Lakes_cci

Product User Guide (PUG)

Reference: : CCI-LAKES2-0021-PUG

Issue 3.0.0 – 09/12/2025

Contract number: 4000125030/18/I-NB –Lakes_cci

CHRONOLOGY ISSUES

Issue	Date	Objective	Written by
1.0.0	06/06/2022	Last version Phase 1	ECV Production teams
2.1.0	11/12/2023	Accompanying CRDP v2	ECV Production teams
3.0.0	09/12/2025	Accompanying CRDP v3	S. Simis, L. Carrea, B. Calmettes, J-F Crétaux, C. Duguay, C. Fatras, X. Liu, A. Mangilli, H. Yésou

Checked by	Stefan Simis – PML	
Approved by	Philippe Mourot - CLS	
Authorized by	Clément Albergel - ESA	

BIBLIOGRAPHIC REFERENCE

This document may be cited as:

Simis, S., Carrea, L., Calmettes, B., Crétaux, J-F., Duguay, C., Fatras, C., Liu, X., Mangilli, A., Yésou, H. 2025. Lakes Climate Change Initiative Product User Guide (PUG), version 3.0.0. European Space Agency. DOI: 10.5281/zenodo.18234648

LIST OF CONTENTS

1 Introduction

1.1 Scope

This Product User Guide (PUG) contains a description of the Lakes_cci Climate Research Data Package (CRDP) version 3.0.0, produced as part of the European Space Agency (ESA) Climate Change Initiative. The documentation set accompanying CRDP v3.0.0 will be valid for any other v3.x.y products unless an updated document version is published. The PUG provides users with practical information regarding the content and recommended use of the Lakes Essential Climate Variable (ECV) products included in this dataset.

The overarching objective of the Lakes_cci project is to produce and validate a consistent data set of the variables grouped under the Lakes ECV. This includes aiming for the longest period of combined satellite observations by operating processing chains for suitable satellite imagery, ultimately featuring in a sustainable production system. This PUG details the contents, format, and standards applied to the files that make up the dataset. It also introduces some software tools that can help new users explore the data contained in the CRDP.

The Lakes ECV covers: Lake Water Level, Lake Water Extent, Lake Surface Water temperature, Lake Ice Cover, Lake Ice Thickness, Lake Water-Leaving Reflectance and Lake Storage Change. The Lakes_cci dataset includes several additional products (both physical and biogeochemical indicators) to aid studies into climate change and its effects on lake systems, either from inspecting the observation data are through their ingestion into numerical models. The full set of products is detailed in the next section.

The specific objectives for the Lakes_cci project are:

- **To assess** the requirements of the climate research community and thereby ensure consistency in the (further) development of the Lakes ECV processing system.
- **To develop**, test and select the best algorithms and standards to produce high quality Lake products for climate applications across sensors.
- **To provide** a specification of the operational production system, aligned with related activities in the Copernicus programme (e.g. Global Land Service, C3S). Algorithms are developed or improved to meet user requirements.
- **To validate** the Lake ECV products through independent climate research groups and use cases.
- **To generate** new interest in the EO climate datasets produced for inland water bodies within the community of limnologists, operating at local to global spatial scales and likely to use varying subsets of the Lakes ECV products.

1.2 Dataset

Lakes are of significant interest to the scientific community, local to national governments, industries and the wider public. A range of scientific disciplines including hydrology, limnology, climatology, biogeochemistry and geodesy are interested in the distribution and functioning of the millions of water bodies (from small ponds to reservoirs, inland seas) loosely referred to as lakes, from the local to the global scale. Remote sensing provides an opportunity to extend the spatiotemporal scale of lake observation, within the observational limitations set by the available satellite sensors.

The Lakes_cci develops products for the following six ECV Products:

- Lake Water Level (LWL): to understand the balance between water inputs and water loss.
- Lake Water Extent (LWE): a proxy for change in glacial regions (lake expansion) and drought in many arid environments, relating to local climate for the cooling effect that water bodies provide.

- Lake Surface Water temperature (LSWT): correlated with regional air temperatures and a proxy for mixing regimes, driving biogeochemical cycling and seasonality.
- Lake Ice Cover (LIC): freeze-up in autumn and advancing break-up in spring are proxies for gradually changing climate patterns and seasonality.
- Lake Ice Thickness (LIT): a proxy indicator of changes in air temperature and on-ice snow mass (depth and density) during the ice growth period.
- Lake Water-Leaving Reflectance (LWLR): a direct indicator of biogeochemical processes and habitats in the visible part of the water column (e.g. seasonal phytoplankton biomass fluctuations) and an indicator of the frequency of extreme events (peak terrestrial run-off, changing mixing conditions).

In this context, Lakes_cci represents a unique framework to provide **consistent and homogenous** data to the multiple communities of lake scientists. The project actively engages with this community to assess the utility and future improvement of Lakes_cci products.

Key considerations for the Lakes ECV dataset are based on the combined **requirements from user communities** including the official Global Climate Observation System (GCOS) targets. These considerations are summarised in the next chapter and detailed in the User Requirements Documentation (URD).

The Lakes_cci Climate Data Research Package extends beyond the GCOS definition of the Lakes ECV by including several **additional products**:

- Lake Storage Change, derived from the LWL/LWE hypsometry and landscape features, a proxy for climate vulnerabilities of water resources.
- Chlorophyll-a concentration, derived from LWLR, as a proxy for phytoplankton (algae and cyanobacteria) biomass.
- Suspended particulate dry weight, derived from LWLR, as an indicator of particle fluxes and turbidity. In CRDP versions prior to v3.0.0, this product was expressed as Turbidity.
- Vertical light attenuation, derived from LWLR, describing diffuse energy absorption of visible and near infra-red light, light availability for primary production, and a habitat for submerged vegetation.
- Ligh absorption by Coloured Dissolved Organic Matter (CDOM), derived from LWLR, a proxy for dissolved organic carbon, the largest pool of aquatic Carbon in lakes.

In addition, a gap-filled LSWT and LIC record, which is derived from the LSWT and LIC observations included in the Lakes_cci CRDP v3.0.0 is generated for general use. The gap-filled dataset is not part of CRDP v3.0.0, but is based on the same observation principles described in the documentation which accompanies the CRDP. Several of these documents explicitly mention the methodology used to generate the gap-filled dataset.

As a rule of thumb, the Lakes ECV makes use of those satellite sensors which can be calibrated to the best available standards, for the longest legacy of sensors possible. Modern sensors generally offer better resolution and sensitivity than legacy sensors, whilst the Lakes ECV datasets are on a harmonized grid resolution. This trade-off determines the size range of observable lakes, which just over 2000 lakes currently targeted at a resolution of 1/120 degree (nominally 1 km), with all available data provided in daily aggregation periods. The water extent, water level and ice thickness products are not derived across the whole lake but derived from narrow satellite tracks.

Details of how each of the products can be used are provided in Section 3, whilst the full methodology for each processing line is described in the Algorithm Theoretical Basis Document (ATBD).

Version 3.0.0 of the Lakes_cci dataset nominally covers the period 1992-2023 where data from suitable sensors were available.

1.3 Requirements

A user requirements analysis was conducted to design the specification of the lakes_cci product to best address the needs of key users. The approach involved a review of existing requirements specified by ESA and the Global Climate Observing System (GCOS). User requirements have been further refined through an online survey, open to both current and potential users of the ECV Lakes for both climate and more general applications.

Table 1 shows the observation target requirements for the Lakes ECV core products. The general method of synthesis for these targets is to adopt the most stringent well-justified statement of requirement. The synthesis is therefore a statement of target requirements and not of what will or can be achieved.

Table 1: Synthesised observation requirements for the Lakes ECV. The source of the requirements is indicated in parentheses as follows: G: GCOS (2022 - Threshold), Q: Lakes_cci questionnaire, P: project team expertise, L: literature review.

Product	Lake Water Level (LWL)	Lake Water Extent (LWE)	Lake Surface Water Temperature (LSWT)	Lake Ice Cover (LIC)	Lake Ice Thickness (LIT)	Lake Water Leaving Reflectance or Lake Colour (LWLR)
Measurement uncertainty	10 cm (G)	5% (relative) (G)	0.2°K (P)	LIC: 10% (G,P)	15 cm (G)	10-30% for peak waveband vs low signal bands (P/L), 0.1 mg m^{-3} chlorophyll-a (L) and 1 g m^{-3} suspended matter.
Stability	10 cm/decade (G)	5% /decade (G)	0.07°K per decade (P)	LIC: 1% /decade (G)	10 cm/decade (G)	1% /decade (G,P,L)
Spatial resolution	N/A : per lake (Q)	N/A : per lake (Q)	1000 m (P)	LIC: 100 m (P)	10000 m (G)	100 m (P)
Temporal resolution	daily ground-based or satellite observations (G)	30 days (G)	Daily (P)	LIC: 3-7 days (G,P)	365 days (G)	Daily observations (Q)
Length of record	>10 years (L)	>10 years (L)	>10 years (L)	>10 years (L)	>10 years (P)	>10 years (L/P)
Maximum delay before availability of data (for climate users)	1 year (P)	1 year (P)	1 year (P)	1 year (P)	1 year (P)	1 year (P)

2 Instruments overview

Observation data from multiple satellite missions are required for the successful generation and validation of all component products for the Lakes ECV.

Table 2 summarises the satellite/sensor used in the estimation of the products that are part of the Lakes ECV core Products. The user is advised to refer to the CRDP metadata to see which satellite/sensors are used for a given observation.

Table 2. Missions and instruments used in the generation of the Lakes_cci ECV Products

Satellite	Sensor	Product						
		LWL	LWE	LSWT	LIC	LIT	LWLR	LSC
Topex/Poseidon	Poseidon-1	X						X
Jason-1	Poseidon-2	X				X		X
Jason-2	Poseidon-3	X				X		X
Jason-3	Poseidon-3B	X				X		X
Sentinel-6A	Poseidon-4	X						X
ENVISAT	Radar Altimeter (RA-2)	X						X
	AATSR			X				
	MERIS						X	
SARAL	AltiKa	X						X
Geosat Follow On (GFO)	Radar Altimeter	X						X
Sentinel-1	C-band SAR		X					X
Sentinel-2	MSI		X					X
Sentinel-3A/B	SRAL	X						X
	OLCI A/B						X	
	SLSTR		X	X				X
Landsat-4	MSS, TM		X					X
Landsat-5	MSS, TM		X					X
Landsat-7	ETM+		X					X
Landsat-8	OLI		X					X
Terra/Aqua	MODIS			X	X		X	
ERS1	AMI		X					X
	SAR		X					X
ERS-2	RA	X						X
	AMI		X					X
	SAR		X					X
	ATSR-2			X				
METOP-A/B	AVHRR			X				

3 Data Description

3.1 Lake Water Level (LWL)

3.1.1 LWL definition and usage

LWL refers to the lake water level above a reference geoid, whereas in situ observations may be more commonly referenced to an average sea level. Radar altimetry from space consists of vertical range measurements between the satellite and the water level. The water level is determined as the difference between the satellite altitude and a reference surface (usually a conventional ellipsoid and then a geoid), determined through precise orbit computation, and satellite-water surface distance. Placed onto a repeat polar orbit, the altimeter satellite passes a given region at regular time intervals (called the orbital cycle), which defines its periodic coverage of the Earth. The time interval, called repetitivity, depends on each mission and is detailed in Table 3. However, a lake can be observed by several missions/orbits, thus the time step for observations per lake varies between one day (for large lakes observed by multiple tracks) and 35 days (for lakes observed only by Envisat, for example).

Table 3. Altimetry mission repetitivity

Mission	Repeat interval (days
Topex/Poseidon, Jason-1, Jason-2, Jason-3, Sentinel-6A	10
Sentinel-3A, Sentinel-3B	27
ERS-1, ERS-2, Envisat, SARAL	35
GFO	17

Water level measurement by satellite altimetry was initially developed and optimised for open oceans. The techniques applied in the Lakes_cci have been adapted to obtain water levels of inland seas, lakes, rivers, floodplains, and wetlands.

3.1.2 LWL data characteristics

The LWL product is composed of three variables: the water level estimation, its associated uncertainty and a quality flag for the measurement. The details for the estimation of those variables are fully described in the Algorithm Theoretical Basis Document (ATBD). The LWL product is generated at irregular time steps, depending on which satellites pass over the target. For harmonisation with the other Lakes ECV products, the LWL result is duplicated onto the grid for the nominal lake area defined by its maximum water extent. Reciprocally, when plotting time-series of LWL of a given lake, it is sufficient to pick one data point within the lake from the gridded dataset.

The LWL product covers a period of more than 30 years, from the first altimetry mission Topex/Poseidon launched in 1992 until the current missions: Sentinel-3A, Sentinel-3B and Sentinel-6A.

Given the characteristics of LWL observation by altimetry satellites, not all the target lakes (2024 targets) in the CRDP 3.0 have time series of water level. There are several reasons for this: (i) some of them are not observed by any mission (ii) the location of the track above the lake, that can vary by up to 1 km, does not allow good quality data to be obtained (iii) the altimetry data is highly impacted by ice, which affect measurement for lakes at hight latitudes and altitudes (iv) the land surrounding may have an impact on the backscattered signal. Figure 1 shows the distribution for lakes in CRDP where LWL is available (571), not observed (260), or rejected due to insufficient data quality (1193). This coverage will improve in future due to new missions such as SWOT, which use new technology with a swath observation instead of the current nadir observation.

