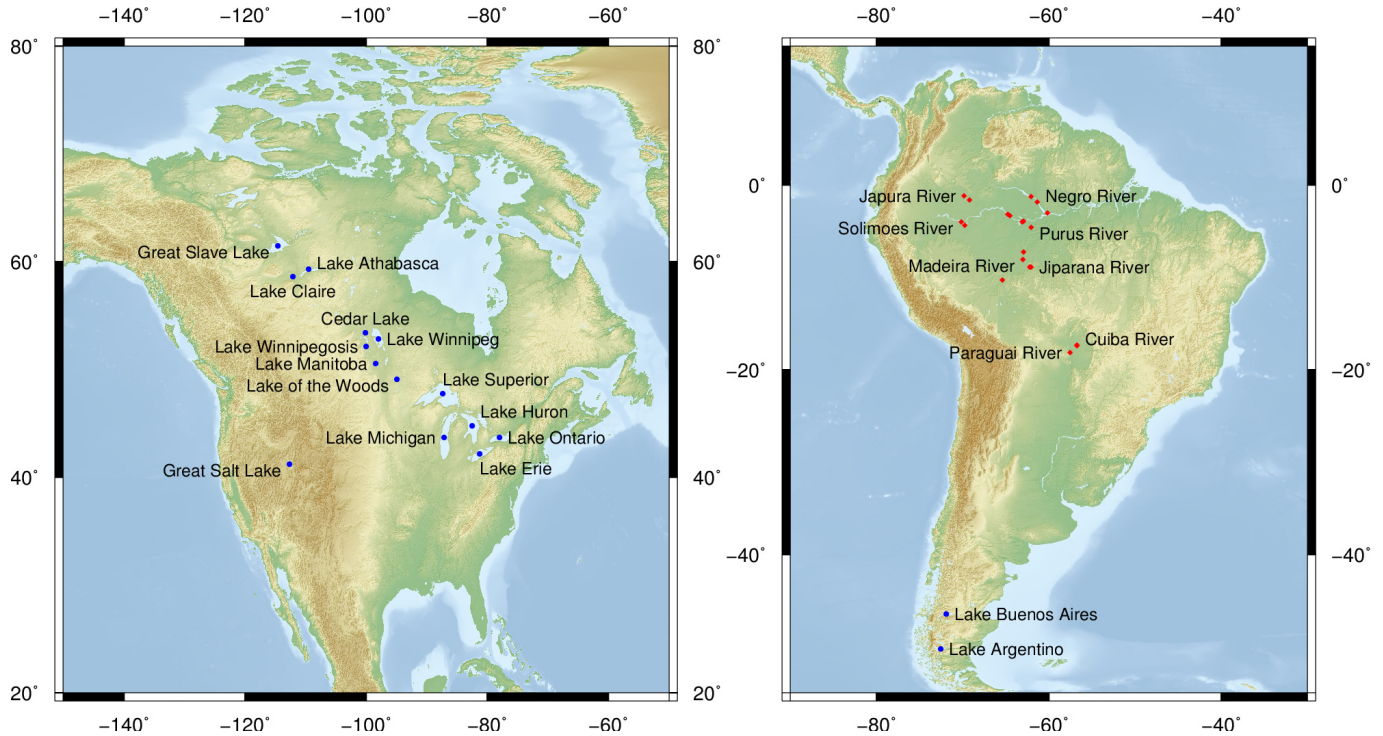
[8]http://waterdata.usgs.gov/

Figure 5: Map of selected study areas of lakes (blue) and rivers (red) in North America (left) and South America (right)

ilar shape with largest extent in across-track direction of the satellites ground track. This leads to rather short track crossings varying between 10 and 15 km. Despite their location in a temperate zone near high mountains the lakes are not frozen in winter. The seasonal variations of both lakes vary between 2.5 m and 3.5 m. For the validation of Lake Argentino and Lake Buenos Aires in situ data from the Ministerio de Planificación Federal, República Argentina[9] are used.

For the analysis of rivers, the Amazon basin is selected as the study area; it is the largest basin in the world and covers about 7,000,000 km². The region is located in the tropics, and the climate is hot and humid throughout the year. Because of the strong precipitation, the resulting seasonal variations of the water level reach peak-to-peak variations up to 15 m. The Amazon basin consists of countless rivers which differ in terms of length, width, meanders, and seasonal variations. This variety is very useful for the quality assessment of water level time series from altimetry. For example, the river widths vary from up to 10 km for the Amazon river and a few hundred metres for the Jiparaná River. Moreover, the Amazon basin is a well-observed area since the Agência Nacional de Águas (ANA)[10] provides data for numerous in situ gauging stations. For the validation, water level time series of gauges at the Japurá River, the Solimões River, the Negro River, the Purus River, Jiparaná River, Paraguai River, and the São Lourenço River are used. Another reason why we chose the Amazon basin is that other groups such as LEGOS and ESA-DMU have also investigated this area.

4.2 Validation data sets

Water level time series from gauges have a high relative accuracy, but some points must be borne in mind in the use of in situ data. The absolute comparison of heights from gauges and satellite altimetry is often very difficult since location, reference height and vertical datum of gauges are not always precisely known or may even be unknown. This leads to height offsets between water level time series from gauge and altimetry which must be considered in the validation step. In particular, the comparison between water levels from altimetry and in situ data over rivers shows in most cases remaining offsets. In general, almost no altimeter satellite track crosses the river at the location of a gauging station, which leads to additional offsets because of the river slope. To avoid handling the uncertainties of in situ data only relative comparisons with water level time series from altimetry are performed.

In order to rank our results with respect to other time series derived from altimeter data, we download water levels from three external inland altimeter data bases, namely Hydroweb, River & Lakes, and GRLM. These results are

---

[9]http://www.hidricosargentina.gov.ar/
[10]http://ana.gov.br/

based on various altimeter missions and diverse approaches were performed to compute the water level time series. As a consequence, these external time series cover different time periods with temporal resolutions between 10 and 35 days. This has to be kept in mind when the time series of the four databases are compared.

4.3 Selected Results

We choose three of the aforementioned water bodies in order to present detailed results of our Kalman filter approach. The targets are selected to represent three diverse inland water body types featuring different characteristics. Lake Superior (Figure 6) is selected as representative of larger lakes with ocean-like conditions. Lake Athabasca (Figure 7) is a smaller lake which has to cope with ice coverage in winter, which is the case for most lakes in North America. Finally, the Madeira River (Figure 8) in the Amazon basin is selected to show the potential of the Kalman filter approach for river monitoring. For all examples, the time series from DAHITI is compared with in situ data and results from Hydroweb, River & Lakes, and GRLM.

4.3.1 Lake Superior

Figure 6 shows the water level time series of Lake Superior between 1992 and 2014: the DAHITI result is plotted in blue (subplot a), the in situ data of station Ontonagon in red, and external altimetry-derived water levels in green (Hydroweb, subplot b), light blue (River & Lakes, subplot c), and orange (GRLM, subplot c). In order to neglect constant offsets between the different solutions, all time series are shifted to the level of in situ data, and only water level changes are compared. The applied offset is estimated by using the average of height differences at all days in which in situ data and time series from altimetry are available. Additionally, differences between water levels from altimetry and in situ data are plotted for each time series. For the DAHITI computation, non-retracked high-frequent altimeter data of Topex, Jason-1, Jason-2, Envisat, ERS-2, and SARAL/AltiKa are used. The Kalman filter provides a continuous time series with an irregular near-daily resolution which shows neither outliers nor inter-mission inconsistencies. In order to achieve reliable water level time series different outlier criteria are applied. Initially, the number of invalid water levels is reduced by using thresholds for latitude (depending on track length over Lake Superior), height (180 m to 185 m) and height error (10 cm). Furthermore, only backscatter coefficients between 10 db and 18 db are selected in order to reject data affected by ice coverage. Then, an SVR using a confidence limit of ± 5 cm is applied along the crossing altimeter track to reject water levels near the shore which are affected by land contamination. Finally, an SVR using a confidence limit of ± 7.5 cm is applied along the final water level time series to reject remaining outliers. Altogether, the time series is com-

posed of 3449 single points, each representing one day with at least one altimeter track crossing the lake. During computation of the final water level time series 24 % of the data are rejected, mostly because of ice coverage.

