Lakes_CCI+ - Phase 2
D2.3. End-to-End ECV Uncertainty Budget (E3UB)

11 rue Hermès, Parc Technologique du Canal — 31520 Ramonville Saint-Agne, France
Tel +33 (0)5 61 39 47 00 Fax +33(0)5 61 75 10 14
www.cls.fr

CHRONOLOGY ISSUES

Issue	Date	Object	Written by
0.1	19/03/2020	Initial submission	S. Simis et al.
1.0	24/04/2020	Update following ESA review	S. Simis et al.
1.1	17/03/2021	LWL/LWE updates LIC updates	J.F. Crétaux C. Duguay
1.2	29/04/2021	Updates following ESA review	J.F. Crétaux C. Duguay
2.0	02/05/2022	Accompanying CRDPv2.0	S. Simis, L. Carrea, B. Calmettes, J-F Crétaux, C. Duguay, X. Liu, H. Yésou
2.1	12/09/2023	Updates accompanying CRDPv2.1. Addition of Lake Ice Thickness.	S. Simis, L. Carrea, B. Calmettes, J-F Crétaux, C. Duguay, X. Liu, A. Mangili, H. Yésou
2.1.1	17/10/2023	Revision after ESA review	S. Simis, L. Carrea, B. Calmettes, J-F Crétaux, C. Duguay, X. Liu, A. Mangili, H. Yésou

Checked by	Stefan Simis – PML J-F Cretaux - LEGOS	*Stefan Simis* *Jean-francois Cretaux*
Approved by	Alice Andral - CLS	*A. Andral*
Authorized by	Clément Albergel - ESA	*Clement Albergel*

DISTRIBUTION

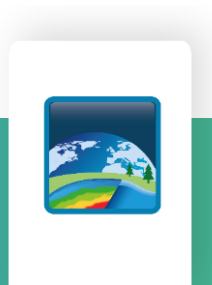| Company	Names	Email
ESA	Clément Albergel	clement.albergel@esa.int
BC	Carsten Brockman	carsten.brockmann@brockmann-consult.de
BC	Dagmar Müller	dagmar.mueller@brockmann-consult.de
BC	Jorrit Scholze	jorrit.scholze@brockmann-consult.de
BC	Kerstin Stelzer	kerstin.stelzer@brockmann-consult.de
BC	Martin Boettcher	martin.boettcher@brockmann-consult.de
BC	Olaf Danne	olaf.danne@brockmann-consult.de
CLS	Alice Andral	aandral@groupcls.com
CLS	Anna Mangilli	amangilli@groupcls.com
CLS	Beatriz Calmettes	bcalmettes@groupcls.com
CLS	Nicolas Taburet	ntaburet@groupcls.com
CLS	Christophe Fatras	cfatras@groupcls.com
CLS	Pierre Thibault	pthibaut@groupcls.com
CNR	Claudia Giardino	giardino.c@irea.cnr.it
CNR	Mariano Bresciani	bresciani.m@irea.cnr.it
CNR	Monica Pinardi	pinardi.m@irea.cnr.it
CNR	Marina Amadori	amadori.m@irea.cnr.it
CNR	Rossana Caroni	caroni.r@irea.cnr.it
CNR	Giulio Tellina	tellina.g@irea.cnr.it
H2O Geo	Claude Duguay	claude.duguay@h2ogeomatics.com
H2O Geo	Yuhao Wu	mark.wu@h2ogeomatics.com
H2O Geo	Jaya Sree Mugunthan	jayasree.mugunthan@h2ogeomatics.com
H2O Geo	Justin Murfitt	justin.murfitt@h2ogeomatics.com
HYGEOS	François Steinmetz	fs@hygeos.com
LEGOS	Jean-François Cretaux	jean-francois.cretaux@cnes.fr
LEGOS	Paul-Gérard Gbektom	paul.gerard.gbetkom@legos.obs-mip.fr
PML	Stefan Simis	stsi@pml.ac.uk
PML	Xiaohan Liu	liux@pml.ac.uk
PML	Nick Selmes	nse@pml.ac.uk
PML	Mark Warren	mark1@pml.ac.uk
Sertit	Hervé Yésou	Herve.yesou@unistra.fr
Sertit	Jérôme Maxant	maxant@unistra.fr
Sertit	Sabrina Amsil	s.amzil@unistra.fr
Sertit	Rémi Braun	remi.braun@unistra.fr
UoR	Chris Merchant	c.j.merchant@reading.ac.uk
Bangor	Iestyn Woolway	iestyn.woolway@bangor.ac.uk
UoR	Laura Carrea	l.carrea@reading.ac.uk
UoS	Dalin Jiang	dalin.jiang@stir.ac.uk
UoS	Evangelos Spyrakos	evangelos.spyrakos@stir.ac.uk
UoS	Ian Jones	ian.jones@stir.ac.uk

LIST OF CONTENTS

LIST OF TABLES AND FIGURES

1 Introduction

The End-to-End ECV Uncertainty Budget (E3UB) identifies and, where possible, characterises each source of product uncertainty, quantifying or estimating its effect on the final retrieved quantity. Sources of uncertainty that are considered include, where applicable:

- The input measurements, including radiometric noise and calibration errors in Level-1 data.
- Retrieval errors and uncertainties, such as errors in auxiliary data (aerosol optical properties, spectroscopic errors, modelled atmospheric profiles, approximated surface characteristics, assumed parameters), and uncertainties in the retrieval model (e.g., radiative transfer approximations or idealised physics).
- Factors in the retrieval process, such as cloud screening failures or representativity errors such as biases in the sampling (e.g., daytime only, clear sky only).
- Uncertainties introduced during transformations between each processing data level.
- Sources of uncertainty that may particularly affect the long-term stability of the CDR (e.g., changes over time in instruments, their capabilities or calibration, diurnal sampling, the effectiveness of cloud clearing, the quality of auxiliary information).

Where possible, the uncertainties are characterised to express the magnitude of the uncertainty at its source, and how these uncertainties propagate into the final retrieved quantity (i.e., their final magnitude). This may make use of estimates or correlations of error between separate observations made at different points in space and time. Biases in the uncertainty distribution and how these effect the final product, are also considered where feasible.

Where it is not feasible to quantitatively estimate each of these aspects, the characteristics of the uncertainty and its likely effect on the final product are described in a qualitative sense. Particular attention should be given to the characterisation of biases which might more severely affect a given environment or season, in the context of climate studies.

Finally, the Lakes_cci E3UB includes recommendations to data users, describing how to employ the provided prognostic uncertainties in their work (e.g., how to treat spatially correlated uncertainties when regridding to different Level-3 resolutions), and advice on conditions leading to low confidence in the data product or its uncertainty estimates (e.g., poor cloud clearing over snow and ice surfaces, poor signal under conditions of low illumination). Recommendations are also made to internal or external validation teams, concerning aspects that may require special attention when the prognostic uncertainty estimates are validated.

It should be noted that, for most of the steps considered above, reference models and data are of crucial importance. For a large part of the Lakes ECV, reference data are known to be scarce, particularly concerning the length of the observation time series considered. In addition, the available reference data rarely consider several of the Lakes ECV product quantities in the same location. Uncertainty estimates are therefore established using the approach that is considered most suitable for each of the Lakes ECV products, as described in the following sections.

2 Lake Water level - LWL

2.1 Uncertainty characteristics

The uncertainties provided with the LWL product result from the standard deviation of the along track altimetry (after including all observation 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:

$$LWL = Alt - R_{corr} - T_E \qquad [2.1]$$

Where LWL 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. Detailed information and 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 the next section.

To quantify the end-to-end uncertainty 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 now available for different combinations of lakes and altimeters.

2.2 Sources of uncertainty

The sources of uncertainty of LWL product vary and depend on several factors. Total 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 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 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. 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 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.
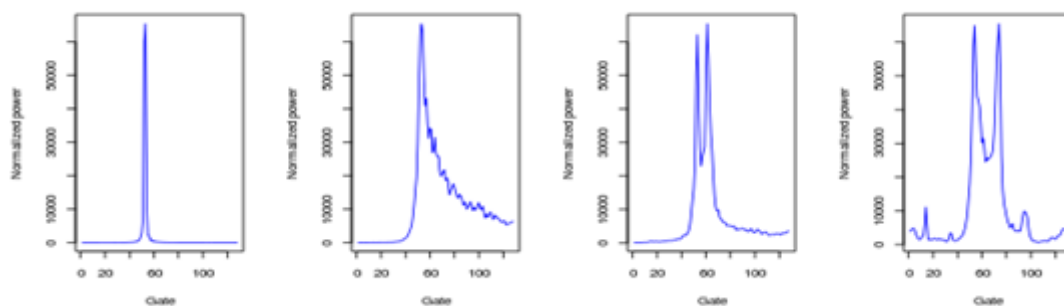
Figure 1: example of typical waveforms, from left to right; specular, ocean like (Brown model), contaminated, and examples of noisy waveforms.

With 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 km^2) 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, in a potentially 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.

Geophysical corrections: Calculating the range is based on the propagation velocity of the electromagnetic wave, i.e., 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

A fourth source of uncertainty to consider is due to reflection on the surface of the water. A correction of the surface state would consider 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 considered 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 (wet tropospheric) 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, 2013, 2018; Quartly et al., 2020) 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.

2.3 End-to-end Uncertainty budget

It is currently not feasible 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, report generic estimates.

For large lakes and for 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 (Ričko et al., 2012).

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 on all 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 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 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 are 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 differing 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.

2.4 Recommendations

2.4.1 To users

Lake Water Level uncertainty, expressed in cm, provides information on the quality of the Lake water level measurements. Based on the uncertainty values, the quality flag of the lake water level values can be classified in three categories:

- Good quality: when the uncertainty is lower than 10 cm
- Medium quality: for LWL with uncertainty between 10 and 30 cm
- Low quality: when the uncertainty is greater than 30 cm

To ensure appropriate use of the data, it is highly recommended that users either apply one of the quality flags to suit their needs or include the uncertainty estimates in their analysis or reporting.

2.4.2 To the validation team

The lake water level is provided with respect to the geoid. For validation purposes, it is important to keep in mind that data coming for external datasets can be referenced to different geoid models which results in a bias between timeseries in the comparison process. Therefore, comparison of water level variation is generally more appropriate to assess product quality.

.

2.5 References

Birkett C. M. (1995). Contribution of TOPEX/POSEIDON to the global monitoring of climatically sensitive lakes, Journal of Geophysical Research, 100, C12, 25, 179-25, 204.

Boy F., Cretaux J-F., Boussaroque M., & Tison C. (2022). Improving Sentinel-3 SAR mode procesisng over lake using numérical simulations, IEEE Transactions on Geoscience and Remote Sensing, 60, 1-18, Art no. 5220518, doi: 10.1109/TGRS.2021.3137034

Brown G. S. (1977). The Average Impulse Response of a rough surface and its applications, IEEE Trans. Antennas Propag, Vol. 25, pp. 67-74. doi: 10.1109/TAP.1977.1141536

Crétaux J-F. and Birkett C. M. (2006). Lake studies from satellite altimetry, C R Geoscience, doi: 10.1016/J.crte.2006.08.002

Crétaux J-F. Calmant S. Romanovski V. Perosanz F. Tashbaeva S. Bonnefond P. Moreira D. Shum C. K. Nino F. Bergé-Nguyen M. Fleury S. Gegout P. Abarca Del Rio R. Maisongrande P. (2011). Absolute Calibration of Jason radar altimeters from GPS kinematic campaigns over Lake Issykkul, Marine Geodesy, 34: 3-4, 291-318, doi: 10.080/01490419.2011.585110

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., 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, 2013

Crétaux J-F, Abarca Del Rio R. Berge-Nguyen M. Arsen A. Drolon V. Clos G. Maisongrande P. (2016). Lake volume monitoring from Space, Survey in geophysics, 37: 269-305. doi: 10.1007/s10712-016-9362-6

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

J-F Cretaux, 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

Quartly, G.D.; Nencioli, F.; Raynal, M.; Bonnefond, P.; Nilo Garcia, P.; Garcia-Mondéjar, A.; Flores de la Cruz, A.; Crétaux, J.-F.; Taburet, N.; Frery, M.-L.; Cancet, M.; Muir, A.; Brockley, D.; McMillan, M.; Abdalla, S.; Fleury, S.; Cadier, E.; Gao, Q.; Escorihuela, M.J.; Roca, M.; Bergé-Nguyen, M.; Laurain, O.; Bruniquel, J.; Féménias, P.; Lucas, B. (2020). The Roles of the S3MPC: Monitoring, Validation and Evolution of Sentinel-3 Altimetry Observations. Remote Sens. 12, 1763. doi:10.3390/rs121111763

Ričko M. C. M. Birkett, J. A. Carton, and J-F. Cretaux. (2012). Intercomparison and validation of continental water level products derived from satellite radar altimetry, J. of Applied Rem. Sensing 6, Art N°: 061710.10.1117/1.JRS.6.061710

Taburet N., Zawadzki., Vayre M., Blumstein D., LE Gac S., Boy F., Raynal M., Labroue S., Cretaux J-F., and Femenias P. (2020). S3MPC: Improvement on inland water tracking and water level monitoring from the OLTC onboard Sentinel-3 altimeters, Remote Sensing 12, 18, AN 3055, doi: 10.3390/rs12183055

Yao F., Ben Livneh B., Rajagopalan B., Wang J., Crétaux J-F., Wada Y., Berge-Nguyen M. (2023). Satellites reveal widespread decline in global lake water storage, *Science* 380, 743-749.

