Consortium Members

D2.3: End-to-End ECV Uncertainty Budget (E3UB)

Reference: CCI-LAKES-0025-E3UB

Issue: 1.2

Date: 29 April 2021

esa European Space Agency

lakes cci

Chronology Issues:			
Issue:	Date:	Reason for change:	Author
0.1	19 March 2020	Initial submission	S. Simis et al.
1.0	24 April 2020	Update following ESA review	S. Simis et al.
1.1	17 March 2021	LWL/LWE updates LIC updates	J.F. Crétaux C. Duguay
1.2	29 April 2021	Updates following ESA review	J.F. Crétaux C. Duguay

People involved in this issue:			Signature
Authors:	Stefan Simis, Xiaohan Liu	Plymouth Marine Laboratory	
	Jean-François Crétaux	LEGOS	
	Hervé Yésou	SERTIT	
	Erik Malnes, Hannah Vickers	NORCE	
	Pablo Blanco	TRE Altamira	
	Chris Merchant, Laura Carrea	University of Reading	
	Claude Duguay	H20 Geomatics	
Internal review:	Stefan Simis	Plymouth Marine Laboratory	Stefan Simis Digitally signed by Stefan Simis Date: 2021.05.06 11:38:34 +01'00'
Approved by:	B. Coulon	CLS	Bruno Coulon Signature numérique de Bruno Coulon Date : 2021.05.06 17:52:48 +02'00'
Authorized by:	C. Albergel	ESA	

Distribution:		
Company	**Names**	**Contact details**
ESA	C. Albergel	Clement.Albergel@esa.int
BC	K. Stelzer	kerstin.stelzer@brockmann-consult.de

Distribution:		
Company	**Names**	**Contact details**
CLS	B. Coulon B. Calmettes P. Thibaut	bcoulon@groupcls.com bcalmettes@groupcls.com pthibaut@groupcls.com
CNR	C. Giardino	giardino.c@irea.cnr.it
Eola	E. Zakharova	zavocado@gmail.com
GeoEcoMar	A. Scrieciu	albert.scrieciu@geoecomar.ro
H2OG	C. Duguay	claude.duguay@h2ogeomatics.com
LEGOS	J.F. Créteaux A. Kouraev	jean-francois.cretaux@legos.obs-mip.fr alexei.kouraev@legos.obs-mip.fr
NORCE	E. Malnes	eima@norceresearch.no
PML	S. Simis	stsi@pml.ac.uk
SERTIT	H. Yésou	herve.yesou@unsitra.fr
TRE-ALTAMIRA	P. Blanco	pablo.blanco@tre-altamira.com
UoR	C. Merchant L. Carrea	c.j.merchant@reading.ac.uk l.carrea@reading.ac.uk
UoS	A. Tyler E. Spyrakos	a.n.tyler@stir.ac.uk evangelos.spyrakos@stir.ac.uk

List of Contents

1. Overview

This document describes the uncertainty characterization, estimation and/or propagation for each product in the lakes ECV included in the lakes cci climate data records.

The lakes ECV covers several domains of hydrology and biogeochemistry with satellites observing, through a range of radar/altimetry, thermal and optical sensors, properties of lakes which are ultimately interpreted as

- Lake water level (LWL)
- Lake water extent (LWE)
- Lake surface water temperature (LSWT)
- Lake Ice Cover (LIC)
- Lake Water-Leaving Reflectance (LWLR)

Details on the methodology to determine per-observation uncertainty products, and how they are presented to users of the lakes cci data record, are provided in this document.

2. General terminology

The uncertainty characterization of the lakes cci follows, in general, the approach and recommendations set out in Merchant et al. (2017), applied widely in the CCI programme, and using metrological principles discussed in Mittaz et al. (2019). It is useful to restate the internationally agreed terminology and concepts for measurement science.

A **measurement** is a set of operations having the object of determining the value of quantity. A **measured value** is the result of a measurement and is an observation of a **measurand**: a quantity subject to measurement. The phrases 'true value of a quantity' and 'value of the measurand' are synonymous. The measured value minus the value of the measurand is the **error**, which is generally unknown.

The measurands in the lake cci are diverse and derived from various satellite sensor measurements (e.g. of radiance, time, etc). In addition to errors associated directly with these sensor measurements (I.e., errors resulting from sensor noise and imperfect calibration) or their operation (detector viewing angles, integration times), errors also result from subsequent stages of data processing, such as classification and inversion of signals to geophysical variables. Errors at a given stage (data transformation step) combine additively, and errors propagate through successive stages.

With increasing complexity of the satellite data processing chain, e.g. including dependencies on external models, it may become difficult identify and quantify the dominant effects causing errors in products, especially as which effects are dominant may depend on the scale of data aggregation in the product.

Uncertainty characterises the dispersion of values it is reasonable to attribute to the measurand having made a measurement. **Standard uncertainty** is practically interpretable as the standard deviation of an estimated distribution of combined errors. Systematic methods for estimating uncertainty in Earth observation by understanding retrieval and data processing have been summarised by Mittaz et al., (2019) and associated mathematical methods by Merchant et al. (2019). Alternatively, it may be possible to characterize the uncertainties associated with a measurement process (such as a satellite retrieval system) using an adequate sample of measured values in comparison to reference measurements, which are trusted sources whose uncertainty characteristics are independently known.

The **end-to-end uncertainty budget** is the result (where possible) of determining the uncertainty associated with measured values in products. Ideally, measured values have associated with them an uncertainty per datum that is obtained by modelling the measurement process and which is validated using comparisons to reference measurements (Merchant et al., 2017). This ideal situation requires significant investment of resources, both by the data producer and via programmes such as ESA's Fiducial Reference Measurements; this is gradually becoming more feasible for an increasing number of essential climate variables, but not all.

In the sections that follow, product uncertainties are described for each of the thematic domains of Lakes cci.

3. Lake Water Level (LWL)

3.1. LWL Uncertainty characterization

The uncertainties provided with the LWL product result from the standard deviation of the along track altimetry (after including all corrections) of all measurements used to calculate the median value for a pass over a lake.

For reference, the calculation of LWL using satellite altimetry follows

$$LWL = Alt - R_{corr} - T_E \quad [3.1]$$

Where L*WL* is considered with respect to a geoid through R_{corr}, *Alt* is the altitude of the satellite above an ellipsoid and T_E is the sum of different corrections in order to account for atmospheric refraction (propagation in the ionosphere and the troposphere), tidal effects (solid Earth, lake and polar), and geoid height above the ellipsoid. For readers who needs more detailed information a full discussion of the computation of LWL can be found in Cretaux et al. (2009).

Each term of the expression of LWL has its own contribution to the final error budget. These sources of uncertainty are introduced in detail in section 3.2.

To quantify the end-to-end error budget both new field work and existing data sources are used. Previous findings are reported in Cretaux et al. (2009, 2011, 2013, 2018).

To determine the impact of some of the altimetry data processing models in different configurations, other studies based on comparison with in situ data have been performed (Ričko et al., 2012, Arsen et al., 2015). This work is continued with new in situ data sets collected during Lakes_cci, through external collaborations.

Based on the quality assessments done over the last years a quantification of the full error budget is available for different combinations of lakes and altimeters.

3.2. Sources of uncertainty in LWL

The sources of uncertainty of LWL product vary and depend on several factors. The uncertainty is calculated from the standard deviation of the distribution of measurements of water height by the satellite altimeter. The different factors affecting this calculation fall into three categories:

Detector noise: For a classic altimeter, the detector noise is sub centimetre and is therefore not the main source of uncertainty for LWL.

Morphology of the lake and its surrounding environment: Radar altimeter measurements involve sending an electromagnetic pulse to the satellite nadir and measuring the propagation time to and from the emitted wave and its echo on the illuminated surface. A received echo is not identical to the initial pulse. Despite how narrow the initial pulse can be, the echo is spread over time by its interaction with the reflecting surface. In practice, the echo spans multiple antenna reception intervals – the "distance doors" – (typically a hundred doors of 2 to 3 nanoseconds each, a length of 30 to 50 cm per door) and the technique of locating the correct time interval to give the distance measurement (range) in this period is called the tracking (if the operation is performed on board) or retracking (if the operation is done on the ground through post-processing techniques). Tracking and retracking are not directly performed on echoes, which are too noisy, but on the echo accumulations, called the waveforms.

As the first altimetry missions were dedicated to oceanic studies, the algorithms developed to address the waveforms were therefore adjusted to the shape of echoes that returned from the ocean surface. These algorithms were based on the fact that the reflective surface is sufficiently isotropic and continuous, and if we accumulate echoes over a longer time span than the decorrelation time of spatial structures from the reflecting surface of the same order of magnitude as the pulse wavelength (2.2 cm for a Ku-band of 13.6 GHz), the waveform could be approximated by a theoretical form from which time A/R can be analytically determined. This theorical form is called it the brown model.

The time for the echo to bounce back to the altimeter can be obtained by finding the best fit between the waveform and the analytical expression. Once established, and if the dating provides the best fit, it is found halfway up the rising edge. In the specific case of continental surfaces, the received echo is very different from the paradigm presented above. The shape of the echo becomes extremely variable (Figure 1) depending on the power backscattered by the water plane environment, in the case small or narrow lakes.

It is no longer possible to find a single analytical expression that can determine a specific A/R time for all waveforms. A key milestone was reached with the ESA Envisat mission, where it was decided to break away from the principle of the single range estimate and offer several range estimates made by very different algorithms for each measurement within its Geophysical Data Records (GDRs) user products.

The morphology of the lake and surrounding terrain affect the final uncertainty in different ways: for large lakes (>100s km2) with a long track coverage (> 30-40 km) it is possible to strictly select the range of measurements that are located far enough from the coastline. Because the selected radar echoes are fully over water, the final uncertainty is in the centimetre range. If, however, the lake is narrow or when the satellite track crosses the lake on an edge or within a short distance from shore, the echoes no longer follow the Brown model and we need to use the released retracking models. If a large proportion of measurements is affected by this issue, their dispersion is much increased with respect to ideal cases, and the uncertainty may reach several decimetres. In practice, the classical OCOG (namely ICE-1) is used for the LWL product since it has been evaluated as the best algorithm for such reflecting surfaces. The same problem arises when specular or quasi specular echoes are registered on lakes with a very calm surface, since the Brown model assumes that the reflecting surface has some small waves rather than specular echoes. For this reason, it is problematic to use Topex / Poseidon or Jason-1 satellite over lakes in such cases because alternative retracking models (like OCOG) are not released in the GDRs. Using retracking models (with Envisat, Jason-3, Saral or Sentinel-3A) does not solve all problems with complex footprints and in the worst cases, the resulting uncertainty may reach several decimetres.

It is important to note that the Ka band (36 GHz) with the Saral/AltiKa mission has reduced the footprint and consequently also decreased the uncertainty for small lakes. It has been shown in Arsen et al. (2015) that accuracy of LWL over small lakes in the Andean chains (with respect to Envisat for the same lakes) improved by a factor of 5 to 10 depending on the lake.

With Sentinel-3A another technical improvement reduces potentially in a significant manner the final uncertainty in some specific geographical contexts. Sentinel-3A operates in SAR mode which brings several advantages for small water bodies. The footprint is reduced by a factor ranging from 10 to 50 compared to Low Resolution Mode (LRM) used with other altimeters. This allows much better selection of the reflecting point and drastically reduces the pollution from the ground when the satellite track approaches the lake shoreline. It is however valid only when the orbital track is almost perpendicular to the shoreline.

.

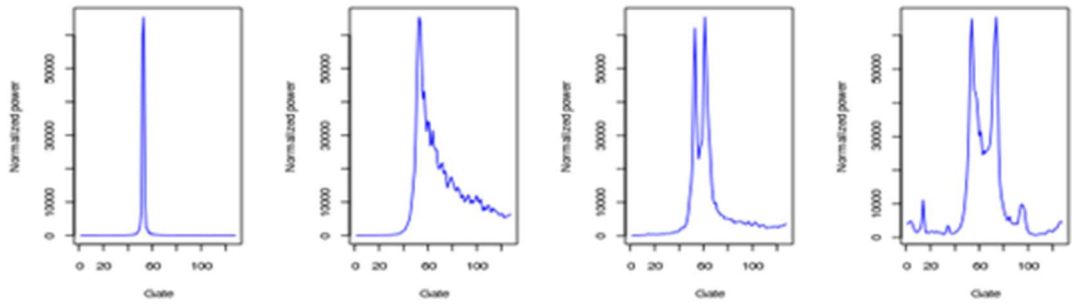

Figure 1: example of typical waveforms, from left to right; specular, ocean like (Brown model), contaminated, and examples of noisy waveforms

Geophysical corrections: Calculating the range is based on the propagation velocity of the electromagnetic wave, which is the speed of light in a vacuum. In practice, as the radar pulse passes through the atmosphere, the actual propagation speed is below this limit value. The impact on the estimated range is an extension of around 2.5 m, mainly linked to the density of the air. Three separate corrections are applied:

- dry tropospheric correction
- wet tropospheric correction
- ionospheric correction

There is actually a fourth source of uncertainty to consider, due to the reflection on the surface of the water. A correction of the surface state would take into account that the height of the facets that reflect the radar pulse may be different from the average water height in the radar beam footprint, and it is also to correct for the fact that the shape of waves changes the height distribution in the surface illuminated by the radar. Whereas it is described for sea state, the effect on lakes is not well known, not modelled and assumed to be relatively weak. In extreme cases, however, it may reach several centimetres and constitutes a source of potential uncertainty. To date this has not been taken into account for inland waters.

Among the above three corrections, the dry tropospheric and the ionospheric ones are well modelled over continental water (using global gridded data sets and the resulting errors are below the centimetre over lakes: see Cretaux et al. (2009). The third correction factor is the most problematic and causes the main measurement uncertainty.

Wet tropospheric correction is related to the water vapour contained in the air column that the electromagnetic wave intersects. This correction can be estimated in two ways: either with an onboard bi or tri-frequency radiometer or from a global meteorological model, as used for dry tropospheric correction. Radiometers measure the instantaneous brightness temperatures at the nadir, a quantity dependent on the atmospheric water vapor content. This measurement has the obvious advantage of time coincidence with the radar altimeter measurements. However, radiometers do not operate properly over the continental areas because the measurement is polluted by the inhomogeneity of the soil emissivity except for very large lakes. To avoid this problem, a WTC based on the ECMWF re-analysis is used. Over lakes, however, the use of local GPS precise positioning near Lake Issykkul in Central Asia (Cretaux et al. 2009, 2011) has shown that the estimated error is in the order of a few centimetres.

Other corrections are applied to the final calculation of LWL, which do not contribute significant (i.e. they are sub-centimetre) sources of uncertainty. These include lake tides, Earth tides and the polar tide, and all are delivered within the GDRs using geophysical models.

3.3. Uncertainty estimation methodology in LWL

The uncertainty estimation of each LWL at a given time is derived from the standard deviation of the individual lake water height along the track. Geoid corrections using the repeat track technique (Cretaux et al. 2016) are applied, and the median value of each of the water height estimation (at

20 Hz for Topex / Poseidon, Jason-1, Jason-2, and Envisat, and 40 Hz for Saral/Altika, Jason-3 and Sentinel3A) is calculated and constitutes the final LWL product. Standard deviations of the distribution of individual measurements are then calculated and reported as LWL measurement uncertainty.

3.4. End to end uncertainty budget in LWL

It is practically impossible to establish a generic error budget for LWL calculation using satellite altimetry because the sources of errors are numerous and they vary strongly from one lake to another, and between altimeters. We can, however, draw some generic conclusions.

For large lakes, whichever satellite mission is considered, and in normal lake state conditions (no specular echoes) the main source of uncertainty comes from the wet tropospheric correction, and depending on the regions it varies from 2-3 cm. The combination of altimeter noise and geophysical corrections then amounts to 8-10 centimetres.

For small and narrow lakes, uncertainty varies between 10 centimetres and 1 meter (above which the data are discarded). In such cases, the uncertainty depends first of all on the form of the echoes (waveform) and the ability of the retracking to analytically interpret it in terms of range between the satellite and the lake surface. Under very poor conditions (very narrow lakes for example) the retracking used (OCOG) may be not robust enough to retrieve the range without large uncertainty. In such cases, another factor influencing final uncertainty is the altimeter itself: with Saral/AltiKa and with Sentinel-3A, the impact of lake morphology on the result is reduced since the footprint is also drastically reduced. With LRM altimeters, the impact is the highest.

With SAR mode, the uncertainty on LWL over small lakes in particular can be significantly reduced. Intrinsically the measurement is more precise, and moreover new approach in water height extraction from the waveform can be developed. This is what has been done at CNES in coordination with LEGOS where a new methodology (call LPP) has been recently developed in order to determine the LWL of lakes using SAR measurements from Sentinel-3 satellites. It is a physical retracker based on simulation of waveforms over the lake.

For each satellite pass over the lake, numerical simulations of the SAR altimetry waveforms over the lake is the core process of the LPP. The simulations are built using a lake contour. All satellite and instrument characteristics are accounted for: radar antenna diagram, radar point target response, tracker information, satellite track position and altitude, a priori WSH.

The novelty of this approach is that the retracking algorithm considers the geometry and the roughness of the surface of each lake. Physical retrackers are essential to model properly the radar signal. With this method, we can reach RMSE below 10cm in SAR mode in very complex situations.

We have applied this processing chain (called LPP) on two set of lakes in two different regions: lakes in Occitanie (South of France) and lakes in Switzerland for which in situ reference data are available.

The size of these lakes varies from hundreds of metres to several kilometres. The surrounding topography is differs between sites with some in mountainous areas and others in open areas.

Figure 2 summarises the accuracy obtained with the LPP, and OCOG. It shows that, even for narrow lakes in Occitanie, the LPP method yields LWL at centimetre-accuracy, while using OCOG the results remain acceptable with decimetre-accuracy. For Lake St Geraud, which is narrow (200 metres wide), LPP still yields results of around 14 cm accuracy while OCOG did not yield accurate LWL. For the Swiss lakes, the complexity of the scene is due to the surrounding relief, but the comparison between LWL resulting from LPP, OCOG and in situ measurements still shows that the new methodology provides accuracy in the order of centimetres while it was in the order of decimetre using OCOG.