The DAHITI water levels coincide very well with the daily in situ data of Ontonagon. The correlation coefficient R^2 is 0.96 and the RMS difference is 4.4 cm. The alternative computation of the water level time series using a median filter instead of the Kalman filter leads to a slightly worse RMS difference of 4.5 cm (see Section 4.3.4). In comparison with the DAHITI time series, the other altimetry-derived water levels show significantly reduced temporal resolutions. In addition, the lengths of the time series differ, depending on the missions used by the different groups. In order to rank the DAHITI result compared with other altimetry-derived water levels, we also compare the three external time series with in situ gauging data within the corresponding time intervals. For all three databases this gives smaller correlations and higher RMS (Hydroweb: RMS=5.7 cm, R^2=0.95, 228 points, River & Lakes: RMS=8.2 cm, R^2=0.82, 82 points, and GRLM: RMS=12.1 cm, R^2=0.74, 760 points). For the validation, the water level time series of the other altimetry-derived water levels are used as they are, without any additional outlier rejection. This leads to higher RMS differences as published in Ričko et al. (2013), who applied an additional outlier rejection based on in situ data. The altimetry derived-solutions differ because of varying input data sets and the different approaches. Hydroweb uses a multi-mission approach with a merged monthly resolution whereas River & Lakes relies purely on Envisat with a temporal resolution of 35 days. GRLM applies a multi-mission approach providing a temporal resolution of about 10 days. The time series of Hydroweb and GRLM still show mission-dependent offsets which can be seen in the differences from the in situ data (mainly positive for ERS-2, mainly negative for Envisat). In contrast, mission-dependent offsets are quite small in the water level time series of DAHITI.

4.3.2 Lake Athabasca

Figure 7 shows the water level time series of Lake Athabasca between 1992 and 2014. Once again, water levels from DAHITI (blue), in situ data of Crackingstone Point (red), Hydroweb (green), River & Lakes (light blue), and GRLM (orange) are plotted. The time series of the four altimeter data bases are shifted to the level of the in situ data. In principle, Lake Athabasca, whose surface covers 7,800 km², should be large enough to provide reliable altimetry-derived water level time series. However, different problems such as ice coverage because of regular freezing in winter, land contamination and off-nadir effects near lake shores have to be considered. For the estimation of the water level time series in DAHITI retracked altimeter data are used, with a 10% Improved Threshold retracker (Hwang et al., 2006). For the computation, altimeter data of Topex, Jason-1, Jason-2, Envisat, ERS-
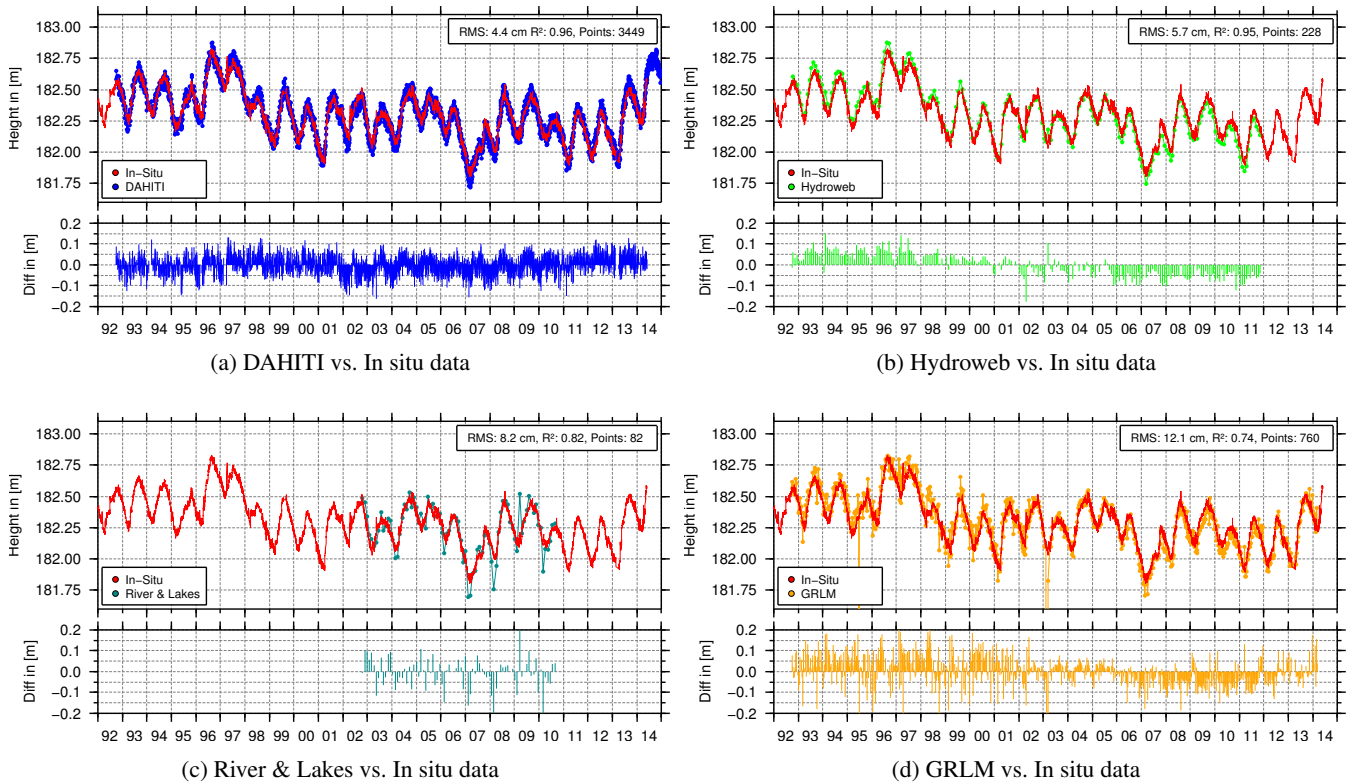
Figure 6: Water level time series of Lake Superior from DAHITI (1992-2014), Hydroweb (1992-2011), River & Lakes (2002-2010) and GRLM (1992-2014) compared with in situ data (Ontonagon, 1992-2014) and shifted to the water level height of the in situ data. Additionally, differences between heights from altimetry and in situ data are plotted for periods in which both data sets are available.

2 and SARAL/AltiKa are used. In order to achieve reliable water level time series the same outlier criteria as for Lake Superior but different thresholds are applied. First, outliers are rejected by using thresholds for latitude (depending on track length over Lake Athabasca), height (208 m to 212 m) and height error (50 cm). Furthermore, water levels affected by ice coverage are rejected if the valid backscatter coefficients are not between 10 db and 18 db. To reject water levels near the shore which are affected by land contamination, an SVR along the crossing altimeter track using a confidence limit of ±5 cm is applied. Finally, an SVR along the final water level time series using a confidence limit of ±50 cm is applied to reject remaining outliers.

The DAHITI water level shows a very good agreement with in situ data in summer and almost no outliers owed to ice coverage are visible in winter compared with time series from Hydroweb and River & Lakes. The overall consistency with the gauge data yields a correlation coefficient of 0.90 and an RMS difference of 15.1 cm using 1279 points in the period between 1992 and 2014. The usage of a median filter leads to slightly worse RMS differences of 15.3 cm for Lake Athabasca. The differences between in situ data and Hydroweb (RMS=32.1 cm, R²=0.79, 224 points), River

& Lakes (RMS=80.5 cm, R²=0.30, 79 points) and GRLM (RMS=55.7 cm, R²=0.27, 76 points) show higher RMS values and smaller correlations. One can clearly see that the problems of altimeter time series occur mostly in winter because of ice coverage. In particular, water level time series of Hydroweb and River & Lakes show strong outliers in winter which are not contained in the time series of DAHITI because of the applied outlier rejection. A new problem with retracker biases arises for time series based on retracked altimeter data. To minimize those effects all altimeter measurements are retracked using with 10% Improved Threshold retracker. However, small retracker biases can also occur if identical retracking algorithms are applied on altimeter missions measuring in different bands such as Ku-band (Envisat) and Ka-band (SARAL/AltiKa).