3.1.3 LWL data sources

Several satellite altimetry missions have operated since the early 1990s (Table 4):) TOPEX/Poseidon (1992-2006), ERS-2/RA (1995-2005), GFO (2000-2008), Jason-1 (2001-2012), ENVISAT/RA-2 (2002-2012), Jason-2 (2008-2018), Jason-3 (2016-2022), Sentinel-6A (2021-), SARAL/AltiKa (2013-), Sentinel-3A (2016-), and Sentinel-3B (2018-).

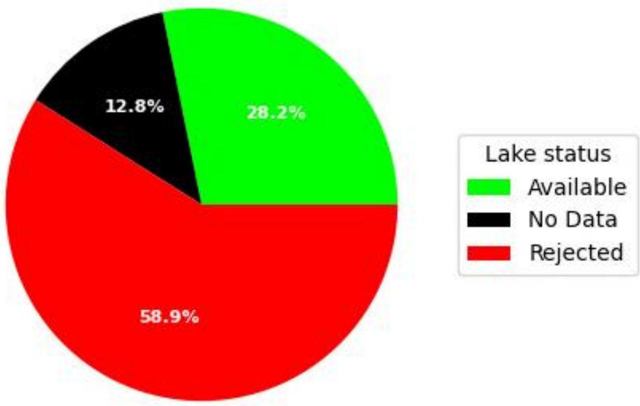

Figure 1. LWL Lake status in CRDP V3.0.0

ERS-2, ENVISAT and SARAL have a 35-day temporal resolution (duration of the orbital cycle) and 70 km inter-track spacing at the equator. TOPEX/Poseidon, Jason-1, Jason-2 and Jason-3 have a 10-day orbital cycle and 350 km equatorial inter-track spacing. GFO has a 17-day orbital cycle and 170 km equatorial intertrack spacing. The Sentinel-3A orbit has a revisit time of 27 days and its inter-tracking separation is 104 km. This has been reduced to 52 km in a two-satellite configuration (Sentinel-3A and B). Lake Water levels are based on merged multi-mission observations. As a result, the combined global altimetry data set has a 30-year history for many lakes and is intended to be continuously updated in the coming decade. Combining altimetry data from several in-orbit altimetry missions increases the space-time resolution of the sensed hydrological variables.

Table 4. Altimetry missions

	1992	1993	1994	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019	2020	2021	2022	2023
Topex/Poseidon	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■																	
Jason-1									■	■	■	■	■	■	■	■	■	■	■	■	■											
Jason-2																	■	■	■	■	■	■	■	■	■	■	■					
Jason-3																											■	■	■	■	■	
Sentinel-6A																															■	■
ERS-2				■	■	■	■	■	■	■	■	■	■	■																		
Envisat											■	■	■	■	■	■	■	■	■	■	■											
SARAL																						■	■	■	■	■	■	■	■	■	■	■
Sentinel-3A																									■	■	■	■	■	■	■	■
Sentinel-3B																											■	■	■	■	■	■

3.1.4 LWL data limitations

An important limitation of the LWL data set is its spatial coverage. 260 of the 2024 lakes targeted in the Lakes_cci are not covered by altimetry mission tracks and can therefore not be resolved. In addition, the quality of observations may be insufficient to optimise processing parameters to generate the timeseries. This activity is time consuming and needs the estimation of multiple inter-mission and inter-track bias, which is not always successful. In some cases, the land surrounding (presence of other nearby lakes, ice or any reflecting surface near the lake) as well as satellite operating mode (Low Resolution mode or Synthetical Aperture Radar) affect the quality of the backscattered signal and the quality of the timeseries may not be sufficient to be included in the dataset as mentioned in previous section.

Additionally, the timestep for LWL product is lake dependent. It is determined by the altimetry missions observing the lake and the repetitivity of these missions. For example, for lakes observed by a single track of Sentinel-3A, a measurement is provided every 27 days. For larger lakes, observed by more than one mission and track, the time step may be daily.

Another limitation concerns the accuracy of the product. The comparison with external data, such as in situ measurements, is carried out to compare the precision rather than the accuracy of the lake water level product. This is because reference levels may differ between different datasets.

Finally, we consider that the water level is the same throughout all the lake. This assumption is true for most of the lakes. Nevertheless, for very large lakes, this value may vary due to factors as wind or tides.

3.2 Lake Water Extent (LWE)

3.2.1 LWE definition and usage

LWE can be expressed as the presence of water (on a map), or as the total areal extent of a waterbody (a single number). The latter approach is used in the Lakes_cci. Studying and monitoring variations and trends in lake area, or lake water extent (LWE), can identify climatic variations because LWE is sensitive to changes in the water cycle and heat balance. LWE, together with LWL, can be used to assess the total volume of water in a lake (Arsen et al., 2014; Cretaux et al., 2016; Nikraftar and Azizia, 2015; Busker at., 2019; Sima et al., 2013; Ryan et al., 2020).

The solution to determining lake water extent variations for many lakes globally is to delineate lake shorelines from satellite imagery. In the Lakes_cci we use high resolution optical imagery, e.g. from Sentinel–2, to determine the lake extent, which is then combined with lake water level observations from radar altimetry to determine a hypsometry relationship. This relationship represents the variation of lake extent with respect to lake water level. When the hypsometry of a lake is established, water height and extent of lakes can be determined simultaneously from altimetry without the need to analyse a prohibitively large volume of high-resolution optical imagery. The results can be provided from daily (for large water bodies) to weekly or monthly, depending on the lake location and satellite track.

For a better understanding of the LWE product, it is worth mentioning some of the characteristics of the optical images and the methodology followed (detailed in the ATBD): The use of optical data for water delineation is based on the analysis of the reflected signal. Depending on the type of lake, i.e. deep versus (very) shallow, the reflected signal will be different, with a greater or lesser influence from the lake bottom. The reflected signal also depends on the presence of dissolved matter and floating or submerged vegetation. In some cases, the presence of such vegetation can lead to a loss of precision in terms of extracting the water surface, as only open water is observed and extracted, and water hidden under flooded vegetation is not detected. In addition, the acquisition conditions also influence the signal, for example in the case of sun reflections.

3.2.2 LWE data characteristics

LWE are initially generated as an intermediate product both as a geocoded binary map (water/non water) in geotiff and shapefile and appear in the CRDP as a single area value (in square kilometres). Information on uncertainties is provided with these products. These are calculated from the ratios of standard deviation to maximum surface area.

3.2.3 LWE data sources

The primary optical sensor used in this project is the Sentinel-2 multispectral instrument (MSI). In addition, Landsat data are exploited as needed to cover periods prior to the launch of the MSI instruments, going back to the mid-1990s with Landsat 5. Landsat data can also be used on a case-by-case basis to characterise specific hydrological conditions, i.e. extreme conditions, in particular intense drought or heavy flooding, in order to have all the altimetric height/water surface pairs available.

In theory, any high resolution optical mission image source could be used to retrieve LWE according to the methodologies developed and employed in the Lakes_cci.

3.2.4 LWE data limitations

The primary limitation to deriving LWE information is the presence of clouds. This is particularly true for some lakes in tropical areas with high cloud cover, where it is challenging to achieve a full set of data points to populate the hypsometry curves, particularly during the wet seasons. Moreover, for some very large lakes, optical image scenes may not cover the whole surface of the lake, requiring the use of several images collected at different times. This is not easily achievable since there must be no changes in water level between these dates.

In addition, the LWE extraction can be limited/disturbed by the presence of elements on the water surfaces, such as ice, floating/§submerged vegetation, as well as acquisition's conditions (sun-glint).

Finally, the procedure adopted for LWE, using hypsometry curves, is time-consuming and can be performed only lake per lake, with visual checking, although many steps of the procedure have been automated.

Another limitation is that the altimeter tracks are relatively loose, so a number of lakes, even large ones, are not covered. It is therefore not possible for these lakes to generate a surface/altitude pair and therefore a hypsometric curve. One possible answer would be to integrate the recent SWOT data, which covers two 50 km wide bands on each pass, and thus a very large number of lakes and reservoirs. However, this would require the SWOT measurements (in terms of surface area and height) to be validated.

3.3 Lake Ice Cover (LIC)

3.3.1 LIC definition and usage

LIC refers to the extent (or area) of a lake covered by ice. Lake-wide ice phenology can be derived from LIC, including freeze onset to complete freeze over (CFO) dates during the freeze-up period, melt onset to water clear of ice (WCI) dates during the break-up period, and ice cover duration between CFO and WCI dates over an ice year (Duguay et al. 2015). For lakes that do not form complete ice cover every year or in some years (e.g. Laurentian Great Lakes of North America), maximum ice cover extent (as a function of time) is also a useful climate indicator that can be determined (Derksen et al. 2019). Similarly, the minimum ice cover extent (as function of time) can be derived for High Arctic lakes that do not completely lose their ice cover in summer, although a recent study suggests that these lakes may be transitioning from perennially to seasonally ice-covered (Surdu et al. 2016). Knowledge of fractional ice coverage (expressed as a percentage of total area of a lake covered by ice) on a daily to weekly basis is also useful for improving numerical weather forecasting in regions where ice cover forms.

LIC is highly sensitive to changes in weather and climate. Documented trends and variability in ice dates have largely been attributed to air temperature changes (e.g. Duguay et al. 2006; Brown and Duguay 2010). Investigations of long-term trends (observable from ground-based records) and short-term (also observable from satellite data records) reveal increasingly later freeze-up and earlier break-up dates, closely corresponding to increasing air temperature trends. Broad spatial patterns in these trends and regime shifts have been associated with changes in major atmospheric circulation patterns originating from the Pacific and Atlantic oceans such as the El Niño-La Niña/Southern Oscillation, the Pacific North American pattern, the Pacific Decadal Oscillation, and the North Atlantic Oscillation/Arctic Oscillation (Bonsal et al. 2006; Prowse et al. 2011). LIC also plays an important role in weather and climate. The presence (or absence) and fractional coverage/concentration of ice cover on large lakes has a significant impact on regional weather and climate (e.g., lake-effect snowfall, thermal moderation effect).

Given the importance of ice cover in lake-atmosphere interactions, the LIC ECV will be of interest to users who wish to: (1) examine short-term trends and interannual variability in ice cover globally (ca. 20 years), (2) investigate the impact of changing ice cover conditions on other variables covered in Lakes_cci, such as Lake Surface Water Temperature (LSWT), (3) conduct data assimilation experiments using state-of-the-art numerical weather prediction systems to demonstrate the impact of better consideration of LIC on, for example, improving predictions of lake-effect snowfall and (4) evaluate lake models (e.g. FLake) used as lake parameterization schemes in numerical weather prediction and climate models. Finally, from a socio-economic perspective, the LIC variable may also serve to examine the impact of changing ice conditions on winter transportation (shipping, ice roads) and food security (access to resources by northern communities via ice roads).

3.3.2 LIC data characteristics

The LIC product consists of nine variables: (1) lake cover class, (2) LIC process flag, (3) LIC class agreement, (4) LIC MODIS overlap, (5) LIC random forest total uncertainty, (6) LIC random forest quality level, (7) fractional ice cover, (8) total cloud cover, and (9) area of the lake covered by ice in km^2**Error! Reference source not found.**. For the lake cover class, each grid cell falling within a lake, as determined by the input lake mask, is assigned one of three possible values: water (value = 1), ice (2), cloud (3). For the total uncertainty and quality level variables, uncertainty is calculated for each pixel where a retrieval is performed, this is weighted by the agreement among overlapping MODIS images to produce a quality level from 0 to 4 (see ATBD document for details).

For CDRP V3.0, the LIC product is generated on a daily basis using MODIS data acquired from multiple Terra and Aqua satellite overpasses on each day s to maximize the number of cloud-free observations. The product, which covers a 24-year period (2000-2023), is merged with the other lakes thematic products on the common (harmonized) grid described in section 4.

3.3.3 LIC data sources

The LIC product is generated from all MODIS observations available from both the Terra (since 24 February 2000) and Aqua (since 4 July 2002) satellite missions. Hence, for the period between 24 February 2000 and 3 July 2002, only Terra observations were ingested for LIC product creation.

Prior to the main processing chain, the Canadian Lake Ice Model (CLIMo) was used to determine which lakes of the Lakes_cci (total 2024 lakes) could have formed ice or have remained ice-free at any time over the 2000-2020 period. Input data to drive CLIMo are from European Centre for Medium-Range Weather Forecasts (ECMWF) ERA5 reanalysis hourly data on single levels (25 km).

The primary observation data source for lake cover class retrieval is the MODIS Terra/Aqua Level 1B Top-Of-Atmosphere (TOA) 5-min L2 swath (MOD02/MYD02), Collection 6.1 product. Six MODIS (Terra/Aqua) TOA reflectance bands, a solar zenith angle (SZA) band, and a TOA emission band are used for feature retrieval (i.e. for labelling as water, ice, or cloud) (Wu et al., 2021). The reflectance bands are MOD02QKM/MYD02QKM at 250 m (band 1: 0.645 μm and band 2: 0.858μm) and MOD02HKM/MYD02HKM at 500 m (band 3: 0.469μm; band 4: 0.555μm; band 6: 1.640 μm; band 7: 2.130 μm) resolutions. The second data source for LIC product generation is the CCI Static Lake Mask at 1/120 degree resolution.

Details regarding the processing steps and the retrieval algorithm are described in the Algorithm Theoretical Basis Document (ATBD). Briefly, the processing steps consist of:

1. Load TOA reflectance and emission bands (1, 2, 3, 4, 6, 7, 31, and SZA) from MODIS Level 1B calibrated radiances product (MOD02/MYD02), Collection 6.1.