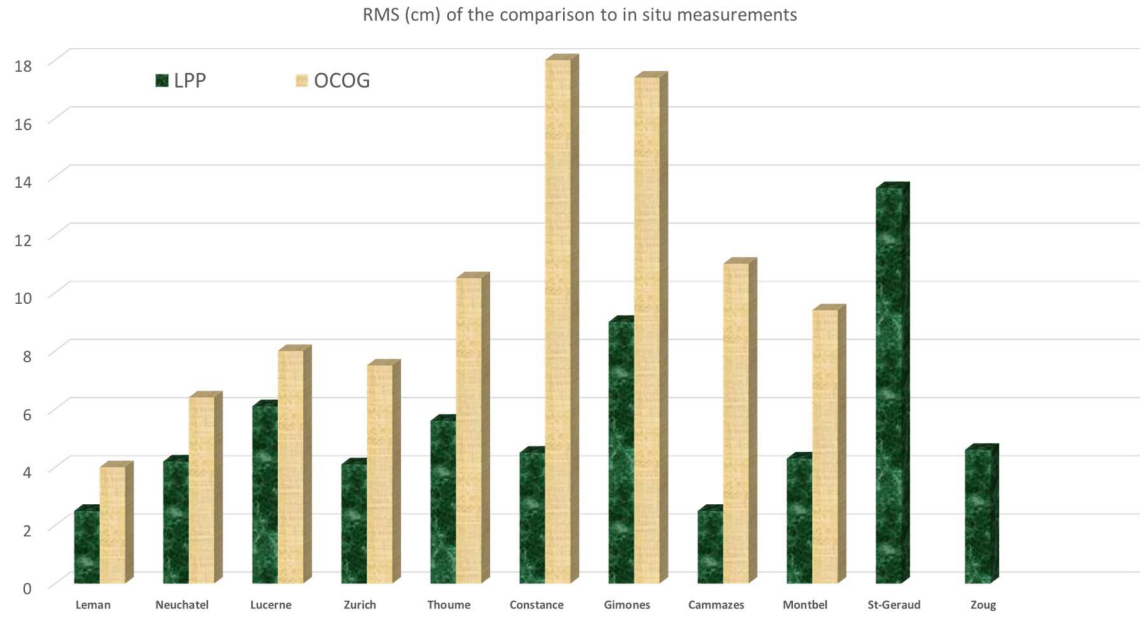

Figure 2: summary of RMS differences between in situ measurements and WSH calculated using LPP and OCOG retracker, for lakes in Occitanie and Swizterland.

4. Lake Water Extent (LWE)

Both optical and SAR data are used to produce LWE. Processing chains being different, candidate algorithms are described independently in this document. A first section focuses on the SAR processing optical approach and the following one presents the optical approach.

A last section describes the comparison of the resulting hypsometry curve for each solution and for a set of few test cases.

4.1. SAR LWE Uncertainty characterization

Sources of uncertainty in SAR data are in general well known. In the context of LWE the main sources of uncertainty derive from the nature of the SAR signal and the classification method employed to retrieve the water pixels, where the latter is currently being studied. Our methodology assumes that lake water pixels have low backscatter compared to their surroundings. As detailed in the next section, this assumption does not always hold, and defining backscatter thresholds introduces uncertainty which propagates to the LWE estimate.

At the present time no uncertainty is calculated with LWE results, because the classification methods to characterize water pixels do not provide error estimates. Some general quality indicators could be proposed as a rough indication of the class separability in an image. At the present time this remains subject to discussion. Ultimately, however, an uncertainty budget of the lake water maps can only be produced using ground truth data.

4.1.1. SAR sources of uncertainty

SAR measurand: SAR backscatter is mainly driven by the geometry of the observed target. This means that SAR is sensitive to roughness on the water body surface, which is assumed to be low proving low backscatter values. Consequently, any modification on the surface modifying this condition will impact on the lake water extension calculation. In the presence of strong wind, the water low

backscatter assumption does not hold, causing errors on the water detection. The presence of ice and snow, not only in the lake surface, but also in the surroundings, cause erroneous lake water extent retrieval.

Other types of land cover e.g. crop fields or wet snow, may have similar backscatter features as water bodies. When they are located in the surroundings of the lakes, especially if the image spatial resolution (nominal or after the filtering) is low compared to the proximity distance, they can mistakenly be classified as part of the lake in the classification process.

Vegetation on the lake can increase the backscatter, so those areas would be classified (at least in a first step) as non-water pixels.

Depending on the resolution of the SAR image, a negative bias can be found with respect to the real water extent measure. Coarse resolution SAR imagery will tend to underestimate the water covered area along the edge of the lake, and depending on the complexity of the lake edge in comparison to the pixel spacing, the negative bias is proportional to the length of the lake circumference.

Detector noise: SAR amplitude images are affected by multiplicative noise known as speckle. To reduce its effect on the water extent estimation de-speckle filtering is applied to the images. Filtering causes a decreasing of the nominal spatial resolution. SAR de-speckle filtering is an open topic so improvements can be achieved on this side.

Observation noise (e.g. related to viewing angle, atmospheric path): Depending on the viewing angle, geometric distortions (layover, shadowing and foreshortening) can be more or less severe. In order to prevent including those areas, a corresponding mask account for the areas presenting those effects is applied to the images.

The viewing angle also conditions the ground range inherent pixel size, therefore it will vary from near to far range. In any case it shall not be a significant variation for this application.

SAR amplitude images are not strongly affected by atmospheric effects. Nevertheless, in case of very severe atmospheric conditions (heavy storms), a mitigation of the signal amplitude could affect the image.

Algorithm effects: as mentioned, de-speckle filtering is a necessary step in the algorithm but at the same time it worsens the spatial resolution of the product.

Calibration effects: GRDH amplitude images are calibrated according to the annotated look-up tables. Consequently, any mismatch on them will impact in the quality of the calibration. SLC amplitude images are calibrated based on the Persistent Scatters behavior among the image stack. The lower the presence of these targets the lower the quality of the calibration. In general terms though no significant errors are expected.

Model effects: A DTM is needed in order to properly make the SAR to Geo coordinates and vice-versa. The more accurate and the larger the spatial resolution of the DTM the better the GRDH re-gridding and the geo-location accuracy of the results. The same impact can be considered for orbit inaccuracies.

Misclassification effects (e.g. cloud, land, water, subpixel variation): This effect is strictly related to the "SAR measurand" and "algorithm effects". The K-means classification employed method uses a finite number of classes so depending on the separability degree on the image, misclassification can have a significant impact on the results. The main sources of misclassification are speckle noise due to waves on the lakes (leading to water erroneously classified as land) and lake ice (ice covered water classified as land). All in all, the majority of the misclassification effect tends to cause a negative bias on the water extent, but the magnitude of the negative bias needs to become better understood.

Therefore, even if SAR system are considered to have all weather acquisition capabilities, in some case some interactions with strong storm cells can be observed, ie during the raining season for example, when strong monsoon rain falls occurred. These interactions between water /ice elements within the clouds and the SAR signal are more often observed on short waves length such as X band, than on longer ones, ie C and or L bands. These interactions are stronger in VV than VH, inducing a

blurring of the texture of the image as well by appearance of shadows areas, which of course involve difficulties to distinguish water from land.

In addition, some artefacts can also occur over large lakes on within there are an intense boat traffic, or sand exploitation. In this case, these boats or dragging engines are acting as strong reflectors, affecting mostly the VV signal forming bright stars, more than on VH signal on which, boats are appearing as bright points.

From the 3 images of Figure 3, the image at the top shows a Sentinel1 color composite, VV, VH, VV/VH, with cloud effects appearing in light blue, middle, VH channel, notice the bright points corresponding to boats, bottom, VV, notice the loss of texture on the channel area, as well as the strong backscattering from boats.

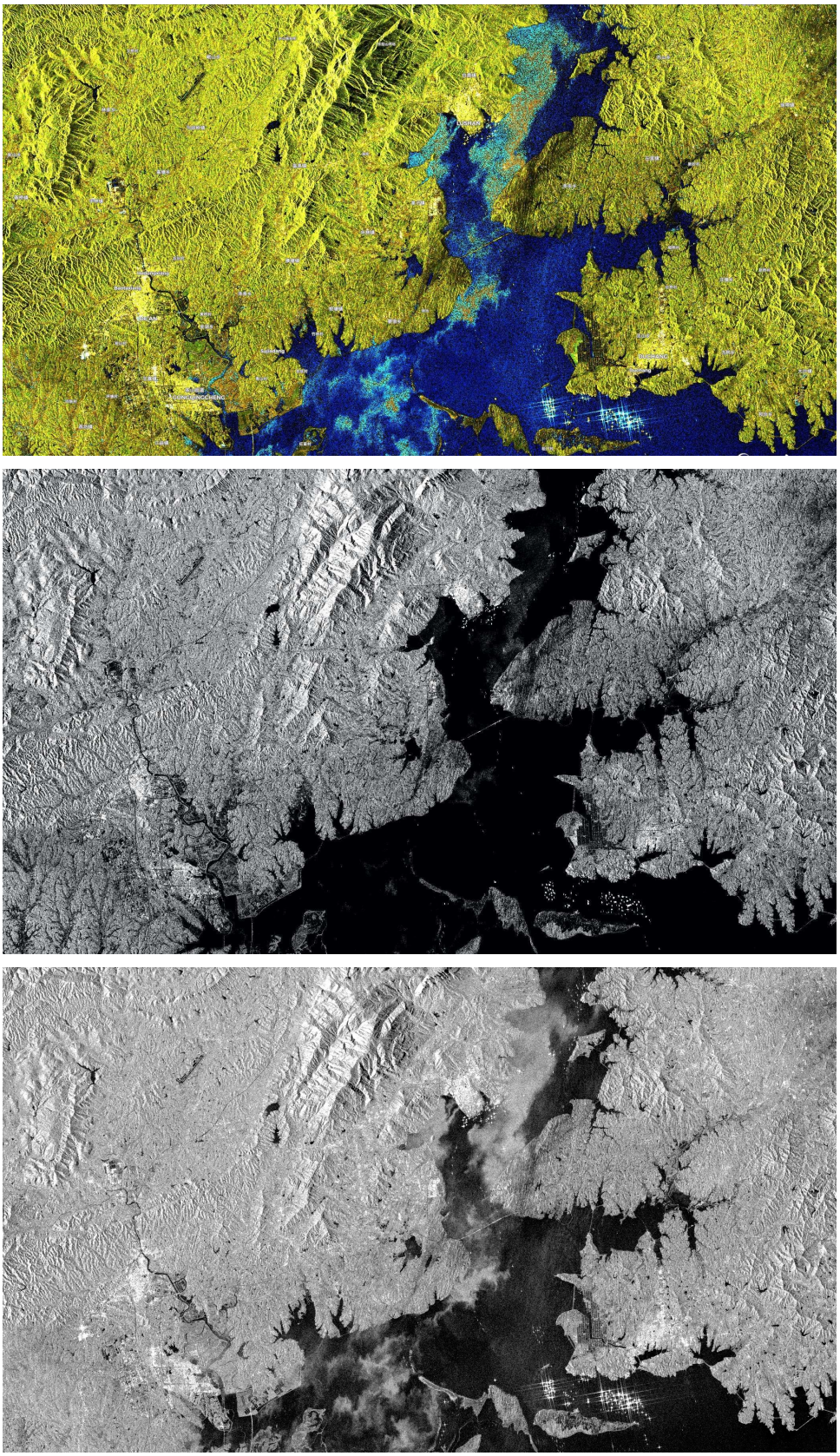

Figure 3: thunderstorm effect on SAR signal, Poyang Lake Sentinel1 image acquired on the 2020-07-02, over the Poyang Lake

4.1.2. Uncertainty estimation methodology in SAR LWE

For a typical 2 class classification problem like separating water from land, we can use the distance from the mean backscatter in the class and the actual backscatter of the pixel as a measure for accuracy. In the current K-means method used by both NORCE and TRE the probability of correct classification of a pixel $X_i = [\sigma^{VV}, \sigma^{VH}]$ can be quantified as:

$$P(Xi, Water) = \left[\frac{|\sigma^{VV} - mean(\sigma^{VV})\ |^2}{2var(\sigma^{VV})} + \frac{|\sigma^{VH} - mean(\sigma^{VH})|^2}{2var(\ ^{VH})}\right]^{-1} \qquad [4.1]$$

where the mean- and var-operator provides the mean/variance of VV/VH-backscatter for the water class. As a typical example we show in Figure 4 the co- and cross-pol channels of an image over Altevatn. In the bottom we show the backscatter distributions of the co- and cross pol-channels for the two classes (water and land) for the same images. It is clearly seen that for both channels there is an overlap between the water and land class that induces lowered probabilities.

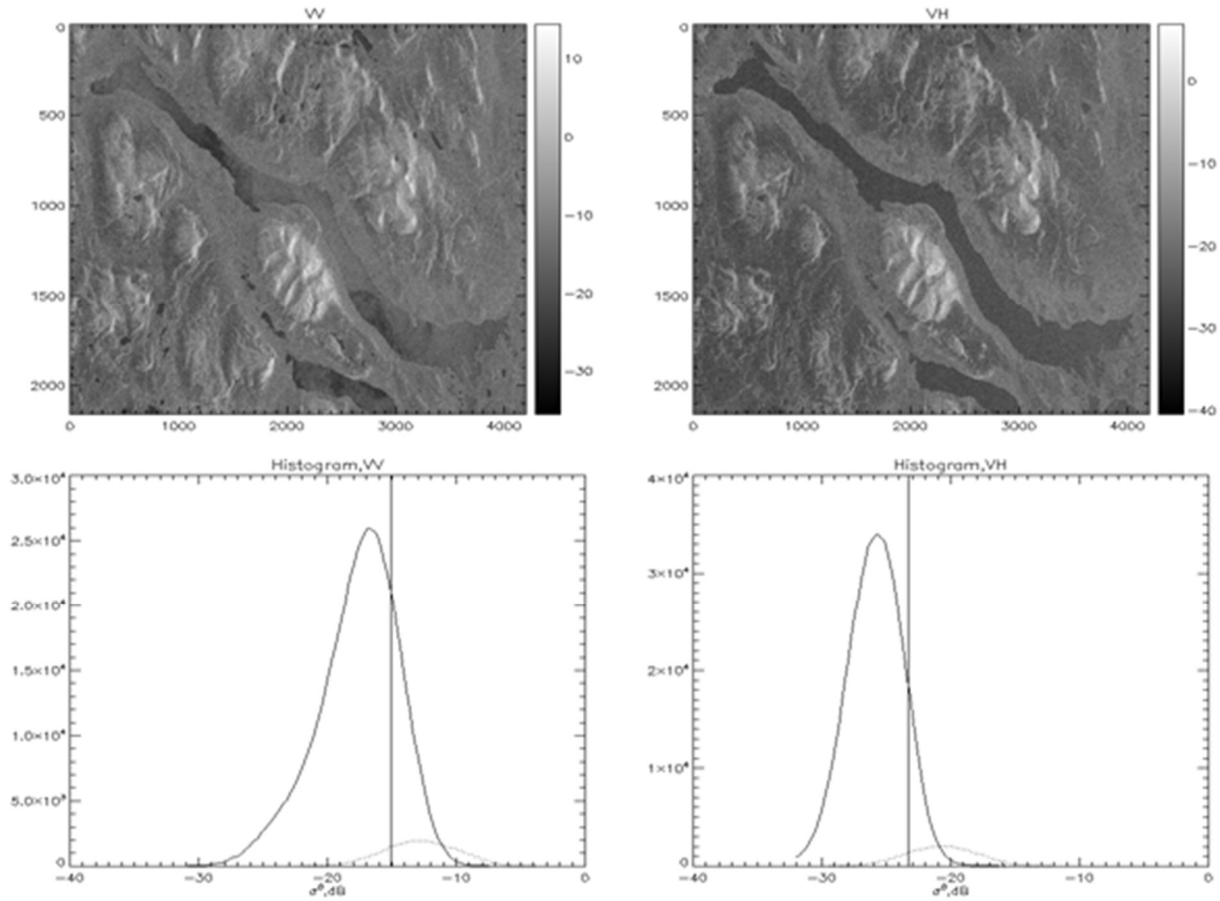

Figure 4: Top left radar backscatter VV, top right radar backscatter VH. Bottom left: Distribution of backscatter values (VV) for the water class (full lines) and the land class (dashed lines).

The resulting per-pixel confidence (P(Xi,Water) for the water class is now displayed in Figure 5 using the error metric provided in eq. [4.1].

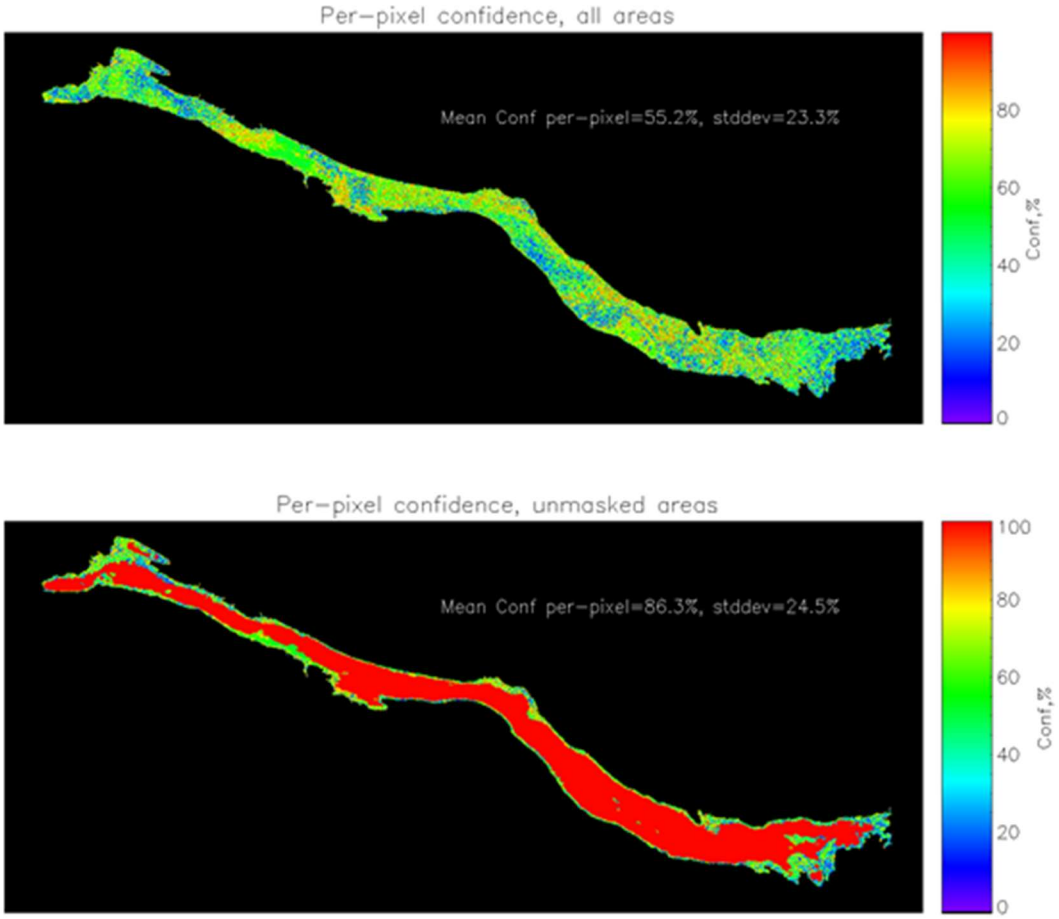

Figure 5: ***Top: Per-pixel confidence for classification of all water pixels. Bottom: using the minimum water mask we may reduce the errors significantly over a large part of the lake, and hence increase the average confidence per pixel.***

A product providing the per-pixel error estimate can hence be provided along with the classified water masks per satellite acquisition.