14/53

Lakes_CCI+ - Phase 2   D2.3. End-to-End ECV Uncertainty Budget (E3UB)
Reference: CCI-LAKES2-0013-E3UB - Issue 2.1.1 – 12/09/2023
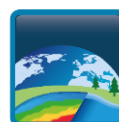
3 Lake water extent -LWE

3.1 Uncertainty characteristics

Optical satellite images are used in combination with LWL to produce LWE hypsometry curves. The LWE product is therefore the result of propagation of uncertainties to the Level-2 processed LWL product (section 2) and the LWE techniques described in this section.

3.2 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:

- – Observation effects
- – Algorithm effects (external sources)
- – Training samples selection
- – Calibration – detector effects
- – Misclassification effects
- – 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 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. The magnitude of the effect depends on the satellite (viewing angles, pixel size) and its orbit with respect to each lake.

For medium resolution satellites (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 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). Examples are shown in Figure 3.

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).
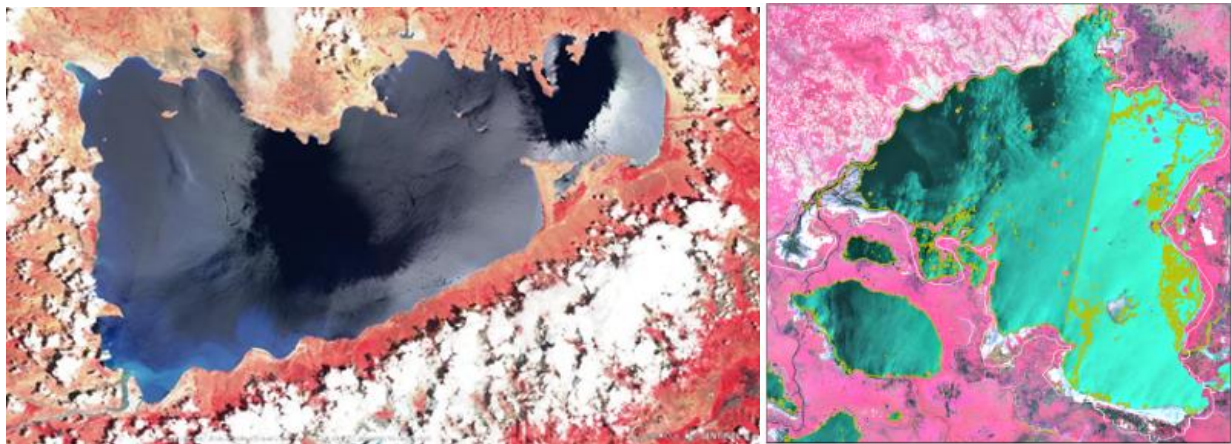
Figure 3: Sun glinted Sentinel-2 MSI image acquired on 14 July 2017 over Namtso Lake (PR China) and Kisale Lake (Congo) on 16/03/2020

Algorithm effects (external sources): 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, delineating the maximum water extent observed over a recent ten-year period. However, this does not always provide sufficient detail when looking at higher resolution imagery (Figure 4).
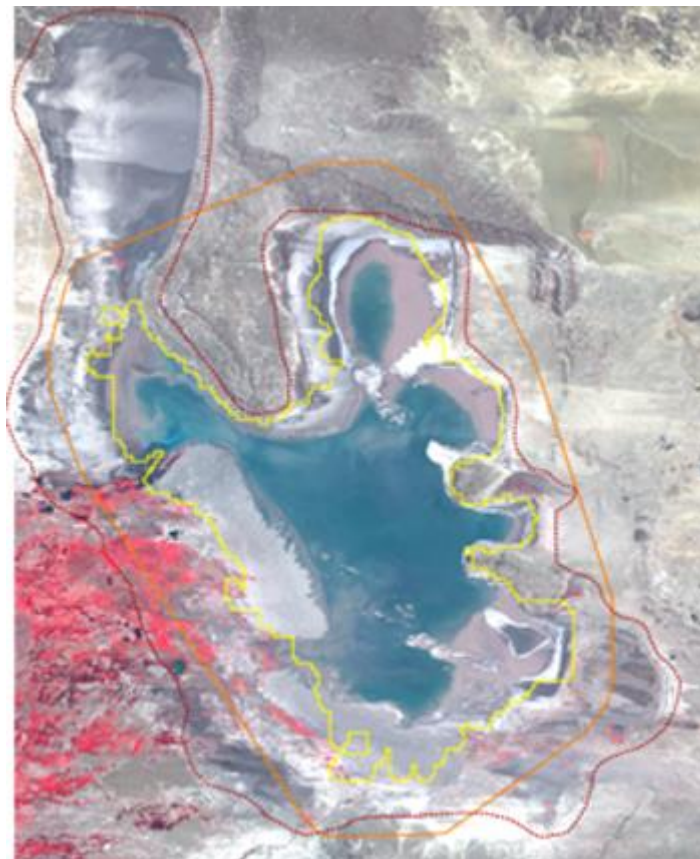
Figure 4: Reference envelope over Colhue Lake (Argentina). Yellow and orange (buffered) envelopes are derived from the ESA CCI Land Cover database and exclude the northern part of the lake. The red contour is based on high resolution satellite imagery.

In some cases, the reference polygons only include part of the lake system. An example is given in Figure 5 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. The same behaviour is observed for Tonle Sap lake, related to the difficulty to determine the extent of the lake with regard to the open water areas versus surrounding wetlands (Figure 6).

Figure 5: Reference envelopes over Sasykol lake (Kazakhstan)
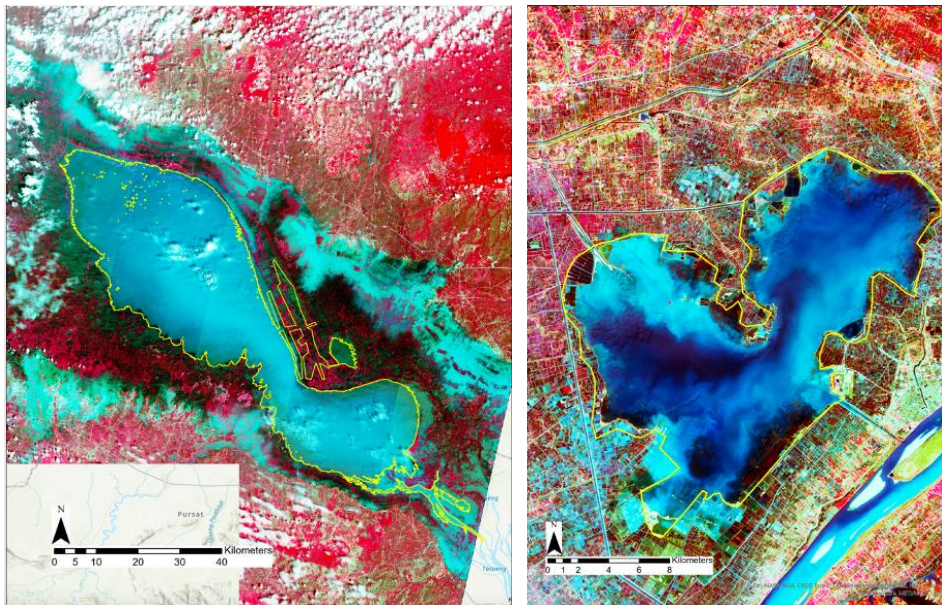
Figure 6: Difficulties to positioning lake limits in cases of complex lake such as Tonle Sap (Cambodia) where open water bodies are mixed with flooded wetlands or on case of Honk lake (China) with inundated paddies fields and wetlands.

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 effect to consider concerns the sun glint. Using a conventional water index (MNDWI, Xu 2006), sun glint reduces the detected water surface area. Other common water index like the AWEI, Feyisa et al. 2014, can be not affected by the sun glint.

Another example to illustrate the potential influence/bias induced by input data is given in Figure 7, showing the Global Surface Water product (GSW, 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.
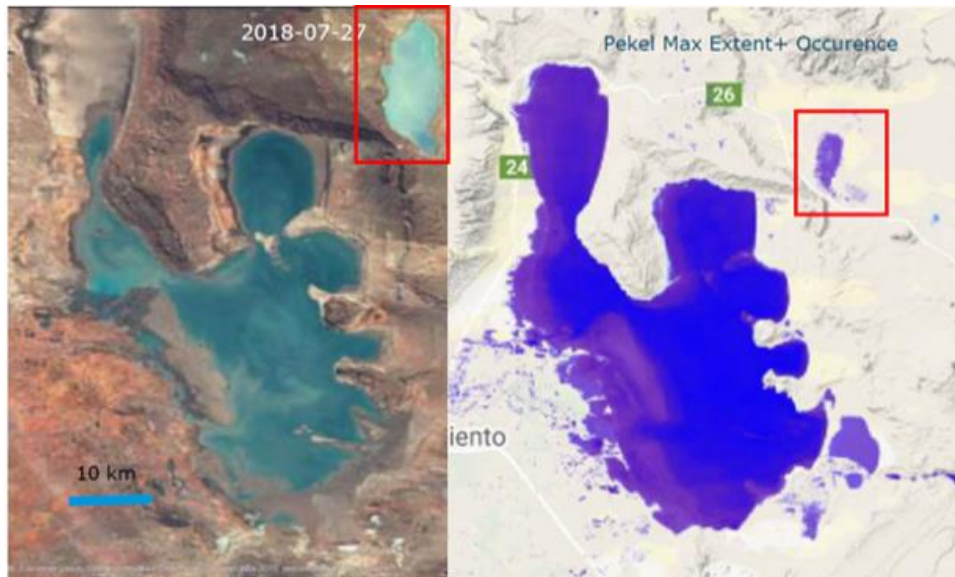
Figure 7: Lake Colhue (Argentina) shows water despite pixels having 0% water occurrence in the GSW data set.

Calibration – detector 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 8). These artefacts are rare for Sentinel-2 sensors.
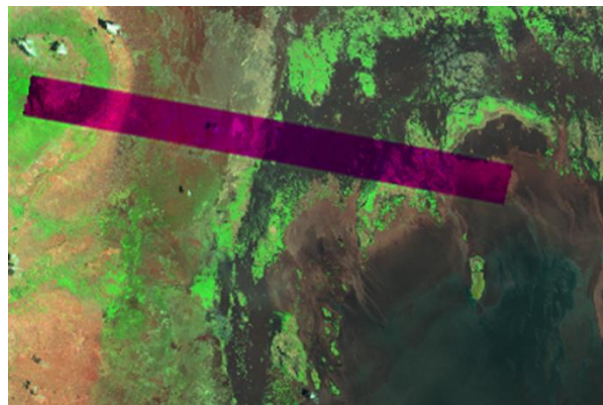
Figure 8: An example of a rare case of radiometric anomaly observed in Sentinel-2 MSI imagery of Lake Chilwa, 22 Nov 2018.

For the Landsat sensors, particularly Landsat 7, there are Scan Line errors due to a problem in forward motion of the satellite (Figure 9).
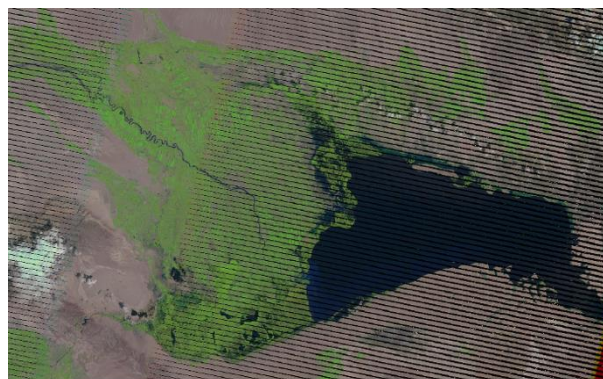
Figure 9: Detector errors in Landsat 7 acquired on the 2003-08-09 over Bosten lake

In some cases, the satellite coverage of a lake is incomplete (Figure 10) and the image is rejected, leaving a potential gap in the time series.

Figure 10: Incomplete lake coverage, case of the lake Naivaska, in Kenya, Landsat image acquired 23 January 2010.

Corrections applied to the products at various processing levels are particularly important with shallow water surrounded by wet mud banks. Two levels of processing of Sentinel-2 imagery are accessible for the LWE procedures:

- Level 1C, providing per-pixel radiometric measurements as Top-of-Atmosphere reflectance.
- Level 2A, atmospherically corrected Surface Reflectance (SR) integrating the atmospheric correction of Sentinel-2 MSI images, which includes correction for scattering by air molecules (Rayleigh scattering), absorbing and scattering effects of atmospheric gases, in particular ozone, oxygen and water vapour, and absorption and scattering due to aerosol particles.