4.3.3 Madeira River

Figure 8 shows the resulting water level derived from an Envisat and SARAL/AltiKa crossing over the Madeira River. The water level time series from DAHITI (blue), Hydroweb (green) and River & Lakes (light blue) are compared with the in situ station Humaitá (red), which is located about 27.6
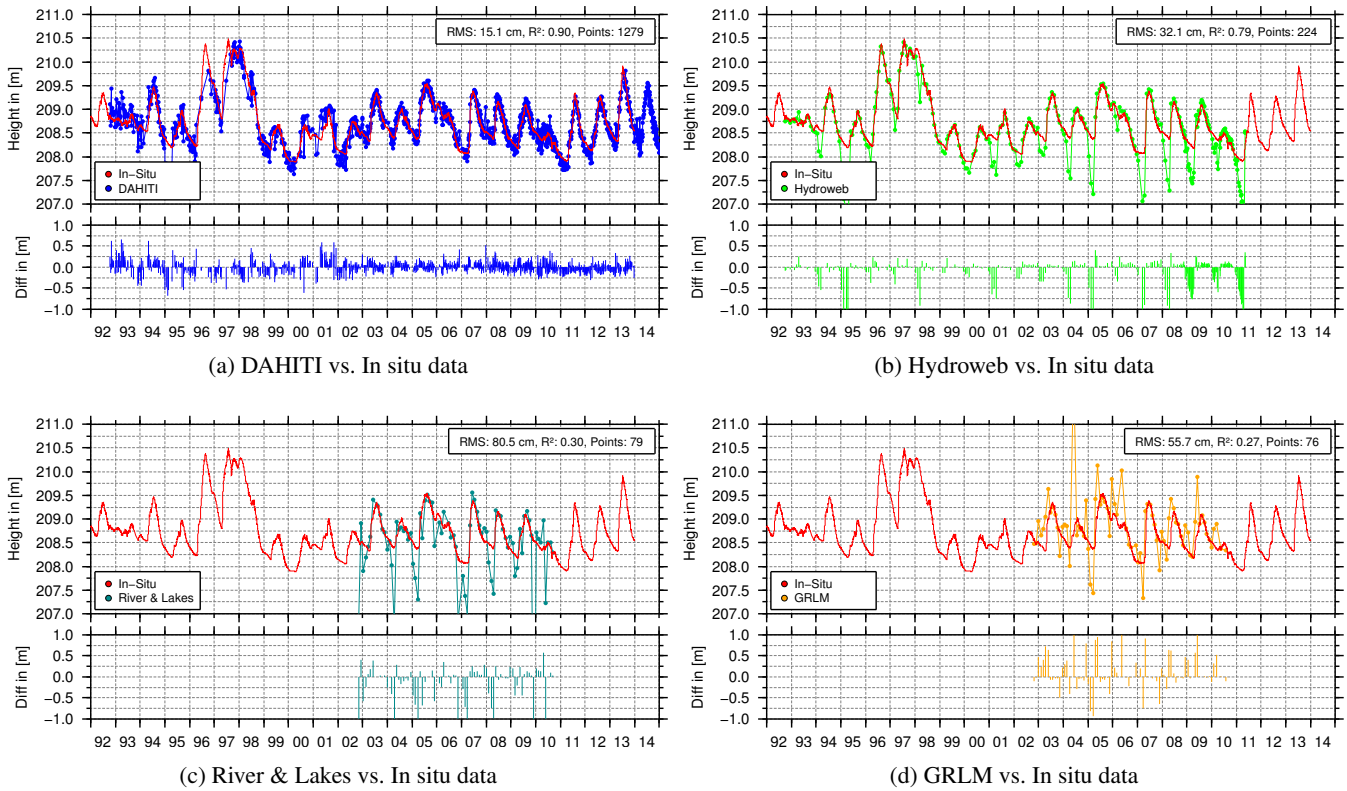
(a) DAHITI vs. In situ data

(b) Hydroweb vs. In situ data

(c) River & Lakes vs. In situ data

(d) GRLM vs. In situ data

Figure 7: Water level time series of Lake Athabasca from DAHITI (1992-2014), Hydroweb (1992-2011), River & Lakes (2002-2010) and GRLM (2002-2014) compared with in situ data (Lake Athabasca, 1992-2013) and shifted to the water level height of the in situ data. Additionally, differences between heights from altimetry and in situ data are plotted for periods in which both data sets are available.

km upstream. All time series from altimetry are shifted to the water level of the in situ station. At this location the Madeira River is about 2.5 km wide. In order to achieve reliable water level time series over the Madeira River different outlier criteria are applied. First, thresholds for latitude (depending on track length over the Madeira River), height (30 m to 50 m) and height error (100 cm) are applied to reduce the number of invalid water levels. Finally, an SVR along the crossing altimeter track using a confidence limit of ± 10 cm and an SVR along the final water level time series using a confidence limit of ± 100 cm are applied to reject remaining outliers. In this case, no limit for the backscatter coefficients is applied because no ice coverage exists in the Amazon basin. In principle, the backscatter coefficient can also be used to distinguish between water and land but this is not considered here. All altimeter time series reach a temporal resolution of about one month since there is only one mission with 35-day temporal resolution at the same time. Altimeter data are available between 2002 and 2014 with a data gap between October 2010 and March 2013. The altimeter data from Envisat on the shifted orbit can not be used between October 2010 and April 2012 for the current water level time series. Gauging information does not start before 2007. Thus, the compari-

son with in situ data only comprises a time period of about 3.5 years. For DAHITI another year of SARAL/AltiKa data is available. The Kalman filter result (blue) shows an RMS difference of 19.4 cm and a correlation coefficient of 1.00 by using 35 points. The estimation of the water level time series using a median filter leads to RMS difference of 19.6 cm. The RMS is comparable to the result for Lake Athabasca, which is even more satisfactory when we take into account the seasonal variations of about 15 m of the Madeira River. The high amplitude is also the reason for the extremely high correlation, which should not be overvalued. The RMS differences of Hydroweb and River & Lakes with respect to the gauge are twice as great, at 45.1 cm (Hydroweb, 29 points) and 53.2 cm (River & Lakes, 28 points) respectively. GRLM does not provide information for this virtual station.

4.3.4 Discussion

The DAHTIT time series show good consistency with in situ observations and clear advances over established approaches. However, some problems remain, especially for smaller lakes and rivers. For larger lakes, the assumption of a uniform surface level may no longer be justified. In addition to height