The total uncertainty of the total area estimate per acquisition should also be provided. This is a highly desired product. Unfortunately, it is not as easy as just calculating the mean error-estimate per pixel over the lake. For the current case (in Figure 5) we notice that the mean confidence per pixel is 86.3%. This is, however, much lower than the overall accuracy that can be found by comparing the SAR classification with simultaneous optical classifications (provided by S2 and SERTIT). For the current case the overall accuracy is 92.6% (see Figure 6)

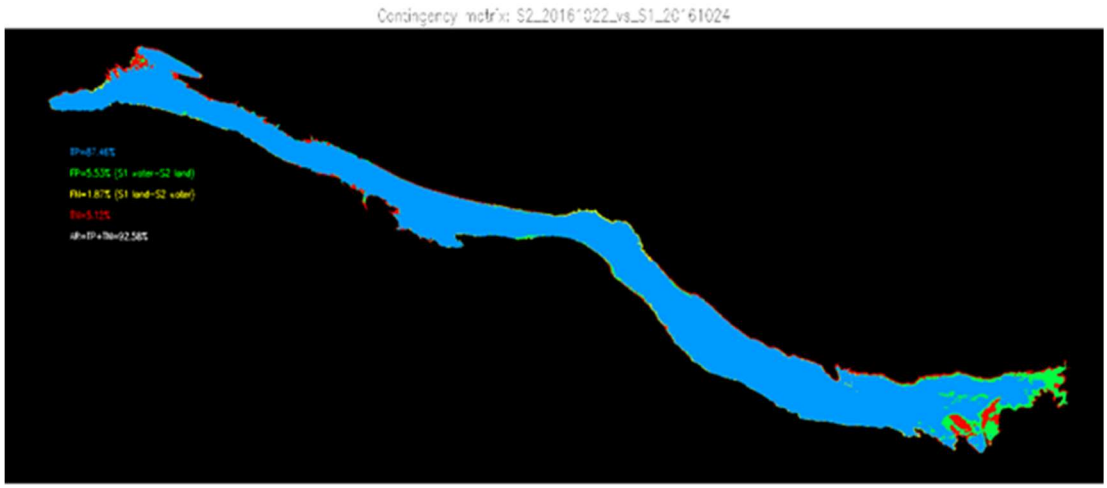

Figure 6: ***S1 /S2 comparison of water classification. The overall accuracy is 92.6% in this case.***

Another means of assessing the accuracy for the total area estimate is to use the hypsometric method. For the case of Altevatn we have used several hundred SAR images and very accurate in situ water level measurements to assess the hypsometric curve which related LWL and LWE (see Vickers et al., 2019). By taking the difference between the LWE measured by SAR and the LWE estimated from the hypsometric curve (using the actual in situ measured LWL) we can see that the accuracy of the area estimate for the SAR image is actually 99.1%. In table 1 we have done similar comparisons between the three methods for several SAR acquisitions where we had both S2 data and in situ water level measurements.

Table 1. ***Cross-comparisons of confidence estimates for different days (Altevatn)***

Date	**K-means, conf (%)**	**S2, conf (%)**	**LWL, conf (%)**
20160629	89.4	93.34	98.5
20160730	88.8	94.76	99.3
20160816	87.9	94.69	99.8
20161024	86.3	92.6	99.1
20170707	88.1	94.80	97.5
20170725	88.4	97.49	98.8
20170910	88.6	96.66	98.9
20170926	86.5	98.24	99.7
20170930	86.0	98.13	99.6
20171008	88.7	97.78	99.9
20171012	86.8	96.37	99.1
Average	**87.8**	**95.9**	**99.1**

In Figure 7 we show the correlation between the different confidence measures. It is clearly seen that they are somewhat uncorrelated, so it is not strait forward just to upscale the accuracy estimate

provided by the per-pixel confidence. A better correlation could have granted an empirical relation between average per-pixel confidence and true confidence (e.g. a linear relationship).

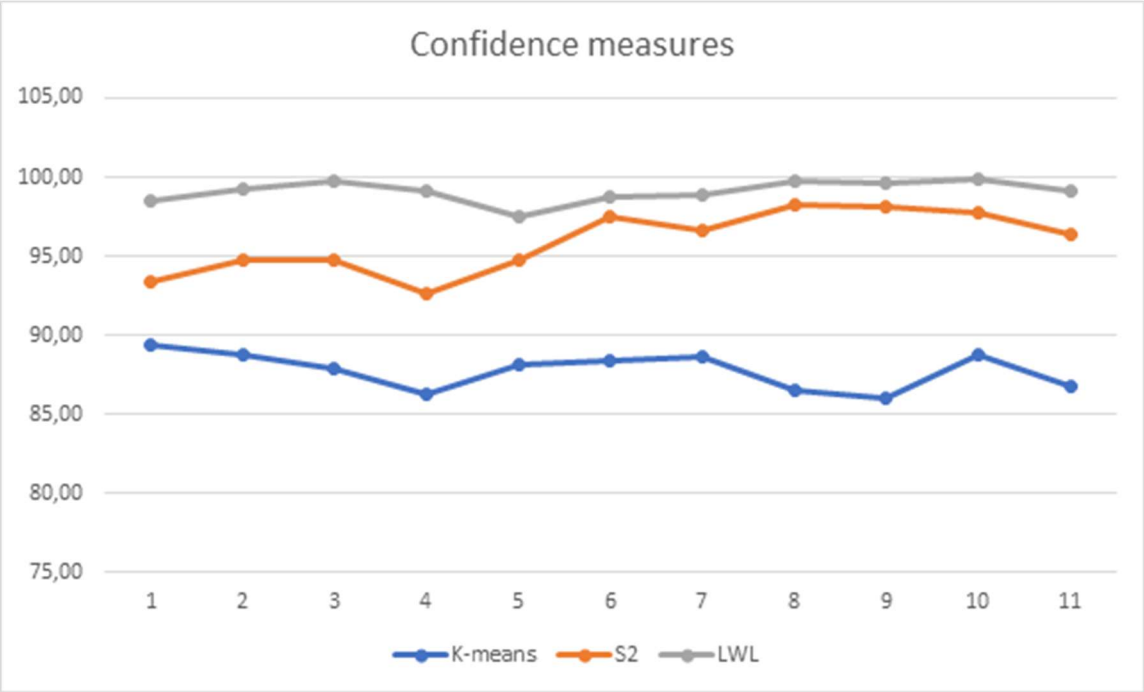

Figure 7. ***Confidence In overall accuracy estimate for 11 different classifications.***

Dense time-series as we have for Altevatn (see Figure 8) can also be used to assess the accuracy of the overall area estimates. By smoothing the time-series and measuring the difference between the actual area estimate and the running-mean estimate we can often obtain good estimates for the accuracy per acquisitions. However, Altevatn, is a particular case where daily satellite coverage is available, in normal cases, images are more sparse, and running averages can often not cope with natural changes in the LWE.

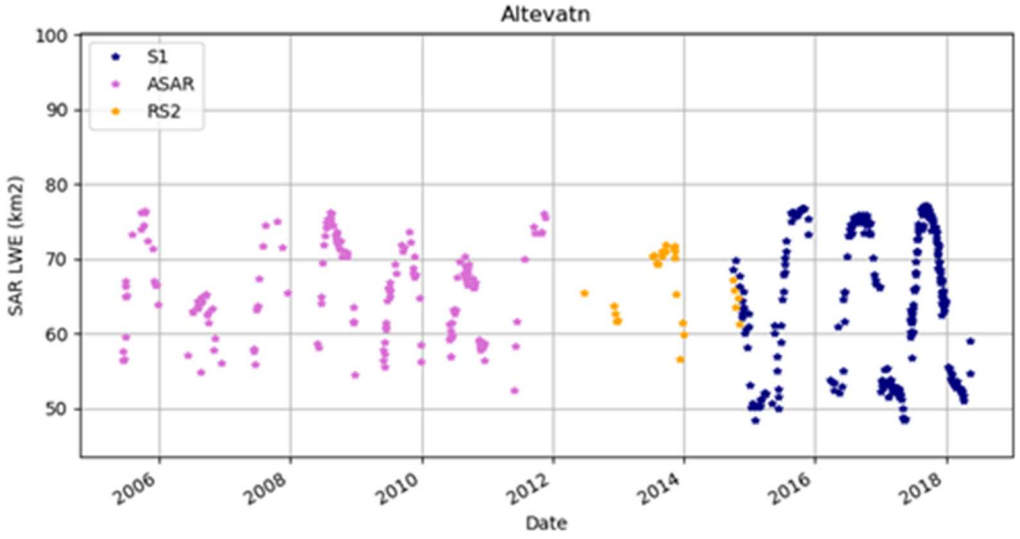

Figure 8. ***Lake Water Extent in km2 per satellite acquisition***

4.1.3. End to end uncertainty budget in SAR LWE

The end to end uncertainty budget in SAR LWE has been described above. The main sources of uncertainties have been identified, and a method to obtain pixelwise classification accuracies based on the image statistics has been described. The main source of inaccuracies for open water lakes is normally wind induced speckle noise causing high backscatter for open water which can be misclassified as land. The K-means methods applies several strategies to mitigate this problem (masking, segmentation), but in the end there will be a fraction of the pixels that has backscatter features that overlaps with the statistics for the land pixels. Improved methods for calculating the uncertainty should be studied further in future CCI Lakes iterations.

4.2. Optical sources of uncertainty

There are numerous sources of uncertainty when deriving water surfaces from images acquired by optical sensors. These can be related to the following effects, as detailed below:

- Observation effects
- Algorithm effects (external sources)
- Training samples selection
- Calibration – detector effects
- Misclassification effects (e.g. cloud, land, water, subpixel variation) and adjacency or proximity effects

Observation effects: Optical remote sensing of inland and marine waters depends on the quality of the retrieval of the water-leaving radiance from the top-of-atmosphere measurements. A notable influence on the retrieval is specular reflection of sunlight on the air-water interface (sun glint) in the direction of the satellite field of view. Depending on the satellite (viewing angles, pixel size) the effects of sun glint are felt differently.

For medium resolution satellites (e.g. MODIS, MERIS, OLCI with pixel sizes of 300 m to 1km) the sun glint can be modelled from the geometry of observation (angle of sight, solar angle and azimuth) and the wind speed (see Cox & Munk, 1954). It is then possible to attempt correction for sun glint or mask the affected pixel.

For high or very high resolution satellites (e.g. sensors on Landsat, Sentinel-2, Pleiades) the spatial resolution (< 30m) no longer allows modelling the sun glint. This effect must then be estimated from the observation itself, in which case Short Wave Infra-Red (SWIR) bands (> 1.5 µm) are particularly useful to delineate glint. The viewing geometry of Sentinel-2 satellites with a near-nadir view makes it particularly vulnerable to sun glint contamination (Harmel et al. 2017) (Figure 9.

The SWIR bands of Sentinel-2 MSI and Landsat-8/OLI can also alleviate limitations associated with using the Near Infra-red (NIR) part of the spectrum. First, light absorption by water is more than an order of magnitude more efficient in the SWIR than in the NIR thus allowing better separation of land and water. Second, the atmosphere is more transparent in the SWIR thus showing less diffuse transmission paths and, consequently, lower contribution of aerosols and air molecules.

Methods to remove or limit the sun glint bias require external input such as aerosol optical thickness, amount of absorbing gas, sea surface pressure and target altitude, which can be populated from the AERONET photometer network or the CAMS dataset (Harmel et al. 2017).

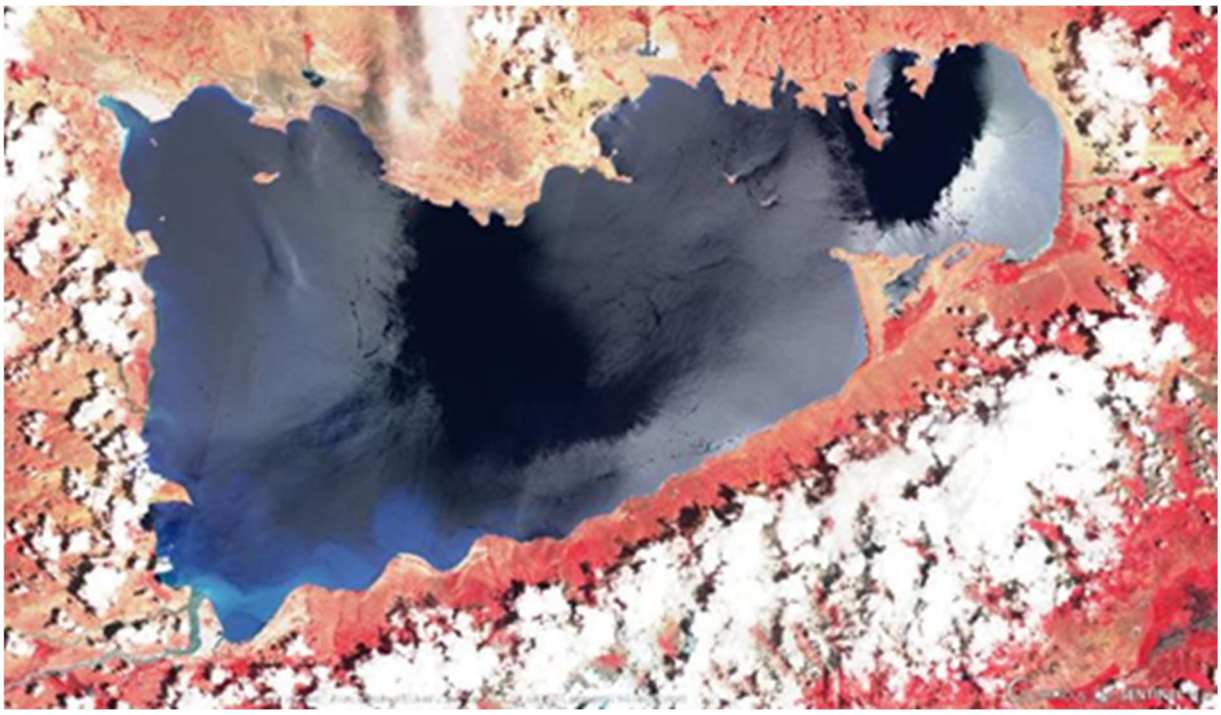

Figure 9: Sun glinted Sentinel-2 MSI image acquired on 14 July 2017 over Namtso lake (PR China)

Algorithm effects (external sources): In order to compute water extent over a lake, a reference area is required to limit the analysis to a spatial buffer around the target area. When working with medium resolution satellite imagery, it may be suitable to adopt these from the ESA CCI Land Cover database. However, this does not always provide sufficient detail when looking at higher resolution imagery **Erreur ! Source du renvoi introuvable.**).

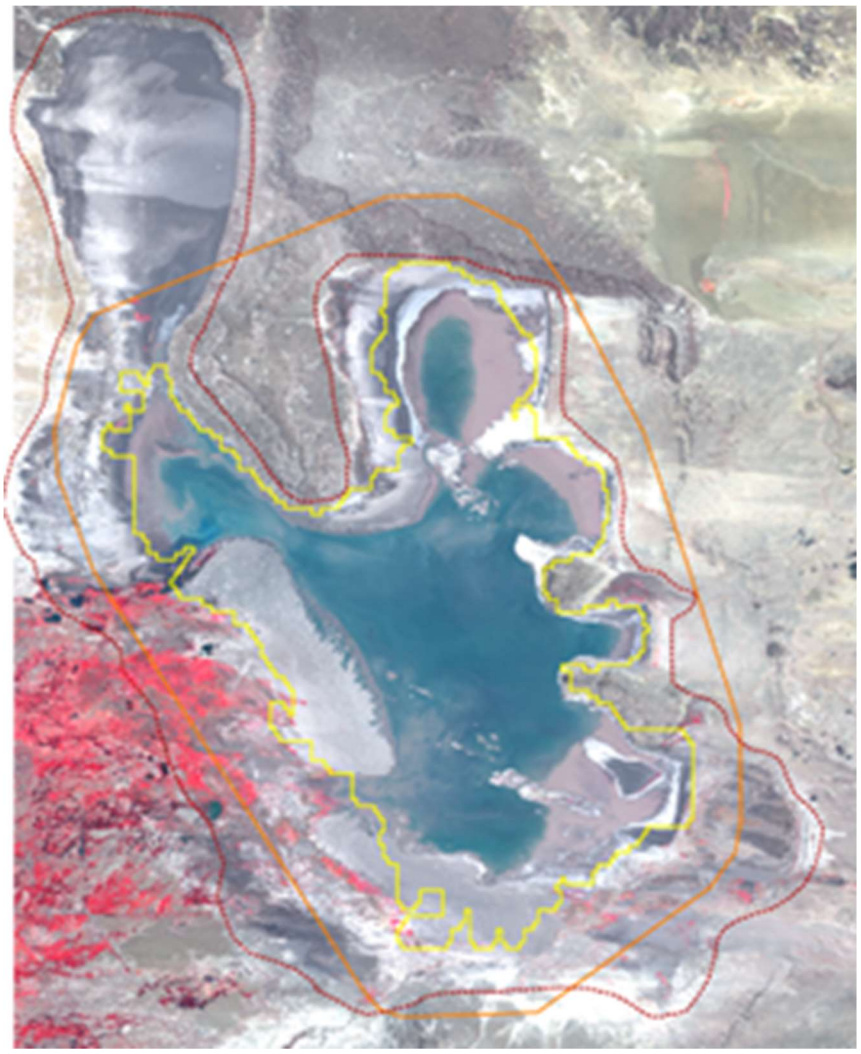

Figure 10: Reference envelope over Colhue Lake (Argentina). Yellow and orange (buffered) envelopes are derived from the ESA CCI Land Cover database and exclude the northern part of the lake. The red contour is based on high resolution satellite imagery.

In some cases, the reference polygons only include part of the lake systems. An example is given in Figure 11 for Sasykkol Lake (Kazakhstan) which is a morphologically complex system. Here, two lakes (Alakol and Sasykkol) are separated by a wetlands complex. A large part of the water surface increase is located within this complex as well as on the North-Western extent. In the medium resolution-based reference database, these surrounding wetlands complex are not included.