In the past, Level 2A images were not produced systematically for all acquired Sentinel-2 tiles, so tests were carried out to quantify the effect induced by lack of atmospheric corrections on image processing, and the particular case of shallow water extraction. These tests were done over lakes of the Grand Est region in France (lac d'Orient and lac du Parc des Pourceaux), which have significant intra-annual water surface dynamics. At low water level period, the open water surfaces are surrounded by sandy/muddy wet banks. When exploiting Level 1C, the wet muddy banks are integrated with open water, whereas using Level 2A, the extracted surfaces correspond to open water, and fit well with a reference database (Figure 11).
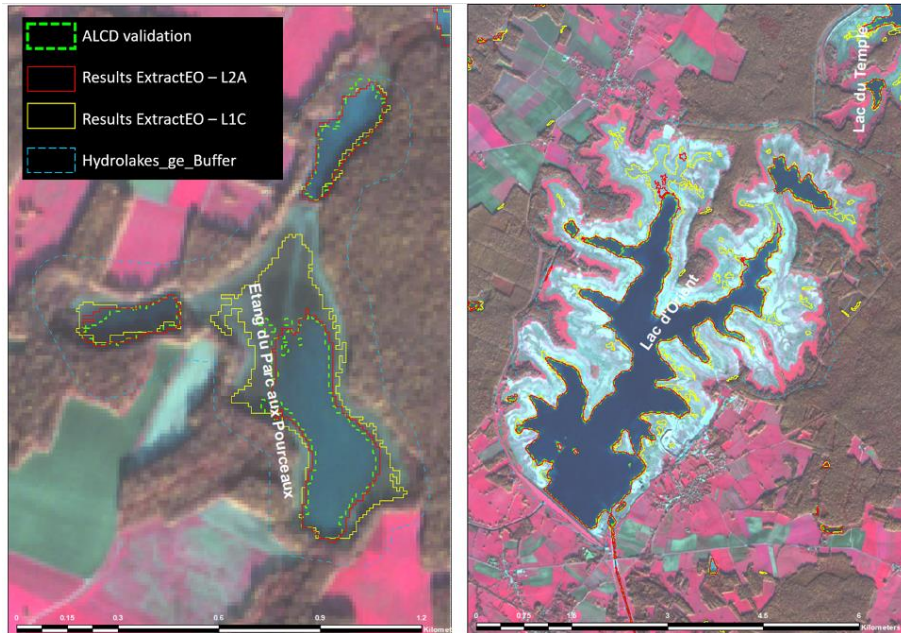
Figure 11: Evaluation of the effect of exploiting LC1 versus LC2 levels products on water recognition : (ALCD reference database provided by Pena Luque et al (2021)).

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 12). Better results were obtained over the Kasba and Baral lakes (Figure 13).
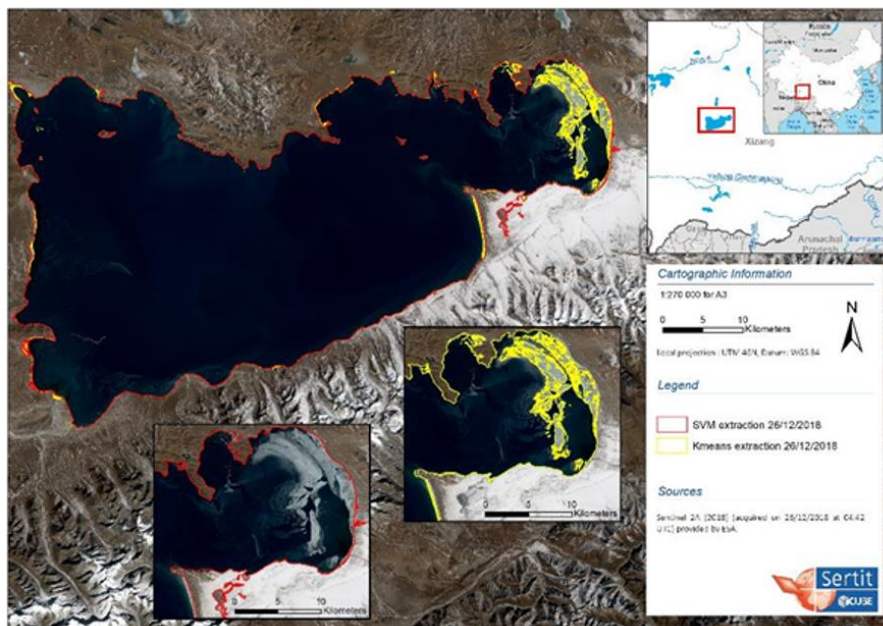
Figure 12: Confusion/commission between Ice and water when exploiting a KMeans approach rather than SVM at Namtso lake (PR China)
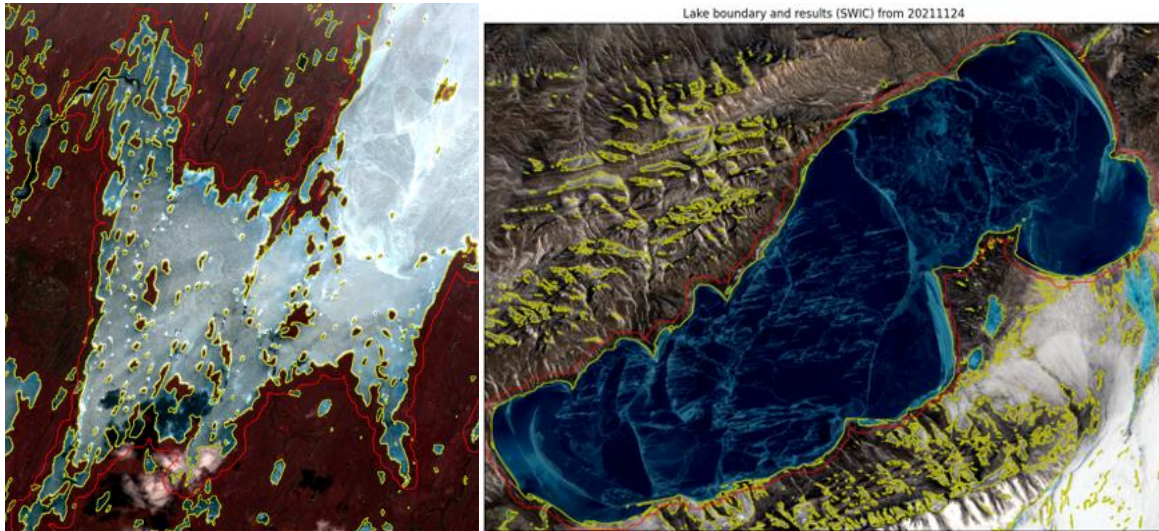
Figure 13: Potential limitations in water recognition purposes related to the ice coverage, over Kasba lake (Canada) and Baral lake (China), observed on Sentinel-2 images acquired respectively on 22-05-2022 and 4-11-2021).

Model parametrization: As presented in the ATBD, the water extraction is done using the ExtractEO processing chain which takes the slope derived from DEM (Maxant et al., 2022) into consideration. Slope and hillshade thresholds derived from HR DEMs are applied to refine the water extraction during post-processing) stages, to limit the effect commission related of the shadows due to the slope. The threshold values have to be adjusted to limit blurring effects leading to water omission (Figure 14).

Tile	Date	Product	resolution	pekel	NDWI thr.	hillshade thr.	sieve thr.	F1 score	Precision	Recall	MCC
			10	20	[0 ; 0.3]	45	20	0.781	0.658	0.960	0.794
UEP	2019.12.04	L1C	10	20	[0 ; 0.3]	60	20	0.816	0.726	0.930	0.821
			10	20	[0 ; 0.3]	70	20	0.810	0.747	0.885	0.812

Figure 14: Evaluation of the hillshade values of water recognition and impact on score's values (ALCD reference database from Pena Luque et al (2021).

Lyons 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 15.
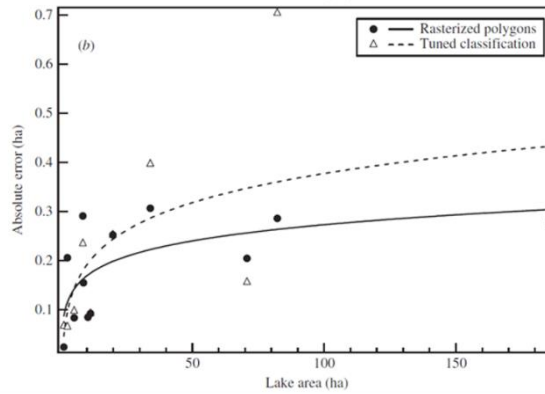
Figure 15: Scatter plots of absolute error in estimated lake area as a function of 'actual' lake area (based on DGPS shoreline measurements). Lines represent power-law fits to the respective data points.

3.3 End-to-end Uncertainty budget

The LWE measurement uncertainty cannot be determined at sufficient scale from in-situ observations. Instead, statistical analysis of the calculated hypsometry curves using each of the (LWL/LWE) vectors that were measured, is used to determine confidence between methods.

We systematically determine the precision of each time series of LWE calculating the RMSE of the difference between LWE inferred from the polynomial function of the hypsometry of the lake, and LWE directly measured by the satellite image, over the whole dataset of images used to calculate the hypsometry coefficient. We have observed that for all lakes where this statistical analysis was done (tens of targets, varying in size) the RMSE always remains between 1 and 10% of average extent of the lake, allowing to compute precise water extent variations (see Figure 16 and Figure 17).

The evaluation of LWE thematic ECV based on the hypsometry curve is a reliable method to examine the relative quality of the LWE calculated from satellite imagery. It does not resolve all inconsistencies, but it does validate good solutions and allows for the uncertainty of the calculation to be determined.
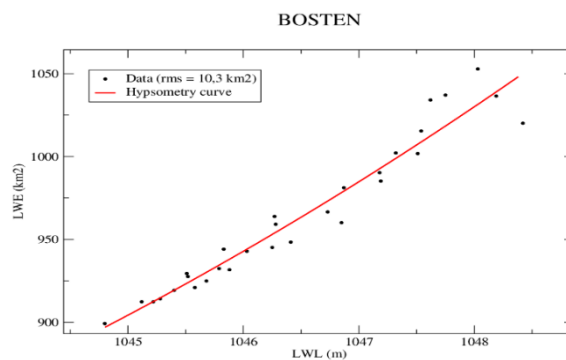
Figure 16:Hypsometric accuracy on Lake Bosten. Red curve represent the theorical hypsometry determined by least square adjustment of a second order polynomial on the datasets (LWL/LWE). With a 10.3 km² of RMSE this represents approximately 1% of total area of the lake.
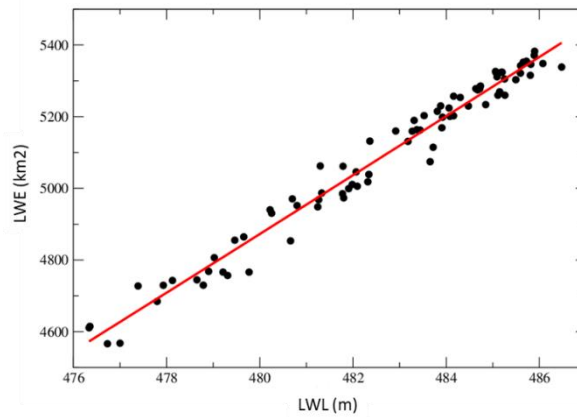
Figure 17: Hypsometric accuracy for Lake Kariba in Africa where the RMSE (37km²) is less than 1% of average extent of 5000 km².

3.4 Recommendations

3.4.1 To users

Lake Water Extent uncertainty, expressed in km², provides information on the quality of the Lake water extent measurements. This surface area expressed in the RMSE must be brought down to the overall surface area of the body in water, which is then obtained as a percentage. Based on these recomputed uncertainty values, the quality flag of the lake water extent values can be classified in three categories:

- Good quality: when the uncertainty is lower than 5 %
- Medium quality: for LWL with uncertainty between 5 to 10%
- Low quality: when the uncertainty is greater than 10%

To ensure appropriate use of the data, it is highly recommended that users either apply one of the quality flags to suit their needs or include the uncertainty estimates in their analysis or reporting.

3.4.2 To the validation team

Validating LWE is not an easy task. Many elements/parameters can disrupt the LWE/LWL relationships used to systematically generate LWE. Similarly, there is no totally reliable database. Firstly, unlike other LWEs, it is not possible to collect data in the field, as can be done for Lake temperature (LWT) or Lake Water Level (LWL) through field measurements or a boy or at gauge stations. Validation can often only be carried out by visual inspection of the resulting hypsometry curve and the calculation of the RMSE of the difference between theorical area (inferred from the hypsometry polynomial adjusted with the (LWL/LWE) vector) and the lake water area used for the calculation of the polynomial's coefficient. We may also use Higher resolution images (like the ones from Pleiades) but this can be done on very limited number of cases.