(a) DAHITI vs. In situ data

(b) Hydroweb vs. In situ data
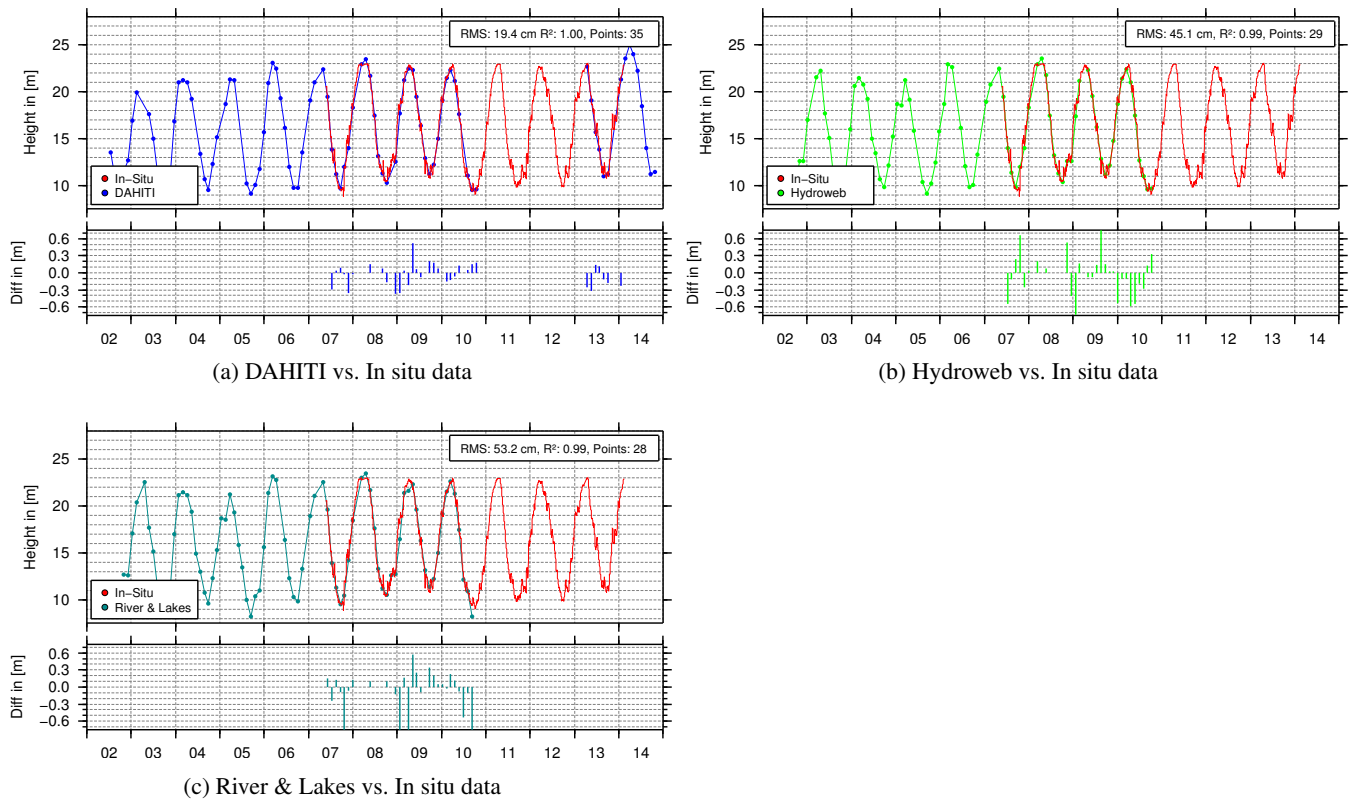
(c) River & Lakes vs. In situ data

Figure 8: Water level time series of the Madeira River from DAHITI (2002-2014), and River & Lakes (2002-2010) compared with in situ data (Humaitá, 2007-2014) and shifted to the water level height of the in situ data. Additionally, differences between heights from altimetry and in situ data are plotted for periods in which both data sets are available.

differences owed to systematic errors in geophysical corrections or the geoid, hydrodynamic effects caused by wind and waves can cause horizontal lake level differences. Currently, these are neglected when combine observations from diverse parts of the lake. Moreover, measurements (altimetry as well as in situ) feature unequal accuracies observed over areas with different surface conditions. This effect can be seen when we compare the DAHITI water level time series of Lake Superior with additional gauging stations. The five possible comparisons lead to RMS differences varying by 2 cm (between 4.4 cm and 6.6 cm; Table 4). The two stations Duluth and Point Iroquois show reduced consistency with altimetry. Both stations are located in smaller bays of the lake and are more affected by wind and waves than the other stations, which leads to more noisy in situ time series.

For small lakes and rivers, land contamination of waveforms is the largest problem because nearly all altimeter measurements are affected. For rivers, almost no nadir measurements may occur and even these can originate from river branches and distort the water level time series from the investigated target. Moreover, the river slope can influence the time series as well as the comparison with in situ data. The crossings between river and altimeter track can vary slightly (up to 1 km) because of orbit instabilities so that the reflec-

tions originate from different areas which do not exhibit the same water level. The most important challenge remaining is the handling of inter-mission biases and retracker biases. The usage of radial errors from a global crossover analysis and the restriction to one common retracker works reasonably well; however, small discrepancies remain in the time series. Moreover, the quality of the single altimeter measurements could surely be further improved by combining different retracking algorithms depending on the waveform shapes. This remains a major challenge and offers enormous potential for future work.

The validation of water level time series of DAHITI for Lake Superior, Lake Athabasca, and the Madeira River compared with in situ data and time series from Hydroweb, River & Lakes, and GRML showed clear improvements. To evaluate the impact of the outlier rejection and Kalman filtering on the improvements of the DAHITI time series, an alternative approach using a simple median filter instead of a Kalman filter was applied.

The resulting RMS differences for three inland waters decreased slightly by 0.1 cm to 0.2 cm, which indicates that the combination strategy has only a moderate effect on the overall accuracy. The strongest improvements are currently owed to rigorous outlier detection and data retracking. However,

the Kalman filter has a considerable potential when upgraded by dynamic modelling and used for real-time applications.

4.4   Quality Assessment

The results for Lake Superior, Lake Athabasca, and the Madeira River presented in Section 4.3 already show the ability of the Kalman filter approach to provide reliable and highly accurate time series of inland water levels. Since three results – even if they do represent different inland water types – are not enough to perform a reliable quality assessment of the method, we extend the validation to a larger sample and include all study targets (16 lakes and 20 river crossings) described in Section 4.1 in the comparison.

Table 3 gives an overview of the different parameters used for the estimation of water level time series in DAHITI. This information is provided for all investigated lakes and rivers. First, the used altimeter missions are shown, followed by the retracking flag which indicates if additional retracking is applied. Then the ice flag shows if the water body is affected by ice coverage in winter. This information originates from external sources, e.g. National Snow and Ice Data Center (*http://nsidc.org/*) for Lake Superior. Table 3 also shows which outlier criteria were applied for the different inland water targets to reject erroneous water levels. Consequently, appropriate thresholds for latitude, height, backscatter coefficient, height error, SVR along the pass and along the final time series can be selected. Finally, the number of data points of the water level time series are shown which is equal to the number of days in which altimeter data are available. The last column describes the percentage of outliers which were rejected during the computation of the water level time series.

Table 4 summarizes the comparisons of lake level time series from DAHITI, Hydroweb, River & Lakes, and GRLM with in situ gauge data. For each target, RMS difference, squared correlation coefficient and the number of points (No) used for the validation are provided. Depending on the availability of in situ time series of the investigated water body, more than one comparison is performed for the larger lakes. The smallest RMS difference for each target is highlighted in bold, the largest one in italics.

DAHITI results show RMS differences with respect to the gauge data between 4 cm and 36 cm. It is obvious that accuracy declines with lake extent and ice coverage. For some lakes, the differences between DAHITI and in situ data vary by more than a factor of two with different lake gauges. Especially for Lake Erie the difference between the RMS values can reach up to 8.1 cm. For most lakes, the relations between the different RMS values are similar for the different altimeter products.

For most lakes the DAHITI water levels are more consistent with in situ data than the results from external altimeter data bases. In addition, the temporal resolutions of the time series are significantly higher, as indicated by the number of points used for the validation. Of course, the different

time periods of the other altimeter data sets have to be taken into account, too. The most notable improvements through the DAHITI Kalman approach with respect to the existing databases can be seen for smaller lakes. For example, for the Lake of the Woods, the DAHITI consistency with in situ data is more than twice as good as the other altimeter products, improving the RMS differences from about 36 cm to approximately 16 cm.

The validation results for different rivers in the Amazon basin are summarized in Table 5. We study eight different rivers with 20 virtual stations altogether. For the computation, data from Jason-2, Envisat, and SARAL/AltiKa are used. Most of the time series are based on only one altimeter track (sometimes from consecutive missions, e.g. Envisat and SARAL/AltiKa). Few locations allow use of more than one track in case of a crossover point between different altimeter tracks. Table 5 shows the comparison results of three altimeter products (DAHITI, Hydroweb, and River & Lakes) with different in situ stations. GRLM does not provide river level time series and is excluded from this investigation. In addition to RMS differences with respect to the gauging time series and correlation coefficients, the number of data points, river width and distance between altimeter crossing and gauge are given. The river width corresponds to altimeter track length crossing the river based on satellite images from Google Maps. Positive distances indicates downstream gauges, negative differences indicate upstream gauges.