Figure 11: Reference envelopes over Sasykol lake (Kazakhstan)

Training sample selection: There are many ways to select training samples for classification including manual/visual effort or automated processes when time series are exploited. In either case, sample selection can introduce bias: local validity of input databases; selection criteria for training samples, and localized observation conditions.

A first example concerns introduction of sun glint (Figure 12**Erreur ! Source du renvoi introuvable.**). Using a conventional water index (MNDWI, Xu 2006), sun glint reduces the detected water surface area. Another common water index (AWEI, Feyisa et al. 2014) is in this case not affected by the sun glint.

Figure 12. Sensitivity of the MNDWI (dark blue) to sun glint, compared to the AWEI (light blue), yielding a reduced water surface area in case of MNDWI for the sun glint affected scene on 14 July 2017 (low point in time series) shown in Figure 9.

Another example to illustrate the potential influence/bias induced by input data is given in Figure 13, showing the GSW product (Pekel et al. 2016) used to define a training set. Where the GSW product yields a 0% occurrence level, water may in fact be (albeit rarely) present. Some pixels associated

with dry areas can even be found inside water bodies. This effect is likely caused by the long observation period underlying the GSW data set, whereas ephemeral water bodies could have very short periods containing water. The opposite effect, where the long-term dataset suggests water but this is not found in current time-series, is also observed. Further screening of the training data set is then essential.

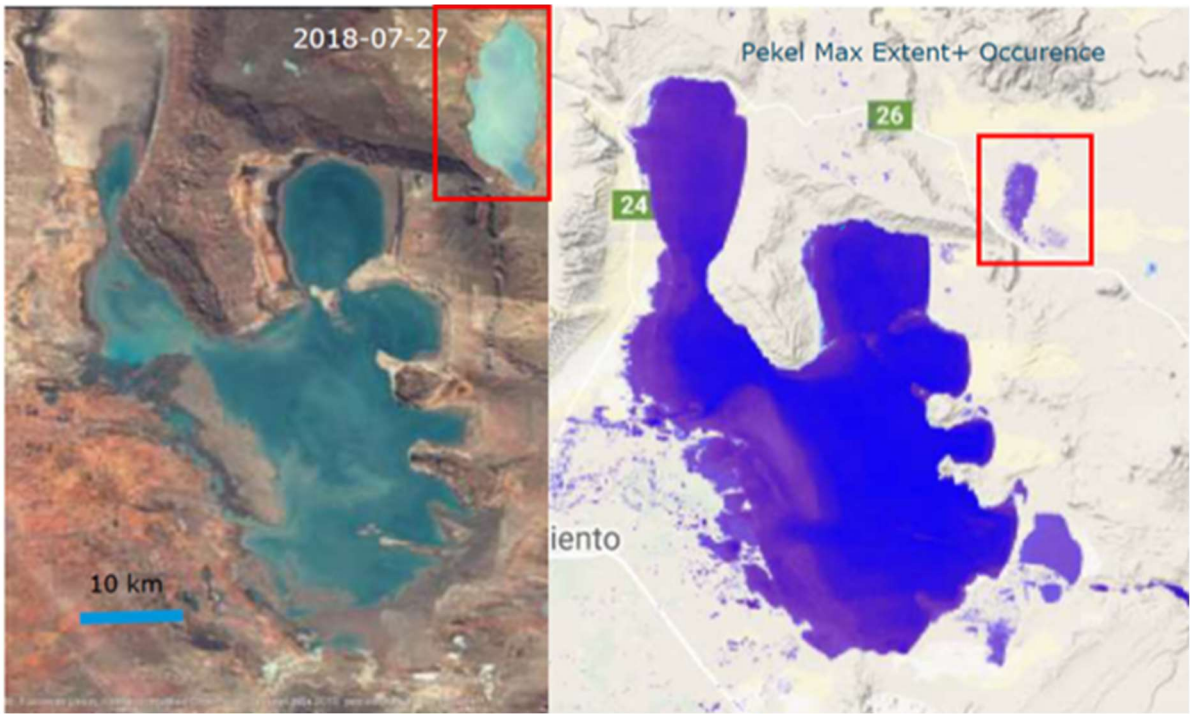

Figure 13: Lake Colhue (Argentina) shows water despite pixels having 0% water occurrence in the GSW data set.

Calibration – detectors effects: From time to time, Sentinel 2 MSI imagery may contain sensor or calibration anomalies, which then affect any use of the imagery including water detection (Figure 14). These artefacts are rare for Sentinel2 sensors.

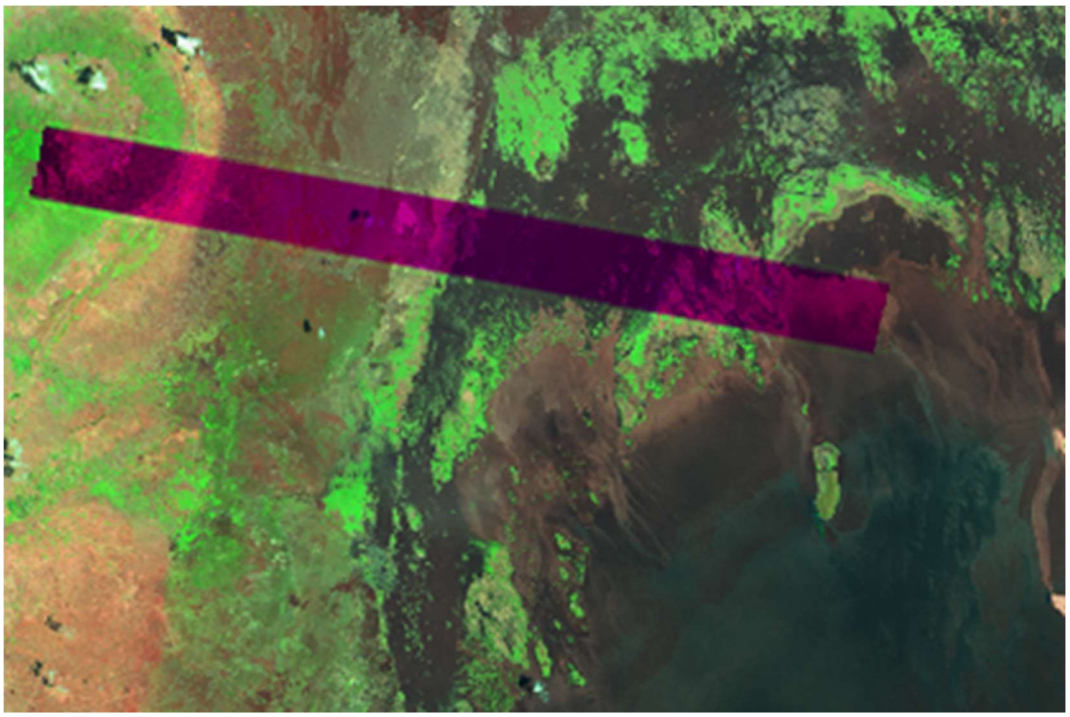

Figure 14: An example of a rare case of radiometric anomaly observed in Sentinel-2 MSI imagery of Lake Chilwa, 22 Nov 2018

Therefore, for Landsat series, particularly for Landsat 7, where the effect of the sensors are limited, as since May 31, 2003, there is Scan Line errors due to a problem in forward motion of the satellite. By the way there is a striping of the values.

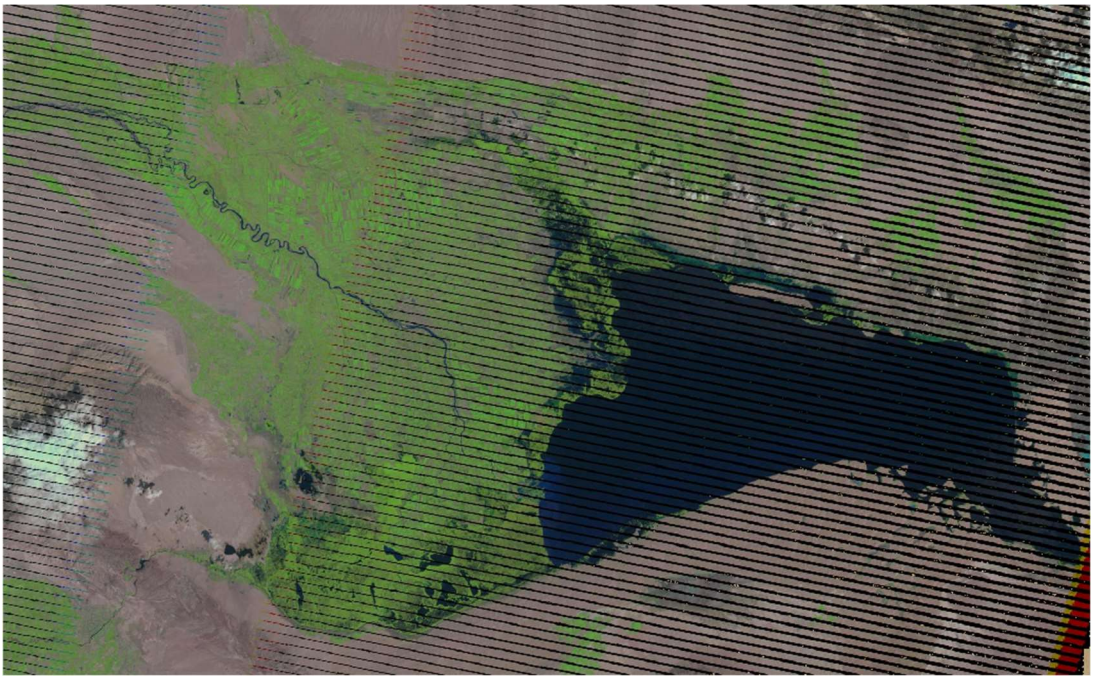

Figure 15: detectors errors on landsat 7 sensors acquired on the 2003-08-09 over Bosten lake

Misclassification effects (e.g. cloud, land, water, subpixel variation) and - adjacency or proximity effects: Confusion and commission errors occur, for example between ice and water or water and snow. The effect is obvious in the below example of Lake Namtso on the Tibetan plain (Figure 16**Erreur ! Source du renvoi introuvable.**). The effect is more evident when using K-means clustering compared to Support Vector Machines.

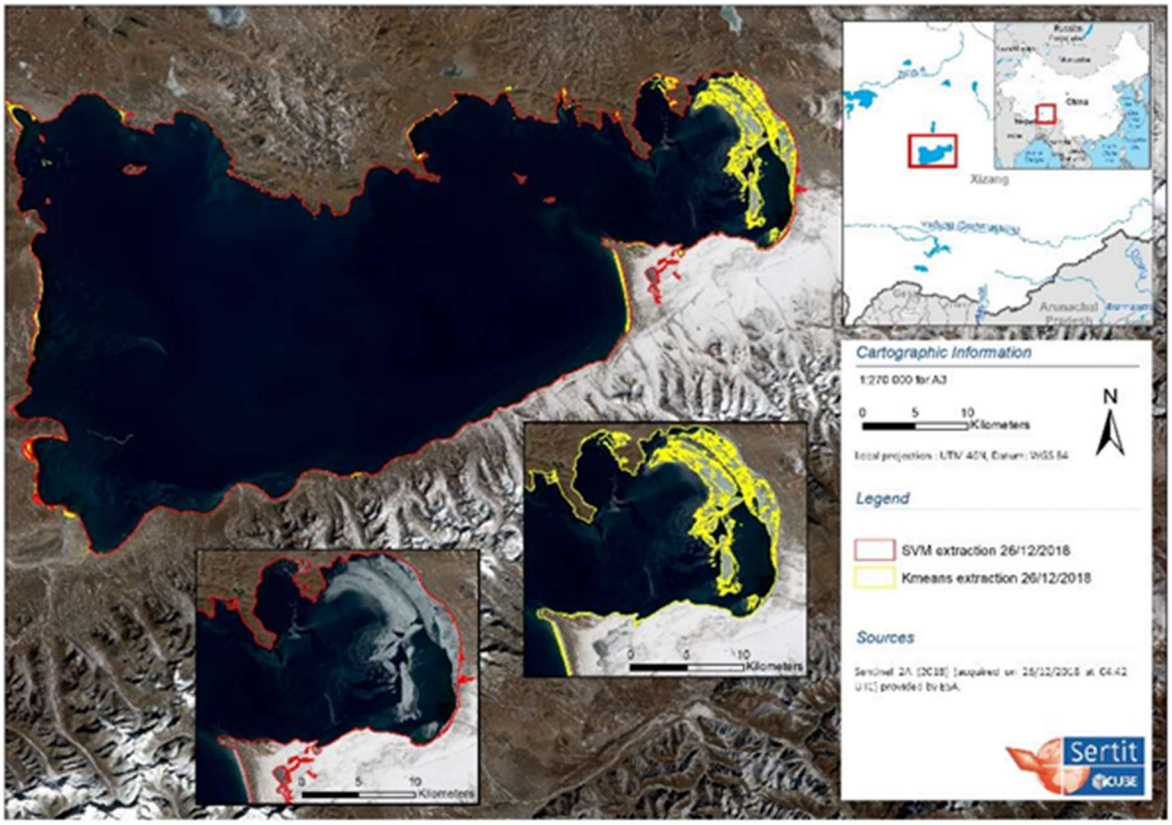

Figure 16: Confusion/commission between Ice and water when exploiting a KMeans approach rather than SVM at Namtso lake (PR China)

Lyon et al. (2013) also identified that increasing lake size and a less complex morphology decreases classification error in terms of observed lake area (relative to actual lake size), as shown in Figure 17 **Erreur ! Source du renvoi introuvable.**.

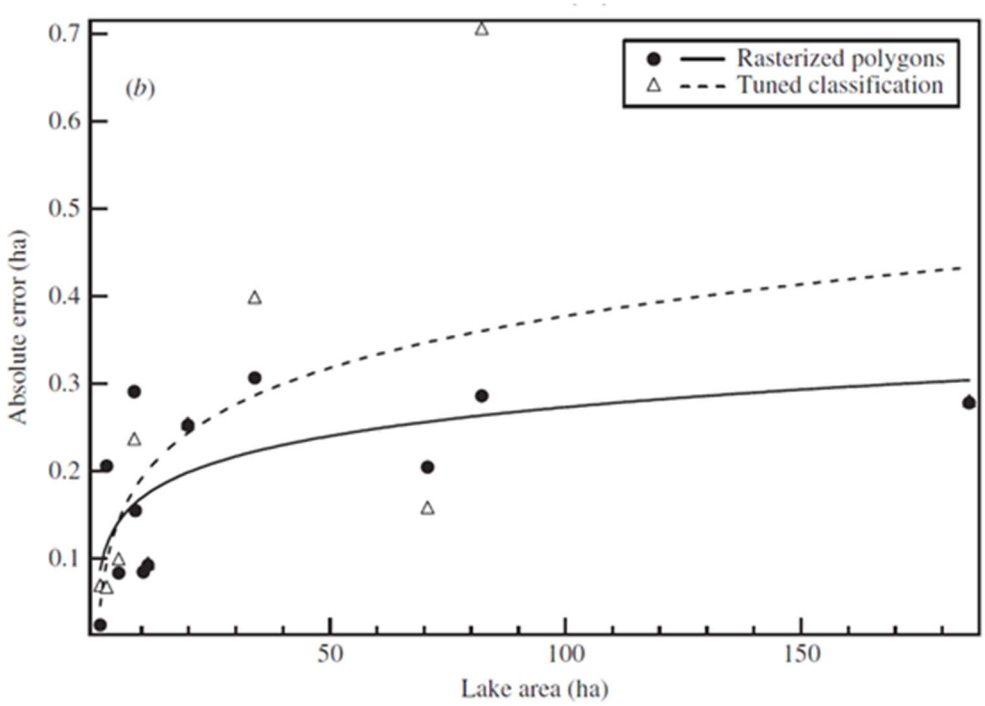

Figure 17: Scatter plots of absolute error in estimated lake area as a function of 'actual' lake area (based on DGPS shoreline measurements). Lines represent power-law fits to the respective data points (from Lyons et al. 2013)

4.3. Analysis of the hypsometry curves

To validate the different solutions, we have calculated the hypsometry curves using each of the (LWL/LWE) vectors that were measured. The results are very irregular. For some lakes like Bosten or Namco, we observed very coherent hypsometry, allowing to compute precise water extent variations (see Figure 18: the hypsometry on Lake Bosten

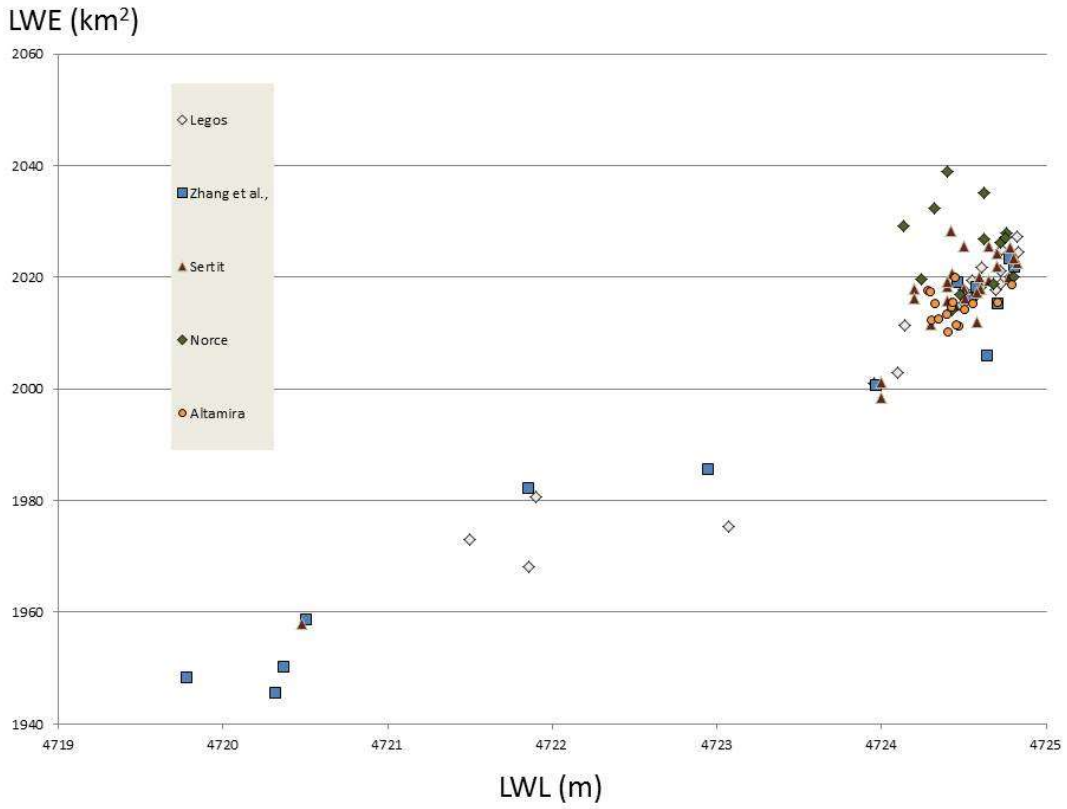

Figure 19 and Figure 19). For such lakes, the solutions agree, the final RMS is approximately 1% of the total extent of the lake and the whole range of variations of level and extent is well covered.