2. Match latitude and longitude coordinates from the TOA products to the closest coordinates from the Lake CCI water mask (1/120 degrees resolution).
3. Label pixels of interest from application of RF algorithm for the detection of clouds, ice, and open water.
4. Filter out pixels with a SZA >85 degrees and apply temperature filters to correct ice and water pixels where high or low temperatures are found.
5. Aggregate pixels by majority vote
6. Determine LIC extent excluding cloud pixels, total cloud cover, and lake ice area for each lake where the RF algorithm was applied.
7. Write and export the daily lake ice cover product in the required format (NetCDF) with metadata.

3.3.4 LIC data limitations

The LIC daily product has been created using the longest possible MODIS time series (2000-2023) and with the intent of maximizing the number of clear-sky observations within each day. This is made feasible through the combination of multiple acquisitions from Terra and Aqua overpasses. While this approach increases the likelihood of detecting a larger number of grid cells containing ice cover or open water, the presence of cloud cover over extended periods of time across the northern hemisphere remains the greatest limiting factor for the generation of a spatially and temporally contiguous lake ice cover product from MODIS or any other optical satellite dataset.

In addition to the impact of cloud cover, users should be aware that other factors may also affect the quality of the LIC CDRP V3.0 (see E3UB document for greater details). They include, in no particular order of importance:

1) High solar zenith during fall freeze-up and early winter at high latitudes (i.e. polar darkness). For example, LIC retrieval is not possible from MODIS above ca. 85 degrees zenith at winter solstice. In such instances the data are flagged as “no retrieval”.
2) The quality and temporal continuity of the MOD02/MYD02 TOA reflectance products used as the primary input data source.
   - Although detector noise/sensor degradation and observation noise have been relatively well characterized (see E3UB document for details), there are few documented cases where noise has been found to lead to false negatives (e.g., detection of ice instead open water).
   - Regarding temporal continuity, a few years have been found not to provide data from either Terra or Aqua on some days (missing days: 12 in 2000, 17 in 2001, 19 in 2002, 7 in 2003, 2 in 2008, and 9 in 2016).

3.4 Lake Ice Thickness (LIT)

3.4.1 LIT definition and usage

Lake ice is a major landscape feature in the winter season at high latitudes and plays a key role in climate moderation and the energy balance (Brown and Duguay 2010). Lake ice conditions, particularly the length of the ice season and ice thickness, have a significant impact on the economy, particularly in northern regions through their influence on transportation, travel, fishing, and recreation (Ghiasi et al. 2020). Therefore, accurate knowledge about lake ice properties, such as lake ice thickness (LIT), is necessary. Furthermore, LIT integrates changes in surface air temperature and on-ice snow mass (depth and density) (Brown and Duguay 2010) and is, therefore, a good proxy of winter severity.

3.4.2 LIT data characteristics

The LIT product consists of LIT time series with the associated uncertainty. These are estimated by performing a retracking analysis of radar altimetry waveform data as described in Mangilli et al. (2022), and further detail is also provided in the ATBD and E3UB.

A quality flag is included in the product with the following values: 0 for good data, 1 for bad or missing data and 2 for poor data, in case the average waveform fit performance in the region of interest is degraded. In general, LIT data with quality flag 2 are not frequent and typically localized at the seasonal transitions. These data can be used, noting lower LIT retrieval performance and increased uncertainty.

The LIT product delivered for CRDP v2.1 consists of three LIT time series generated from three different regions of interest (RoI) over the Great Slave Lake (Canada). The LIT time series are computed from December 2001 to April 2022 with a temporal resolution 10 days, corresponding to a data cycle (that is, the time of revisit of the same ground pass) of the radar altimetry satellites used to generate the LIT estimations. , The LIT dataset has been expanded for CRDPv3.0 to include 13 additional lakes: Amadjuak Lake (passes 43 and 72), Lake Athabasca (pass 95), Baker Lake (pass 19), Dubawnt Lake (pass 19), Gauer Lake (pass 152), Great Bear Lake (passes 24, 100, and 225), Reindeer Lake (passes 76 and 145), Schultz Lake (pass 121), South Henik Lake (pass 202), Southern Indian Lake (pass 152), Tulemalu Lake (pass 24), Wharton Lake (pass 176), and Lake Winnipeg (pass 195). This expansion broadens the spatial coverage of the LIT dataset. LIT products are provided in separate files with respect to the other ECVs.

3.4.3 LIT data sources

The data used to generate the LIT time series product are Low Resolution Mode (LRM) Ku-band radar altimetry waveform data at 20 Hz from Jason-1, Jason-2 and Jason-3 missions over Great Slave Lake (GSL). The satellites are on the same orbit, so they cover the same area of GSL, allowing for consistent LIT estimations over time. The LIT time series have been generated from three different Regions of Interest (ROI) of the lake, corresponding to three along-track segments of the three satellite overpasses (pass 45, 178 and 254 respectively). Each of the ROI along-track segment used for the LIT analysis spans ~0.1° in latitude, including more than one hundred radar waveforms, within a distance of ≃ 35 km along the altimeter track. The location of each ROI has been chosen to be far enough from the lakeshore, so that the radar echoes are not contaminated by spurious signals from land, while the length of the segment has been chosen to provide enough statistics for the LIT estimation while limiting the spatial coverage and, thus, the spatial evolution of LIT within the segment in order to obtain consistent constraints within the RoI, in particular at the seasonal transitions. The LIT time series product has been generated by analysing the radar waveform data with the LIT retracker described in Mangilli et al. (2022).

3.4.4 LIT data limitations

A few limitations of the LIT product have been identified, mainly related to LIT estimations at the seasonal transitions:

1. The LIT retracker faces challenges in precisely estimating ice thickness in the initial days of ice formation and once melt begins. This issue has been reported in Mangilli et al. (2022).
2. In the absence or minimal presence of snow on the ice surface, the step-like feature required to estimate LIT from the analytical approach (Mangilli et al. 2022) does not appear in the radar waveforms, leading to underestimated or no LIT retrievals.
3. The presence of leads or footprints with mixed open water and ice results in poor estimation of LIT values.

The LIT retracker (Mangilli et al. 2022) can capture the seasonal transitions of ice forming and melting but cannot precisely follow the ice evolution at the transitions because of the difficulty of retracking heterogeneous surfaces when the ice is too thin and when snow on the ice surface begins to melt.

3.5 Lake Surface Water Temperature (LSWT)

3.5.1 LSWT definition and usage

LSWT is the surface expression of the thermal structure of lakes which is changing in response to climatic trends. LSWT is needed for climate change studies, water budget analysis (linked to evaporation), lake physical and ecological modelling.

Lake surface water temperature (LSWT) is the temperature of the upper layer of lake water. In the case of a satellite observation of LSWT, the obtained value is sensitive to the skin temperature of the water, which is the temperature of a layer <0.1 mm thick from which thermal radiation is emitted by the lake. In Lakes_cci products, the LSWT is an estimate of this skin temperature, which may differ from the temperature as measured by a thermometer a few centimetres below the water-air interface. Typically, the temperature difference between skin and sub-skin LSWT is of order –0.2 K (meaning, the skin temperature is on average cooler). However, the difference depends on meteorological conditions and on the lake. Although the skin effect is variable, the satellite LSWT is nonetheless tightly coupled to the LSWT as measured conventionally, and satellite LSWT has been used to quantify worldwide aspects of lake thermal dynamics such as seasonal cycles (Maberly et al. 2020), onset of summer stratification (Woolway and Merchant 2018), lake mixing dynamics (Woolway and Merchant 2019), over-turning behaviour (Fichot et al. 2019) and several other aspects of lake-climate interactions.

3.5.2 LSWT data characteristics

Users of the Lakes products will note that the LSWT fields are not, in general, spatially complete. This is because the LSWT estimation process (MacCallum and Merchant 2012) is valid only for cloud-free views of the water surface from space, since the satellite sensors rely on infrared wavebands to which clouds are opaque. The degree of unobserved surface is therefore variable in time and space, according to weather conditions. In regions with persistent cloud cover observations may be rare.

The primary LSWT variable is the lake surface water temperature field itself. Two other LSWT variables should, however, be considered by users: (1) an estimate of uncertainty is provided per datum. The provided uncertainty field is an evaluation of total standard uncertainty (so-called “1-sigma”) and is estimated within the LSWT retrieval process. The uncertainty arises from a range of sources, some of which are independent between data, while other sources cause errors that are correlated to other nearby data. (2) A quality level is provided per datum. This is an index ranging from 2 (suspect/marginal quality) to 5 (best quality). “Quality” here means the level of confidence that the LSWT value and its evaluated uncertainty are both valid (Merchant et al. 2017). For climate applications, we recommend use of quality levels 4 and 5. However, LSWT with quality levels 2 and 3 are present in the product, and users can assess their usefulness for their own application. Quality levels 1 (bad) and 0 (unprocessed) are never valid and their use should be avoided.

The provided data are on the same regular latitude-longitude grid as the other Lakes ECV products, which means they have been regridded from the less regular pixel locations originally observed by the satellite sensors.

3.5.3 LSWT data sources

As background information, the basis on which LSWT data are obtained is summarised in this section. For full details of the basis of the data, refer to the Algorithm Theoretical Basis Document.

The algorithms to derive LSWT products aim to retrieve the LSWT from the observed reflectance and brightness temperature for only inland water pixels. The core retrieval is the Optimal Estimation (OE) of LSWT, which is a form of Bayesian inference of the LSWT given the difference between the satellite observations and simulations of those observations made for an assumed LSWT (the prior value). The other components of the algorithm prepare the inputs for the retrieval, classifying a satellite pixel as water or non-water. Finally, the observations are gridded and a cross-sensor adjustment is estimated and applied in order to obtain a harmonized result across the several sensors that contribute.

Preparatory processing: this includes orbit file reading, validity checks, association of auxiliary information to the orbit file being processed (including prior fields from numerical weather prediction, where relevant), and any pre-processing adjustment to the data themselves.

Classification: this step identifies valid pixels for LSWT retrieval. Although sometimes referred to as cloud detection, this also involves identifying which image pixels cover only lake water (no coast or islands within the pixel), and exclusion of pixels affected by ice (for which LSWT cannot be obtained). Valid LSWT is estimated only for pixels that are fully water and free of cloud. The algorithm for the discrimination of water and non-water pixels in presence of clouds is based on threshold tests on the visible, near-infrared, and short-wave-infrared channels of the ATSR and AVHRR instruments. The water detection algorithm is applied only to candidate pixels identified as potential inland water in the water-bodies identifier mask (Carrea et al. 2015) built from the ESA CCI Land Cover project. The water detection algorithm can be used only during daytime since the visible and short-wave infrared channels are available during daytime only.

Retrieval of LSWT (geophysical inversion): for pixels classified as water, LSWT is calculated dynamically given prior information with the Optimal Estimation technique (MacCallum and Merchant 2012). The prior information comprises NWP fields as inputs to a radiative transfer model, whose simulations in comparison to the observations are used in the retrieval. The LSWT is estimated for each (clear-sky) water pixel using joint optimal estimation of surface temperature and Total Column Water Vapour given the simulations and observations. The form of OE used is to return the Maximum A-posteriori Probability assuming Gaussian error characteristics. OE also gives an uncertainty estimate for each retrieval. Quality levels are also estimated which reflects the degree of confidence in the validity of the uncertainty estimate (not the magnitude of data uncertainty).

Gridding/averaging: the algorithm grids the irregular swath-based imagery into the regular grid.

Daily collation: the complete 14-15 orbits each day per sensor are collated to produce one data layer for each 24-hour period, corresponding to day-time observations. The average of the best quality observations from all available sensors is used as the combined gridded LSWT for each grid cell.

Inter-sensor adjustment: to stabilise the record for changes in satellite sensor, an adjustment using overlaps of sensors is made, using as the (unadjusted) reference the LSWTs from the AVHRR on MetOpA.

3.5.4 LSWT data limitations

The classification algorithm relies on threshold tests, using one generic set of thresholds for all the lakes (although some in reality have different reflectivity). For any classification scheme, some water pixels may have not been detected as water and some non-water pixels may have been included in the set of pixel where the retrieval has been applied. The classification scheme is “fuzzy”: the confidence of the water detection is captured in a water detection score which is used (together with other parameters) to set the value of the LSWT quality levels.

The emissivity assumed in the LSWT retrieval is always set to that of fresh water, and for highly saline lakes, this may introduce some small bias (whose magnitude is yet to be assessed and may be negligible). The retrieved LSWT corresponds to the skin temperature of the lake (the radiating layer of surface water), and a cool offset of order 0.2 K should be expected relative to sub-surface water temperature measurements.

The temporal density of observations of any particular quality varies greatly between lakes. Lakes that are narrow (only a couple of kilometres across) generally produce few water-only pixels with these sensors (whose best resolution is 1 km), even if the lake is extensive and its area overall is large. Some lakes that are targeted in the products, but whose geometry is unfavourable, may have few or no high quality LSWT results.

Prior to the availability of global 1-km resolution MODIS (Terra) observations (2000 onwards), the temporal density of observations is generally lower because of the lesser coverage from the earlier ATSR series instruments used. After 2006, the data density increases further because of the availability of AVHRR (MetOpA/B) observations.