The RMS differences between altimeter time series and in situ data vary between 8 cm and 114 cm in the case of DAHITI. For most virtual stations, the consistency with the gauge is considerably lower than for lakes. It is not possible to prove a dependence between river width and distance to the gauge, not only because of the altimeter time series but also because of the accuracies of the in situ data which also contain measurement errors. Also, the angle in which the satellite track crosses the river has a strong impact on the quality of the water level time series. Furthermore, distances of tenths of kilometres between the in situ station and the nearest crossing altimeter track make it more difficult to prove dependences owed to unpredictable river flow effects.

Compared with time series from Hydroweb and River & Lakes, the new DAHITI approach can improve the gauge consistency for most of the targets. The improvement can reach several decimetres. Many correlation coefficients in Table 5 are close to one. This is not necessarily an indication of optimal consistency between altimeter water level and gauging observations but is significantly influenced by the large absolute water level variations (more than 10 m).

5   Conclusions

This paper presents a new method for estimating water level time series over inland water using multi-mission satellite altimetry data. It is based on careful data preprocessing (in-

Table 3: Relevant parameters for the estimation of the water level time series for DAHITI. This table shows information about the used altimeter missions for selected lakes and rivers, applied retracking (Retr.), ice coverage (Ice), and applied outlier criteria which are used for the processing of water level time series. Finally, the number of data points and percentage of outliers of the final water level time series are given.

Target name (DAHITI Id)	Used Missions						Retr.	Ice	Applied Outlier Criteria						Data Points	Outlier
	TP	J1	J2	E2	EN	SA			Lat.	Height	Sig.0	Error	SVR Pass	SVR Series		
Superior, Lake (3)	✓	✓	✓	✓	✓	✓	−	✓	✓	✓	✓	✓	✓	✓	3614	24 %
Huron, Lake (33)	✓	✓	✓	✓	✓	✓	−	✓	✓	✓	✓	✓	✓	✓	2132	23 %
Michigan, Lake (11)	✓	✓	✓	✓	✓	✓	−	✓	✓	✓	✓	✓	✓	✓	2132	37 %
Erie, Lake (6)	✓	✓	✓	✓	✓	✓	−	✓	✓	✓	✓	✓	✓	✓	1968	41 %
Ontario, Lake (35)	✓	✓	✓	✓	✓	✓	−	✓	✓	✓	✓	✓	✓	✓	2174	29 %
Athabasca, Lake (100)	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	1398	45 %
Great Slave, Lake (99)	✓	✓	✓	✓	✓	✓	−	✓	✓	✓	✓	✓	✓	✓	1396	39 %
Claire, Lake (578)	−	✓	✓	−	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	593	23 %
Winnipeg Lake (101)	−	−	−	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	816	13 %
Manitoba, Lake (191)	−	−	−	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	−	535	4 %
Cedar, Lake (200)	✓	✓	✓	−	✓	✓	−	✓	✓	✓	✓	✓	✓	✓	647	18 %
Winnipegosis, Lake (281)	−	✓	✓	−	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	529	20 %
Lake of the Woods (73)	✓	✓	✓	−	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	687	49 %
Great Salt, Lake (72)	−	−	−	✓	✓	✓	✓	−	✓	✓	−	✓	✓	✓	147	9 %
Argentino, Lake (182)	✓	✓	✓	−	✓	✓	−	−	✓	✓	−	✓	✓	✓	880	49 %
Buenos Aires, Lake (139)	−	−	−	−	✓	✓	−	−	✓	✓	−	✓	✓	−	116	3 %
Solimões, River (405)	−	−	✓	−	✓	✓	✓	−	✓	✓	−	✓	✓	−	323	3 %
Solimões, River (406)	−	−	−	−	✓	✓	✓	−	✓	✓	−	✓	✓	−	90	12 %
Solimões, River (389)	−	−	−	−	✓	✓	✓	−	✓	✓	−	✓	✓	−	86	12 %
Solimões, River (581)	−	−	✓	−	−	−	✓	−	✓	✓	−	✓	✓	✓	178	26 %
Solimões, River (384)	−	−	−	−	✓	−	✓	−	✓	✓	−	✓	✓	−	82	0 %
Solimões, River (582)	−	−	✓	−	−	−	✓	−	✓	✓	−	✓	✓	−	198	17 %
Purus, River (583)	−	−	−	✓	✓	−	✓	−	✓	✓	−	✓	✓	−	86	10 %
Jiparaná, River (584)	−	−	−	✓	✓	−	✓	−	✓	✓	−	✓	−	−	91	9 %
Jiparaná, River (585)	−	−	✓	−	−	−	✓	−	✓	✓	−	✓	−	✓	235	2 %
Japurá, River (579)	−	−	−	✓	✓	−	✓	−	✓	✓	−	✓	−	−	99	0 %
Japurá, River (580)	−	−	−	−	✓	−	✓	−	✓	✓	−	✓	−	−	81	2 %
São Lourenço, River (1093)	−	−	✓	−	−	−	✓	−	✓	✓	−	✓	✓	✓	233	3 %
São Lourenço, River (1094)	−	−	✓	−	−	−	✓	−	✓	✓	−	✓	✓	✓	232	3 %
Madeira, River (371)	−	−	−	−	✓	✓	✓	−	✓	✓	−	✓	✓	−	90	10 %
Madeira, River (360)	−	−	✓	−	−	−	✓	−	✓	✓	−	✓	✓	✓	227	5 %
Madeira, River (575)	−	−	−	−	✓	−	✓	−	✓	✓	−	✓	✓	−	81	4 %
Negro, River (161)	−	−	−	−	✓	✓	✓	−	✓	✓	−	✓	✓	−	89	13 %
Negro, River (352)	−	−	−	−	✓	✓	✓	−	✓	✓	−	✓	−	−	99	2 %
Negro, River (346)	−	−	−	−	✓	−	✓	−	✓	✓	−	✓	✓	−	81	2 %
Paraguai, River (1095)	−	−	−	−	✓	✓	✓	−	✓	✓	−	✓	−	−	99	0 %

Missions: Topex/Poseidon (TP), Jason-1 (J1), Jason-2 (J2), ERS-2 (E2), Envisat (EV), SARAL/AltiKa (SA)

cluding waveform retracking), a Kalman filter approach, and a rigorous outlier detection. The introduced method is the basis of DAHITI, an online database for inland water level time series from satellite altimetry observations operated by the Deutsches Geodätisches Forschungsinstitut der Technischen Universität München (DGFI-TUM).

The study demonstrates the performance of the new method for numerous lakes and rivers in North and South America. A comprehensive validation is performed by comparison with time series of water level variations from in situ gauging stations. Moreover, a comparison with external altimetry-derived water level variations is presented based on data from Hydroweb (LEGOS), the River & Lakes database (ESA-DMU), and the Global Reservoir and Lake Monitor (GRLM).

The lake level data sets computed with the presented approach yield accuracies between 4 cm and 36 cm depending on the surface extent of the lake and climate conditions (i.e. ice coverage). For rivers, the performance is consider-

ably lower with RMS differences varying between 8 cm and 114 cm. Here the accuracy mainly depends on the crossing angle of the altimeter track and the surrounding conditions. Also other surrounding conditions such as topography, quality of waveforms and their retracked water heights can influence the resulting water level time series. Especially in the Amazon basin the river meander can also change over the years because of strong seasonal variations.

For most study cases, the new approach yields significant accuracy improvements compared with water level variations provided by established inland altimeter databases, especially for smaller lakes and rivers. In addition, the temporal resolution of the DAHITI lake time series is significantly improved compared with other data sets, allowing for the detection of sub-monthly temporal changes.