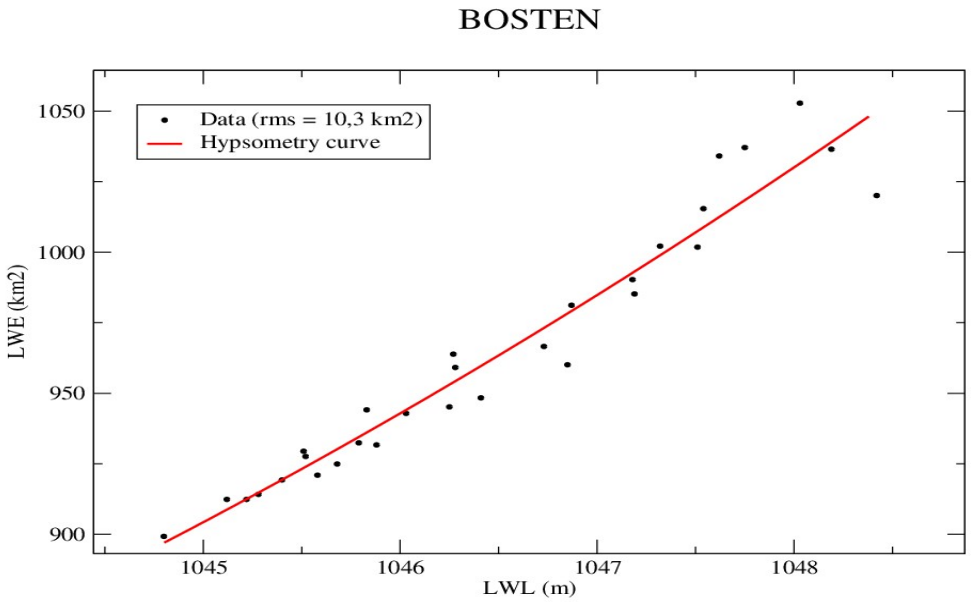

Figure 18: the hypsometry on Lake Bosten

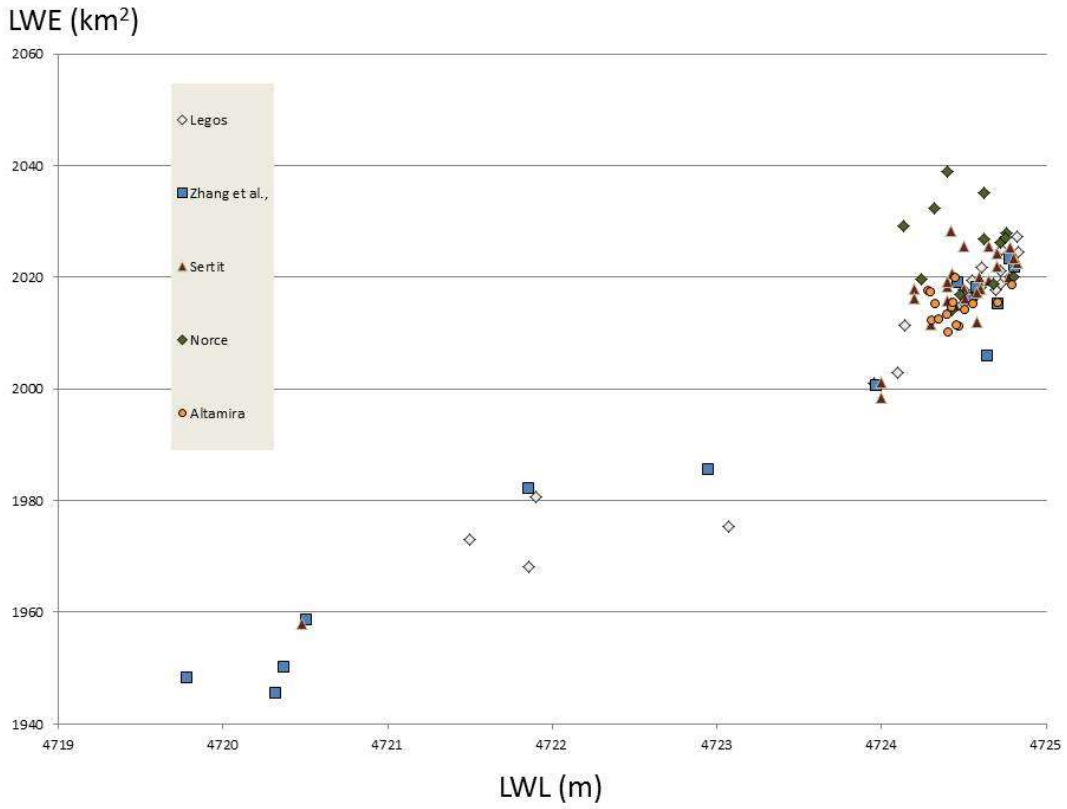

Figure 19: the hypsometry on Lake Namco

In other cases (Figure 20), it turned out that it is nearly impossible to determine which solution is best and which one to use for the final production of LWE time series. In Lake Khanka, for example, the hypsometry curves using SAR or optical are fully different and there is no coherence between the two SAR solutions (NORCE and TRE-Altamira) and the two optical solutions (Icube-SERTIT and LEGOS).

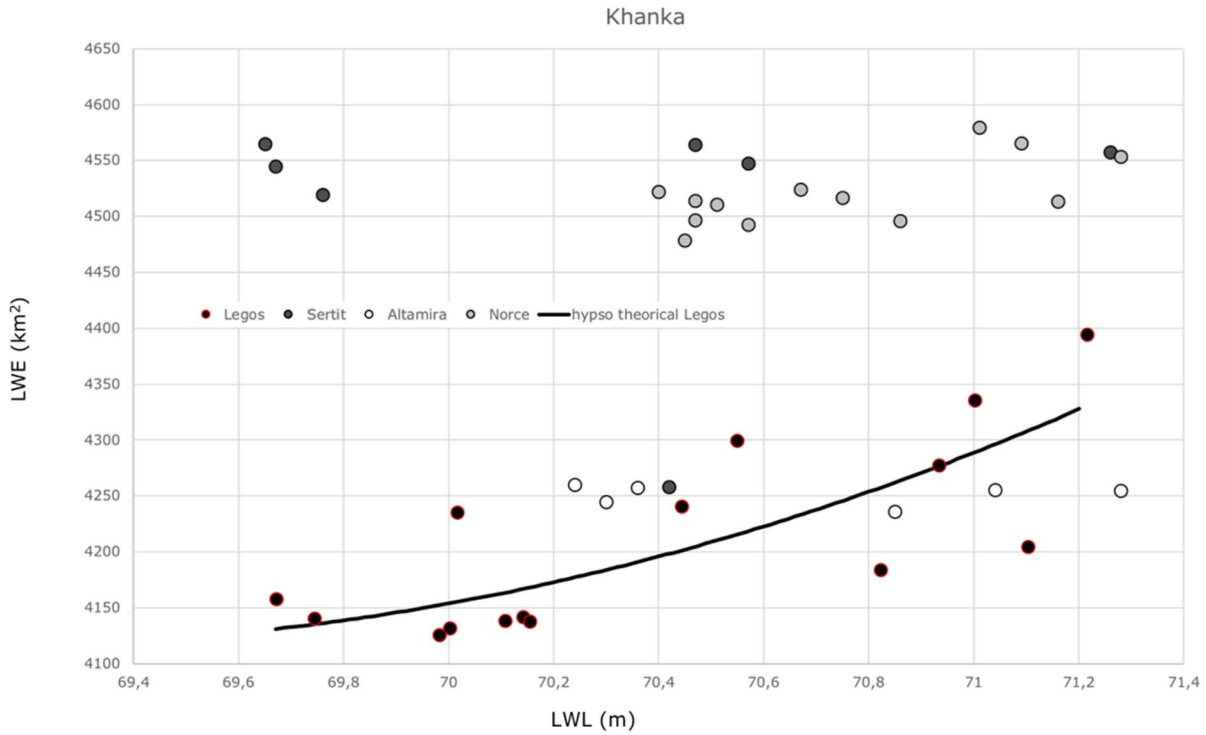

Figure 20: attempt to calculate hypsometry for the lake Khanka. We observe two groups of solutions, (Sertit/NORCE & LEGOS/Altamira) with strong bias. It also turns out that for the radar solutions LWE does not increase when LWL does.

The evaluation of the thematic variable LWE based on the Hypsometry curve is a reliable method to examine the quality of the LWE calculated from satellite images. It does not resolve all inconsistencies, but it does validate good solutions and it also allows the uncertainty of the calculation to be determined.

5. Lake Surface Water Temperature (LSWT)

5.1. LSWT Uncertainty characterization

5.1.1. Basis of LSWT uncertainty characterisation

An evaluation of uncertainty in LSWT products is given per datum, and is obtained by calculation of propagated, retrieval and sampling uncertainties.

The propagated part addresses the amplification of error in the satellite observations (the brightness temperatures, BTs) through the retrieval process, using standard equations for the type of retrieval used (optimal estimation, see below).

The retrieval uncertainty expresses the range of possible LSWTs compatible with the observations even if they were error free, since the intervening atmosphere produces some ambiguity in the relationship between the surface LSWT and the top of atmosphere satellite BTs. Another component of retrieval uncertainty is the influence of the prior value used in the optimal estimate, and error in which also (slightly) affects the result. The retrieval uncertainty component is also expressed using standard optimal estimation equations.

Sampling uncertainty at level 3 (gridded data) arises is only part of the lake within the grid cell is observable. This is well parameterised as a function of the fraction and properties of the grid cell that is observable.

The uncertainty evaluation includes all sources of uncertainty (sensor errors, modelling errors, prior errors, retrieval indeterminacy and L3 sampling) expected for a valid retrieval of LSWT.

5.1.2. Limitations of LSWT uncertainty characterisation

The standard equation for the uncertainty evaluation (and the retrieval) uses error covariance parameters that are, for LSWT, still relatively poorly known (this is quite common in optimal estimation approaches). Therefore, the uncertainty evaluation is still expected to improve in realism with further research and development.

An aspect of uncertainty that is "unexpected" (in the sense of not being accounted for in the retrieval process) and not included in the quoted LSWT uncertainty is related to any errors that may arise in cloud detection / water-surface identification. This is addressed via a quality indicator approach, whereby quality level (QL) 5 means high confidence that the retrieval assumptions are fully met and that the uncertainty provided is valid. Lower QL is attributed where circumstances suggest that retrieval assumptions are less closely met and the uncertainty evaluation may be less valid (usually underestimated). For climate studies, use of QL 4 and 5 is recommended, and for these data the uncertainty evaluations are credible.

Further aspects of LSWT uncertainty depend on the usage of the data. Where, for example, a user wishes to form a wider spatio-temporal average of LSWT, there is sampling uncertainty arising from the fact that the space-time box for the average is not fully sampled by satellite data (which may have cloud-related gaps and are obtained only when a satellite passes overhead). This is aspect goes beyond what can be provided in a product with per datum uncertainty, since it is usage dependent.

5.2. Sources of uncertainty in LSWT

The sources of uncertainty in any retrieval are the effects that cause errors in either the observations used or in other parameters that influence the retrieval result.

Detector noise: Brightness temperatures for the sensors used for LSWT are noisy at levels between about 0.03 and 0.12 K. The combination of BTs into an LSWT estimate generally amplifies the noise by a factor of between 2 and 4 (for single-view retrievals as used here). In the LSWT CCI processing, sensor-specific noise estimates are used, and the standard equations evaluate the noise amplification.

Forward model errors: The fast radiative transfer model RTTOV is used within the optimal estimation framework. Errors in simulation of BTs also propagate to LSWT. These errors are not noise in the retrieval of LSWT from a given overpass because they are largely in common between lake pixels nearby to each other, and therefore the effect is locally systematic. The magnitude of RTTOV uncertainty is relatively poorly known, but appears to be comparable to the BT noise. The combined effect of detector and forward model errors is represented by an error covariance matrix in the optimal estimation framework.

The effects of instrument calibration and forward model errors are indistinguishable in the uncertainty budget of optimal estimation: the simulation should ideally represent the instrument (with its calibration characteristics) and therefore we don't distinguish forward model error and calibration error in practice (when doing this sort of retrieval).

The combined LSWT uncertainty from observation noise and forward model errors is typically of order 0.3 K.

Prior error: Optimal estimation retrieval starts from a prior estimate of the state (here, of the LSWT and atmospheric water vapour) which is then updated using the new observations (here, BTs). The prior estimate is subject to uncertainty (otherwise, we would not need to do the retrieval) and therefore in a given instance there is a prior error that is unknown. Because of the indeterminacy of the retrieval (an intrinsic feature of inverse problems) some of the prior error persists in the retrieved LSWT. For LSWT, the prior uncertainty is around 1 K and typically 10% to 20% of any prior error persists. (The degree of propagation of prior uncertainty can be quantified in an optimal estimation framework using standard theory.) Thus, the uncertainty associated with this component is of order

0.1 K to 0.2 K. If the prior is biased, this effect will be systematic on the same spatio-temporal scales as the prior errors.

Retrieval error: The optimal estimation framework uses the error covariance information supplied to the retrieval along with the Jacobians (sensitivities) of the observations to LSWT (calculated by RTTOV) to evaluate the degree to which the retrieval result is uncertain, using standard equations (that also account for all the elements described above).

Sampling error: All of the above sources relate to single pixel retrieval. Sampling errors arise when combining the swath pixels to gridded products (level 3). Not all cells are fully observed because of swath edges and intervening clouds that obscure the lake in the infrared. However, users interpret gridded data as representing the average across the entire cell. Therefore, when the cell is not fully observed, there is a sampling error which is equal to the difference the unobserved parts of the cell would make to the cell mean if they could be observed. This is generally considered to be a random effect (although biased circumstances can be imagined). Sampling uncertainty is greater when less of the cell is observed, and when the LSWT in the cell is more variable. The parameterisation of this in terms of the cell fraction observed and the variability observed follows that proposed by Bulgin et al (2016) for sea surface temperature.

Somewhat balancing the "extra" sampling uncertainty is the effect of averaging down the noise when combining several LSWTs into a cell mean (the familiar "$1/\sqrt{n}$" uncertainty reduction).

Classification error: The uncertainty model is complete and valid for situations where the retrieval assumptions are met: cloud-free low-aerosol skies over ice-free lake surfaces. Where residual cloud or heavy aerosol is present, the retrieval may be affected to a degree not captured by the uncertainty evaluation. Likewise, if part of the surface is not water but is ice, land or flotsam, the obtained temperature interpreted as LSWT will have an additional error.

As is apparent from the example in Figure 21 below, clouds may have fuzzy edges that make some degree of residual cloud contamination likely in some pixels used for LSWT.

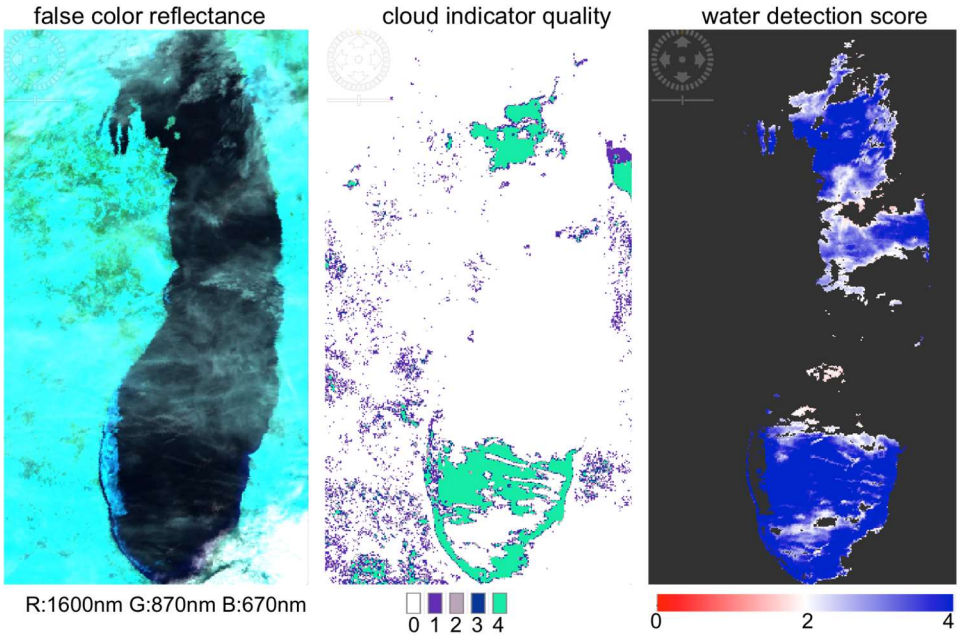

Figure 21 (Left) False colour composite image of reflectance at 1600 nm (R), 870 nm (G) and 670 nm (B). (Middle) the co-incident MERIS cloud indicator. (Right) The water detection score: QL 5 retrievals are made where this score reaches 4 or more. All panels for lake Michigan in USA on the 15-Feb-2011.

There being no known means to evaluate classification uncertainty on a per datum basis, this aspect is handled via quality levels, as mentioned above.

5.3. Uncertainties estimation methodology in LSWT

Optimal Estimation retrieval has been used for LSWT for all the sensors since it is based on physics, and can be applied where no in situ data for retrieval tuning are available. This gives good reason to expect stable performance across domains in time and space.

For single-view instruments, the LSWT is retrieved using an optimal estimation (OE) scheme (MacCallum and Merchant, 2012):

$$\hat{x} = x_a + G(y - F(x_a)) \quad [5.1]$$

$$G = (K^T S_\varepsilon^{-1} K + S_a^{-1})^{-1} K^T S_\varepsilon^{-1} \quad [5.2]$$

The retrieved state is the prior state plus an increment of

$$G(y - F(x_a)) \quad [5.3]$$

F is the forward model and the matrix K expresses how the observations change for departures from the prior state, i.e., it is a matrix where a given row contains the partial derivatives of the BT in a particular channel with respect to each element of the state vector in turn. The partial derivatives are the tangent linear outputs from the forward model. S_ε is the error covariance of the differences between the model and observed BTs. This error covariance matrix is the sum of the radiometric error covariance in the observations (S_o) and estimated error covariance of the forward model (S_m). S_a is the error covariance matrix for the prior state variables.