3.5 References

Harmel T. Malik Chami, Thierry Tormos, Nathalie Reynaud, Pierre-Alain Danis. (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, https://doi.org/10.1016/j.rse.2017.10.022.

Lyons, E.A.; Sheng, Y.; Smith, L.C.; Li, J.; Hinkel, K.M.; Lenters, J.D.; Wang, J. (2013). Quantifying Sources of Error in Multitemporal Multisensor Lake Mapping. Int. J. Remote Sens. 34: 7887–7905.

Maxant, J.; Braun, R.; Caspard, M.; Clandillon, S. (2022). ExtractEO, a Pipeline for Disaster Extent Mapping in the Context of Emergency Management. Remote Sens. 14: 5253. https://doi.org/10.3390/rs14205253

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

Peña-Luque, S.; Ferrant, S.; Cordeiro, M.C.R.; Ledauphin, T.; Maxant, J.; Martinez, J.-M. (2021). Sentinel-1&2 Multitemporal Water Surface Detection Accuracies, Evaluated at Regional and Reservoirs Level. Remote Sens 13, 3279. https://doi.org/10.3390/rs13163279

Tavares M.H. , Lins, R.G., Harmel, T. Fragoso Jr. C.R., Martínez, J-M., Motta-Marques, D. (2021). Atmospheric and sunglint correction for retrieving chlorophyll-a in a productive tropical estuarine-lagoon system using Sentinel-2 MSI imagery, ISPRS Journal of Photogrammetry and Remote Sensing 174, 215-236, https://doi.org/10.1016/j.isprsjprs.2021.01.021.

4 Lake surface water temperature – LSWT

4.1 Uncertainty characteristics

4.1.1 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 processing 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.

4.1.2 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.

4.2 Sources of uncertainty

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/√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 18 below, clouds may have fuzzy edges that make some degree of residual cloud contamination likely in some pixels used for LSWT.
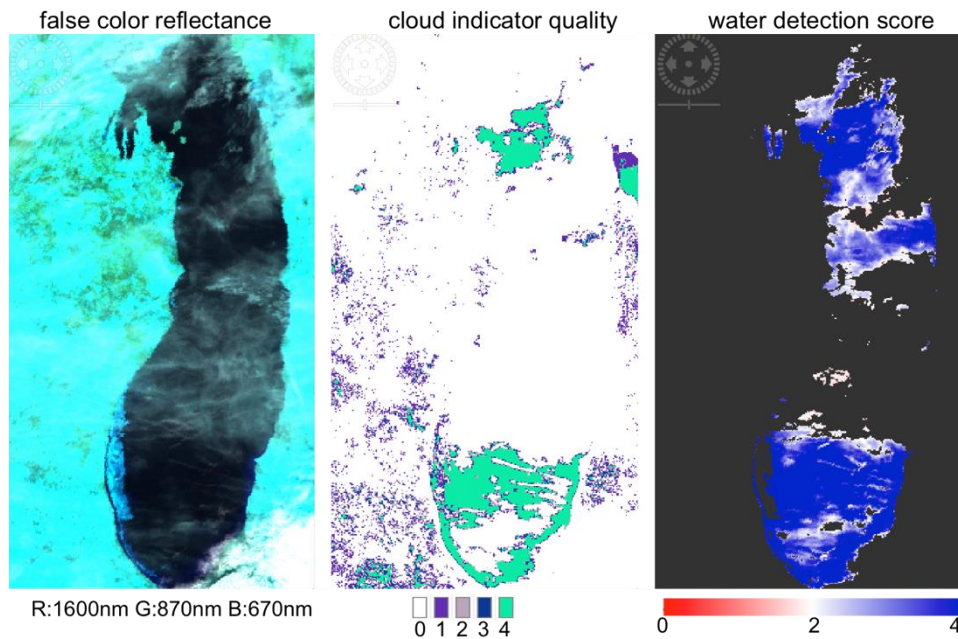
Figure 18: (Left) False colour composite image of reflectance at 1600 nm (R), 870 nm (G) and 670 nm (B). (Middle) the co-incident MERIS cloud indicator. (Right) The water detection score: QL 5 retrievals are made where this score reaches 4 or more. All panels for lake Michigan in USA on the 15-Feb-2011.

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

4.3 End to end Uncertainty budget

4.3.1 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):

$$\hat{x} = x_a + G(y - F(x_a)) \tag{4.1}$$

$$G = (K^T S_\varepsilon^{-1} K + S_a^{-1})^{-1} K^T S_\varepsilon^{-1} \tag{4.2}$$

The retrieved state is the prior state plus an increment of

$$G(y - F(x_a)) \tag{4.3}$$

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, $z(x) = \begin{bmatrix} x \\ w \end{bmatrix}$ 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 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:

$$\sqrt{GS_oG^T} \qquad\qquad\qquad [4.4]$$

and:

$$\sqrt{GS_mG^T} \qquad\qquad\qquad [4.5]$$

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.

4.3.2   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 19.
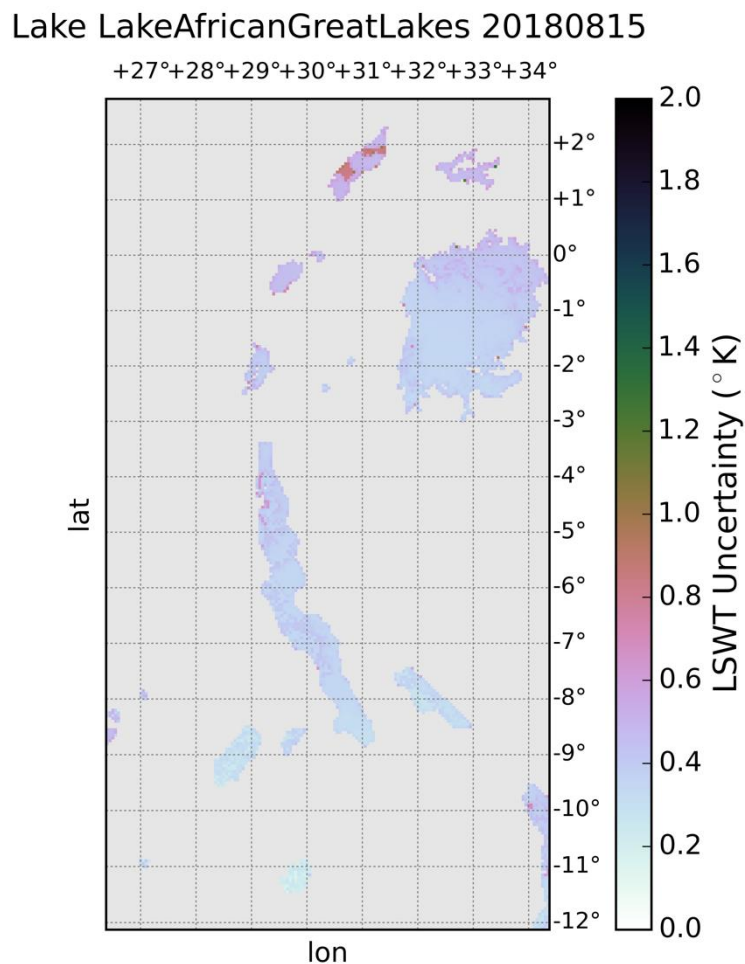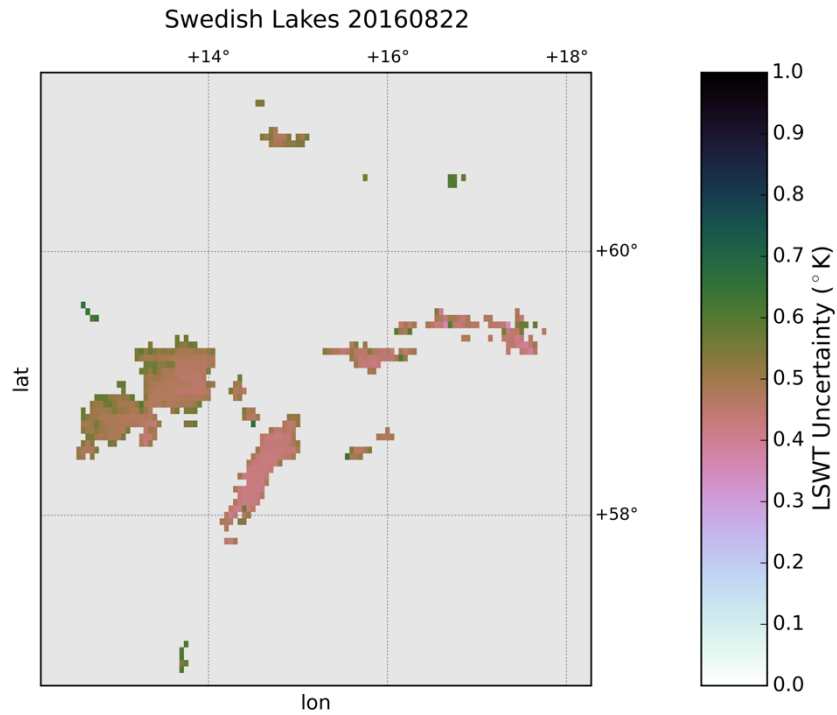
Figure 19 Uncertainty fields for L3 product for (top) Swedish lakes (the largest being Vanern and Vattern) and (lower) east African lakes (the largest being Lake Victoria).

As Figure 19 shows, uncertainty tends to increase towards lake edges and may be greater in cells where cloud has partly obscured the lake surface.

Figure 20 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).
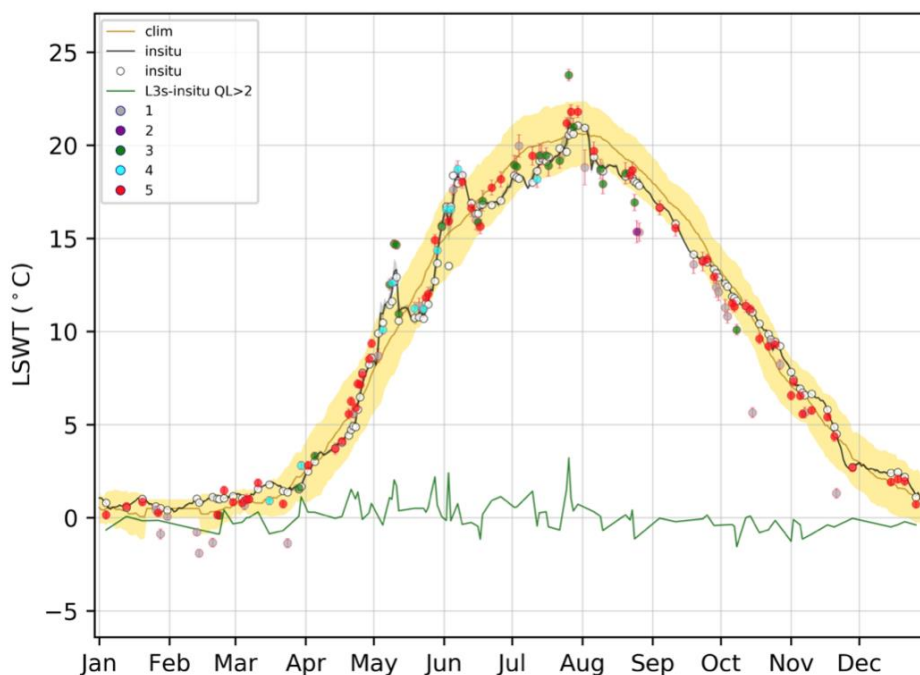
Figure 20 Satellite observations (dots), in-situ matches (white dots), in-situ measurements (black line), satellite minus in-situ temperature difference for quality levels 3,4,5 (green line) and climatology (golden line) for lake Erken in Sweden in 2008.

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 across all the lakes (sample of opportunity, not geophysically representative), the statistics of satellite-minus-in-situ difference are as in Table 1 .

Table 1 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.

4.4 Recommendations

4.4.1 To users

We strongly suggest to use the LSWT together with the uncertainty and the quality levels which are provided for each of the pixels. LSWT of quality levels 4 and 5 are recommended to be used especially in climate related applications.

4.4.2 To the validation team

It is recommended to carry out the validation of the LSWT and the LSWT uncertainty taking into account the provided quality levels. For the validation of uncertainty, it is very important to have knowledge of the uncertainty of the in-situ measurements, which is most of the time not known.

4.5 References

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

MacCallum, S. N. and Merchant, C. J. (2012) Surface water temperature observations for large lakes by optimal estimation. Canadian Journal of Remote Sensing, 38(1). pp. 25-45. ISSN 1712-7971 doi:10.5589/m12-010

5 Lake water leaving reflectance – LWLR

5.1 Uncertainty characteristics

The uncertainty characterization of the LWLR products, including water-leaving reflectance, chlorophyll-a and turbidity, is based solely on end-to-end validation of the satellite product against in-situ matchups with satellite measurements. These results are separated per Optical Water Type (OWT) 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. Sensitivity of the system to errors such as noise can be explored separately but is not in scope of this work.

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. This propagation of uncertainties is only as representative as the collection of in-situ reference observations, and the set of modelled responses. For example, adjacency to land or atmospheric composition are expected sources of uncertainty, but not currently modelled due to lacking (meta)data in the in-situ reference datasets.