3.6 Lake Water Leaving Reflectance (LWLR)

3.6.1 LWLR definition and usage

Lake Water-Leaving Reflectance (LWLR), also referred to as water colour, is the quantity of sunlight reaching the remote detector after interaction with the water column. The maximum depth from which the reflected signal is observed depends on the optical properties of the water column, is dependent on the colour band (waveband) considered and, in natural waters, can range from tens of meters (up to nearly 100 m in the clearest ocean waters) to just centimetres in highly absorbing and/or turbid waters. The colour of water is retrieved using imaging or line-scanning optical detectors on satellites. Each sensor offers a specific trade-off between the observation time (longer periods yielding lower instrument noise) and the spatial resolution as well as the number of discrete wavebands in which reflectance is measured. Because relatively small changes in absorption by, for example, phytoplankton pigment need to be distinguishable, an adequate signal-to-noise of an ocean-colour sensor for the signal received at the top of the atmosphere should be at least 1000:1 (IOCCG 2012). Correspondingly, the spatial footprint of such sensitive detectors on modern polar-orbiting satellites which scan the entire Earth every 2-3 days (a speed-over-ground exceeding 25,000 km/h) is around 300m at the equator. For previous sensors with less advanced detectors, which may also have been limited by downlink capacity, the spatial resolution tends to be 1 km or coarser while the number of discrete wavebands has increased from eight (at 10 bits digitization) on SeaWIFs (1997-2010) to 21 (using 12-bit precision) on OLCI (2016-present).

Lake Water-Leaving Reflectance (LWLR) is the result of atmospheric correction of top-of-atmosphere radiance over water pixels. This correction is the result of model optimization and subject to the possibility of ambiguous solutions. The main effects that introduce uncertainty are mixing of reflectance from water and nearby land in the atmosphere, bottom reflectance, in-water bio optical model ambiguities and limited waveband configurations to help bound the mentioned numerical optimisation.

Once LWLR is estimated, several optical-biogeochemical characteristics of the lake may be determined from its colour. The main quantities of interest are:

- The concentration of phytoplankton pigment, particularly chlorophyll-*a*, which is found in all species as the major photosynthetic pigment.
- Vertical transparency, for submerged vegetation habitat mapping or primary production models when combined with chlorophyll-a and temperature observations or models
- The concentration of (coloured) dissolved (organic) matter as a proxy for the dissolved organic Carbon pool, as well as the quality of underwater light.
- The total amount of suspended sediment (TSM), either expressed as equivalent particulate dry weight or as the Turbidity caused by it.

3.6.2 LWLR data characteristics

The daily observations used for CDRP v3.0.0 are obtained from Envisat-MERIS (2002-2012), Aqua-MODIS (2002 – present) and Sentinel-3 OLCI (2016 – present) satellite sensors. They are provided on a common grid with the other Lakes Essential Climate Variables. MERIS and OLCI sensors offer a native 300 m resolution, whilst for MERIS the 1km reduced resolution data provides better spatio-temporal coverage and is used when full resolution products are not available. MODIS observations are obtained at 1 km resolution and only used to fill the gap between MERIS and OLCI (April 2012 - April 2016) in lakes were good consistency with the other sensors is observed. The use of MODIS is typically limited to large lakes, as correcting for atmospheric effects including those exacerbated by nearby land is more difficult due to the absence of key wavebands used in the correction, in the near infrared. MERIS and OLCI also offer better retrieval capabilities for chlorophyll-*a* in relatively turbid water, through a 709 nm waveband in the near infrared. The consistency criteria for MODIS versus OLCI or MERIS are also influenced by the ability of the processing chain to consistently resolve the full seasonality and occurrence of phytoplankton blooms or turbid regions. As MODIS is less capable of resolving high-turbidity conditions, the resultant observation bias favouring clear periods may result in the removal of the MODIS observation timeseries from the CRDP. Users wishing to investigate results for individual lakes more closely, are welcome to request additional subsets of the dataset from the production team.

Data are not available for each of the included lakes at every datum. This is due to a combination of satellite overpass limitations (more frequent with OLCI than with MERIS, particularly for 2019-present when two OLCI sensors were in simultaneous operation) and removal of pixels affected by cloud (including edges and shadows) or ice.

The LWLR is presented in a number of discrete wavebands corresponding to the frequency or wavelength of light across the visible to near infra-red spectrum. The number of wavebands differs per sensor and each sensor has variations in where the bands are centred. No attempts have been made to 'shift' the signal of any sensor to match that of others, because the underlying reflectance form is neither known nor modelled, owing to the wide optical variability of inland water bodies (compared to oceans for example). However, bands that are within 6 nm of an OLCI waveband are presented as belonging to the same position along the spectrum. This has been done to reduce the number of variables in the dataset, and to allow easier ingestion and analysis of time-series data. For details on the waveband definitions of the individual sensors, please refer to the ATBD, and refer to the metadata of each data file to see which sensor was used for a particular datum.

In addition to LWLR the CDRP V3.0 includes estimates of chlorophyll-*a* (mg m^{-3}) and TSM (g m^{-3}), $a_{CDOM}(440)$ (m^{-1}), the vertical diffuse attenuation coefficient K_d (m^{-1}), and phycocyanin (mg m^{-3}) which are derived from the LWLR. For a detailed overview of the algorithms and correction used to obtain these estimates please refer to the ATBD. With some exception, each product comes with an associated per-pixel product uncertainty estimate, as long as sufficient in situ reference observations were available to model the uncertainty for the specific variable, observed value range and lake optical type. The procedure used to determine product uncertainties is defined in the E3UB document. There are a few exceptions to note. Firstly, the phycocyanin product is derived from a single algorithm source (MDN-PC), which is different from the other products which have algorithms selected and tuned for each of thirteen optical water types. Neither the phycocyanin nor the K_d product come with specific product uncertainty estimates in v3.0.0. For the latter, the LWLR uncertainty propagated for the equivalent K_d wavebands, as these will be indicative of K_d accuracy. The phycocyanin product may be considered a prototype CCI product, included for convenience to assist researchers with initial analysis of global trends in phytoplankton composition. The main difference with the other LWLR-derived products is that inter-sensory consistency has not yet been studied to the same extent. The PVASR document describes the performance of each product against in situ reference observations. We note that the latest product additions (K_d, a_{CDOM}, PC) will remain subject to further scrutiny, e.g. in product consistency assessments that are still underway. Some observation artefacts may not be present in the subset of the global dataset corresponding to in situ observations, and user feedback on unexpected variations is always welcome.

3.6.3 LWLR data sources

The full observation archive from Envisat-MERIS and Sentinel-3 OLCI-A/B is used to assess the full set of Lakes_cci targets (2024 in total). These archives contain observations from April 2002 until April 2012 for MERIS and April 2016 – present for OLCI. CRDP v3.0 includes all years up to 2023. MERIS satellite passes identified as invalid were omitted. The MERIS dataset consists of a combination of Level 1B Full-Resolution and Reduced-Resolution data from ESA's 4th reprocessing cycle. Full-resolution passes are prioritized and selected whenever available to ensure the highest possible spatial detail. The OLCI source data are L1B at full resolution. From v2.0 of the CRDP onwards, Aqua-MODIS data are processed and included in the dataset to fill the gap between April 2012 and April 2016. These data are only included where MODIS-Aqua time-series are consistent with MERIS and OLCI-A/B during a period of overlapping observations of three-years before and after the gap. The number of lakes for which is the case is much smaller (order of 5%) than for the periods covered by MERIS and OLCI sensors.

Table 5 Lakes with uninterrupted LWLR coverage

Region name	Lake name	Continents	Area(km²)	Latitude	Longitude
GLWD00000012	Erie	North America	25938	42.144	-81.238
GLWD00000015	Ontario	North America	19842	43.826	-77.635
GLWD00000020	Titicaca	South America	7753	-15.916	-69.303
GLWD00000040	Zaysan	Asia	4334	48.717	83.432
GLWD00000041	Qinghai	Asia	4250	36.891	100.047
GLWD00000051	Van	Asia	3594	38.656	42.807
GLWD00000053	Uvs	Asia;Europe	3630	50.338	92.762
GLWD00000058	Alakol	Asia	2970	46.119	81.685
GLWD00000064	Boeng Tonle Chhma	Asia	2526	12.830	104.070
GLWD00000103	Williston	North America	1675	56.049	-124.075
GLWD00000121	Khyargas	Asia	1400	49.145	93.449
GLWD00000131	Qapshaghay Bogeni	Asia	1279	43.830	77.635
GLWD00000135	Sevan	Asia	1246	40.394	45.349
GLWD00000143		North America	1354	67.873	-97.719
GLWD00000146	Saint Clair	North America	1235	42.494	-82.703
GLWD00000179	Zhari Namco	Asia	985	30.908	85.607
GLWD00000181	Bosten	Asia	1162	41.983	87.064
GLWD00000215	Tangra	Asia	837	31.060	86.576
GLWD00000239	Ulungar	Asia	865	47.219	87.212
GLWD00000241	Sarykamyshskoye	Asia	3853	41.961	57.360
GLWD00000244	Dead Sea	Asia	644	31.534	35.478
GLWD00000249	Hirakud	Asia	616	21.658	83.734
GLWD00000250	Manych oudilo	Europe	753	46.274	42.890
GLWD00000253	Guillaume-Delisle	North America	700	56.323	-76.303
GLWD00000296	Chardarinskoye	Asia	387	41.138	68.122
GLWD00000310	Balaton	Europe	584	46.883	17.846
GLWD00000312	Ayakkum	Asia	738	37.543	89.399
GLWD00000327	Geneva	Europe	581	46.362	6.394
GLWD00000352	Bodensee	Europe	476	47.648	9.344
GLWD00000375	Flathead-Lake	North America	498	47.884	-114.125
GLWD00000382	Qyaring-Lake	Asia	481	31.135	88.355
GLWD00000383	Taro-Lake	Asia	479	31.132	84.122
GLWD00000390	Egirdir	Asia	463	38.060	30.894
GLWD00000396	Mcalpine-Lane	North America	475	66.515	-102.646
GLWD00000403	Markakol	Asia	456	48.741	85.735
GLWD00000409	Primrose	North America	435	54.892	-109.744
GLWD00000414	Cienaga-Grande-De-Santa-Marta	South America	741	10.863	-74.462

GLWD00000437	Mica	North America	401	52.147	-118.435
GLWD00000447	Kluane	North America	415	61.315	-138.783
GLWD00000469	Netsilik	North America	413	69.248	-93.116
GLWD00000488	Scutari	Europe	364	42.193	19.296
GLWD00000511	Aqqikkol	Asia	374	37.069	88.428
GLWD00000520	Chibzhang-Lake	Asia	485	33.490	90.362
GLWD00000536	Bangong	Asia	354	33.722	79.341
GLWD00000545	Cormorant	North America	349	54.222	-100.910
GLWD00000546	Pend-Oreille	North America	380	48.134	-116.375
GLWD00000549	Dorgon	Asia	346	47.747	93.408
GLWD00000554	Clark	North America	338	60.233	-154.299
GLWD00000609	Hoh Xil	Asia	306	35.583	91.092
GLWD00000649	Puma-Yumco	Asia	291	28.560	90.401
GLWD00000690	Prespa	Europe	265	40.885	21.035
GLWD00000718	Gozha	Asia	259	35.033	81.087
GLWD00000744	Ta-tse	Asia	249	31.901	87.540
GLWD00000812	Lixi'Oidain	Asia	230	35.753	90.174
GLWD00000817	Great Bitter	Africa	218	30.297	32.392
GLWD00000892	Fuxian	Asia	215	24.497	102.886
GLWD00000893	Neuchatel	Europe	216	46.898	6.843
GLWD00000913	Hsu-ju	Asia	209	30.277	86.409
GLWD00000982	Sangiyn Dalay	Asia	184	49.222	99.108
GLWD00000983	Rinqin-Xub-Lake	Asia	187	31.276	83.467
GLWD00001066	Bam	Asia	236	31.258	90.587
GLWD00001076	Dukan	Asia	116	36.103	44.925
GLWD00001079	Cuona	Asia	192	32.026	91.474
GLWD00001110	Hirfanli	Asia	220	39.145	33.673
GLWD00001128	Walker	North America	142	38.698	-118.717
GLWD00001130	Toson	Asia	137	37.145	96.942
GLWD00001192	Kyebxang-Lake	Asia	155	32.459	89.985
GLWD00001196	Sea of Galilee	Asia	164	32.799	35.589
GLWD00001354	Memar	Asia	139	34.210	82.315
GLWD00001529	Trasimeno	Europe	121	43.137	12.103
GLWD00001534	Dawa	Asia	118	31.237	84.965
GLWD00001627	Azuei	North America	118	18.572	-72.001
HYLA00001704	Hawea	Oceania	148	-44.445	169.318

The observed area per lake is defined by the maximum water extent following the ESA Land Cover CCI v4.0 water mask. Within this water extent, land/water/ice/cloud classification is performed. Areas outside the maximum water extent are not considered here.

3.6.4 LWLR data limitations

The LWLR data set is provided for every pixel recognised as water from the optical observations. However, at the spatial resolution of the satellite sensor (nominally 300-1km) there is a risk of 'mixed pixels': water in which small fractions of land, cloud, ice or vegetation are included. The water classification is set up to be restrictive: when classified as water, other influences are likely to be minor. However, given the difference in optical contrast between water and other features even minor differences can lead to large estimation errors, which are then most likely to be found near land (and islands), cloud fringes and floating or shoreline vegetation. Users are advised to consider their application of the data with care, and e.g. consider removing data near shorelines to answer questions about long-term change.

Mixed-pixel effects are also observed during periods of freeze-up and ice melting, when thin or partial ice cover may not be efficiently recognised. From v2.1 onwards, outlier values that are associated with (partial) ice cover, observed during periods of low water temperature, are flagged as 'lwlr_poor_consistency' or 'lwl_low_consistency' respectively when either all or some of the observations within a pixel were affected.

The major source of uncertainty in lake optical water quality estimation is the separation of water-leaving radiance from atmospheric effects. The latter increase in severity nearby land and this 'adjacency effect' can, depending on the state of the atmosphere, extend several kilometres. This has two major influences on the observations:

- Longer wavebands tend to be more affected by mixing of reflectance from land and water, because the contrast in reflectance from these features is largest in these wavebands. This affects particularly the retrieval of TSM/turbidity (which relies on near-infrared reflectance) and chlorophyll-a at concentrations > 10 mg m^{-3} (which uses red and near-infrared wavebands).
- Elevated near-infrared reflectance introduced from nearby land may lead to over-correction of LWLR at shorter wavebands.