It has been shown that a reduced state vector, $\mathrm{z(x)} = \begin{bmatrix} \mathrm{x} \\ \mathrm{w} \end{bmatrix}$ where x is the LSWT and w the total column water vapour can be used in the retrieval instead of the full prior state vector $\mathrm{x_a}$. However, the full prior state vector is used in the forward model (see (MacCallum and Merchant, 2012).

Given the above, in LSWT processing the uncertainty due to noise (assumed uncorrelated between pixels) and due to uncertainty from retrieval-related factors (assumed correlated on synoptic scales) are estimated by:

$$\sqrt{\mathrm{GS_oG^T}} \quad [5.4]$$

and:

$$\sqrt{\mathrm{GS_mG^T}} \quad [5.5]$$

These equations are not a complete decomposition of uncertainty estimation using an optimal estimation framework, and will be refined in a future version.

To determine the uncertainty of the gridded product, the components are combined according to their correlation structure. Over a grid cell, the uncertainty from propagated noise is treated as random between pixels, the uncertainty from retrieval-related factors is assumed to be systematic across all pixels and the sampling uncertainty is combined as a parameterised value obtained at the whole-cell level.

The total uncertainty is available for users in the gridded product.

5.4. End to end uncertainty budget in LSWT

The evaluated uncertainty for a L3 LSWT is typically in the range 0.4 to 0.7 K. (All statements here are made on the basis of LSWT v4.0.) Two example uncertainty fields are shown in Figure 22.

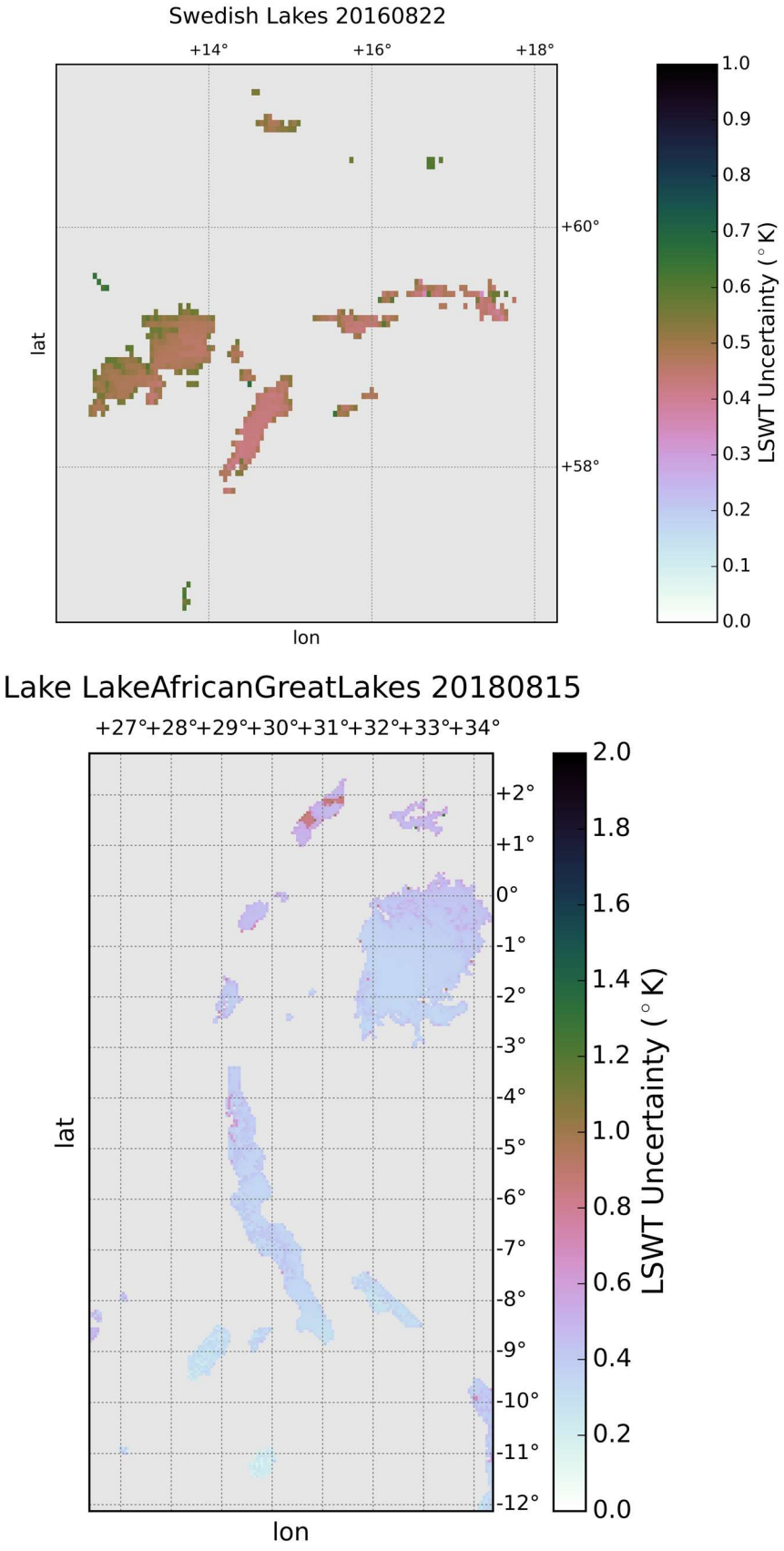

Figure 22 Uncertainty fields for L3 product for (top) Swedish lakes (the largest being Vanern and Vattern) and (lower) east African lakes (the largest being Lake Victoria).

As Figure 22 shows, uncertainty tends to increase towards lake edges and may be greater in cells where cloud has partly obscured the lake surface.

Figure 23 shows a validation of satellite LSWT against in situ measurements of the sort that is routinely done for the 56 lakes for which (so far) intensive personal collection efforts have yielded usable data. The satellite data are coloured by QL (1 to 5, of which 4 & 5 are recommended for use), and uncertainty bars are plotted (not always visible beyond the circles).

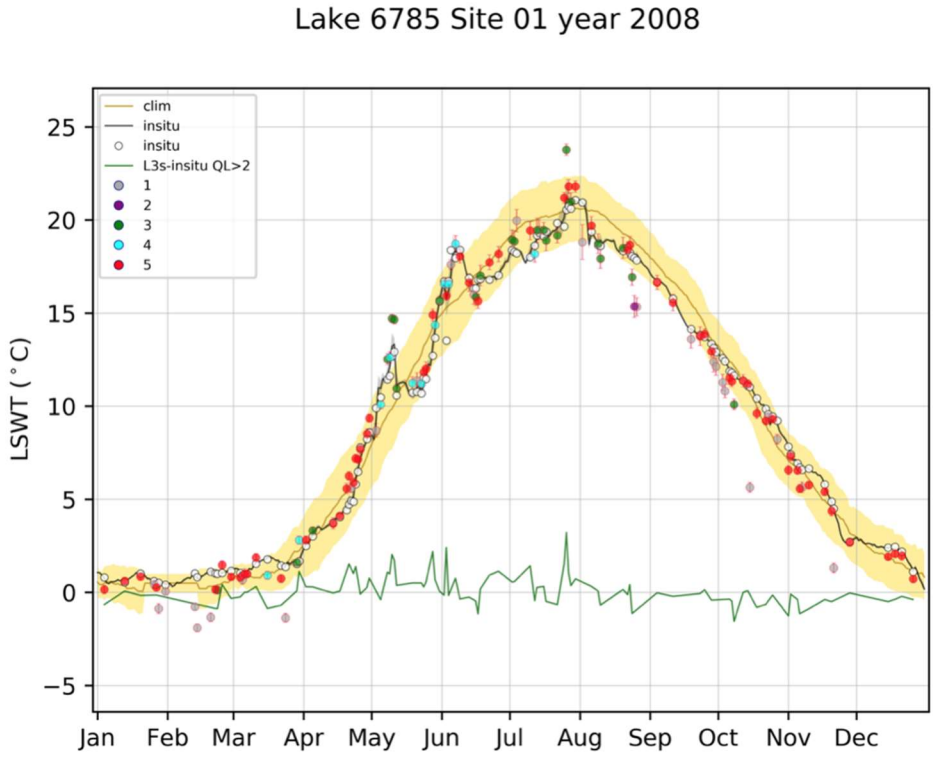

Figure 23 Satellite observations (dots), in situ matches (white dots), in situ measurements (black line), satellite minus in situ temperature difference for quality levels 3,4,5 (green line) and climatology (golden line) for lake Erken in Sweden in 2008.

The LSWT variability is, for a given time of year, shown by the width of the climatological band (representing 1 standard deviation of variability around the climatological mean), and is about 2 K. Clearly, the satellite data tracks the LSWT fluctuations of this specific year (2008) around the climatological expectation, which are also seen in in situ data. The "signal to noise", using the variability as the measure of signal, is thus ~ 2 K / 0.4 K = 5. Using the climatological range to represent the signal would give an SNR an order of magnitude larger.

Using all the matches of the sort shown in Fig 4.3 across all the lakes (sample of opportunity, not geophysically representative), the statistics of satellite-minus-in situ difference are as in Table 2.

Table 2 Statistics of satellite minus in situ LSWT per quality level.

QL	N	Median/K	Standard Deviation/K	Robust Standard Deviation/K
5	29819	-0.13	0.91	0.52
4	11361	-0.28	1.18	0.79
3	12618	-0.25	1.39	0.95

Because of the water-temperature skin effect, the expected offset between these datasets is of order -0.17 K, and therefore we see that the QL 5 relative mean difference is well within an accuracy target

of 0.1 K. For QL 5, the robust standard deviation is compatible with a closed uncertainty budget, assuming that in situ uncertainty is of order 0.2 K (a number based on experience of ocean-going buoys). The higher standard deviation indicates the presence of a minority of larger outliers (I.e., the distribution is not normal, but has heavy tails). Thus, for the majority of QL 5 data, the evaluated uncertainty appears to be of the right magnitude; the reasons for the heavy tails in the distribution, where discrepancies are larger than expected, can include in situ data where 0.2 K is overly optimistic as an estimate of uncertainty. The QL system is working as expected, as shown by the fact that the SD and RSD increase in size for the lower quality levels, for which the evaluated uncertainty is likely to be underestimated.

Overall, despite the limitations pointed out earlier, the uncertainty budget used and applied to the LSWT products is quantitatively realistic for QL 5 data, and is useful in discriminating more and less certainty LSWT data points.

6. Lake Ice Cover (LIC)

6.1. LIC Uncertainty characterization

The assessment of uncertainty in the LIC product is currently performed through computation of a confusion matrix built on independent statistical validation. Thus, uncertainties are not assessed at a per-pixel level, but rather from classification error calculated from multiple samples/images. The reference data for validation are collected from the visual interpretation of imagery from several ice seasons (freeze-up and break-up periods) by skilled ice analysts. Classification error (%) derived from the confusion matrix is the metric used to report total uncertainty for each class (ice, water, cloud). In Lakes_cci LIC (L3) product v1.1, pixels belonging to the same class are simply assigned the same % error value in the uncertainty band provided.

6.2. Sources of uncertainty in LIC

The sources of uncertainty in the Lakes_cci LIC product are provided below along with a brief description of the expected effects from each source. It is important to note that the random forest algorithm implemented for LIC v1.1 production uses MODIS Terra Level 1B (TOA Reflectance) 5-Min Swath (MOD02), Collection 6.1 (C6.1), data as input data. On-orbit noise characterization of MODIS reflective and thermal bands is regularly monitored to ensure the fidelity of the on-orbit calibration to the prelaunch measurement (Angal et al., 2015; Sun et al., 2014). The primary source of uncertainty in the LIC product is expected to come from the retrieval algorithm (including the selection of sampled sites/pixels used to develop the algorithm).

Detector noise/sensor degradation: The MODIS sensors (launched on the Terra and Aqua satellites in December 1999 and May 2002, respectively) are currently operating well beyond their life expectancy of 6 years; their detectors are degrading. However, MODIS was designed with stringent requirements on the sensor's calibration accuracy and data product quality. It is equipped with a set of on-board calibrators (OBCs), including a solar diffuser (SD) and a solar diffuser stability monitor (SDSM) for the reflective solar bands (RSB) calibration and a blackbody (BB) for the thermal emissive bands (TEB) calibration (Xiong et al., 2005). MODIS RSB on-orbit calibration is reflectance based using the on-board SD plate with its bi-directional reflectance factor (BRF) determined pre-launch. MODIS TEB are calibrated on-orbit using the onboard BB.

For technical reasons, not elaborated upon herein (see MODIS Characterization Support Team website, https://mcst.gsfc.nasa.gov), nonphysical trends in MODIS Terra data products, which result from calibration drift, have been observed and are well documented. On-orbit calibration procedures to mitigate long-term calibration drift, particularly at the shorter wavelengths, have been applied to MODIS C6 L1B data.

Low illumination conditions: Low solar illumination conditions occur when solar zenith angles (SZA) are larger than 70 degrees and when a swath is near the day/night terminator. Those situations usually occur in the lakes located in high latitudes or during the period of ice formation just before or after polar darkness. With solar zenith angle increasing, the diffuse radiation flux increases,

resulting in the attenuation of the radiation flux of incident sunlight at the surface (Coakley, 2003). Therefore, top-of-the-atmosphere (TOA) reflectance is very low due to the lack of solar radiation reflected by the surface. The low solar reflectance conditions may increase the uncertainly in detection of ice cover. According to the research presented by Wu et al. (2021), due to the low illumination conditions resulting in low TOA reflectance from the Earth, more classification errors could appear in the period of ice freeze-up compared to the period of ice break-up.

Geometry errors: MODIS Level 1B product provides images in 36 spectral bands and corresponding latitude and longitude data of WGS84 without projection information. In the pre-processing stage, MODIS pixels are resampled based on their coordinate to output grids by applying the nearest neighbour approach. Since MODIS uses the cross-track scanning mirror, pixels near the edge of a swath could be mis-matched to the output grids due to the bowtie effect and data errors of latitude and longitude. The mis-matched pixels highlighted in Figure 24(b) could present wrong surface features in the LIC product as Figure 24(a) shows. Moreover, because of the bowtie effect, the spatial resolution degrades from nadir to the scan edge. Therefore, as a result, the coarse MODIS pixels near the swath edge could raise uncertainty in the LIC product. For example, compared to Figure 24 (c), the MODIS image acquired from the same day by Terra shown in Figure 24 (b) shows less details of surface condition.

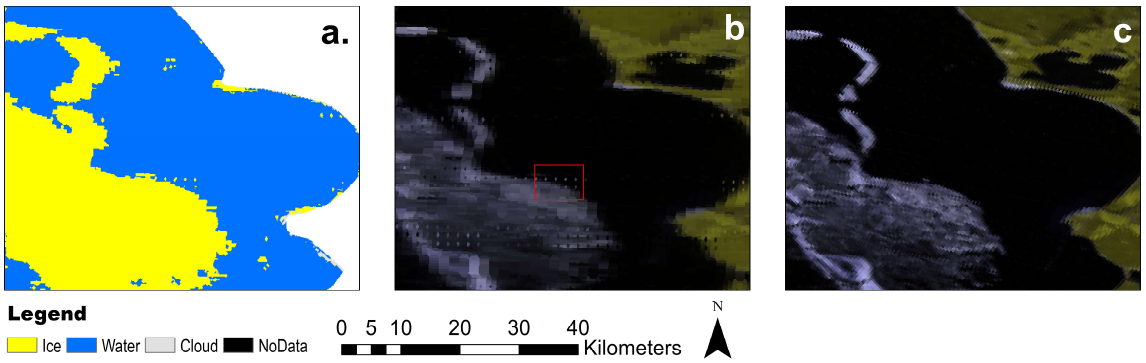

Figure 24 Example of impact of geometry errors on LIC product, Lake Onega, 28 April 2019. (a) LIC daily product from MODIS Terra, (b) false colour composite of MODIS Terra image (bowtie), and (c) false colour composite of MODIS Terra image (non-bowtie).

Algorithm: The algorithm implemented for Lakes_cci LIC v1.1 production follows a machine learning approach; here random forest was selected. Wu et al. (2021) presented and assessed the lake ice classification performance of random forest and other algorithms by using MODIS Terra Level 1B product. Sampled data (AOIs) from 17 lakes across the Northern Hemisphere were used to train and validate the random forest algorithm. Lake sites (sampled pixels) for algorithm development and validation must be carefully selected to capture the broadest suite of conditions (e.g. clear-sky, overcast, partly cloudy, low to high solar zenith angles, open water from low to high turbidity, clear and thin ice, snow-covered ice) that have an impact on the magnitude and variations of TOA reflectance (MOD02) for the classes of interest; here ice, open water and cloud. Compared to the classification accuracy calculated from a random k-fold cross-validation approach, lower classification accuracy was obtained through temporal and spatial cross-validation (Wu et al., 2021). The accuracy differences between cross-validation methods, therefore, reveal that spatial and temporal heterogeneity of MODIS observations could result in classification errors by applying random forest or any other classifier. Unexpected classification errors may be present in the global lake ice maps. In the LIC v2.0 product, additional scenario would be developed and applied to correct the misclassification pixels.

Misclassification: In the processing chain implemented for LIC v1.1 production, the maximum water extent mask derived from ESA CCI Land Cover v4.0 at 150-m resolution is used to determine which pixels within a lake will be processed with the threshold-based algorithm to determine class belonging. Since the mask represents maximum water extent, it can introduce some misclassification errors along the shoreline of lakes (i.e. mask spilling over land so that some MODIS land pixels will be incorrectly flagged as lake pixels). A cursory look reveals that this may indeed be the case (**Figure 25Erreur ! Source du renvoi introuvable.**), but the full impact of this mask remains to be quantified.

In addition, it has been found that when applying the algorithm at the global scale, some pixels may be misclassified as ice covered where lakes are dried-up in summer due to the high reflectance of sediments from exposed lakebeds. The presence of aquatic vegetation and algae blooms during the ice-free season change the reflectance characteristics and may also cause erroneous ice cover detection in the spring or summer (Riggs and Hall, 2015). To reduce uncertainty introduced from these varied sources, validation is ongoing over a larger number of lakes globally and training pixels are being added to increase the performance of the random forest classifier in these more difficult cases) prior to the release of LIC v2.0 product.