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 underlying the per-pixel uncertainty in CDRP v2.1 is 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.

The uncertainty characterization approach described here is further detailed in Liu et al. (2021).

5.2 Sources of uncertainty

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 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

to isolate this effect. Due to the non-linear optimisation methodology of POLYMER, the atmospheric correction algorithm used for MERIS, OLCI and 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 with several 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 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 or 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.

5.3 End-to-end uncertainty budget

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 derived 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 21.
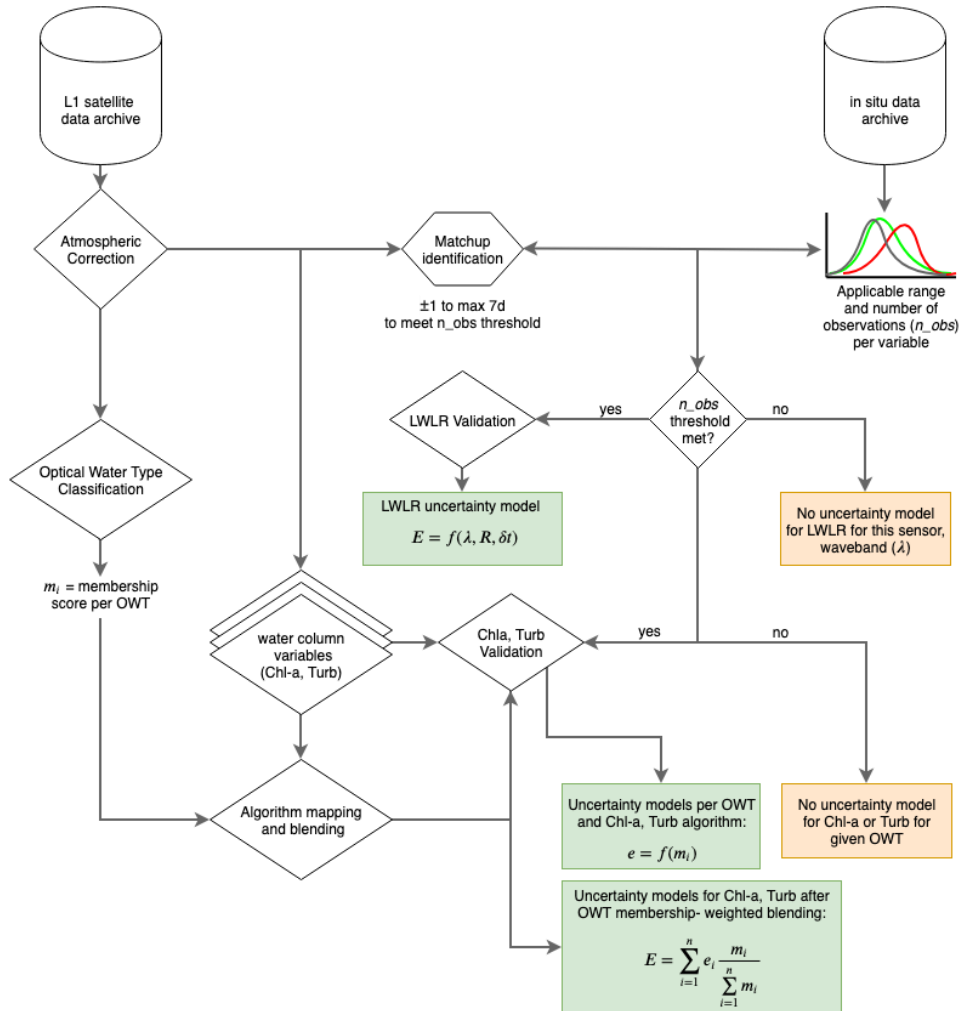
Figure 21: Flow chart of the end-to-end validation of LWLR products against in-situ observations, resulting in uncertainty models.

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:

$$Diff = R_{sat} - R_{IS} \qquad [5.1]$$

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

$$Y = a*R_{IS} + b \qquad [5.2]$$

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

$$UD = R_{sat} - Y \qquad [5.3]$$

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 22 and Figure 23 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

$$RU\ (\%) = (\ Diff_{pix} / R_{pix}\ )*100\% \qquad [5.4]$$

Where $Diff_{pix}$ is the retrieved difference of each pixel using the models shown in Figure 22**Erreur ! Source du renvoi introuvable.** and R_{pix} is the corresponding R_w on the satellite image.

Relative Unbiased Uncertainty (RUU)

$$RUU\ (\%) = (\ UD_{pix} / R_{pix}\ )*100\% \qquad [5.5]$$

Where UD_{pix} is the retrieved Unbiased Difference of each pixel using the models shown in Figure 23.

Please note that the RU and RUU calculated for the LWLR are both not absolute, in order to show the systematic under-estimation of LWLR (Figure 22, Figure 23). 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 22: 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).
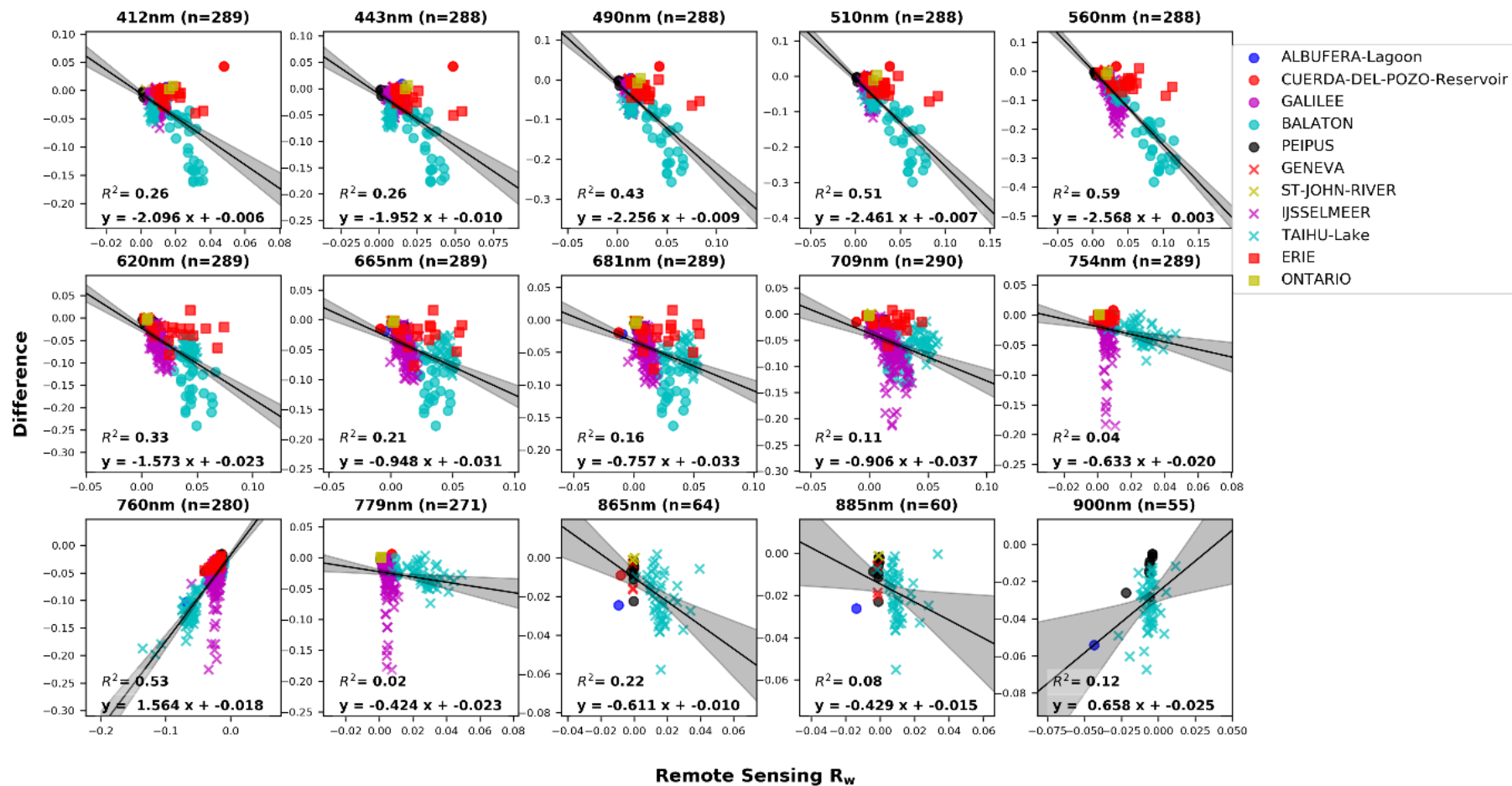
Figure 23: Linear correlation between Remote sensing Rw and the Unbiased Difference between in-situ and remote sensing Rw matchups (among which the wavebands of 865, 885 and 900 nm are not included in the Rw uncertainty products because of the sparse matchups).
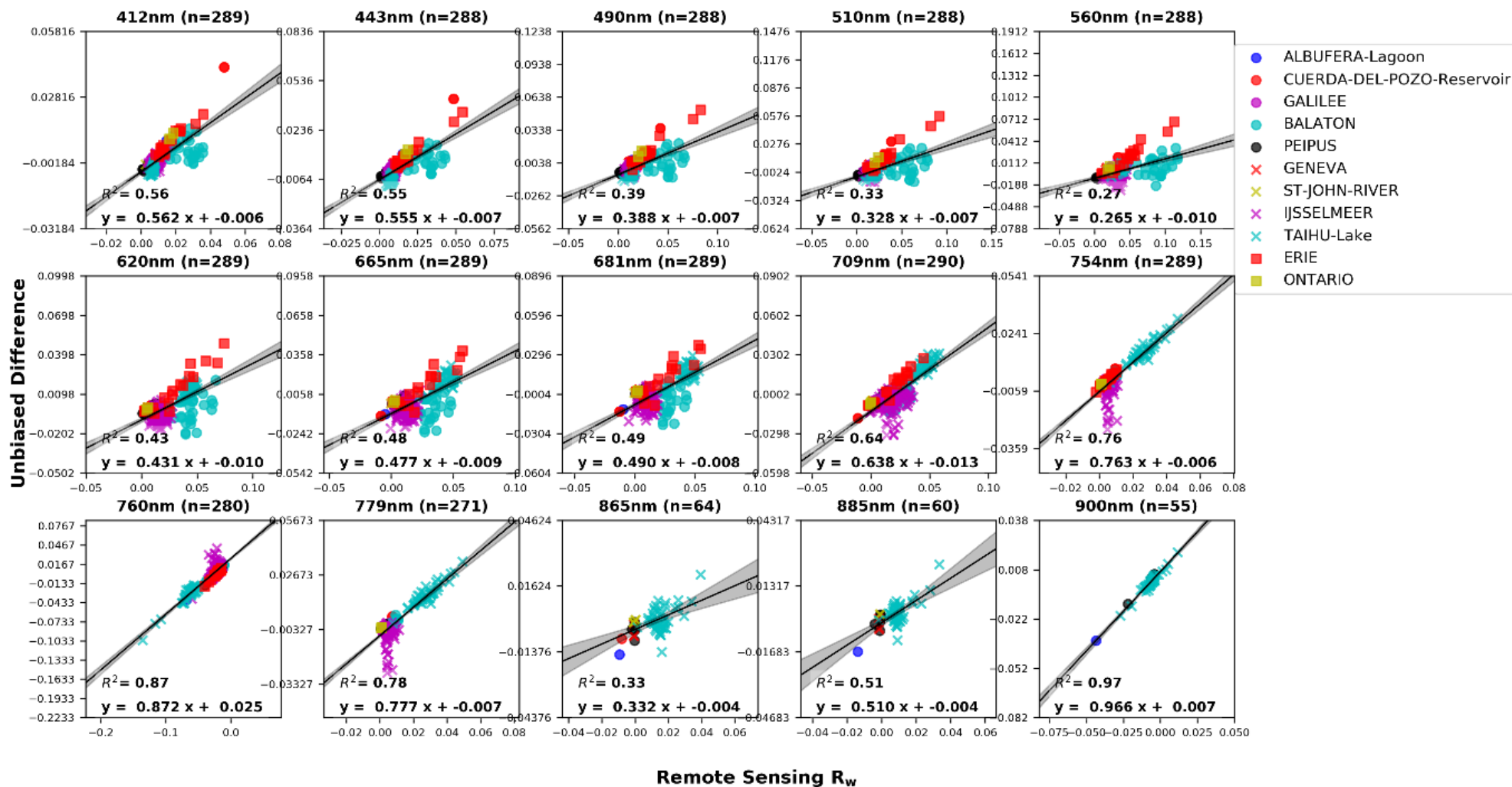
The OWTs are not taken into account during validation of LWLR due to the typically low number of in-situ reference measurements that are available. The OWTs are, however, considered in the validation of algorithms for chlorophyll-a and turbidity since these vary considerably in their suitability to describe 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.