The shape of the LWLR spectrum has been analysed to identify likely influence of nearby land. When this is detected, a quality flag (lwlr_land_contaminated) is raised, to be used at the discretion of the data user. Particularly in small lakes this flag may be raised for a considerable fraction of observations, depending on the thickness of the atmosphere.

Per-pixel uncertainty estimates are currently modelled as a function of algorithm performance per optical water type (see the E3UB document for details). Other effects such as proximity to land have not been included in this model, but will likely explain a significant fraction of the overall product uncertainty. This implies that the uncertainties in open water are likely to be a worst-case estimate at present.

3.7 Lake Storage change (LSC)

3.7.1 LSC definition and usage

Lake Storage Change – also noted LSC - refers to the fluctuation in the amount of water stored in a lake or reservoir over time. This change can be caused by various factors, like precipitation, evaporation, seasonal variation, climate change or anthropic pressure. A change over the years in the LSC might thus be a marker of the impact of one or more of these factors. In short, lake storage change is the net result of all these factors combined and is often monitored in the context of water resource management, hydrology, and environmental studies.

3.7.2 LSC data characteristics

The LSC product is a time series representing the water volume variation of a given water body with time, along with an information on its reliability. That is why the LSC product is composed of two variables: the lake storage change estimation and its confidence index. The details for the estimation of those variables are fully described in the Algorithm Theoretical Basis Document (ATBD) and the End-to-End ECV Uncertainty Budget (E3UB). The LSC time frequency depends on Lake Water Level (LWL) frequency, itself depending on the altimetry satellite(s) orbit(s) overpassing the water body of interest.

Satellite altimeters and optical instruments are used to determine the water level and surface extent time-series of lakes and reservoirs with a high precision over time. The time-series are produced on a lake-by-lake basis, with volume changes provided synchronously to the LWL measurement date.

3.7.3 LSC data sources

The methodology used to derive the LSC time-series relies on two different approaches, and hence on slightly different data sources. Indeed, the lakes are split between those that are surface-varying and those that are or not.

For most surface-varying lakes, LSC is derived using the related hypsometric curve. This curve is established from LWL that use satellite altimetry, and from the Lake Water Extent (LWE) that use optical or SAR data. In case LWL and LWE are not or cannot be determined, the combined information from water body borders from optical imagery and Digital Elevation Model data can be an alternative to estimate the hypsometric curve, and hence the LSC. Combining this information with topographic data allows us to extrapolate changes in the volume of water in a lake over even longer time spans (up to 1984 if needed) along with height information when altimetry is not available.

For lakes that are considered as “static” (not surface-varying), the LSC is estimated from a constant reference surface (in general taken from the HydroLakes database, Messager et al. 2016) and the LWL time-series product.

3.7.4 LSC data limitations

Derived from LWL and LWE, LSC time series inherits the data limitations from these two datasets exposed in previous sections. For example, for LWL, we consider that the water level is the same throughout all the lake. This assumption is true for most of the lakes.

Otherwise, the main limitation of this product relies on the existence of LWL time-series on the water body, which drives in almost all cases the LSC time-series production. In case of surface-unvarying lakes, the absence of LWL time-series means that no long times-series of LSC can be produced.

When water borders are used on DEM for LSC estimation, the quality of the DEM is obviously one limitation. But more than that, the dates of the data acquisitions used for DEM production is key in case the water level rises or drops with time. For instance, if a water body has a rising water level trend since the year 2000 (case of SRTM dataset), water contours extracted after this date would be usable, but not before this date.

3.8 LSWT and LIC gap filled product

3.8.1 Gap-filled product definition and usage

The LSWT and LIC observations of the Lakes_cci product contain gaps mainly due to cloud cover and incomplete satellite track coverage. The gap-filled product is a spatial and temporal reconstruction of LSWT and LIC for the 2024 Lakes_cci lakes covering the period from 2000 to 2023. Gap-filled LSWT/LIC are typically more useful for data analysis as dealing with missing data can be challenging. In addition, LSWT and LIC products are used for calibrating, training and validating models which generally require gap-free data.

3.8.2 Gap-filled product data characteristics

The primary variables are the reconstructed LSWT and LIC fields. For the reconstructed LSWT three other LSWT variables are provided: (1) an estimate of uncertainty is provided per datum; (2) A flag indicating the source of LSWT reconstruction; (3) the fraction of the lake with LSWT observations from the LSWT Lake_cci v3.0.0 product that have been used for the reconstruction. For the reconstructed LIC there will be four other variables provided: (1) the probability for the reconstructed data, (2) fractional ice cover extent per lake, and (3) the area of the lake covered by ice in km^2, (4) the fraction of the lake that used the reconstructed LIC product.

The provided LSWT and LIC data are on the same regular latitude-longitude grid as the Lakes_cci v3.0.0 products, and they have been harmonised to be consistent in space.

3.8.3 Gap-filled product data sources

The data source for both LSWT and LIC are the observations of the Lakes_cci v3.0.0 products. The LSWT reconstructions are tested using two algorithms in different configurations: DINEOF (Data Interpolating Empirical Orthogonal Function) which has been used as a gap-filling technique in the precursor ESA ARC-Lake project, and DINCAE (Data Interpolating Convolutional Auto-Encoder) which is a recent neural network algorithm developed and optimized for geophysical data. For LIC reconstructions, simple spatial (3x3 window) and temporal (3-day window) inference methods are deployed to reconstruct roughly 20% of obstructions. The remaining reconstruction is tested using a spatiotemporal weighted distance method which has been used for gap-filling MODIS snow cover products, and a custom 3D CNN (Convolutional Neural Network) trained to remove cloud-cover.

4 Lakes ECV Dataset

4.1 Definition

The Climate Research Data Package (CRDP) intended to fulfil the Lakes ECV observation challenge is a merged ('L3S') product composed of the thematic products described in the previous sections:

- Lake water level (LWL)
- Lake water extent (LWE)
- Lake Ice Cover (LIC)
- Lake Surface Water Temperature (LSWT)
- Lake Water Leaving Reflectance (LWLR)
- Lake Storage Change (LSC)

In addition, a Lake Ice Thickness product for 14 lakes is provided in separate files.

The global coverage of CRDP v3.0.0 is shown in Figure 2. A list of lakes including their location and thematic ECV data availability, can be found on the project website in comma-separated text format.

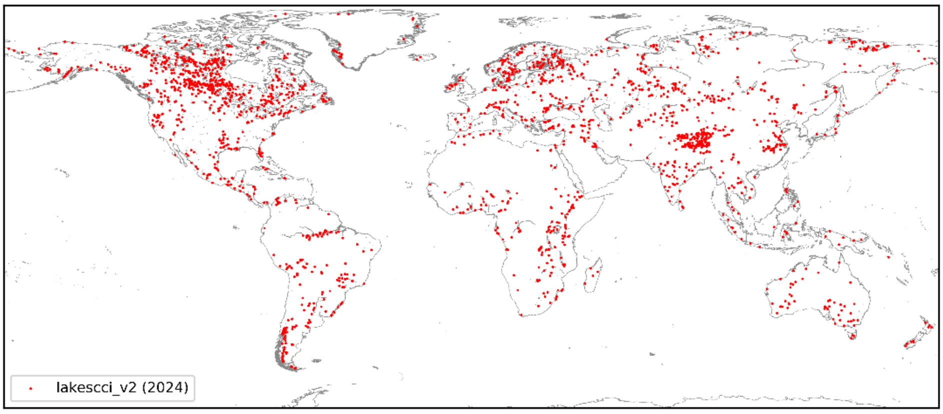

Figure 2: Inland waterbodies included in CRDP v3.0

4.2 Main characteristics

Data generated in the Lakes_cci project are derived from multiple sensors and satellites (for details see the Product Specification Document (PSD)) and, consequently, different temporal and spatial resolutions. One of the objectives in Lakes_cci project is the harmonisation of the different products as a single dataset with the following characteristics:

- Daily aggregation interval (products are specified as 12:00:00 UTC).
- Grid format with spatial resolution of 1/120 degrees (near 1 km at the equator).
- Variables not produced in grid format (LWL and LWE) are duplicated in the grid for the area given under the nominal spatial delineation of that lake.
- Common 1/120 degree grid (latitude and longitude)
- Common regions of interest. The full set of lake definitions is made available as a set of polygons and on the global grid equivalent to the CRDP. The definitions are based on the maximum water extent V4.0 maps from ESA Land Cover. The grid representation of the lake definitions also contains the distance to the nearest land for each lake pixel.

- Extent: -180 to 180 degrees longitude, -90 to 90 degrees latitude, where positive signs point north and east. The pixel coordinate is the centre of the pixel.

Uncertainty estimates are provided for each product. Procedures to derive uncertainty estimates are provided in the End-to-End Uncertainty Budgets document (E3UB).

The sequential identifiers of the lake regions follows the Global Lakes and Wetlands Database and HydroLakes databases, respectively. Where these databases overlap or where their delineations of water bodies is incomplete, the numbering follows the provenance of projects where they were initially defined (Table 6), so that users can combine old and new datasets.

Table 6 Provenance of lake identifiers

ID	Alphanumeric ID	Provenance
000000000 – 099999999	GLWD00000000 – 99999999	Global Lakes and Wetlands Database
100000000 – 199999999	GLBL00000000 – 99999999	GloboLakes project
200000000 – 299999999	CGL200000000 – 99999999	Copernicus Land Monitoring Service
300000000 – 399999999	HYLA00000000 – 99999999	Hydrolakes

4.3 Nomenclature

File naming in the Lakes_cci follows the CCI standard formatting:

ESACCI-LAKES-<Processing Level>-<Data Type>-<Product String>-<Indicative Date>-fv<Version>.nc

Where:

Processing Level: L3S, meaning reprojected and super-collated. Observations from multiple instruments and observation times are combined into a common spatiotemporal grid.

Data type: LK_PRODUCTS

Product String: MERGED, meaning data combined from more than one platform / sensor.

Indicative Date: YYYYMMDD format, the date of the observations.

Version: 3.0.0

Thus, an example file name in the first data release is:

ESACCI-LAKES-L3S-LK_PRODUCTS-MERGED-20190214-fv3.0.0.nc

For LIT products, delivered in separated files, the nomenclature is as follow:

ESACCI-LAKES-L3S-LK_PRODUCTS-LIT-<LakeName>-Pass<Pass number>.nc with LakeName and Pass number as specified in Section 3.4

4.4 Format

The Lakes_cci dataset is stored in the NetCDF 4 classic format (Network Command Data Form) using the CF (Climate and Forecast) metadata convention (v1.8) and CCI Data Standards (v2.3).

The following sections describe the components of each NetCDF for the merged product and for the LIT files.

4.4.1 Merged product

4.4.1.1 Global attributes

The global attributes provide general information about the product. The Lakes_cci global attributes are those recommended in the CF standards (Table 7)

Table 7. CF Global Attributes

Attribute Name	Attribute description
title	description of the dataset
institution	where the data were produced
source	original data source
history	processing history of the dataset
references	reference website

Additionally, the NetCDF files contain the recommended global attributes for dataset discovery and additional attributes defined in the CCI data standards v2.3 (Table 8).

Table 8. CCI Global Attributes

Attribute Name	Attribute description
tracking_id	Universal Unique Identifier (randomly generated)
conventions	CF and CCI Data Standards
product_version	Lakes_cci merged product version
summary	description of the dataset
keywords	list of keywords
keywords_vocabulary	science keywords
id	filename
naming authority	lakes.esa-cci
cdm_data_type	Grid
date_created	Date of file creation
creator_name	ESA Lakes_cci
creator_url	http://cci.esa.int/lakes
creator_email	lakes_cci@groupcls.com
project	Climate Change Initiative - European Space Agency
geospatial_lat_min	-90.0
geospatial_lat_max	90.0
geospatial_lon_min	-180.0
geospatial_lon_max	180.0
geospatial_vertical_min	NA
geospatial_vertical_max	NA
time_coverage_start	Start of observations in ISO8601 format
time_coverage_end	End of observations in ISO8601 format
time_coverage_duration	File time coverage duration in ISO8601 format

Attribute Name	Attribute description
standard_name_vocabulary	NetCDF Climate and Forecast (CF) Metadata Convention
License	ESA CCI Data Policy: free and open access
Platform	list of satellites used in this data file
Sensor	list of sensors used in this data file
spatial_resolution	1 km at Equator
key_variables	water_surface_height_above_reference_datum, lake_surface_water_extent, lake_ice_cover_class, lake_surface_water_temperature, chla, tsm, acdom440, phycocyanin, Kd490, Kd560, Kd665, KdPAR, and Rw[xxx] where xxx is one of 400, 412, 443, 469, 490, 510, 531, 547, 560, 620, 645, 665, 674, 681, 709, 754, 779, 859, 885, 900, 1020nm
geospatial_lat_units	degrees north
geospatial_lon_units	degrees east
geospatial_lat_resolution	0.0083333
geospatial_lon_resolution	0.0083333
doi	Digital Object Identifier of the dataset

Appendix A -contains an example of these global attributes.

4.4.1.2 Dimensions

Following the CCI data standards, the gridded products of the Lakes_cci have three dimensions: time, latitude and longitude. As indicated in chapter 4.2, for reasons of consistency, all the included variables share the same dimensions.