The reasons for the improved performance of the presented approach are multiple: first, a larger observation data set is used as input as a multi-mission concept is realized. All available altimeter missions are cross-calibrated and in-

Table 4: Water level time series of selected lakes and reservoirs from DAHITI, Hydroweb, River & Lakes and GRLM compared with in situ data. For each comparison of water level time series from altimetry with in situ data an RMS difference and squared correlation is computed. The number of points from the final water level time series which were used for the validation is given in the third column (No) of each altimeter dataset.

Lake name - Station name - (DAHITI Id)	DAHITI			Hydroweb			River & Lakes			GRLM		
	RMS [cm]	R²	No	RMS [cm]	R²	No	RMS [cm]	R²	No	RMS [cm]	R²	No
Superior - Duluth[1] (3)	**5.4**	0.92	3485	6.6	0.92	229	9.3	0.73	82	_12.4_	0.72	767
Superior - Grand Marais[1] (3)	**4.4**	0.95	3483	5.2	0.95	279	8.5	0.80	82	_11.8_	0.75	767
Superior - Marquette[1] (3)	**4.8**	0.94	3485	5.5	0.95	229	8.5	0.80	82	_11.9_	0.75	767
Superior - Ontonagon[1] (3)	**4.4**	0.95	3449	5.7	0.95	228	8.2	0.82	82	_12.1_	0.74	760
Superior - Point Iroqouis[1] (3)	6.6	0.84	2619	**6.5**	0.84	141	9.5	0.75	82	_12.4_	0.62	510
Huron - Essexville[1] (33)	9.1	0.93	2048	11.2	0.93	230	**8.9**	0.80	76	_11.5_	0.90	772
Huron - Harbor Beach[1] (33)	**5.2**	0.98	2049	_7.7_	0.97	230	6.4	0.89	76	6.8	0.96	773
Huron - Lakeport[1] (33)	**6.3**	0.96	1960	_8.4_	0.96	215	7.2	0.86	75	7.6	0.95	737
Huron - Mackinaw City[1] (33)	**4.9**	0.97	1925	6.2	0.98	208	6.7	0.88	75	_7.7_	0.94	711
Huron - De Tour Village[1] (33)	**4.7**	0.98	2007	_6.9_	0.98	222	6.2	0.89	76	_6.9_	0.96	749
Michigan - Calumet Harbor[1] (11)	7.6	0.95	2045	_10.5_	0.94	228	**7.5**	0.87	76	8.7	0.94	765
Michigan - Holland[1] (11)	5.6	0.91	1464	7.2	0.84	131	**5.3**	0.93	76	_8.3_	0.82	481
Michigan - Kewaunee[1] (11)	5.4	0.92	1403	6.7	0.86	124	**5.0**	0.94	73	_8.7_	0.80	459
Michigan - Ludington[1] (11)	5.4	0.93	1448	6.8	0.85	124	**5.0**	0.94	73	_8.7_	0.80	459
Michigan - Milwaukee[1] (11)	5.9	0.97	2075	_8.8_	0.96	230	**5.3**	0.93	78	8.4	0.95	774
Michigan - Port Inland[1] (11)	**5.8**	0.91	1448	7.4	0.83	130	6.1	0.91	69	_10.1_	0.74	480
Erie - Buffalo[1] (6)	**11.5**	0.79	1892	16.4	0.72	221	17.2	0.50	81	_18.7_	0.61	714
Erie - Cleveland[1] (6)	**5.9**	0.94	1859	9.0	0.92	213	13.2	0.70	81	_13.6_	0.78	136
Erie - Fairport[1] (6)	**5.2**	0.95	1858	8.6	0.93	213	12.6	0.74	81	_13.5_	0.79	694
Erie - Fermi Power Plant[1] (6)	**10.5**	0.84	1819	14.9	0.78	212	15.3	0.63	81	_17.0_	0.70	693
Erie - Marblehead[1] (6)	**8.8**	0.88	1827	13.4	0.81	210	15.3	0.62	80	_15.9_	0.73	684
Erie - Sturgeon Point[1] (6)	**10.3**	0.83	1870	14.8	0.77	218	16.3	0.55	80	_17.5_	0.65	706
Erie - Toledo[1] (6)	**13.3**	0.77	1888	18.6	0.69	220	17.2	0.57	81	_19.5_	0.64	714
Ontario - Cape Vincent[1] (35)	5.3	0.96	2089	6.5	0.95	227	**4.5**	0.97	75	_10.8_	0.85	729
Ontario - Olcott[1] (35)	**4.5**	0.97	1976	6.1	0.96	210	4.9	0.96	72	_11.0_	0.85	681
Ontario - Oswego [1] (35)	5.2	0.96	2098	6.6	0.95	229	**4.6**	0.97	75	_10.8_	0.85	732
Ontario - Rochester[1] (35)	4.6	0.97	2099	6.1	0.96	229	**4.4**	0.97	75	_10.8_	0.85	732
Athabasca - Crackingstone Point[2] (100)	**15.1**	0.90	1279	32.1	0.79	224	_80.5_	0.30	79	55.7	0.27	76
Great Slave - Hay River[2] (99)	**13.3**	0.68	1209	_31.2_	0.37	246	-	-	-	-	-	-
Claire - Prairie Point[2] (578)	**19.6**	0.37	404	-	-	-	_37.9_	0.25	70	-	-	-
Winnipeg - George Island[2] (101)	**11.8**	0.87	778	28.6	0.66	146	_41.9_	0.49	77	33.0	0.59	397
Winnipeg - Gimli[2] (101)	**15.6**	0.79	758	30.1	0.61	147	_42.4_	0.48	76	36.2	0.50	394
Winnipeg - Pine Dock[2] (101)	**12.6**	0.86	694	29.7	0.67	139	_42.7_	0.51	74	34.6	0.56	381
Manitoba - Westbourne[2] (191)	**13.4**	0.85	499	34.2	0.42	100	34.2	0.33	73	_46.0_	0.11	71
Manitoba - Steep Rock[2] (191)	**13.3**	0.85	499	36.4	0.40	101	35.5	0.33	75	_47.3_	0.11	72
Cedar - Oleson Point [2] (200)	**35.9**	0.86	545	_76.7_	0.20	252	54.8	0.54	78	-	-	-
Winnipegosis - Winnipegosis[2] (281)	**16.5**	0.91	469	_36.7_	0.63	136	34.2	0.61	70	36.2	0.53	67
Lake of the Woods - Clearwater Bay[2] (73)	**16.8**	0.72	648	32.3	0.46	206	_36.6_	0.40	77	-	-	-
Lake of the Woods - Cyclone Island[2] (73)	**16.1**	0.74	642	31.0	0.51	206	_35.9_	0.41	77	-	-	-
Lake of the Woods - Hanson Bay[2] (73)	**16.3**	0.73	642	30.4	0.53	207	_36.0_	0.40	77	-	-	-
Great Salt - Saltair Boat Harbor[3] (72)	**7.4**	0.91	44	20.0	0.38	35	-	-	-	_29.4_	0.21	27
Argentino - Calafate[4] (182)	**14.6**	0.97	856	_21.9_	0.93	185	-	-	-	-	-	-
Buenos Aires - Los Antiguos[4] (139)	**19.0**	0.73	47	_29.4_	0.70	19	-	-	-	-	-	-

Source of in-situ data: [1]NOAA Tides and Currents, [2]Canada Wateroffice, [3]U.S. Geological Survey (USGS), [4]Ministerio de Planificación Federal, República Argentina

corporated in the computations. Second, the applied prepro-1115 cessing consists of a robust outlier elimination and optional retracking. This ensures that only highly accurate data will be used. Moreover, the Kalman filter approach permits the

optimal combination of all data sets and also includes the accuracies of the input data for weighting. This also enables rigorous error propagation and the computation of formal errors for each water level height. Further comparisons for

Table 5: Water level time series of selected rivers of the Amazon basin from DAHITI, Hydroweb and River & Lakes compared with in situ data. For each comparison of water level time series from altimetry with in situ data an RMS difference and squared correlation is computed. The number of points from the final water level time series which were used for the validation is given in the third column (No) of each altimeter dataset. Additionally, the distance to the nearest in-situ station (upstream (+), downstream (-)) and the river width at the crossing altimeter track is shown.