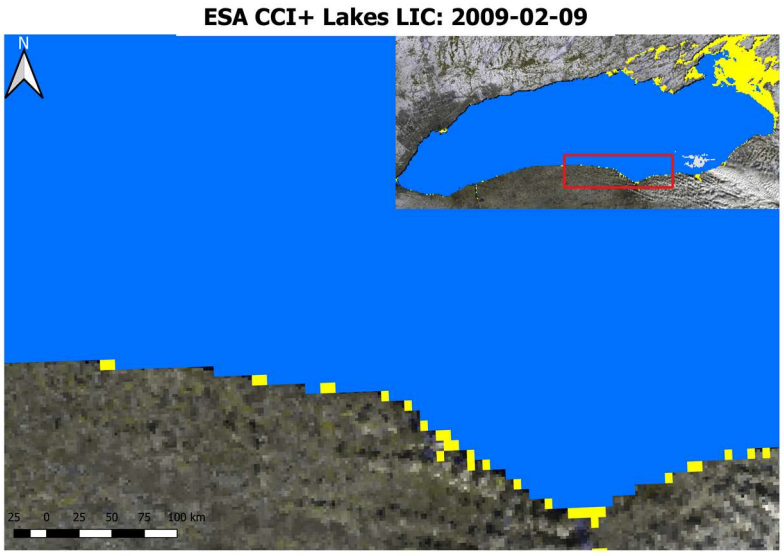

Figure 25. Example of pixels along the shoreline of Lake Ontario mislabelled as ice (in yellow), 9 February 2019.

6.3. Uncertainties estimation methodology in LIC

For Lakes_cci LIC product v1.1, class uncertainty is computed from a confusion matrix built on an independent statistical validation process. The confusion matrix shown in Table 3**Erreur ! Source du renvoi introuvable.** was produced from AOIs (n = 10,075,081 pixels) collected through visual interpretation of 219 Terra images over an ice season at Great Slave Lake (GSL) and Lake Ladoga (LL) (GSL: 2018-2019, LL: 2019).

Table 3 shows the overall accuracy (97.34%) as well as the accuracy of individual classes (97.77% for ice cover, 99.17% for water, and 96.93% for cloud cover). Uncertainty is reported as % error from classification (100 - % class accuracy) for each class. Currently, pixels belonging to the same class are given the same error value.

Table 3: Confusion matrix with class accuracies for Lakes_cci LIC product v1.1.

Lakes_cci LIC		Retrieval Algorithm			class accuracy (recall)	Uncertainty
		Ice	Water	Cloud		
User-defined	Ice	1,514,517	4,518	30,057	97.77%	2.23%
	Water	1,540	1,286,093	9,213	99.17%	0.83%
	Cloud	199,996	22,265	7,006,882	96.93%	3.07%
	Overall Accuracy: 97.34%					

Following the release of the LIC v1.1 product, efforts will be placed in identifying additional metrics for consideration and possible implementation leading to product v2.0. For example, probability, which informs about the confidence in the classification, is considered as a valuable proxy for the

uncertainty associated with machine learning (ML) algorithms; algorithms that are currently being considered for LIC v2.0 production.

6.4. End to end uncertainty budget in LIC

The evaluated uncertainty for LIC L3 product is 0.83% for open water, 3.07% for cloud cover and 2.23% for ice cover, as derived from accuracy assessment (confusion matrix) of each class through independent statistical validation. An example of the uncertainty map for the Caspian Sea is shown in Figure 26**Erreur ! Source du renvoi introuvable.**.

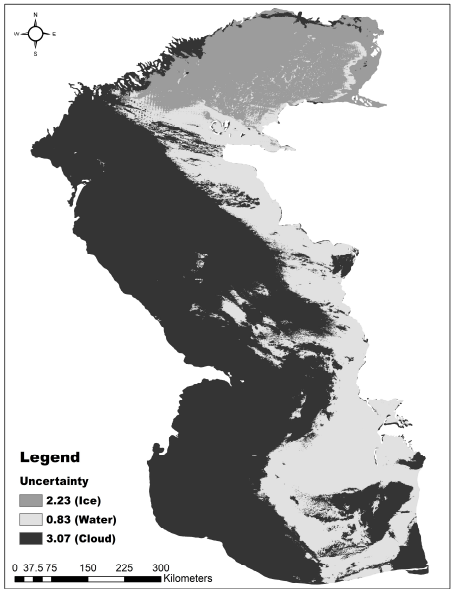

Figure 26. Uncertainty map of Caspian Sea, 1 February 2004.

7. Lake Water-Leaving Reflectance (LWLR)

7.1. LWLR Uncertainty characterization

The uncertainty characterization of the LWLR products, including water-leaving reflectance, chlorophyll-a and turbidity, is based on end-to-end validation of the satellite product against in situ matchups with satellite measurements. These results are separated per optical water type and extrapolated over the applicable range of each individual algorithm used to compute a single measurement. Uncertainties are thus simultaneously characterized for the whole system including satellite sensor, atmospheric correction, and algorithms for water column substance concentrations.

End-to-end characterization of uncertainty is more feasible than the propagation of uncertainty associated with each part of the system, because the processing chain includes elements of non-linear optimization in algorithms with multiple free variables introducing a high degree of complexity.

The characterization of uncertainty based on in situ validation yields an uncertainty function for each desired uncertainty metric, which is either the relative uncertainty (%), relative unbiased uncertainty (%, not including systematic effects) or absolute relative uncertainty (%) depending on the specific product. Where these uncertainty functions are based on a sufficient number and range of in situ observations, they are used to propagate product uncertainty through each contributing algorithm and for each measurement or pixel.

Lakes_cci marks the first introduction of the per-pixel uncertainty products in the *Calimnos* multi-sensor satellite data processing chain. The evaluation of uncertainty is entirely based on the Lake Bio-optical Measurements and Matchup Date for Remote Sensing (LIMNADES) data set of in situ optical-

biogeochemical observations, contributed by various research groups and curated at the University of Stirling. Over time, it is expected that the number of recent observations in this database will grow and this would lead to more accurate estimates of uncertainty for currently under-sampled water types and their corresponding algorithms.

7.2. Sources of uncertainty in LWLR

The following is an overview of sources of uncertainty and their expected (or unknown) effect on the LWLR product or derived water column products. For each element it is indicated whether the uncertainty is planned to be studied in detail or only part of the end-to-end uncertainty characterization.

Detector noise: this effect is expected to be random and variable over time. It is included in the end-to-end uncertainty characterisation insofar as it is present in the match-up data set used during validation of individual algorithms. It is possible, for some sensors that the magnitude of this effect changes over time. We expect this effect to be accounted for (if present) during analysis of inter-sensor bias, when the observation periods of two sensors overlap.

Observation noise: certain observation effects, notably the optical pathlength to the target under variable viewing and illumination angles, are expected to influence the magnitude in uncertainties resulting from atmospheric correction of the LWLR signal. However, due to limited reference data we do not anticipate that we will be able to isolate this effect. Due to the non-linear optimisation methodology of POLYMER, the atmospheric correction algorithm used for at least MERIS and OLCI and candidate for MODIS, it is not possible to predict whether the effect is likely positive or negative, this will equally depend on the shape and magnitude of non-atmospheric components, including LWLR and sun glint.

Algorithm effects: It is commonly understood that the dominant source of error in deriving LWLR stems from the separation of atmospheric path radiance and LWLR. This is an optimisation problem occasionally suffering a relatively large number of free variables, only bounded by supplying specific wavebands which respond to either atmospheric or water optical features. The algorithm uncertainty also relates to *adjacency effects*, since the atmospheric correction error is expected to increase with proximity to land and specifically when LWLR departs from zero in the near infra-red and shortwave infra-red, where highly efficient light absorption by water is a dominant optical feature.

Algorithms for the retrieval of optical-biogeochemical water column properties have limited ranges of applicability, e.g. to relatively clear or turbid waters. Candidate algorithms in Lakes_cci are calibrated against in situ data to counteract biases in both LWLR and the algorithms themselves. This takes place within a framework of Optical Water Types to limit the application of any given algorithm outside of its calibrated range. For some water types, calibration, validation and resulting characterisation of uncertainty still relies on the same data set, due to limited availability of satellite match-ups. Where possible, the optimisation and characterisation of uncertainties are separated.

Calibration effects: the methodologies and accuracy of in-flight sensor calibration have evolved over several generations of ocean-colour sensors. Degradation of both sensor and on-board calibration materials and the need to calibrate the response of a (usually) 2-dimensional sensor array which may further include multiple detectors, add to the complexity of this challenge. Two activities contribute to achieving consistent sensor response between satellite missions. Sensor-to-sensor comparisons cannot solely rely on simultaneous observations, because variability across and between detectors needs to be accounted for, while different viewing and illumination angles will introduce real divergence between sensors. Alternatively, system vicarious calibration aligns the system of sensor response plus atmospheric correction between sensors and optionally against reference measurements at ground level. Within Lakes_cci we will primarily consider vicarious calibration during the overlap of sensor missions. This approach does not specifically attribute uncertainty (specifically, bias) to the sensor but prioritises harmonized retrievals of LWLR despite differences in sensor capabilities such as radiometric sensitivity or waveband configurations.

***Misclassification effects** (e.g. cloud, land, water, subpixel variation):* For inland waters, the presence of cloud or land may introduce an adjacency effect (usually a brightening in the near infra-red) on nearby water pixels. This adjacency effect leads to misinterpretation of the contribution of the atmosphere to signal at the sensor, usually resulting in over-correction for atmospheric effects

and low or negative water-leaving reflectance values. This effect propagates to the derived products (chlorophyll-a and turbidity) depending on specific algorithm sensitivity to these effects. This effect is not systematic in nature and thus expected to be reflected in product uncertainty.

Moreover, inland waters are highly dynamic and optically complex waters, where vegetation (including both algae and macrophytes) occur in irregular shapes which can fully or partially cover image pixels. The uncertainties on chlorophyll-a/turbidity estimates would then be introduced by mixed pixels containing both algae and algae free surfaces. For some shallow waters, the presence of submerged or emergent macrophytes may also introduce uncertainties in chlorophyll-a retrieval since the current algorithms do not necessarily separate vegetation (particularly with sub-pixel cover). We do not anticipate to isolate this effect because such cases are not typically included in in situ reference observations.

7.3. Uncertainty estimation methodology in LWLR

LWLR per-pixel product uncertainties are extrapolated from end-to-end algorithm validation. For LWLR, the validation results are specific to each combination of satellite sensor and algorithm. For products generated from LWLR (chlorophyll-a, turbidity) they are further separated by Optical Water Type (OWT). A flow chart of the validation procedure yielding the uncertainty models from which per-pixel uncertainties are generated, is provided in Figure 29.

The uncertainty of LWLR is always expressed as a function of waveband. Additionally, the amplitude of reflectance (in case the response is non-linear) and the time difference between satellite and in situ reference observations can be evaluated. For the propagation of the uncertainty model to individual satellite observations (pixels) only the waveband and optionally the amplitude are taken into account.

If the number of matchups between in situ and satellite observations is low, the matchup time window may be extended at the expense of higher uncertainties. The time window will not exceed more than 7 days from the satellite observation. The minimum number of matchups to derive statistically robust estimates of uncertainty will be in the order of 50-100 observations distributed over multiple lakes and spanning a range of reflectance amplitudes representative of the global variability in LWLR. When such conditions are not met for a given sensor-waveband combination, the uncertainty model is not used; corresponding pixels are flagged as having unknown product uncertainty.

Ultimately, the validation procedure for LWLR yields uncertainty models per sensor-waveband for the relative uncertainty (RU, %) and relative unbiased uncertainty (RUU, %).

The uncertainty estimation of RU and RUU for LWLR is conducted by first calculating the difference and unbiased Difference (see below) between the matchups of the in situ and remote sensing R_w for each band.

Difference

The difference is expressed as:

$$Diff = R_{IS} - R_{sat} \quad [7.1]$$

Where R_{IS} is the in situ observation of R_w, and R_{sat} is the remotely sensed R_w.

Unbiased Difference (UD)

The UD is defined as the distance between the remote sensing R_w and the regression line (between the in situ and satellite R_w matchups). This removes systematic effects from the uncertainty estimate which is desirable when the product is subject to further calibration (as is done with downstream chl-a and turbidity algorithms). The linear regression line (Y) between in situ and remote sensing R_w matchups at each waveband can be written as

$$Y = a*R_{IS} + b \quad [7.2]$$

Where a and b are coefficients from the regression. The UD can be then obtained by subtracting the Y from the satellite R_w

$$UD = R_{sat} - Y \quad [7.3]$$

Secondly, linear regression relationships of each band are obtained between the satellite R_w and the uncertainty (Difference and Unbiased Difference) using the in situ and satellite R_w matchups (Figure 27 and Figure 28 below).

Finally, the linear models of Difference and Unbiased Difference are then applied to each pixel. The Relative Uncertainty and the Relative Unbiased Uncertainty of each pixel can be calculated as follows:

Relative Uncertainty (RU)

This metric is expressed as

$$RU\ (\%) = (\ Diff_{pix}\ /\ R_{pix}\ )*100\% \qquad [7.4]$$

Where $Diff_{pix}$ is the retrieved difference of each pixel using the models in Figure 27 and R_{pix} is the corresponding R_w on the satellite image.

Relative Unbiased Uncertainty (RUU)

$$RUU\ (\%) = (\ UD_{pix}\ /\ R_{pix}\ )*100\% \qquad [7.5]$$

Where UD_{pix} is the retrieved Unbiased Difference of each pixel using the models on the Figure 28.

Please note that the RU and RUU calculated for the LWLR are both not absolute, in order to show the systematically under-estimation of LWLR (Figure 27, Figure 28) due to the challenges that are faced with accurately performing atmospheric correction in optically complex inland waters. The application boundary constraints of uncertainty models for each waveband are determined based on the matchup dataset. If any waveband has no associated uncertainty model (e.g. in situ data were too sparse) or the LWLR is out of the application range, the pixel will be flagged as having unknown uncertainty in the uncertainty product.

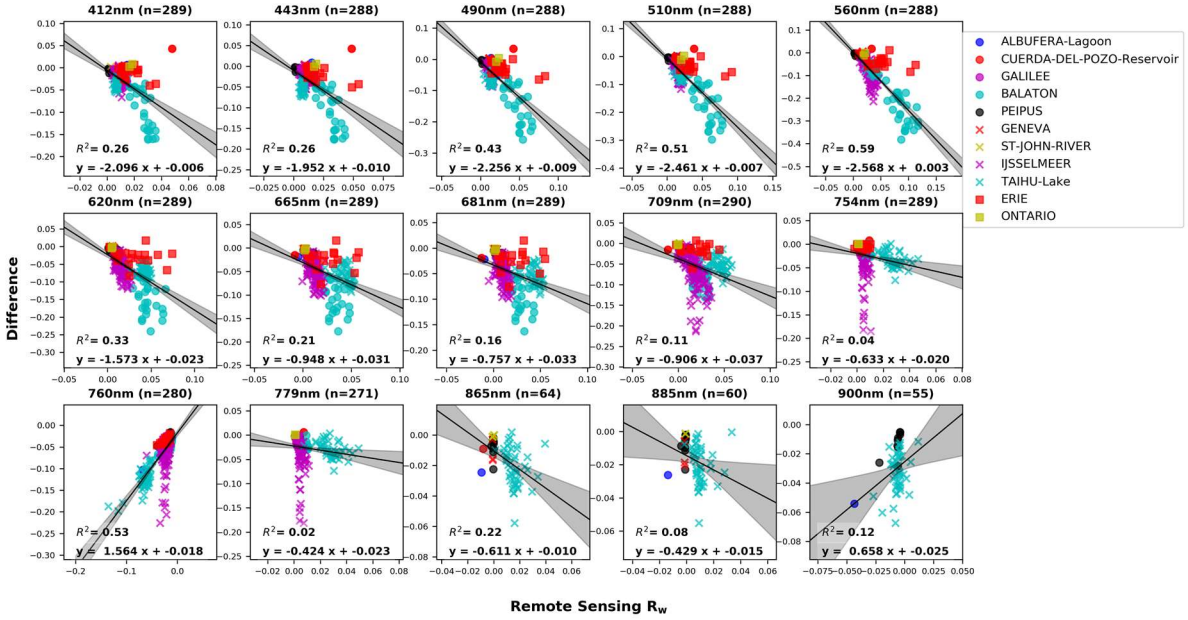

Figure 27 Linear correlation between Remote sensing Rw and the Difference between in situ and remote sensing Rw matchups (among which the wavebands of 865, 885 and 900 nm are not included in the Rw uncertainty products because of the sparse matchups).

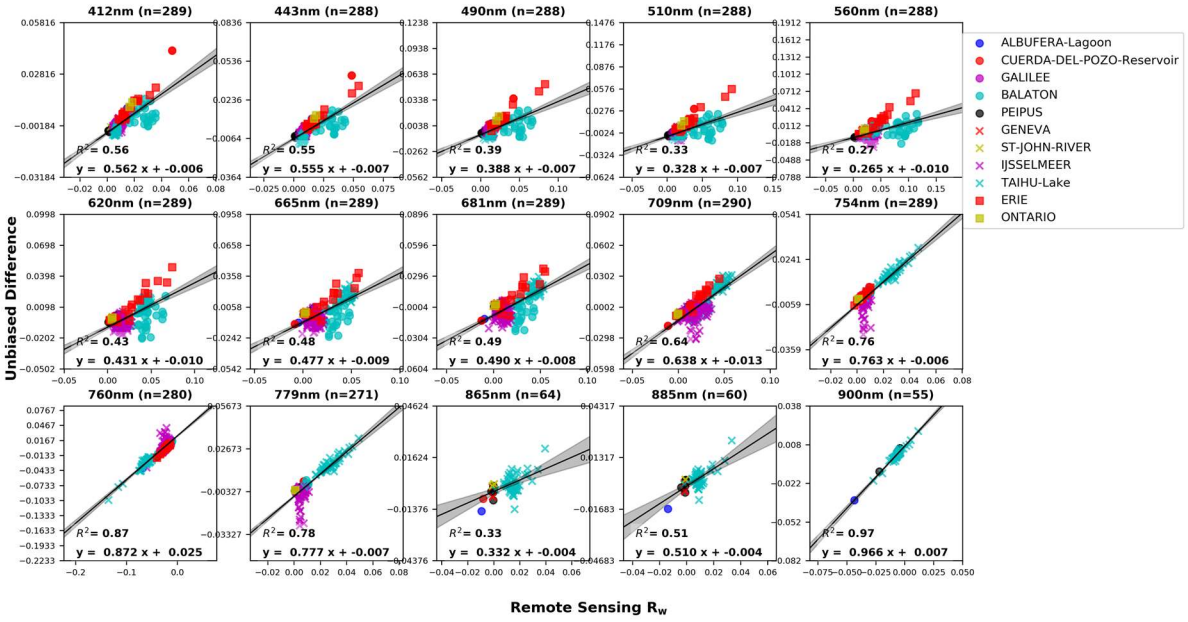

Figure 28 Linear correlation between Remote sensing Rw and the Unbiased Difference between in situ and remote sensing Rw matchups (among which the wavebands of 865, 885 and 900 nm are not included in the Rw uncertainty products because of the sparse matchups).