4.4.1.3 Variables

The attributes of the variables in the NetCDF files follow the CCI data standards guidelines and consequently, the CF recommendations:

- standard_name: standard name if it exists in the CF convention
- long_name: description of the variable in human-readable format
- units: units of the variable
- valid_min: smallest value to be considered valid
- valid_max: largest value to be considered valid
- _FillValue: the value used to indicate lack of data
- scale_factor (optional): multiplicative factor for packing data
- add_offset (optional): additive offset for packing data
- comment (optional): Miscellaneous information for the user, such as the meaning of product quality flags
- grid mapping
- ancillary_variables: indicated in the primary variable
- flag_values and flag meanings: defined for flag variables

The variables included in the dataset are listed in Table 9

Table 9. List of variables in the NetCDF file

Variable Name	Lakes_cci product
water_surface_height_above_reference_datum	LWL
lwl_uncertainty	LWL
lwl_quality_flag	LWL
lake_surface_water_extent	LWE
lwe_uncertainty	LWE
lwe_quality_flag	LWE
Lake_ice_cover_flag	LIC
lake_ice_cover_class	LIC
lake_ice_cover_uncertainty	LIC
lake_surface_water_temperature	LSWT
lswt_uncertainty	LSWT
lswt_quality_level	LSWT
Rw[xxx]*	LWLR
Rw[xxx]_uncertainty_bias*	LWLR
Rw[xxx]_uncertainty_rmsd*	LWLR
lwlr_quality_flag	LWLR
chla	LWLR
chla_uncertainty	LWLR
tsm	LWLR
tsm_uncertainty	LWLR
acdom440	LWLR
acdom440_uncertainty	LWLR
Kd[xxx]**	LWLR
phycoyanin	LWLR
lake_storage_change	LSC
lsc_quality_flag	LSC

* where xxx is one 400, 412, 443, 469, 490, 510, 531, 547, 560, 620, 645, 665, 674, 681, 709, 754, 779, 859, 885, 900, 1020nm.

** where xxx is one of 412, 560, 674 or PAR

4.4.1.4 Data availability

As described in section 1.2, data used for the generation of the lakes_cci product come from multiple instruments in different missions. Thus, the temporal resolution and spatial coverage are not the same for all variables. Figure 3 to Figure 5 show the spatial coverage of the data for LWL, LWE and LSWT respectively. Temporal time step for LWL and LWE is satellite dependant, ranging from daily for lakes observed by multiple missions/tracks to 27 days for lakes being observed by sentinel-3A/B missions. LIC data is available since 2001 for all lakes with a flag indicating whether a lake do not form ice. LWLR data is available for all lakes since 2002 with a reduction in cover during 2012-2016. The data availability is classified in five categories:

- Lakes with data available at least 50% percent of the temporal coverage (16 years) – dark green dots
- Lakes with data available less than 50% of the temporal coverage – light green dots
- Lakes no covered by satellite instruments – black dots
- Lakes analysed with rejected data because the low quality of the measurements – red dots

- Lakes still under analysis – orange dots

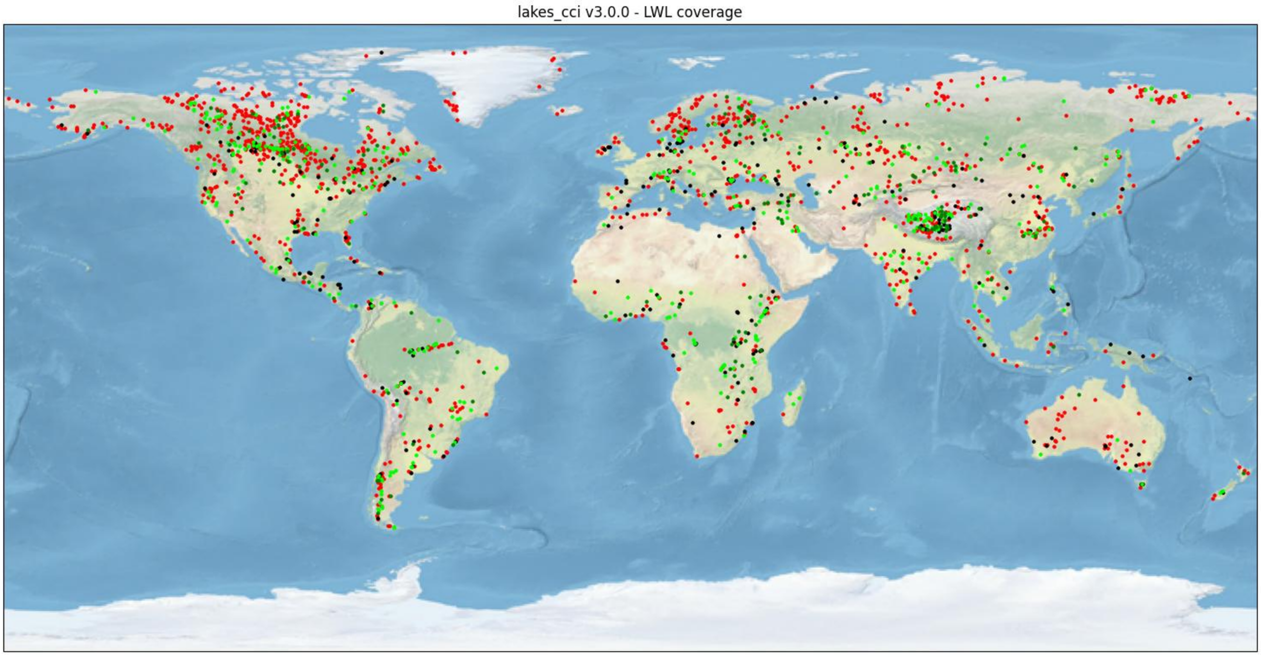

Figure 3. Lake Water Level spatial coverage (starting in 1992)

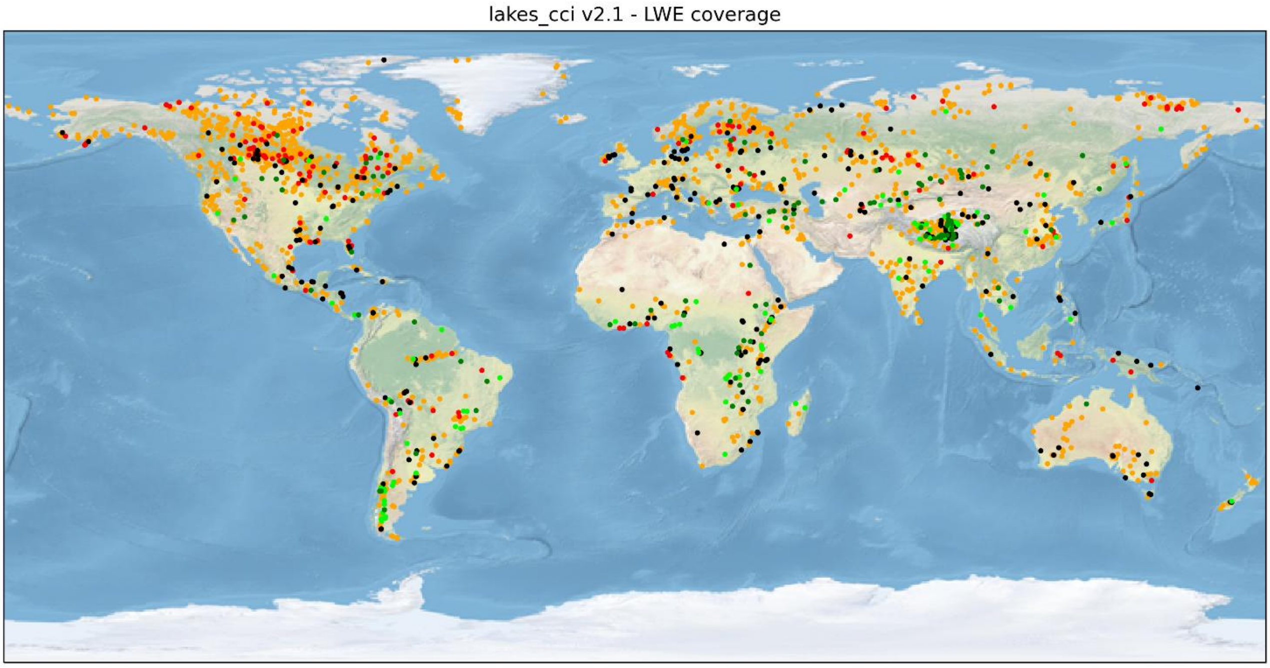

Figure 4. Lake Water Extent- Spatial coverage (starting in 1992)

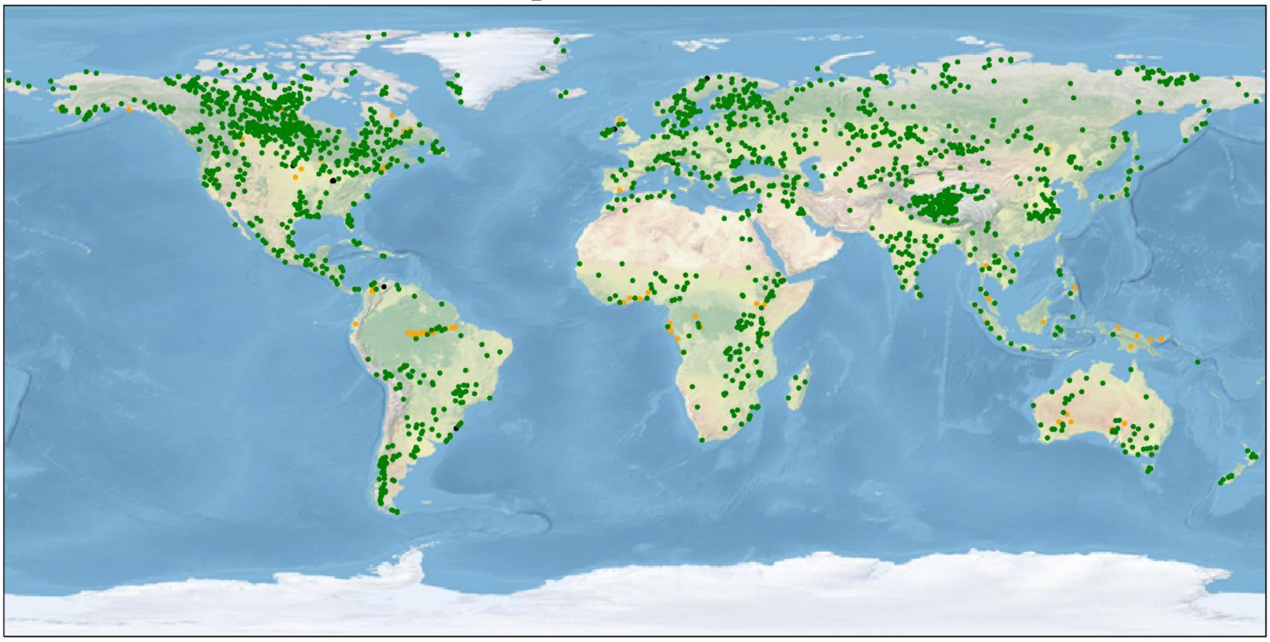

Figure 5. Lake Surface Water Temperature – Spatial coverage (starting in 1995)

4.4.2 LIT Product

4.4.2.1 Global attributes

As for the merged products, the LIT files contain global attributes recommended in the CF standards (Table 10) as well as those recommended by CCI data standards (DataStandards_v2.3)

Table 10. LIT CF Global Attributes

Attribute Name	Attribute description
title	description of the dataset
institution	where the data were produced
source	original data source
history	processing history of the dataset
references	reference website
summary	description of the dataset
keywords	list of keywords
id	filename.nc
naming_authority	lakes.esa-cci
comment	These data were produced for the ESA Lakes_cci project
creator_name	ESA Lakes_cci
creator_email	lakes_cci@groupcls.com
project	Climate Change Initiative - European Space Agency
standard_name_vocabulary	NetCDF Climate and Forecast (CF) Metadata Convention
license	ESA CCI Data Policy: free and open access
platform	Jason1, Jason2, Jason3

Attribute Name	Attribute description
key_variables	LIT
latitude	latitude of the center of the LIT ROI:
longitude	longitude of the center of the LIT ROI
time_coverage_start	First time available in the file in ISO8601 format
time_coverage_end	Last time available in the file in ISO8601 format
doi	Digital Object Identifier of the dataset

4.4.2.2 Dimensions

The only dimension for LIT products is time. Each file contains measurements for a complete time series.

4.4.2.3 Variables

There are three variables in the LIT files:

- Lit: Lake Ice Thickness
- Lit_uncertainty expressed in meters.
- Lit_quality_flag: good, bad_no_data, degraded_fit_performance

For the first two variables, the units and valid range are in the variable attributes. For the quality_flag variable, the values and the meaning of the flags are indicated in the CF recommendations.

5 Gap-filled LSWT and LIC Dataset

5.1 Definition

The gap-filled product is composed of two of the thematic products in the CRDP Lakes_cci:

- Lake Ice Cover (LIC)
- Lake Surface Water Temperature (LSWT)

The global coverage of the gap-filled product is the same as for the CRDP as shown in Figure 1. A list of lakes including their location can be found on the project website in comma-separated text format.

5.2 Main characteristics

The gap-filled data are generated from the Lakes_cci v3.0 LSWT and LIC dataset. One of the objectives of the production is the harmonisation of the LSWT and LIC such that LSWT will be available only for water classified pixels and not for land or ice classified pixels.

The gap-filled products are offered as netCDF files:

- daily global files containing the LSWT+LIC variables for the 2024 lakes on a given day
- per-lake files containing LSWT+LIC variables for a specific lake for each day between 2000 and 2023.