Target name - Station name (DAHITI-ID)	Distance [km]	River width [km]	DAHITI RMS [cm]	DAHITI R²	DAHITI No	Hydroweb RMS [cm]	Hydroweb R²	Hydroweb No	River & Lakes RMS [cm]	River & Lakes R²	River & Lakes No
Solimões, River - Tabatinga[1] (405)	+ 28.8	~3.8	39.6	0.99	222	*39.9*	0.99	86	**29.5**	1.00	28
Solimões, River - Tabatinga[1] (406)	- 23.2	~2.8	**17.4**	1.00	48	-	-	-	*119.9*	0.88	20
Solimões, River - Tefé[1] (389)	- 14.0	~2.6	**12.3**	1.00	35	-	-	-	*14.8*	1.00	29
Solimões, River - Tefé[1] (581)	+ 23.1	~3.7	**24.5**	0.99	95	*53.9*	0.98	84	-	-	-
Solimões, River - Itapéua[1] (384)	- 13.9	~4.4	**33.9**	0.99	40	*110.9*	0.91	39	-	-	-
Solimões, River - Itapéua[1] (582)	+ 8.9	~2.6	**31.3**	0.99	137	*61.2*	0.97	97	-	-	-
Purus, River - Aruma-Jusante[1] (583)	- 12.5	~1.4	**20.0**	1.00	16	24.1	1.00	7	*318.9*	0.61	6
Jiparaná, River - Tabajara[1] (584)	- 14.3	~0.4	**113.8**	0.87	47	*335.5*	0.29	33	-	-	-
Jiparaná, River - Tabajara[1] (585)	+ 2.4	~0.3	**46.7**	0.97	93	-	-	-	-	-	-
Japurá, River - Vila Bittencourt[1] (579)	- 40.1	~2.6	34.0	0.99	24	67.2	0.90	24	**31.2**	0.99	25
Japurá, River- Vila Bittencourt[1] (580)	+ 47.5	~1.9	**41.0**	0.98	26	61.3	0.93	25	*115.1*	0.80	14
São Lourenço, River - Posada Taiama[1] (1093)	+ 0.8	~0.3	**25.4**	0.91	160	-	-	-	-	-	-
São Lourenço, River - Posada Taiama[1] (1094)	+ 4.8	~0.3	**21.7**	0.94	157	-	-	-	-	-	-
Madeira, River - Humaitá[1] (371)	- 27.6	~2.5	**19.4**	1.00	35	45.1	0.99	29	*53.2*	0.99	28
Madeira, River - Humaitá[1] (360)	+ 70.5	~1.5	**36.3**	0.99	173	*50.2*	0.99	91	-	-	-
Madeira, River - Guajará-Mirim[1] (575)	- 48.7	~2.4	**75.5**	0.91	36	87.4	0.88	35	*134.3*	0.77	36
Negro, River - Porto de Manaus[1] (161)	+ 15.5	~10.0	**7.6**	1.00	88	25.2	1.00	79	*72.0*	0.96	78
Negro, River - Moura[1] (352)	- 64.8	~4.5	**60.0**	0.97	62	*70.9*	0.96	43	-	-	-
Negro, River - Moura[1] (346)	+ 51.8	~18.5	**43.6**	0.98	42	*46.3*	0.97	45	44.1	0.98	40
Paraguai, River - Sao Francisco[1] (1095)	- 32.3	~0.8	**22.5**	0.96	46	-	-	-	-	-	-

Source of in-situ data: [1] Agência Nacional de Águas (ANA)

the three selected areas show that using the Kalman filter approach instead of a median approach leads to slightly decreased RMS differences. This indicates that the major improvements in the water level times of DAHITI are owed to the extended outlier rejection. In future, the Kalman filter approach will also be used for (near) real-time analysis and integration of altimeter data (with the so-called Operational Geophysical Data Record, OGDR). This enables daily actualization of the water level time series and may also be used for short-time predictions. Furthermore, the introduction of a dynamic model in the Kalman filter will cause an increase in the temporal resolution of the water level time series. For the development of the dynamic model external data sets such as GRACE, precipitation, etc. can be used.

In spite of the improved water level time series of DAHITI compared with results from Hydroweb, River & Lakes and GRLM, there are still some challenging tasks which have to be taken into account to make further improvements. Retracking is the most challenging task in using altimeter data for smaller water bodies. The mixture of different waveform shapes such as ocean-like, specular, etc. makes it difficult to choose a suitable retracking algorithm. Each retracker is optimized for special waveform shapes, but switching the retracking algorithm to achieve the best ranges will lead to retracker biases which have to be taken into account. Furthermore, inter-mission offsets can also arise because of the different characteristics of the measurement systems (e.g. Ku-band (Envisat) and Ka-band (SARAL/AltiKa)).

All presented water level time series as well as results for many additional targets are freely available in the 'Database for Hydrological Time Series over Inland Water' (DAHITI) at http://dahiti.dgfi.tum.de.

References

Berry, P., Jasper, A., and Bracke, H.: Retracking ERS-1 altimeter waveforms over land for topographic height determination: an expert system approach, ESA Pub. SP-414, 1, 403–408, 1997.

Berry, P. A. M., Garlick, J. D., Freeman, J. A., and Mathers, E. L.: Global inland water monitoring from multimission altimetry, Geophysical Research Letters, 32, n/a–n/a, doi:10.1029/2005GL022814, http://dx.doi.org/10.1029/2005GL022814, 116401, 2005.

Birkett, C., Reynolds, C., Beckley, B., and Doorn, B.: From Research to Operations: The USDA Global Reservoir and Lake Monitor, in: Coastal Altimetry, edited by Vignudelli, S., Kostianoy, A. G., Cipollini, P., and Benveniste, J., pp. 19–50, Springer Berlin Heidelberg, doi:10.1007/978-3-642-12796-0_2, 2011.

Birkett, C. M.: The contribution of TOPEX/POSEIDON to the global monitoring of climatically sensitive lakes, Journal of Geophysical Research: Oceans, 100, 25 179–25 204, doi:10.1029/95JC02125, 1995.

Boehm, J., Kouba, J., and Schuh, H.: Forecast Vienna Mapping Functions 1 for real-time analysis of space geodetic observations, Journal of Geodesy, 83, 397–401, doi:10.1007/s00190-008-0216-y, 2009.

Bosch, W., Dettmering, D., and Schwatke, C.: Multi-Mission Cross-Calibration of Satellite Altimeters: Constructing a Long-Term Data Record for Global and Regional Sea Level Change Studies, Remote Sensing, 6, 2255–2281, doi:10.3390/rs6032255, 2014.

Boser, B. E., Guyon, I. M., and Vapnik, V. N.: A Training Algorithm for Optimal Margin Classifiers, in: Proceedings of the Fifth Annual Workshop on Computational Learning Theory, COLT '92, pp. 144–152, ACM, New York, NY, USA, doi:10.1145/130385.130401, 1992.

Brown, G.: The average impulse response of a rough surface and its applications, Antennas and Propagation, IEEE Transactions on, 25, 67–74, doi:10.1109/TAP.1977.1141536, 1977.

Challenor, P. and Srokosz, M.: The extraction of geophysical parameters from radar altimeter return from a non-linear sea surface, Mathematics in remote sensing, pp. 257–268, 1989.

Chelton, B., Ries, J., Haines, B., Fu, L., and P.S., C.: Satellite Altimetry, in: Satellite Altimetry and Earth Sciences: A Handbook of Techniques and Applications, edited by Fu, L. and Cazenave, A., pp. 1–131, Academic Press, 2001.

Crétaux, J.-F. and Birkett, C.: Lake studies from satellite radar altimetry, Comptes Rendus Geoscience, 338, 1098 – 1112, doi:10.1016/j.crte.2006.08.002, la Terre observée depuis l'espace Observing the Earth from space, 2006.