The OWTs are not taken into account during validation of LWLR due to the typically low number of in situ reference measurements that are available. The OWTs are, however, considered in the validation of algorithms for chlorophyll-a and turbidity since these vary considerably in their suitability to describe water column properties of the different water types.

The uncertainty of the algorithms for chlorophyll-a and turbidity is provided for each of the statistical metrics listed above. Each algorithm is evaluated against the full matchup data set available for the observation period of the satellite sensor. This includes samples with a value considered outside of the applicable range of the individual algorithm. This is done because OWT membership is a fuzzy property (one observation belongs to multiple classes, with varying degrees of similarity) so that a clear separation of in situ data and a subset of algorithms cannot be made objectively. This approach also ensures that the uncertainty model captures the reduction of uncertainty with increasing OWT class membership, which is ultimately how the per-pixel uncertainty is generated. Finally, this has the added benefit that the number of data points on which the analysis is based is higher than if any arbitrary thresholds were used, which results in a more robust statistical model of product uncertainty.

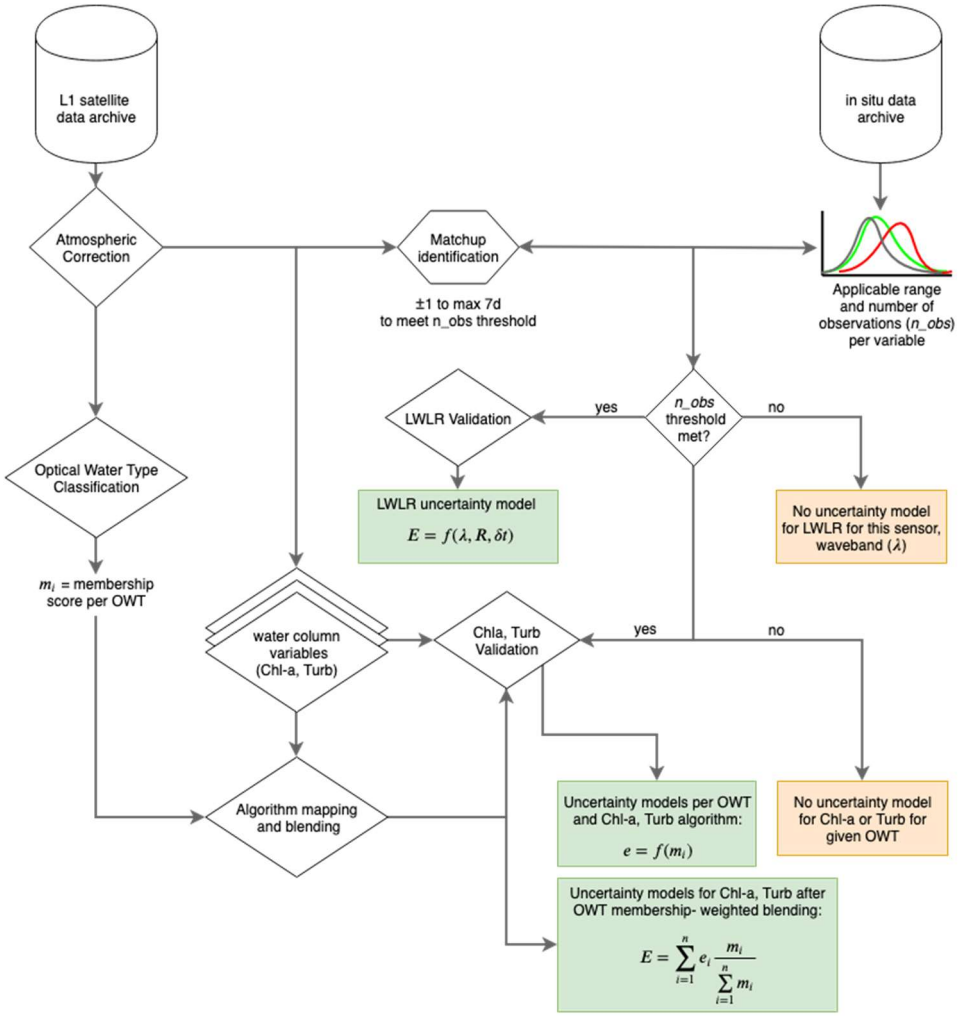

Figure 29 Flow chart of the end-to-end validation of LWLR products against in situ observations, resulting in uncertainty models.

As shown in Figure 29, the uncertainty model per OWT and chlorophyll-a or turbidity algorithm is expressed as a function of OWT class membership. An algorithm that is suitable for a given OWT and water column property is expected to show a linear response over its applicable range. The uncertainty metric used for the chlorophyll-a or turbidity algorithm is absolute relative uncertainty (ARU, %), which is defined using the matchup dataset as follows:

$$ARU\ (\%) = (|x_d - x_m| \ / \ |x_m|) * 100\% \qquad [7.6]$$

where x_d is the satellite derived value from the chlorophyll-a/turbidity algorithms and x_m is the in situ measured chlorophyll-/turbidity.

In the Lakes_CCI processor, the top-3 ranking OWT scores and their corresponding algorithm results are included in the algorithm blending procedure, which provides a weighted average of the algorithm results corresponding to those 3 OWTs. An example of the relationship between an ARU metric and OWT class membership is given in Figure 30 which shows the weighted chlorophyll-a by the top 3 OWT class membership in relation to the membership of OWT 9 using matchup data for the entire MERIS observation period. This shows a robust and relatively low ARU over a wide range of class membership values, indicating a wide application range of this scheme. There is a slight skew of the RU towards lower values which increases with class membership score of OWT 9, suggesting that the blended algorithm is optically suited for this OWT. A linear regression fit of this relationship provides the ARU uncertainty *e* of this algorithm-OWT combination as

$$e_9 = am_9 + b \qquad [7.7]$$

where *a* and *b* are the slope and intercept of the linear fit and subscript (4) is the OWT class.

Finally, to compute chlorophyll-a product uncertainty for each satellite observation, the mapping of different algorithms corresponding to OWT membership scores are taken into account. The uncertainties associated with each of the selected algorithms are weighted in the same fashion, taking the sum of each of the top-3 OWT uncertainties multiplied by membership score for that OWT relative to the sum of membership scores of the number (*n*=3) of classes considered, yielding the per-pixel uncertainty *E*:

$$E = \sum_{i=1}^{n} e_i \frac{m_i}{\sum_{i=1}^{n} m_i} \qquad [7.8]$$

Upper and lower boundary of the OWT membership score for the application of each OWT uncertainty model is determined based on the matchup dataset. If the OWT membership score is out of the application range in the uncertainty calculation procedure, then the pixel will be flagged as having unknown product uncertainty.

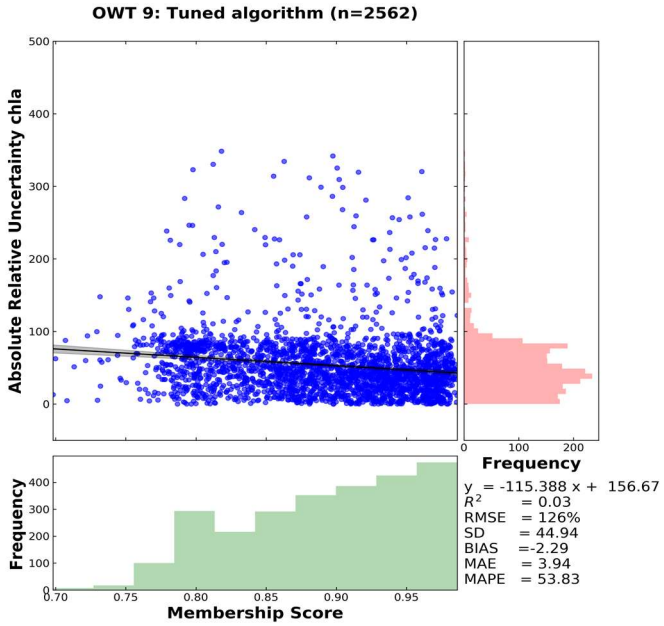

Figure 30 Correlation of Absolute Relative Uncertainty (ARU) of the top-3 weighted chlorophyll-a by the OWT membership score in relation to the membership score for optical water type 9. Reproduced from Liu et al. (in prep).

7.4. End to end uncertainty budget in LWLR

Applying the uncertainty models to MERIS observations result in uncertainty maps such as those shown in Figure 31. Maps are shown for each of the uncertainty products provided in the CDRPv1.0 over lakes Vänern and Vättern in Sweden on 26th July 2006. All uncertainties are relative product uncertainty, with Turbidity and chlorophyll-*a* relative uncertainty (RU) as their absolute values. The uncertainty products for LWLR are not absolute in order to show the systematic underestimation of reflectance. For the latter, both relative uncertainty (RU) and relative unbiased uncertainty (RUU) show negative uncertainties in the two lakes, which illustrates 'over-correction' for atmospheric effects (Figure 31a and b). Positive uncertainty values for Rw665 are observed near land, as a result of the land-adjacency effect. In terms of turbidity, patches with high uncertainties are observed both in open water and near land (Figure 31c). For chlorophyll-a, generally lower uncertainties are found

in lake Vänern (in the north-west) compared to lake Vättern (south-east), as seen in Figure 31d. This is consistent with smaller uncertainties in $R_w(665)$ in lake Vänern.

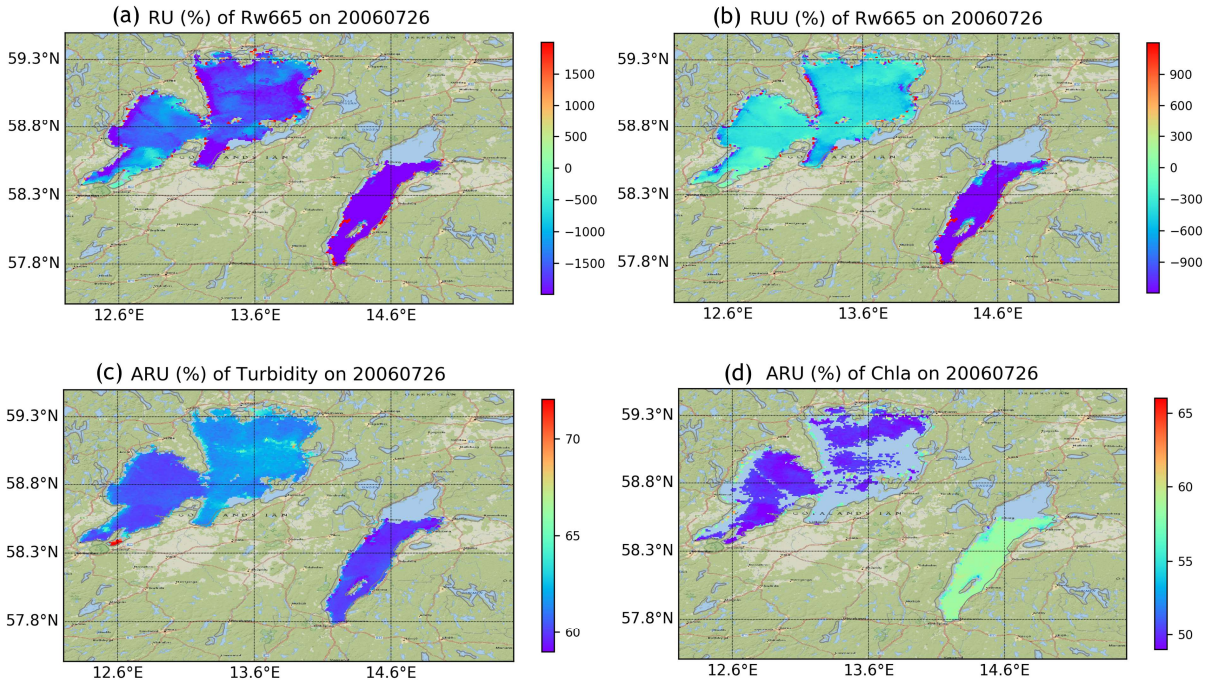

Figure 31 Per-pixel uncertainty products. (a) Relative uncertainty of $R_w(665)$, (b) Relative unbiased uncertainty of $R_w(665)$, (c) Absolute relative uncertainty of Turbidity and (d) Absolute relative uncertainty of chlorophyll-*a*.

8. References

Angal, A., Xiong, X., Sun, J. and Geng, X. (2015) On-orbit noise characterization of MODIS reflective solar bands. Journal of Applied Remote Sensing 9(1), 094092. https://doi.org/10.1117/1.JRS.9.094092

Arsen, A., JF. Cretaux, and R. Abarca-Del-Rio. 2015. Use of SARAL/AltiKa over mountainous lakes, intercomparison with Envisat mission *J. of Adv. Space Res*. The Saral/ALtiKa satellite Altimetry Mission, 38, 534-548,2015, DOI: 10.1080/01490419.2014.1002590

Bulgin, C. E., Embury, O. and Merchant, C. J. (2016) Sampling uncertainty in gridded sea surface temperature products and Advanced Very High Resolution Radiometer (AVHRR) Global Area Coverage (GAC) data. Remote Sensing of Environment, 117. pp. 287-294. ISSN 0034-4257 doi:10.1016/j.rse.2016.02.021

Cox, C., Munk, W., 1954a. Measurement of the roughness of the sea surface from photographs of the Sun's glitter. J. Opt. Soc. Am. 44 (11), 838–850. http://dx.doi.org/10.1364/JOSA.44.000838.

Cox, C., Munk, W.H., 1954b. Statistics of the sea surface derived from sun glitter. J. Mar. Res. 13 (2), 198–227.

Crétaux J-F., Bergé-Nguyen M., Calmant S., Romanovski V.V., Meyssignac B., Perosanz F., Tashbaeva S., Arsen A., Fund F., Martignago N., Bonnefond P., Laurain O., Morrow R., Maisongrande P., 2013 Calibration of envisat radar altimeter over Lake Issykkul, J. Adv. Space Res., Vol 51, 8, 1523-1541, doi: 10.1016/j.asr.2012.06.039

Crétaux, J.F., S. Calmant, V. Romanovski, et al. 2009. An absolute calibration site for radar altimeters in the continental domain: lake Issykkul in Central Asia, *Journal of Geodesy* 83 (8) 723-735 DOI: 10.1007/s00190-008-0289-7

Crétaux, J.F., S. Calmant, V. Romanovski, et al. 2011 Absolute Calibration of Jason radar altimeters from GPS kinematic campaigns over Lake Issykkul, *Marine Geodesy*, 34 : 3-4,291-318, DOI: 10.1080/01490419.2011.585110

Feyisa, G.L., Meilby, H., Fensholt, R., Proud, S.R., 2014. Automated Water Extraction Index: A new technique for surface water mapping using Landsat imagery. Remote Sens. Environ. 140, 23–35. doi:10.1016/j.rse.2013.08.029

Harmel, T., Chami, M., Tormos, T., Reynaud, N., & Danis, P.-A. (2018). Sunglint correction of the Multi-Spectral Instrument (MSI)-SENTINEL-2 imagery over inland and sea waters from SWIR bands. Remote Sensing of Environment, 204, 308–321. doi:10.1016/j.rse.2017.10.022

J-F Crétaux, M. Bergé-Nguyen, S. Calmant, N. Jamangulova, R. Satylkanov, F. Lyard, F. Perosanz, J. Verron, A.S. Montazem, G. Leguilcher, D. Leroux, J. Barrie, P. Maisongrande and P. Bonnefond, 2018, Absolute calibration / validation of the altimeters on Sentinel-3A and Jason-3 over the lake Issykkul, Remote sensing, 10, 1679,; doi:10.3390/rs10111679

Lyons E. A., Sheng Y., C. Smith L.C., Li J., Hinkel K.M. , Lenters J.D. & Wang J.(2013) Quantifying sources of error in multitemporal multisensor lake mapping, International Journal of Remote Sensing, 34:22, 7887-7905

Merchant CJ, Paul F, Popp T, Ablain M, Bontemps S, Defourny P, et al. Uncertainty information in climate data records from Earth observation. Earth Syst Sci Data. 2017;9(2):511-27.

Merchant, C. J., Holl, G., Mittaz, J. P. D. and Woolliams, E. R. (2019) Radiance uncertainty characterisation to facilitate climate data record creation. Remote Sensing, 11 (5). 474. ISSN 2072-4292 doi: https://doi.org/10.3390/rs11050474

Mittaz, J., Merchant, C. J. and Woolliams, E. R. (2019) Applying principles of metrology to historical Earth observations from satellites. Metrologia, 56 (3). ISSN 0026-1394 doi: https://doi.org/10.1088/1681-7575/ab1705

Mortimer, C. A. and Sharp, M. (2018) Spatiotemporal variability of Canadian High Arctic glacier surface albedo from MODIS data, 2001–2016. The Cryosphere, 12, 701–720, https://doi.org/10.5194/tc-12-701-2018.

Pekel, J.F., Cottam, A., Gorelick, N., Belward, A.S., 2016. High-resolution mapping of global surface water and its long-term changes. Nature 540, 418–422. doi:10.1038/nature20584

Ričko M., C.M. Birkett, J.A. Carton, and J-F. Cretaux, Intercomparison and validation of continental water level products derived from satellite radar altimetry, *J. of Applied Rem. Sensing*, Volume 6, Art N°: 061710, DOI: 10.1117/1.JRS.6.061710, 2012.

Riggs, G. A. and Hall, D. K. (2015) MODIS snow products Collection 6 user guide. 66 pp.

Sun, J., Xiong, X., Angal, A., Chen, H., Wu, A., and Geng, X. (2014) Time-dependent response versus scan angle for MODIS reflective solar bands. IEEE Transactions on Geoscience and Remote Sensing, vol. 52, no. 6, pp. 3159-3174, doi: 10.1109/TGRS.2013.2271448

Vermote, E. F. and El Saleous, N.Z. (2006) Operational atmospheric correction of MODIS visible to middle land surface data in the case of an infinite lambertian target. Earth Science Satellite Remote Sensing Book, Volume 1, Chapter 8, p.123-153.

Vermote, E. F. and Vermeulen, A. (1999) Atmospheric correction algorithm: spectral reflectances (MOD09). NASA contract NAS5-96062. 107 pp.

Vermote, E. F., Roger, J. C., and Ray, J. P. (2015) MODIS surface reflectance user's guide collection 6. Maryland: MODIS Land Surface Reflectance Science Computing Facility. 35 pp.

Vickers, H., Malnes, E., & Høgda, K. A. 2019. Long-Term Water Surface Area Monitoring and Derived Water Level Using Synthetic Aperture Radar (SAR) at Altevatn, a Medium-Sized Arctic Lake. *Remote Sensing*, *11*(23), 2780.

Xu, H., 2006. Modification of normalised difference water index (NDWI) to enhance open water features in remotely sensed imagery. Int. J. Remote Sens. 27, 3025–3033. doi:10.1080/01431160600589179