An auxiliary file containing the list of the 2024 lakes with some information to locate files and lakes will be provided.

The dataset has the following characteristics valid for both the global daily and per-lake files:

- Daily aggregation interval pinned to 12:00:00 UTC.
- Grid format with spatial resolution of 1/120 degrees (near 1 km at the equator)
- Some of the auxiliary variables are not produced in grid format but contains a value per each time.
- Extent: -180 to 180 degrees longitude, -90 to 90 degrees latitude, where positive signs point north and east. The pixel coordinate is the centre of the pixel.

An uncertainty estimates are provided for each product and the procedures to derive uncertainty estimates will be provided.

5.3 Nomenclature

File naming in the Lakes_cci follows the CCI standard formatting:

ESACCI-LAKES-<Processing Level>-<Data Type>-<Product String>-<Indicative Date>-fv<Version>.nc

Where:

Processing Level:	L4, meaning analysis data, a product with no gaps which have been filled with interpolation/statistical techniques.
Data type:	LK_PRODUCTS
Product String:	MERGED, meaning data combined from more than one platform / sensor.
Indicative Date:	YYYYMMDD format, the date of the observations.
Version:	1.0.0

Thus, an example file name for the global daily file:

ESACCI-LAKES-L4-LK_PRODUCTS-MERGED-20080808-fv1.0.0.nc

An example file name for the per-lake file:

ESACCI-LAKES-L4-LK_PRODUCTS-PERLAKE_ID000000002-2000_2023-fv1.0.0.nc

5.4 Format

The gap-filled dataset like the main Lakes_cci dataset is stored in the NetCDF 4 classic format (Network Command Data Form) using the CF (Climate and Forecast) metadata convention (v1.8) and CCI Data Standards (v2.3).

The following sections describe the components of each NetCDF for the daily global and for the per-lake files.

5.4.1 Global attributes

The global attributes provide general information about the product. The gap-filled file global attributes are those recommended in the CF standards (Table 5).

Additionally, the NetCDF files contain the recommended global attributes for dataset discovery and additional attributes defined in the CCI data standards v2.3 (Table 6).

Table 10. CCI gap-filled file Global Attributes

Attribute Name	Attribute description
tracking_id	Universal Unique Identifier (randomly generated)
conventions	CF and CCI Data Standards
product_version	Lakes_cci merged product version
summary	description of the dataset
keywords	list of keywords
keywords_vocabulary	science keywords
id	filename
naming authority	lakes.esa-cci
cdm_data_type	Grid
date_created	Date of file creation
creator_name	ESA Lakes_cci
creator_url	http://cci.esa.int/lakes
creator_email	lakes_cci@groupcls.com
project	Climate Change Initiative - European Space Agency
geospatial_lat_min	-90.0 *[For the per-lake file it is the lat min of the box around the lake]*
geospatial_lat_max	90.0 *[For the per-lake file it is the lat max of the box around the lake]*
geospatial_lon_min	-180.0 *[For the per-lake file it is the lon min of the box around the lake]*
geospatial_lon_max	180.0 *[For the per-lake file it is the lon max of the box around the lake]*
geospatial_vertical_min	NA
geospatial_vertical_max	NA
time_coverage_start	Start of observations in ISO8601 format
time_coverage_end	End of observations in ISO8601 format
time_coverage_duration	File time coverage duration in ISO8601 format
standard_name_vocabulary	NetCDF Climate and Forecast (CF) Metadata Convention
License	ESA CCI Data Policy: free and open access
spatial_resolution	1 km at Equator
key_variables	analysis_lswt, analysis_lake_cover_class
geospatial_lat_units	degrees north
geospatial_lon_units	degrees east
geospatial_lat_resolution	0.0083333
geospatial_lon_resolution	0.0083333
doi	Digital Object Identifier of the dataset

5.4.2 Dimensions

Following the CCI data standards, the gridded products of the Lakes_cci have three dimensions: time, latitude and longitude. Some of the variables in the gap-filled product have only the time dimension.

5.4.3 Variables

The attributes of the variables in the NetCDF files follow the CCI data standards guidelines and consequently, the CF recommendations:

- standard_name: standard name if it exists in the CF convention
- long_name: description of the variable in human-readable format
- units: units of the variable
- valid_min: smallest value to be considered valid
- valid_max: largest value to be considered valid
- _FillValue: the value used to indicate lack of data
- scale_factor (optional): multiplicative factor for packing data
- add_offset (optional): additive offset for packing data
- comment (optional): Miscellaneous information for the user, such as the meaning of product quality flags
- grid mapping
- ancillary_variables: indicated in the primary variable
- flag_values and flag meanings: defined for flag variables

The variables included in the dataset are listed in Table 11.

Table 11. List of variables in the NetCDF file for the gap-filled product

Variable Name	Lakes_cci product
analysis_lake_cover_class	LIC
analysis_lic_probability	LIC
fractional_ice_cover	LIC
ice_cover_area	LIC
analysis_lic_fraction_reconstructed	LIC
analysis_lswt	LSWT
analysis_lswt_flag	LSWT
analysis_lswt_uncertainty	LSWT
analysis_lswt_fraction_obs	LSWT
lake_cci_id	lake mask
distance_to_land	lake mask

6 Supporting software

Lakes_cci data are stored in the NetCDF4 classic file format. A wide choice of software packages can be used to visualise or manipulate the NetCDF data. A list of software is provided on the Unidata web site[1].

There are several ways in which you can explore, download, and analyse the Lakes_cci datasets.

- The data are hosted at CEDA, where you find the full global datasets as well as the lake mask with the maximum extent. CEDA offers multiple download mechanisms and ways to extract a set of variables, or data for a specific region.
- To explore some of the ways in which you can extract data for a specific lake or region, we offer a series of python scripts and a jupyter notebook. These are intended to help new users familiarise themselves with data extraction for further analysis: https://github.com/cci-lakes/lakes_cci_tools
- For interactive visualisation of the dataset you can use the WebGIS. The light-weight GIS provides visualisation of most variables (Figure 6), and has functionality to extract small sections of the data for download or plotting (for larger requests, please use the CEDA tools). This is also a useful resource for training and education, e.g. by sharing links to your visualisation with others: https://lakescci.eofrom.space

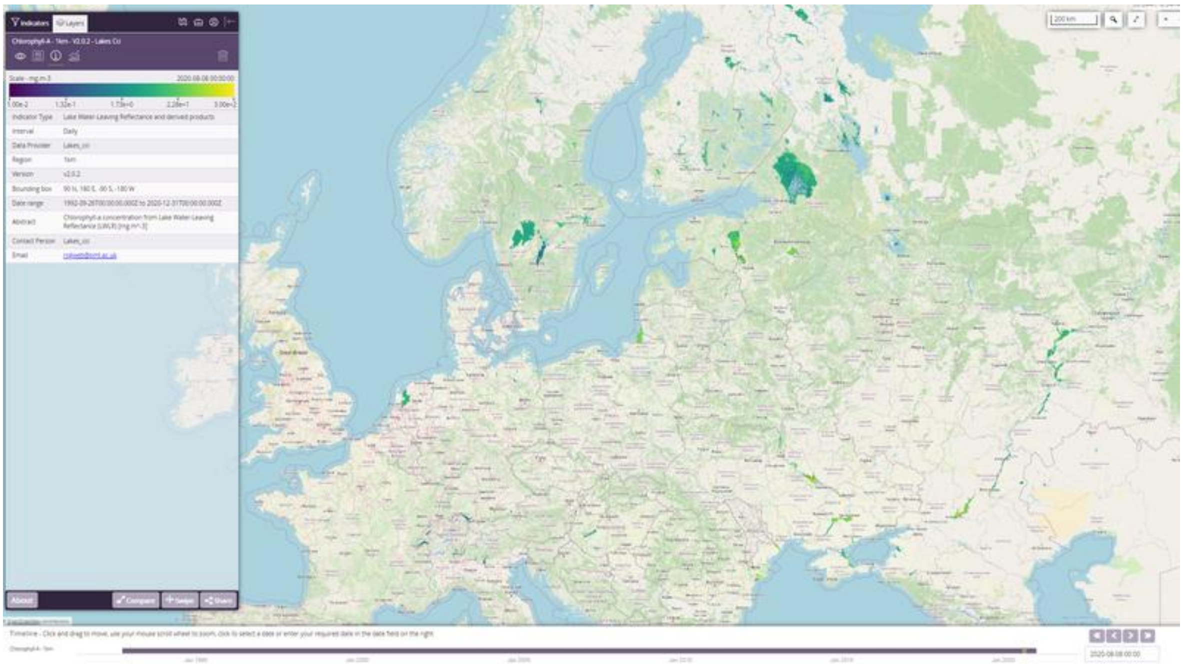

Figure 6. Lakes_cci light-weight GIS

The Lakes_cci Global Climate Indicators Dashboard (Figure 7) provides intuitive visualization of change in lakes using a range of aggregation methods (e.g. country, continent, altitude, population pressure). Current indicators include the Lake Surface Water Temperature anomaly, lake heat wave occurrences and change in ice-covered area. The dashboard is hosted at https://lakescci.climate-indicators.brockmann-consult.de and will be extended with additional indicators and functionalities. Anyone interested in additional climate indicators is welcome to contact the developers at info@brockmann-consult.de

[1] https://www.unidata.ucar.edu/software/netcdf/software.html

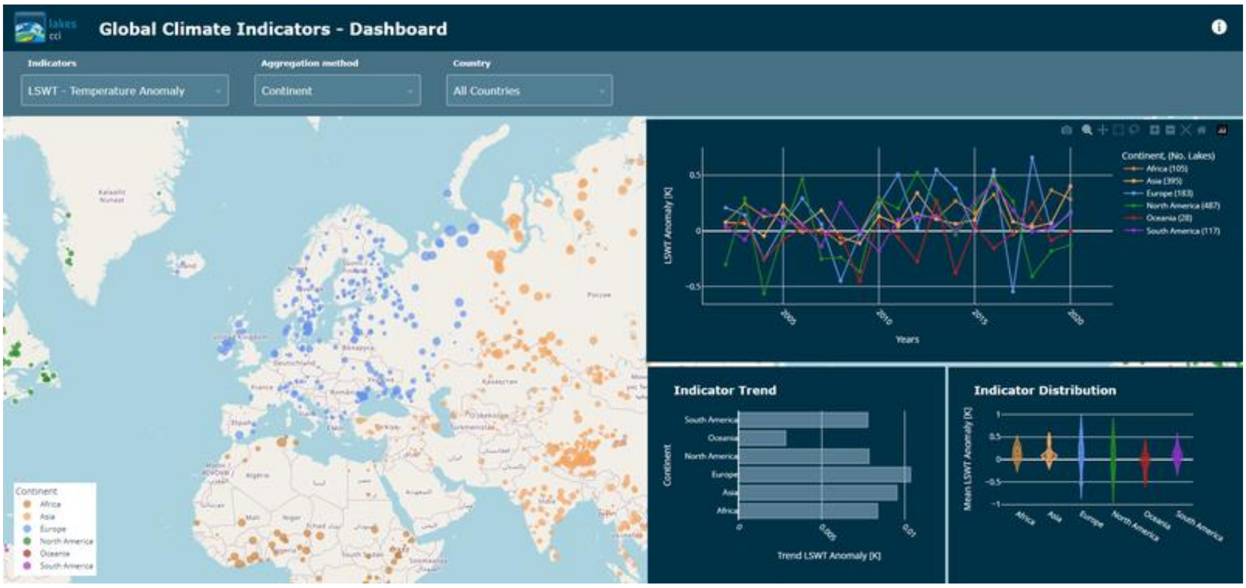

Figure 7. Lakes_cci Global Climate Indicators Dashboard

7 References

Arsen, A.; Crétaux, J.-F.; Berge-Nguyen, M.; Del Rio, R.A. Remote Sensing-Derived Bathymetry of Lake Poopó. Remote Sens.2014, 6, 407-420. https://doi.org/10.3390/rs6010407

Bonsal, B.R., T.D. Prowse, C.R. Duguay, and M.P. Lacroix, 2006. Impacts of large-scale teleconnections on freshwater-ice duration over Canada. *Journal of Hydrology*, 330, 340-353.

Brown, L.C. and C.R. Duguay, 2010. The response and role of ice cover in lake-climate interactions. *Progress in Physical Geography*, 34(5), 671-704, doi: 10.1177/0309133310375653.

Busker, T., de Roo, A., Gelati, E., Schwatke, C., Adamovic, M., Bisselink, B., Pekel, J.-F. & Cottam, A. (2019) « A global lake and reservoir volume analysis using a surface water dataset and satellite altimetry », Hydrol. Earth Syst. Sci., vol. 23, no. 2, pp. 669–690,https://doi.org/10.5194/hess-23-669-2019, 2019.

Crétaux, JF., Abarca-del-Río, R., Bergé-Nguyen, M. Arsen A., Drolon V., Clos G & Maisongrande P., 2016 , Lake Volume Monitoring from Space. Surv Geophys 37, 269–305 (2016). https://doi.org/10.1007/s10712-016-9362-6

Derksen, C., D. Burgess, C. Duguay, S. Howell, L. Mudryk, S. Smith, C. Thackeray, and M. Kirchmeier-Young, 2019. Changes in snow, ice, and permafrost across Canada; Chapter 5 in *Canada's Changing Climate Report*, (ed.) E. Bush and D.S. Lemmen; Government of Canada, Ottawa, Ontario, p. 194-260.

Duguay, C.R., M. Bernier, Y. Gauthier, and A. Kouraev, 2015. Remote sensing of lake and river ice. In *Remote Sensing of the Cryosphere*, Edited by M. Tedesco. Wiley-Blackwell (Oxford, UK), pp. 273-306.