As shown in the processing flowchart (Figure 21), 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:

$$ARU\ (\%) = (|x_d - x_m| / |x_m|) * 100\% \tag{5.6}$$

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 24 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

$$e_9 = am_9 + b \tag{5.7}$$

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:

$$E = \sum_{i=1}^{n} e_i \frac{m_i}{\sum_{i=1}^{n} m_i} \tag{5.8}$$

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.

OWT 9: Tuned algorithm (n=2562)

y = -115.388 x + 156.67
R^2 = 0.03
RMSE = 126%
SD = 44.94
BIAS = -2.29
MAE = 3.94
MAPE = 53.83
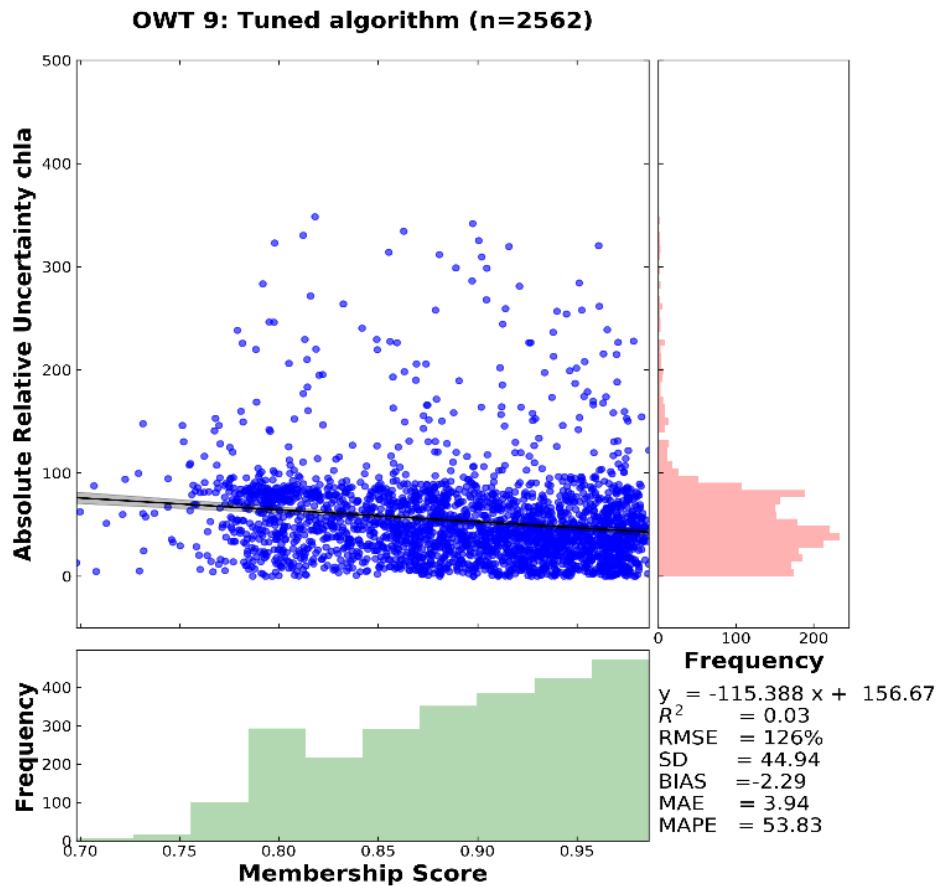
Figure 24: Correlation of Absolute Relative Uncertainty (ARU) of the top-3 weighted chlorophyll-a by the OWT membership score in relation to the membership score for optical water type 9. Reproduced from Liu et al. (2021).

Applying the uncertainty models to MERIS observations result in uncertainty maps such as those shown in Figure 25. Maps are shown for each of the uncertainty products provided in CDRP v1.0 over lakes Vänern and Vättern in Sweden on 26th July 2006. The procedure for CRDP v2.1 is unchanged, but includes separate matchup-based analysis and uncertainty models for MODIS-Aqua observations. Uncertainties are reported as relative product uncertainty, with Turbidity and chlorophyll-*a* relative uncertainty (RU) as their absolute values. The uncertainty products for LWLR are provided as absolute and relative-unbiased uncertainty 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 25a and b). Positive uncertainty values for $R_w(665)$ 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 25c). 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 25d. This is consistent with smaller uncertainties in $R_w(665)$ in lake Vänern.
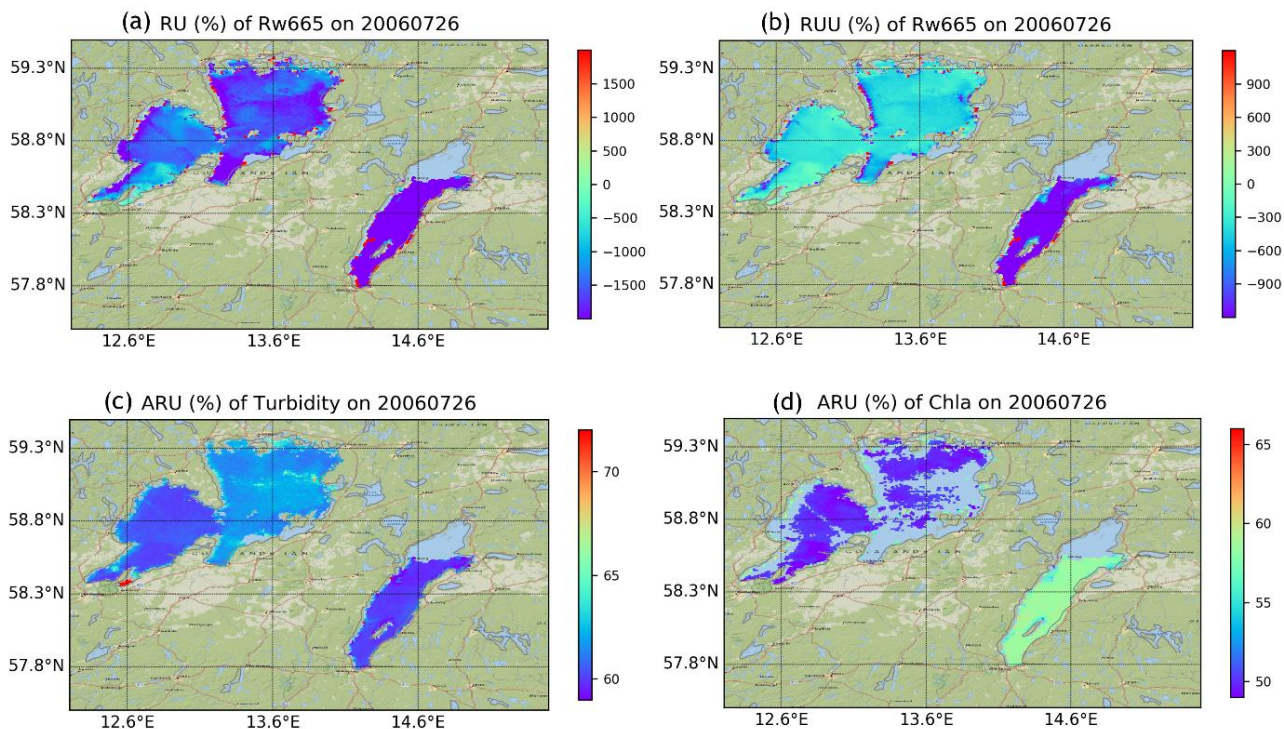
Figure 25: Per-pixel uncertainty products. (a) Relative uncertainty of $R_w(665)$, (b) Relative unbiased uncertainty of $R_w(665)$, (c) Absolute relative uncertainty of Turbidity and (d) Absolute relative uncertainty of chlorophyll-a.

5.4 Recommendations

5.4.1   To users

The provided uncertainty products for LWLR (Land and Water Regression) are calculated on a per-pixel basis for the variables of Rw (per-band), chlorophyll-a, and Turbidity.

For each Rw band, two uncertainty products are provided: Relative Uncertainty (RU) and Relative Unbiased Uncertainty (RUU). Both of these are expressed as percentages (%). The RUU is specifically calculated to eliminate the systematic differences between satellite and in-situ measurements. By doing so, it addresses the issue of systematically underestimating LWLR due to challenges associated with accurately performing atmospheric correction in optically complex inland waters.

Users should consider RU to be the most representative uncertainty product for LWLR, describing uncertainties stemming from the sensor, sensor calibrations, radiometric corrections, viewing effects and atmospheric correction. However, when using products derived from LWLR (chlorophyll-a, turbidity) which have been individually tuned against in-situ reference observations of those quantities, RUU should be considered more representative of the relative effect of LWLR uncertainties on those products, as this takes into account the effect of tuning the downstream algorithms.

The derived products have their own uncertainty models, which may differ significantly from the RUU of the Rw products. This is the result of tuning these products per OWT against larger in-situ datasets, and using specific Rw wavebands which are most diagnostic of changes in the targeted quantity. When using LWLR to derive new products without such tuning steps, the RU of the Rw should again be used as primary reference.

In cases where a specific waveband lacks an associated uncertainty model (for example, due to sparse in-situ data) or when the LWLR falls outside the application range, the pixel will be flagged as having an unknown level of uncertainty in the uncertainty product.

The uncertainty products for chlorophyll-a and turbidity are both presented as relative uncertainties, expressed as a percentage (%). It is important to note that high uncertainties associated with low estimates do not automatically indicate poor estimations. To illustrate, consider a pixel with a chlorophyll-a estimate of 0.1 mg m^{-3} and an uncertainty of 70%, where the uncertainty range is 0.03 – 0.17 mg m^{-3}.

In CRDP v2.1.0, a new quality flag band (lwlr_quality_flag) is included for LWLR. The quality indicators included in this band serve to provide users with valuable information regarding the reasons behind missing observations and assist in the appropriate utilization of the data. Among the flags available, two "consistency" flags have been incorporated. These flags highlight potential inconsistencies between detected ice and LWLR products. Whilst (sub-pixel) ice cover and open-water lake colour may occur on the same observation day, the occurrence of thin ice may lead to unrealistic values of LWLR and particularly turbidity. Such cases are flagged based on the value distribution of the turbidity and chlorophyll-a products against per-lake skin temperature (LSWT) climatology. To ensure the reliability and accuracy of the data, it is highly recommended that users consider these product flags, namely "lwlr_poor_consistency" and "lwlr_low_consistency." Pixels flagged with these indicators should be excluded, particularly in studies that focus on climatology or phenology including the ice season.

5.4.2   To the validation team

In addition to the aforementioned recommendations for users, it is advised that the validation team examines the various flags present in the LWLR L2 dataset. These flags encompass the L1B product flags, pixel identification flags, and atmospheric correction flags. These flags provide valuable information regarding the processing and quality assessment of the data. By analysing these flags, the validation team can gain insights into potential issues or limitations within the LWLR L2 dataset. The LWLR L2 dataset is not publicly accessible. Interested parties can obtain access to this dataset by contacting the LWLR team directly. This ensures that the dataset is distributed to teams capable of analysing the unfiltered data, fostering accurate validation and analysis.

An upgrade to POLYMER (4.15+) is now available which allows for a wider range of initialization conditions to make appropriate use of the mineral absorption component in the bio-optical model, which was previously unstable. This version also makes use of the SWIR band to better resolve highly turbid water conditions. It is highly recommended to test this potential improvement at scale, across all sensors.

5.5 References

Liu X., Steele C., Simis S., Warren M., Tyler A., Spyrakos E., et al. (2021). Retrieval of Chlorophyll-a concentration and associated product uncertainty in optically diverse lakes and reservoirs. Remote Sens Environ: 267:112710.

6 Lake Ice cover – LIC

6.1 Uncertainty characteristics

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 v2.1, pixels belonging to the same class are simply assigned the same % error value in the uncertainty variable.

6.2 Sources of uncertainty

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 v2.1 production uses MODIS Terra/Aqua Level 1B (TOA Reflectance) 5-min swath (MOD02), collection 6.1 (C6.1), 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 six years; their detectors are degrading. However, MODIS was designed with stringent requirements on the sensor 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. 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 C6.1 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. 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 uncertainy 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 26b could present wrong surface features in the LIC product as Figure 26a 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 26c, the MODIS image acquired from the same day by Terra shown in Figure 26b shows less details of surface condition.

Legend
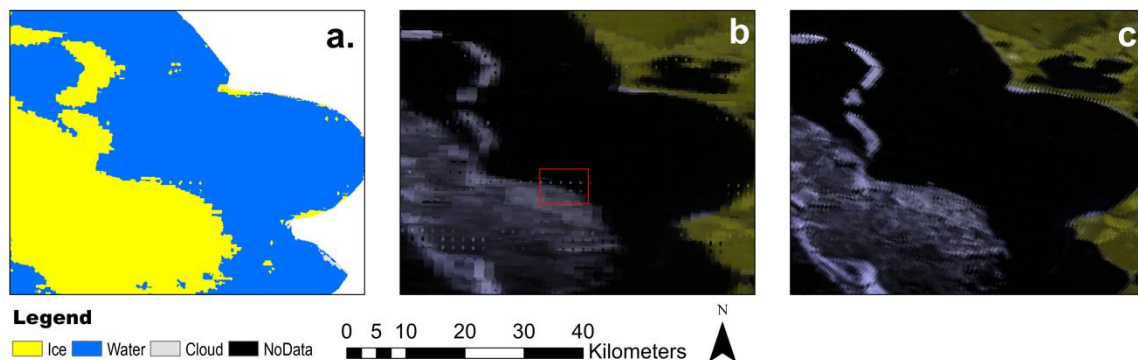Ice   Water   Cloud   NoData
0 5 10   20   30   40 Kilometers   N

Figure 26: 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 v2.1 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 preparation for the delivery LIC v.3.0 product, additional training sites (i.e., from more lakes and conditions) will be collected to further improve accuracies for each of the classes.