Crétaux, J.-F., Calmant, S., Romanovski, V., Shabunin, A., Lyard, F., Bergé-Nguyen, M., Cazenave, A., Hernandez, F., and Perosanz, F.: An absolute calibration site for radar altimeters in the continental domain: Lake Issykkul in Central Asia, Journal of Geodesy, 83, 723–735, doi:10.1007/s00190-008-0289-7, http://dx.doi.org/10.1007/s00190-008-0289-7, 2009.

Crétaux, J.-F., Jelinski, W., Calmant, S., Kouraev, A., Vuglinski, V., Bergé-Nguyen, M., Gennero, M.-C., Nino, F., Rio, R. A. D., Cazenave, A., and Maisongrande, P.: SOLS: A lake database to monitor in the Near Real Time water level and storage variations from remote sensing data, Advances in Space Research, 47, 1497 – 1507, doi:10.1016/j.asr.2011.01.004, 2011.

da Silva, J. S., Calmant, S., Seyler, F., Filho, O. C. R., Cochonneau, G., and Mansur, W. J.: Water levels in the Amazon basin derived from the ERS 2 and ENVISAT radar altimetry missions, Remote Sensing of Environment, 114, 2160 – 2181, doi:10.1016/j.rse.2010.04.020, 2010.

de Oliveira Campos, I., Mercier, F., Maheu, C., Cochonneau, G., Kosuth, P., Blitzkow, D., and Cazenave, A.: Temporal variations of river basin waters from Topex/Poseidon satellite altimetry. Application to the Amazon basin, Comptes Rendus de l'Académie

des Sciences - Series IIA - Earth and Planetary Science, 333, 633 – 643, doi:10.1016/S1251-8050(01)01688-3, 2001.

Eicker, A., Schumacher, M., Kusche, J., Döll, P., and Schmied, H.: Calibration/Data Assimilation Approach for Integrating GRACE Data into the WaterGAP Global Hydrology Model (WGHM) Using an Ensemble Kalman Filter: First Results, Surveys in Geophysics, 35, 1285–1309, doi:10.1007/s10712-014-9309-8, http://dx.doi.org/10.1007/s10712-014-9309-8, 2014.

Förste, C., Bruinsma, S., Flechtner, F., Marty, J., Lemoine, J., Dahle, C., Abrikosov, O., Neumayer, H., Biancale, R., Barthelmes, F., and Balmino, G.: A new release of EIGEN-6C, presented AGU Fall Meeting 2012, San Francisco 2012, 2012.

Global Runoff Data Center: Long-Term Mean Monthly Discharges and Annual Characteristics of GRDC Station / Global Runoff Data Center, Koblenz, Germany: Federal Institute of Hydrology (BfG), 2013.

Gruber, C., Moon, Y., Flechtner, F., Dahle, C., Novák, P., König, R., and Neumayer, H.: Submonthly GRACE Solutions from Localizing Integral Equations and Kalman Filtering, in: Earth on the Edge: Science for a Sustainable Planet, edited by Rizos, C. and Willis, P., vol. 139 of International Association of Geodesy Symposia, pp. 383–389, Springer Berlin Heidelberg, doi:10.1007/978-3-642-37222-3_51, http://dx.doi.org/10.1007/978-3-642-37222-3_51, 2014.

Guha-Sapir, D. and Vos, F.: Quantifying Global Environmental Change Impacts: Methods, Criteria and Definitions for Compiling Data on Hydro-meteorological Disasters, in: Coping with Global Environmental Change, Disasters and Security, edited by Brauch, H. G., Oswald Spring, U., Mesjasz, C., Grin, J., Kameri-Mbote, P., Chourou, B., Dunay, P., and Birkmann, J., vol. 5 of Hexagon Series on Human and Environmental Security and Peace, pp. 693–717, Springer Berlin Heidelberg, doi:10.1007/978-3-642-17776-7_40, 2011.

Hwang, C., Guo, J., Deng, X., Hsu, H.-Y., and Liu, Y.: Coastal gravity anomalies from retracked Geosat/GM altimetry: improvement, limitation and the role of airborne gravity data, Journal of Geodesy, 80, 204–216, 2006.

Kalman, R. E.: A New Approach to Linear Filtering and Prediction Problems, Transactions of the ASME–Journal of Basic Engineering, 82, 35–45, 1960.

Maillard, P., Bercher, N., and Calmant, S.: New processing approaches on the retrieval of water levels in Envisat and SARAL radar altimetry over rivers: A case study of the São Francisco River, Brazil, Remote Sensing of Environment, 156, 226 – 241, doi:10.1016/j.rse.2014.09.027, 2015.

Martin, T. V., Zwally, H. J., Brenner, A. C., and Bindschadler, R. A.: Analysis and retracking of continental ice sheet radar altimeter waveforms, Journal of Geophysical Research: Oceans, 88, 1608–1616, doi:10.1029/JC088iC03p01608, 1983.

McCarthy, D. and Petit, G.: IERS Conventions 2003 (IERS Technical Note ; 32), Frankfurt am Main: Verlag des Bundesamtes für Kartographie und Geodäsie, p. 127 pp., 2004.

Ponchaut, F. and Cazenave, A.: Continental lake level variations from Topex/Poseidon (1993–1996), Comptes Rendus de l'Académie des Sciences - Series IIA - Earth and Planetary Science, 326, 13 – 20, doi:10.1016/S1251-8050(97)83198-9, 1998.

Ričko, M., Birkett, C. M., Carton, J. A., and Crétaux, J.-F.: Publisher's Note: Intercomparison and validation of continental water level products derived from satellite radar altime-

try, Journal of Applied Remote Sensing, 6, 060 103–060 103, doi:10.1117/1.JRS.6.060103, http://dx.doi.org/10.1117/1.JRS.6.060103, 2013.

Scharroo, R. and Smith, W. H. F.: A global positioning system-based climatology for the total electron content in the ionosphere, Journal of Geophysical Research: Space Physics, 115, n/a–n/a, doi:10.1029/2009JA014719, 2010.

Schwatke, C., Dettmering, D., and Boergens, E.: Potential of SARAL/AltiKa for inland water applications, Marine Geodesy, n/a, n/a–n/a, doi:10.1080/01490419.2015.1008710, 2015.

Seyler, F., Calmant, S., da Silva, J. S., Moreira, D. M., Mercier, F., and Shum, C. K.: From TOPEX/Poseidon to Jason-2/OSTM in the Amazon basin, Advances in Space Research, 51, 1542 – 1550, doi:10.1016/j.asr.2012.11.002, satellite Altimetry Calibration and Deformation Monitoring using GNSS, 2013.

Smola, A. J. and Schölkopf, B.: A Tutorial on Support Vector Regression, Statistics and Computing, 14, 199–222, doi:10.1023/B:STCO.0000035301.49549.88, 2004.

Stakhiv, E. and Stewart, B.: Needs for Climate Information in Support of Decision-Making in the Water Sector, Procedia Environmental Sciences, 1, 102 – 119, doi:http://dx.doi.org/10.1016/j.proenv.2010.09.008, http://www.sciencedirect.com/science/article/pii/S1878029610000095, world Climate Conference - 3, 2010.

Wingham, D., Rapley, C., and Griffiths, H.: New techniques in satellite altimeter tracking systems, in: ESA Proceedings of the 1986 International Geoscience and Remote Sensing Symposium(IGARSS'86) on Remote Sensing: Today's Solutions for Tomorrow's Information Needs,, vol. 3, 1986.

Yang, Y. and Gao, W.: An Optimal Adaptive Kalman Filter, Journal of Geodesy, 80, 177–183, doi:10.1007/s00190-006-0041-0, http://dx.doi.org/10.1007/s00190-006-0041-0, 2006.

Zakharova, E. A., Kouraev, A. V., Cazenave, A., and Seyler, F.: Amazon River discharge estimated from TOPEX/Poseidon altimetry, Comptes Rendus Geoscience, 338, 188 – 196, doi:10.1016/j.crte.2005.10.003, 2006.