Duguay, C.R., T.D. Prowse, B.R. Bonsal, R.D. Brown, M.P. Lacroix, and P. Ménard, 2006. Recent trends in Canadian lake ice cover. *Hydrological Processes*, 20, 781-801.

Ghiasi, Y., C.R. Duguay, J. Murfitt, A. Thompson, J. van der Sanden, H. Drouin, and C. Prévost, 2020. Application of GNSS interferometric reflectometry for the estimation of lake ice thickness. *Remote Sensing*, 12, 2721; doi:10.3390/rs12172721.

Mangilli, A., P. Thibaut, C.R. Duguay, and J. Murfitt, 2022. A new approach for the estimation of lake ice thickness from conventional radar altimetry. *IEEE Transactions on Geoscience and Remote Sensing*, 60: 1-15, 10.1109/TGRS.2022.3186253.

Prowse, T., K. Alfredsen, S. Beltaos, B. Bonsal, C. Duguay, A. Korhola, J. McNamara, W.F. Vincent, V. Vuglinsky, and G.A. Weyhenmeyer, 211. Past and future changes in lake and river ice. *Ambio*, 40 (S1), 53-62, DOI 10.1007/s13280-011-0216-7.

Surdu, C.M., C.R. Duguay, and D. Fernández Prieto, 2016. Evidence of recent changes in the ice regime of high arctic lakes from spaceborne satellite observations. *The Cryosphere*, 10, 941-960, doi: doi:10.5194/tc-10-941-2016.

SOLS: A Lake database to monitor in Near Real Time water level and storage variations from remote sensing data, J. Adv. Space Res. Vol 47, 9, 1497-1507, doi:10.1016/j.asr.2011.01.004, 2011 Crétaux J-F , A. Arsen, S. Calmant , A. Kouraev , V. Vuglinski , M. Bergé Nguyen , M-C. Gennero, F. Nino, R. Abarca Del Rio , A. Cazenave , P. Maisongrande

Lake studies from satellite altimetry, C R Geoscience, Vol 338, 14-15, 1098-1112, doi: 10.1016/j.crte.2006.08.002, 2006 Crétaux J-F and C. Birkett

A new Global Lakes Database for a remote sensing programme studying climatically sensitive lakes, J. Great Lakes Res., 21 (3), 307-318. Birkett, C.M., Mason, I.M., 1995.

Lake volume monitoring from space. *Surveys in Geophysics*, *37*(2), 269-305. Crétaux, J. F., Abarca-del-Río, R., Berge-Nguyen, M., Arsen, A., Drolon, V., Clos, G., & Maisongrande, P., 2016.

C. Birkett 1995. The global remote sensing of lakes, wetlands and rivers fro hydrological and climate research. International Geoscience and Remote Sensing Symposium, IGARSS'95. Quantitative Remote Sensing for Science and Applications.

C. Birkett 1998. Contribution of the TOPEX NASA Radar Altimeter to the global monitoring of large rivers and wetlands. AGU1000. Doi: https://doi.org/10.1029/98WR00124

F. Mercier et Al. 2002. Interannual lake level fluctuations (1993-1999) in Africa from Topex/Poseidon: connections with ocean-atmosphere interactions over the Indian Ocean. Global and Planetary Change.

Maberly, S. C., O'Donnell, R. A., Woolway, R. I., Cutler, M. E. J., Gong, M., Jones, I. D., Merchant, C. J., Miller, C. A., Politi, E., Scott, E. M., Thackeray, S. J. and Tyler, A. N. (2020) Global lake thermal regions shift under climate change. Nature Communications, 11. 1232. ISSN 2041-1723 doi: https://doi.org/10.1038/s41467-020-15108-z

Nikraftar Z. et Azizia A., 2015: EVALUATION AND ANALYSIS OF URMIA LAKE WATER LEVEL FLUCTUATIONS BETTWEN 1998-2006 USING LANDSAT IMAGES AND TOPEX ALTIMETRY DATA The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, Volume XL-1/W5, 2015 International Conference on Sensors & Models in Remote Sensing & Photogrammetry, 23–25 Nov 2015, Kish Island, Iran

Ryan, J.C., L.C Smith, S.W. Cooley, L.H. Pitcher, T.M. Pavelsky (2020). Global characterization of inland water reservoirs using ICESat-2 altimetry and climate reanalysis. Geophysical Research Letters. https://doi.org/10.1029/2020GL088543

Sima, S. & Tajrishy, M. (2013) « Using satellite data to extract volume–area–elevation relationships for Urmia Lake, Iran », Journal of Great Lakes Research, vol. 39., No. 1, pp. 90-99. 10.1016/j.jglr.2012.12.013

Woolway, R. I. and Merchant, C. J. (2019) Worldwide alteration of lake mixing regimes in response to climate change. Nature Geoscience, 12. pp. 271-276. ISSN 1782-0908 doi: https://doi.org/10.1038/s41561-019-0322-x

Woolway, R. I. and Merchant, C. J. (2018) Intralake heterogeneity of thermal responses to climate change: a study of large northern hemisphere lakes. Journal of Geophysical Research: Atmospheres, 123 (6). pp. 3087-3098. ISSN 2169-8996 doi: https://doi.org/10.1002/2017JD027661

Fichot, CG ; Matsumoto, K; Holt, B; Gierach, MM; Tokos, KS (2019) "Assessing change in the overturning behavior of the Laurentian Great Lakes using remotely sensed lake surface water temperatures" View Web of Science Researcher ID and ORCID REMOTE SENSING OF ENVIRONMENT Volume: 235 Article Number: 111427 DOI: 10.1016/j.rse.2019.111427

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

IOCCG (2012). Mission Requirements for Future Ocean-Colour Sensors. McClain, C. R. and Meister, G. (eds.), Reports of the International Ocean-Colour Coordinating Group, No. 13, IOCCG, Dartmouth, Canada

Merchant, C. J.,Paul, F., Popp, T., Ablain, M., Bontemps, S., Defourny, P., Hollmann, R., Lavergne, T., Laeng, A., de Leeuw, G., Mittaz, J., Poulsen, C., Povey, A. C., Reuter, M., Sathyendranath, S., Sandven, S., Sofeiva, V. F. and Wagner, W. (2017) Uncertainty information in climate data records from Earth observation. Earth System Science Data, 9 (2). pp. 511-527. doi: https://doi.org/10.5194/essd-9-511-2017

Appendix A - Example of Global attributes for merged files

Attribute	Value
title	ESA Lakes_cci product
institution	LWL: Laboratoire d'Etudes en Geodesie et Oceanographie Spatiales, Collecte Localisation Satellites;
	LWE: Laboratoire d'Etudes en Geodesie et Oceanographie Spatiales, Collecte Localisation Satellites
	LSWT: University of Reading
	LIC: H2O Geomatics
	LWLR: Plymouth Marine Laboratory
source	LWL: European Space Agency (ESA), National Aeronautics and Space Administration (NASA), European Organisation for the Exploitation of Meteorological Satellites (EUMETSAT), National Oceanic and Atmospheric Administration (NOAA), Indian Space Research Organisation (ISRO)
	LWE: European Space Agency (ESA), National Aeronautics and Space Administration (NASA)
	LSWT: European Space Agency (ESA), European Organisation for the Exploitation of Meteorological Satellites (EUMETSAT), European Centre for Medium-Range Weather Forecasts (ECMWF), National Aeronautics and Space Administration (NASA)
	LIC: European Space Agency (ESA), National Aeronautics and Space Administration (NASA)
	LWLR: European Space Agency (ESA), National Aeronautics and Space Administration (NASA)
history processing history of the dataset	LWL: Generated by Laboratoire d'Etudes en Geodesie et Oceanographie Spatiales, Collecte Localisation Satellites
	LWE: Generated by Laboratoire d'Etudes en Geodesie et Oceanographie Spatiales, Collecte Localisation Satellites
	LSWT: University of Reading LSWT processor version v2.6.1-146-gfe50b81_RES120
	LIC: Lake ice cover processor by H2O Geomatics
	LWLR: Calimnos processor by Plymouth Marine Laboratory, including calls to Idepix (SNAP) and POLYMER (Hygeos) algorithms
references	http://cci.esa.int/lakes
tracking_id	f5170cc0-0da7-4b2b-b829-93e485aa451a
conventions	CF-1.8

Attribute	Value
product_version	V 3.0.0
Format_version	CCI Data Standards v2.3
Summary	This dataset contains L3S daily ECV Lakes products: Water Level (LWL), Water Extent (LWE), Ice cover (LIC), Surface Water Temperature (LSWT) and Water Leaving Reflectance (LWLR). L3S data are observations combined from multiple instruments into a common spatiotemporal grid
keywords	Satellite, Lake, Climate Change, Lake Water Level, Lake Water Extent, Lake Surface Water Temperature, Lake Ice Cover, Lake Water Leaving Reflectance
id	ESACCI-LAKES-L3S-LK_PRODUCTS-MERGED-20190214-fv3.0.0.nc
naming_authority	lakes.esa-cci
keywords_vocabulary	inspire: INSPIRE spatial data themes, gcmd: NASA Global Change Master Directory (GCMD) Science Keywords, gemet: GEMET keywords
keywords	Satellite, Lake, Climate Change, Lake Water Level, Lake Water Extent, Lake Surface Water Temperature, Lake Ice Cover, Lake Water Leaving Reflectance, Orthoimagery, EARTH_SCIENCE-OCEANS-OCEAN_OPTICS-WATER-LEAVING RADIANCE, EARTH_SCIENCE-TERRESTRIAL_HYDROSPHERE- WATER_QUALITY_WATER_CHEMISTRY-CHLOROPHYLL, SUSPENDED_SOLIDS, TURBIDITY, water, algal bloom, aquatic environment, freshwater, freshwater quality, ice, inland water, lagoon, lake; dam; phytoplankton; turbidity, water monitoring, water quality, water reservoir, climate; seasonal variation, environmental data, environmental monitoring, monitoring, remote sensing,
cdm_data_type	Grid
comment	These data were produced for the ESA Lakes_cci project
date_created	2022-02-22
creator_name	ESA Lakes_cci
creator_url	https://climate.esa.int/en/projects/lakes/
creator_email	lakes_cci@groupcls.com
project	Climate Change Initiative - European Space Agency
geospatial_lat_min	-90.0
geospatial_lat_max	90.0
geospatial_lon_min	-180.0
geospatial_lon_max	180.0
geospatial_vertical_min	NA
geospatial_vertical_max	NA
time_coverage_start	19920101T120000Z

Attribute	Value
time_coverage_end	20191231T120000Z
time_coverage_duration	P1D
time_coverage_resolution	P1D
standard_name_vocabulary	CF Standard Name Table v78
license	ESA CCI Data Policy: free and open access
platform	LWL: TOPEX/Poseidon, Jason-1, Jason-2, Jason-3, Sentinel-6A, Envisat, SARAL, GFO, Sentinel-3A
	LWE: Landsat-<4,5,7,8>, Sentinel-1A
	LSWT: ERS-2, Envisat, Metop-A, Metop-B, Sentinel3A, Sentinel3B, Terra
	LIC: Aqua, Terra
	LWLR: Aqua, Envisat, Sentinel-3A, Sentinel-3B, Orbview-2, Suomi NPP
sensor	LWL: Poseidon-1, Poseidon-2, Poseidon-3, RA, RA-2, AltiKa, SRAL
	LWE: MSS, TM, OLI
	LSWT: ATSR-2, AATSR, AVHRR-3, SLSTR, MODIS
	LIC: MODIS
	LWLR: SeaWifs, MODIS, MERIS, VIIRS, OLCI
spatial_resolution	1 km at Equator
key_variables	water_surface_height_above_reference_datum, lake_surface_water_extent, lake_ice_cover, lake_surface_water_temperature, chla_mean, turbidity_mean, Rw[xxx] where xxx is one of 400, 412, 443, 469, 490, 510, 531, 547, 560, 620, 645, 665, 674, 681, 709, 754, 779, 859, 885, 900, 1020nm
geospatial_lat_units	degrees_north
geospatial_lon_units	degrees_east
geospatial_lat_resolution	0.008333333
geospatial_lon_resolution	0.008333333
doi	10.5285/7fc9df8070d34cacab8092e45ef276f1

Appendix B - Example of Global attributes for LIT files

Attribute	Value
title	CCI-LAKES Lake Ice Thickness dataset. LRM_LIT data analysis over the Great Slave Lake for pass 45 and JASON-1, JASON-2 and JASON-3 missions
institution	Collecte Localisation Satellites
source	CNES, NASA, EUMETSAT, NOAA
history	"Generated by Collecte Localisation Satellites
references	https://climate.esa.int/en/projects/lakes/
Conventions	CF-1.8
product_version	V1.0
summary	this dataset contains LIT information over Great Slaves lake
keywords	Satellite, Lake, ice thickness
id	ESACCI-LAKES-L3C-LIT-GreatSlave_Pass45.nc
naming_authority	lakes.esa-cci
comment	These data were produced for the ESA Lakes_cci project
creator_name	ESA Lakes_cci
creator_email	lakes_cci@groupcls.com
project	Climate Change Initiative - European Space Agency
standard_name_vocabulary	CF Standard Name Table v78
license	ESA CCI Data Policy: free and open access
platform	Jason1, Jason2, Jason3
key_variables	LIT
latitude	latitude of the center of the LIT ROI: 61.49 degrees_north
longitude	longitude of the center of the LIT ROI: 115.04 degrees_west
time_coverage_start	20020116T231528
time_coverage_end	20220330T050143
doi	10.5285/7fc9df8070d34cacab8092e45ef276f1