Misclassification: In the processing chain implemented for LIC v2.1 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 27, but the full impact of this mask remains to be quantified. High turbidity and the presence of algae blooms during the ice-free season change the reflectance characteristics of open water and may also cause erroneous ice cover detection in the spring or summer (Riggs and Hall, 2015). There are also cases when thin clouds are

classified as ice cover and snow-free melting ice in the advanced stage as open water. 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 v3.0 product.
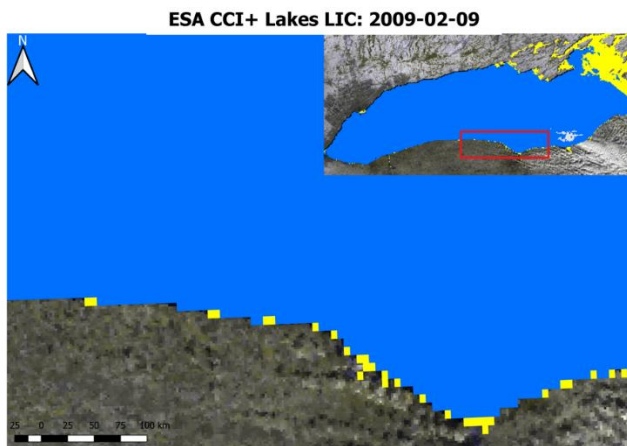
Figure 27: Example of pixels along the shoreline of Lake Ontario mislabelled as ice (in yellow), 9 February 2019

6.3 End to end Uncertainty budget

For Lakes_cci LIC product v2.1, class uncertainty is computed from a confusion matrix built on an independent statistical validation process. The confusion matrix shown in Table 2 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 2 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. 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 Lake Baikal is shown in Figure 29.

Table 2: Confusion matrix with class accuracies for Lakes_cci LIC product v2.1

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%					
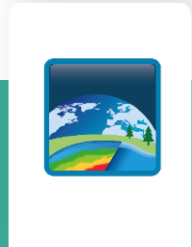
Figure 28: Uncertainty map of Lake Baikal, 19 December 2016

6.4 Recommendations

6.4.1 To users

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 v2.1, all pixels belonging to the same class are assigned the same % error value in the uncertainty band provided; that is all pixels on any day are given a value of either 0.83% for open water, 3.07% fo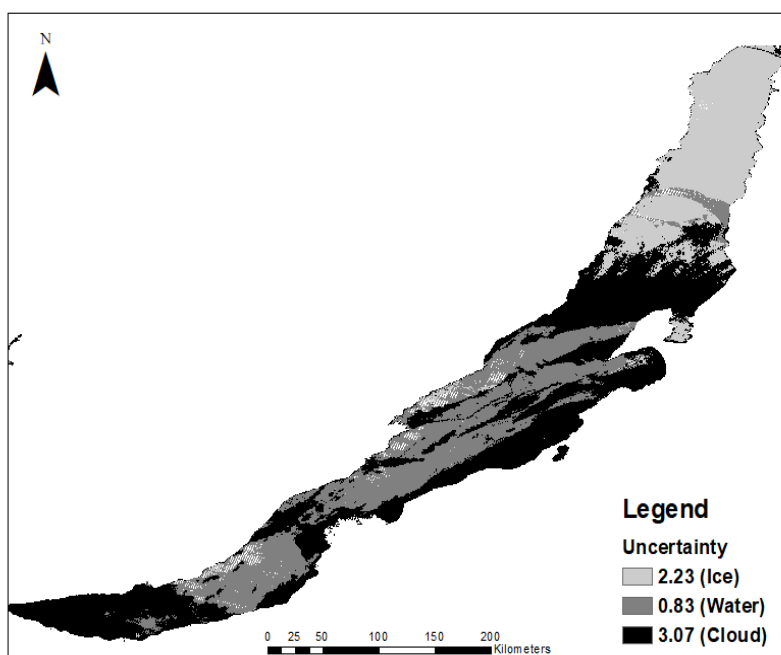r cloud cover and 2.23% for ice cover. Users should be aware that while the % classification errors are relatively low, they are only indicative of the overall performance (not per-pixel level) of the random forest (RF) algorithm. Through extensive visual examination of the product for many lakes over several years, there are instances where pixels on a given day or over a few days will have been misclassified (see algorithm and misclassification under Section 6.2).

6.4.2 To the validation team

As indicated above, the uncertainty band provided for LIC v2.1 contains only three possible values for any pixel on any given day (overall % error for water, ice and cloud cover). The uncertainty band is only indicative of the overall performance of the RF classifier. Validation of the uncertainty currently provided requires the use of colour composites of the original MODIS images used as input to the RF algorithm to identify cases (pixels) where classification is incorrect. The identified cases could be added to the training dataset to further improve the RF performance towards the release of LIC v3.0. It should also be noted that efforts are underway to implement per-pixel uncertainty (aleatoric, systematic and total) estimates within the RF classifier which will be provided in LIC v3.0.

6.5 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
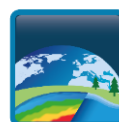
Riggs, G. A. and Hall, D. K. (2015) MODIS snow products Collection 6 user guide. 66 pp.

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

Wu, Y., Duguay, C.R. & Xu, L. (2021). Assessment of machine learning classifiers for global lake ice cover mapping from MODIS TOA reflectance data. Remote Sensing of Environment, 253, 112206, https://doi.org/10.1016/j.rse.2020.112206

7 Lake Ice Thickness- LIT

7.1 Uncertainty characteristics

The reported LIT estimates are, for each satellite overpass on a target lake, the mean and standard deviation of the LIT measurements obtained by fitting the Ku radar waveforms in the Region of Interest (RoI) with a tailored analytical model for LIT (LRM_LIT retracking analysis). The RoI is a segment of the satellite ground track over the target lake chosen to minimize land contamination, thus consisting in the part of the track furthest from the shore, and large enough to include enough statistics (typically ~150+ waveforms) to have robust and reliable LIT estimations. The LIT uncertainty reported in the LIT product includes contributions related to the retracked radar waveform's noise and the spatial variation of the LIT estimations within the RoI.

7.2 Sources of uncertainty

The observed radar waveform is a backscattered radar wave signal corrupted by a multiplicative speckle noise. The LIT estimations are obtained by fitting with an analytical waveform model the (noisy) Ku radar waveforms within the Region of Interest of a target lake. The waveform model is an analytical five parameters model, and the estimator is a Weighted Least Square Estimator where the weights are defined as the data variance within the RoI, ensuring unbiased and optimal estimation of the parameters given the noise (see Mangilli et al. 2022). The LIT mean and standard deviation, for each flyover and pass, is obtained from the Gaussian fit of the histogram of the LIT measurements in the RoI.

In the case of uniform ice cover, the dispersion of the LIT measurements (i.e. the width of the histogram) is due to the waveform noise, as is typically the case in the middle of the ice season when the ice is well formed and the LIT spatial variations are small. This is illustrated in the top row plots of Figure 29 that show the LIT spatial evolution in the RoI superimposed to the MODIS image on the same date (left) and as a function of the latitude (middle) in February 2016. The ice coverage is homogenous leading to a stable spatial evolution of the LIT measurements around a mean LIT of 0.97m with an uncertainty of 11cm (see the histogram in top right plot). However, at the seasonal transitions, when the ice is forming or melting, the lake surface is no longer homogeneous, leading to a significative variability of the LIT signature in the radar waveforms. This translates in a significative spatial variation of the retrieved ice thickness which dominates the error budget, resulting in an increased LIT uncertainty with respect to the uncertainty expected from the waveform noise contribution. This is illustrated in the bottom row plots of Figure 29 in the case of the beginning of the snow/ice melting over the Great Slave Lake at the end of April 2016. The MODIS image shows in fact an irregular ice surface with melting pond, cracks and open water channels that explains the higher variability of the LIT spatial evolution in the RoI. As shown in the middle and bottom right plots, many LIT measurements are 0, but non-zero LIT is still detected in some part of the RoI. The mean LIT in the RoI is compatible with 0 and, because of this spatial variability, the associated uncertainty is a factor of ~2.5 bigger than the case of an uniform ice coverage.
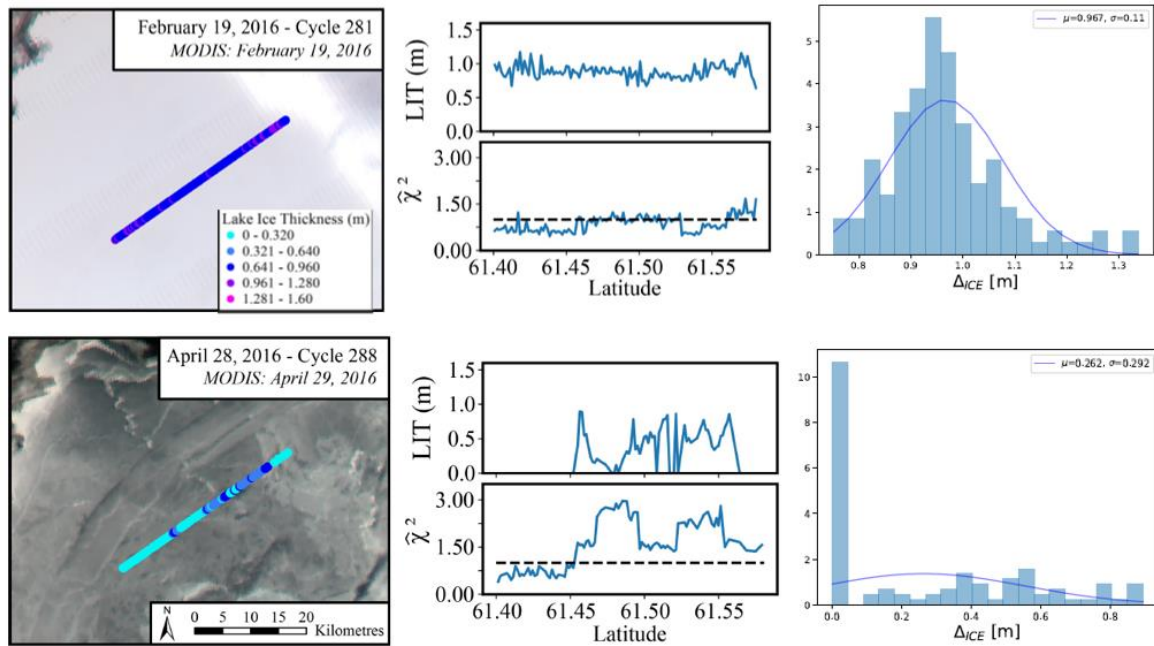
Figure 29: Evolution of Jason-2 LIT estimates over Great Slave Lake. The left column plots show the along-track LIT estimates superimposed on MODIS images. Plots in the middle column show the LIT spatial evolution. The right column plots show the histograms of the LIT measurements in the Region of Interest over the lake, and the associated Gaussian fit. The figures in the top row refer to February 2016 while the figures at the bottom to end of April 2016.

7.3 End to end Uncertainty budget

We generated two sets of Jason-like waveform simulations to assess the LIT uncertainty due to the waveform noise, and, by comparison with the results obtained with real data, to better quantify and understand the contribution linked to the LIT spatial variation:

- SIM1: set of 100 simulations representative of Jason winter-like waveforms where the ice thickness signature is present (top left panel of Figure 30)
- SIM2: set of 100 representative of Jason summer-like waveforms without the ice thickness signature (bottom left panel of Figure 30)

The right column of Figure 30 shows the distribution of the retrieved LIT normalized to the total number of counts for the winter-like simulations SIM1 (top panel) and for the summer-like simulations SIM2 (bottom panel). The LIT estimation from the winter-like simulations is 1.046±0.124 m, which is fully consistent with the input value used to generate the simulations of 1.049 m, shown as a black dashed line in the top right histogram plot. In the case of the summer-like simulations SIM2, the LIT estimation is 0.034 ± 0.054 m, which is compatible with the ice thickness of zero used as input value to generate the SIM2 simulations.

As expected, the LIT uncertainty derived from the SIM1 simulations is fully compatible with the uncertainty derived from real data in the case of uniform ice coverage (top right plot of Figure 29), confirming that in this case the LIT uncertainty is entirely due to the waveform noise. The LIT uncertainty derived from the SIM2 simulations is smaller than the uncertainty obtained with the SIM1 simulations, likely because there is not the step-like LIT signature in the SIM2 waveforms, allowing for a smaller dispersion of the parameters that, as LIT, are linked to the waveform leading edge that has no noise in the case of SIM2. Also, the accuracy of the SIM2 LIT estimates is significantly smaller than the accuracy estimated from real data, when, during the seasonal transitions, the LIT measurements are compatible with zero (see e.g. the bottom right plot of Figure 29). As discussed in the previous section, this is mainly due to the significative LIT spatial variability in real data, linked to the fact that the ice coverage is no longer uniform

because of the melting conditions. Related to this, the fractional ice coverage (i.e. altimeter footprint containing both open water and ice) can also contribute to increase the LIT spatial variability, as in this case the LIT signature could disappear from the waveforms making it more difficult to detect the ice in some part of the RoI. Finally, it should be emphasized that the LRM_LIT retracker is specifically designed for the analysis of LIT and it is not the optimal tool for the analysis of Brownian type echoes, in particular with a slope due to the presence of waves (SWH is not a parameter of the LRM_LIT waveform model). This could lead to an increased uncertainty when retracking open water waveforms.
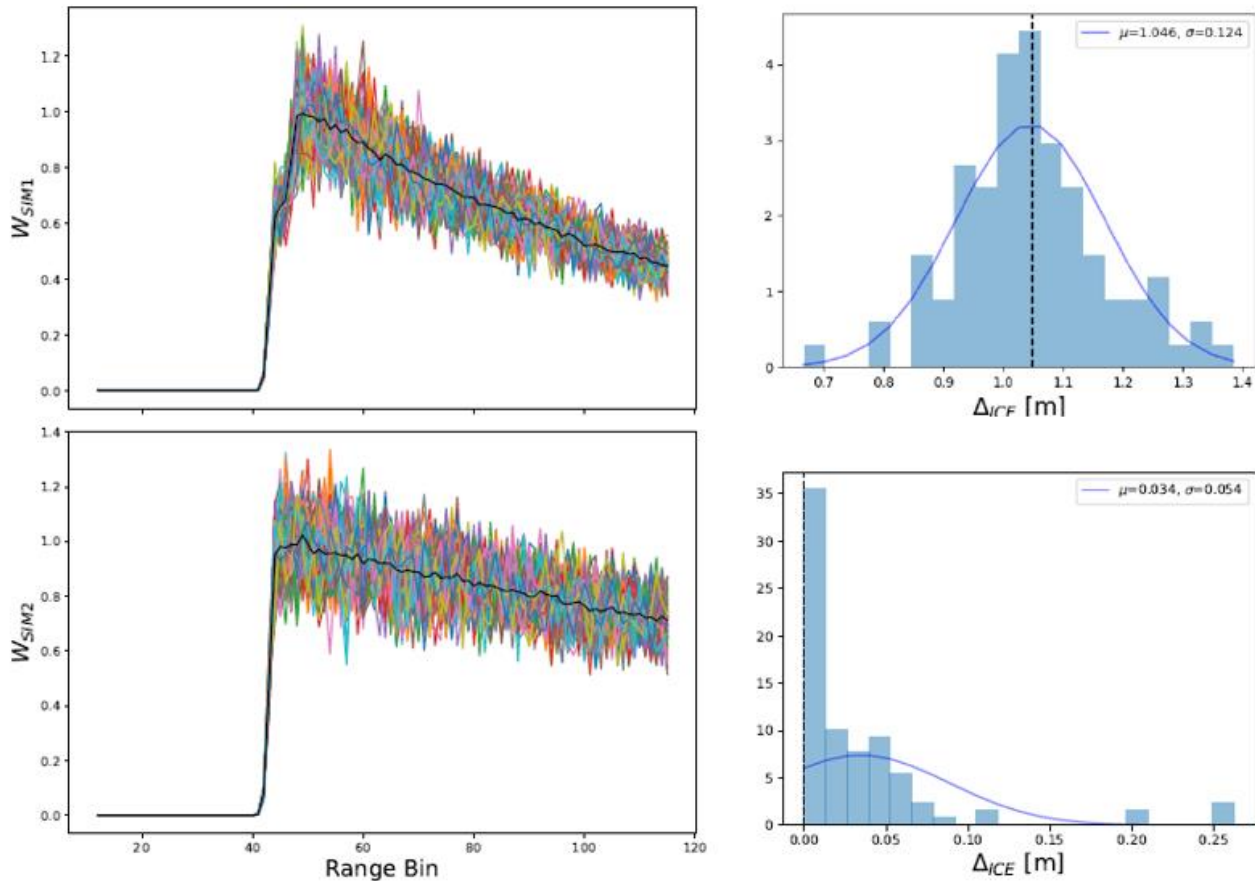
Figure 30: Validation of the LIT estimation with the LRM_LIT retracker on Jason-like simulations. In the left column are shown Jason-like waveform simulations corresponding to the winter-like SIM1 waveforms (top) and to the summer-like SIM2 waveforms. In the right column are shown LIT histograms computed for the winter-like simulations SIM1 (top panel) and for the summer-like simulations SIM2 (bottom panel). The blue lines correspond to the Gaussian fit of the histograms. The input values used to generate the simulations are also shown as dashed black lines.

7.4 Recommendations

7.4.1   To users

The reported LIT estimates are, for each satellite overpass on a target lake, the mean and standard deviation of the LIT measurements in the Region of Interest (RoI) chosen to minimize land contamination, thus consisting in the part of the track furthest from the shore, and large enough to include enough statistics to have robust and reliable LIT estimations. The uncertainty reported in the LIT product includes therefore contributions related to the retracked radar waveform's noise and the spatial variation of the LIT estimations within the RoI. For this reason, the LIT uncertainty increases at the seasonal transitions as during ice forming and melting, the lake surface is not homogeneous with a higher spatial variation of the ice (and snow) properties and thickness. As the LIT measurements are taken from a specific region

on the lake, the reported LIT and the associated uncertainty cannot be generalized to the whole lake. The spatial information of the LIT measurement should be always taken into account when using the LIT products, in particular when comparing the results with LIT estimates obtained with other datasets.

7.4.2   To the validation team

We recommend not to use in-situ data to validate the LIT measurements provided in the products, unless in situ data are taken from a region of the lake near the satellite ground track (which is usually not the case). In fact, in situ data are generally collected near the shore, while satellite data are taken from the middle of the lake to avoid land contamination. This difference of measurement induces two different environments in terms of bathymetry, wind exposure, snow type and quantity. All these parameters play a key role on ice formation and thickness and they can lead to significant LIT differences, making in situ data not suitable for a robust validation as these differences are significatively bigger than the measured LIT uncertainties. A lake ice model, as CLIMo (Duguay et al 2003), is a more suitable tool for validation. In the case in situ data near the analysed RoI should be collected and made publicly available, they would be the preferred dataset to be used for validation.

7.5 References

A. Mangilli, P. Thibaut, C. R. Duguay and J. Murfitt. (2022). A New Approach for the Estimation of Lake Ice Thickness From Conventional Radar Altimetry. IEEE Trans Geoscience and Rem Sens 60: 1-15, 4305515. https://doi.org/10.1109/TGRS.2022.3197109

Duguay, C.R., Flato, G.M., Jeffries, M.O., Ménard, P., Morris, K. & Rouse, W.R. (2003). Ice cover variability on shallow lakes at high latitudes: Model simulations and observations. Hydrological Processes, 17(17), 3465-3483.

Appendix A - List of acronyms

AATSR	Advanced Along Track Scanning Radiometer
AATSR	Advanced Along Track Scanning Radiometer
AERONET-OC	Aerosol Robotic NETwork – Ocean Color
AMI	Active Microwave Instrument
AMSR-E	Advanced Microwave Scanning Radiometer for EOS
APP	Alternating Polarization mode Precision
ASAR	Advanced Synthetic Aperture Radar
ASLO	Association for the Sciences of Limnology and Oceanography
ATBD	Algorithm Theoretical Basis Document
ATSR	Along Track Scanning Radiometer
AVHRR	Advanced very-high-resolution radiometer
BAMS	Bulletin of the American Meteorological Society
BC	Brockman Consult
C3S	Copernicus Climate Change Service
CCI	Climate Change Initiative
CDR	Climate Data Record
CDOM	Coloured Dissolved Organic Matter
CEDA	Centre for Environmental Data Archival
CEMS	Centre for Environmental Monitoring from Space
CEOS	Committee on Earth Observation Satellites
CGLOPS	Copernicus Global Land Operation Service
CIS	Canadian Ice Service
CLS	Collecte Localisation Satellite
CMEMS	Copernicus Marine Environment Monitoring Service
CMUG	Climate Modelling User Group
CNES	Centre national d'études spatiales
CNR	National Research Council of Italy
CORALS	Climate Oriented Record of Altimetry and Sea-Level
CPD	Communication Plan Document
CR	Cardinal Requirement
CRG	Climate Research Group
CSWG	Climate Science Working Group
CTOH	Center for Topographic studies of the Ocean and Hydrosphere
DOC	Dissolved Organic Carbon
DUE	Data User Element
ECMWF	European Centre for Medium-Range Weather Forecasts
ECV	Essential Climate Variable
ELLS-IAGRL	European Large Lakes Symposium-International Association for Great Lakes Research
ENVISAT	Environmental Satellite
EO	Earth Observation
EOMORES	Earth Observation-based Services for Monitoring and Reporting of Ecological Status
ERS	European Remote-Sensing Satellite
ESA	European Space Agency
ESRIN	European Space Research Institute
ETM+	Enhanced Thematic Mapper Plus
EU	European Union
EUMETSAT	European Organisation for the Exploitation of Meteorological Satellites
FAQ	Frequently Asked Questions
FCDR	Fundamental Climate Data Record

FIDUCEO	Fidelity and Uncertainty in Climate data records from Earth Observations
FP7	Seventh Framework Programme
GAC	Global Area Coverage
GCOS	Global Climate Observing System
GEMS/Water	Global Environment Monitoring System for freshwater
GEO	Group on Earth Observations
GEWEX	Global Energy and Water Exchanges
GloboLakes	Global Observatory of Lake Responses to Environmental Change
GLOPS	Copernicus Global Land Service
GTN-H	Global Terrestrial Network – Hydrology
GTN-L	Global Terrestrial Network – Lakes
H2020	Horizon 2020
HYDROLARE	International Data Centre on Hydrology of Lakes and Reservoirs
ILEC	International Lake Environment Committee
INFORM	Index for Risk Management
IPCC	Intergovernmental Panel on Climate Change
ISC	International Science Council
ISO	International Organization for Standardization
ISRO	Indian Space Research Organisation
JRC	Joint Research Centre
KPI	Key Performance Indicators
LEGOS	Laboratoire d'Etudes en Géophysique et Océanographie Spatiales
LIC	Lake Ice Cover
LIT	Lake Ice Thickness
LSC	Lake Storage Change
LSWT	Lake Surface Water Temperature
LWE	Lake Water Extent
LWL	Lake Water Level
LWLR	Lake Water Leaving Reflectance
MERIS	MEdium Resolution Imaging Spectrometer
MGDR	Merged Geophysical Data Record
MODIS	Moderate Resolution Imaging Spectroradiometer
MSI	MultiSpectral Instrument
MSS	MultiSpectral Scanner
NASA	National Aeronautics and Space Administration
NERC	Natural Environment Research Council
NetCDF	Network Common Data Form
NOAA	National Oceanic and Atmospheric Administration
NSERC	Natural Sciences and Engineering Research Council
NSIDC	National Snow & Ice Data Center
NTU	Nephelometric Turbidity Unit
NWP	Numerical Weather Prediction
OLCI	Ocean and Land Colour Instrument
OLI	Operational Land Imager
OSTST	Ocean Surface Topography Science Team
PML	Plymouth Marine Laboratory
PP	Payment Plan
PRISMA	PRecursore IperSpettrale della Missione Applicativa
Proba	Project for On-Board Autonomy
QSR	Quarterly Status Report
R	Linear Correlation Coefficient
RA	Radar Altimeter
RMSE	Root Mean Square Error

SAF	Satellite Application Facility
SAR	Synthetic Aperture Radar
SeaWIFS	Sea-viewing Wide Field-of-view Sensor
SIL	International Society of Limnology
SLSTR	Sea and Land Surface Temperature Radiometer
SoW	Statement of Work
SPONGE	SPaceborne Observations to Nourish the GEMS
SRD	System Requirements Document
SSD	System Specification Document
SST	Sea Surface Temperature
STSE	Support To Science Element
SWOT	Surface Water and Ocean Topography
TAPAS	Tools for Assessment and Planning of Aquaculture Sustainability
TB	Brightness Temperature
TM	Thematic Mapper
TOA	Top Of Atmosphere
TR	Technical Requirement
UNEP	United Nations Environment Programme
UoR	University of Reading
UoS	University of Stirling
US	United States
VIIRS	Visible Infrared Imaging Radiometer Suite
WCRP	World Climate Research Program
WHYCOS	World Hydrological Cycle Observing Systems
WMO	World Meteorological Organization
WP	Work Package