D3.x: System Description Document

Reference:	CCI-LAKES-SDD-0020
Issue:	1.1
Date:	Apr. 24, 20

Chronology Issues:				
Issue:	Date:	Reason for change:		Author
0.1	22/05/19	Draft version		B. Calmettes
1.0	24/03/20	Version 1.0		B. Calmettes, S Simis, C Merchant, C Duguay, H Yésou, J-F Crétaux
1.1	24/04/20 20	Revision following ESA review		B. Calmettes, S Simis, C Merchant, C Duguay, H Yésou, J-F Crétaux

People involved in this issue:			Signature
Authors:	Stefan Simis,	Plymouth Marine Laboratory	
	Jean-François Crétaux	LEGOS	
	Hervé Yésou	SERTIT	
	B Calmettes	CLS	
	Chris Merchant,	University of Reading	
	Claude Duguay	H20 Geomatics	
Internal review:	Stefan Simis,	Plymouth Marine Laboratory	
	B. Calmettes	CLS	
Approved by:	B. Coulon	CLS	
Authorized by:	C. Albergel	ESA	

Distribution:		
Company	Names	Contact Details

Distribution:		
Company	**Names**	**Contact Details**
ESA	A.M. Trofaier	Anna.Maria.Trofaier@esa.int
	C. Albergel	Clement.Albergel@esa.int
	P. Cipollini	Paolo.Cipollini@esa.int
BC	K. Stelzer	kerstin.stelzer@brockmann-consult.de
CLS	B. Coulon	bcoulon@groupcls.com
	B. Calmettes	bcalmettes@groupcls.com
CNR	C. Giardino	giardino.c@irea.cnr.it
GeoEcoMar	A. Scrieciu	albert.scrieciu@geoecomar.ro
H2OG	C. Duguay	claude.duguay@h2ogeomatics.com
LEGOS	J.F. Crétaux	jean-francois.cretaux@legos.obs-mip.fr
NORCE	E. Malnes	eima@norceresearch.no
PML	S. Simis	stsi@pml.ac.uk
SERTIT	H. Yésou	herve.yesou@unsitra.fr
TRE-ALTAMIRA	P. Blanco	pablo.blanco@tre-altamira.com
UoR	C. Merchant	c.j.merchant@reading.ac.uk
	L. Carrea	l.carrea@reading.ac.uk
UoS	A. Tyler	a.n.tyler@stir.ac.uk
	E. Spyrakos	evangelos.spyrakos@stir.ac.uk

Reference documents

RD- 1: Algorithm Theoretical Basis Document
CCI-LAKES-0024-ATBD_v1.2

RD- 2: Data Access Requirement Document
CCI-LAKES-0017-DARD-v1.1

RD- 3: Product Specification Document
CCI-LAKES-0016-PSD-v1.1

RD- 4: Product User Guide
CCI-LAKES-0029-PUG_V1.0

RD- 5: Product Validation and Intercomparison Report
CCI-LAKES-0032-PVIR_V1.0

RD- 6: Theia/Hydroweb Project
http://hydroweb.theia-land.fr/

List of Contents

1. Introduction

The purpose of this document is to describe the Lakes_cci system which is designed as a system of systems. The Lakes CCI system is based on existing processing systems used for scientific studies. These processing systems generate each product in the ECV lakes which are:

- Lake Water Level (LWL)
- Lake Water Extent (LWE)
- Lake Surface Water Temperature (LSWT)
- Lake Ice Cover (LIC)
- Lake Water-Leaving Reflectance (LWLR)

To make a consistent Lakes CCI system and to be able to generate products that meet user needs as much as possible, existing processing chains will be updated, and additional processing systems will be developed to generate a consistent data set.

This document contains the description of the existing systems. Each system is already implemented following user community requirements, and validation tests are performed. The systems will be updated to meet new needs/improvements that will arise all along the project life. This document content covers the 3 documents addressing to the system requirements, specification and verification.

Section 2 defines main development objectives for the Lakes CCI project from a system point of view and its distinction as a system of systems. It also describes the work performed the harmonisation of the product in order to deliver a consistent data set.

It is followed by the description of each one of the systems making part of the Lakes CCI system including the harmonisation step (section 3). It includes the harmonisation of each individual system needed to generate an ECV lake dataset following the requirements identified in the User Requirements Document (URD). These requirements are based on user needs, GCOS recommendations, team experience and literature review.

The section 4 gives the compliance of the system regarding the system requirements from the SOW and the URD.

2. Processing system objectives and harmonisation process

2.1. Objectives

The objective of the Lakes_cci project is to produce and validate a consistent data set of the variables grouped under the Lakes Essential Climate Variable. Following an assessment of the requirements of the climate research community, the different algorithms will be tested to select the most relevant ones. This involves both R&D activity and the application of rules to meet CCI data Standards to generate high quality products. At the system level, the objective is to define the process and rules to generate a consistent data set so that the user can easily manipulate the 5 parameters of the lake ECV.

In the Lakes CCI context, the specification of the production system will be useful for related activities in the Copernicus programme: the C3S Project implementing these algorithms for a periodic release and the Global Land Project for an operational release.

2.2. Harmonisation process

The consistency of the product is met if each product follows the rules below:

- Product format and time step
- Grid common format,
- Common land mask
- Common lake IDs

The harmonisation process is a three steps process to generate consistent product.

First step:

To define the rules with all parameter leaders and science leaders.

1. Product format and time step: there will be one product per day, containing all parameters. If a parameter is missing the field will be filed with a default value.
2. Grid common format: the product will provide dada over a grid with spatial resolution of 1/120 degrees. For parameter like LWL and LWE which provides one value per lakes, this value will pit on each grid point according to the land mask.
3. Common land mask: a land mask has been defined and is taken into account by the processing system of LSWT, LIC and LWLR. For LWL and LWE it is only used during the merge process.
4. Common lake IDs: each lake has an ID coming from existing database or created for the project.

These rules are entries for each processing system and are identified as configuration parameters in the following figures in the document (for harmonisation).

Second step:

For each parameter, a processing system is already existing. The second step consists in an update of the current system to take into account the 4 rules defined during the first step.

Third step:

Once each system has generated its own products, a merge process is done to generate a daily product containing all parameters for all lakes. In case of missing value, a default value is used.

There is no value transformation like interpolation for instance. This last step does not implement any algorithm that could change the value generated and qualified by the processing system in charge of parameters.

This process verifies on the input side that the files for each product are available and the corresponding variables are included, and on the output side that the corresponding file, with the correct size and the output variables were generated. These output files are compliant with CCI Data Standards V2.

The outputs from this 3rd steps are fully described by the Product Specification Document (RD-3) and the Product User Guide (RD-4).

2.3. General overview

The function of the system is the production of a consistent, long-term, multi-mission and multi-sensor global data set.

Two teams are cooperating within the organizational framework in order to fulfil the mission objective of the system.

- **The Science Team** oversees the physical reliability and validity of the products generated by the system. Consequently, the function of the Science Team is to design scientific algorithms capable to produce long-term consistent, multi-mission global products.
- **The System Development Team** develops the scientific algorithms based on the design of the science team and integrate them in the processing baselines.

In the Lakes_cci context, each contributing product (LWL, LWE, LSWT, LIC and LWLR) is generated separately (Figure 1), with similar temporal and spatial resolutions, to be integrated thereafter into the final Lakes ECV (Figure 2).
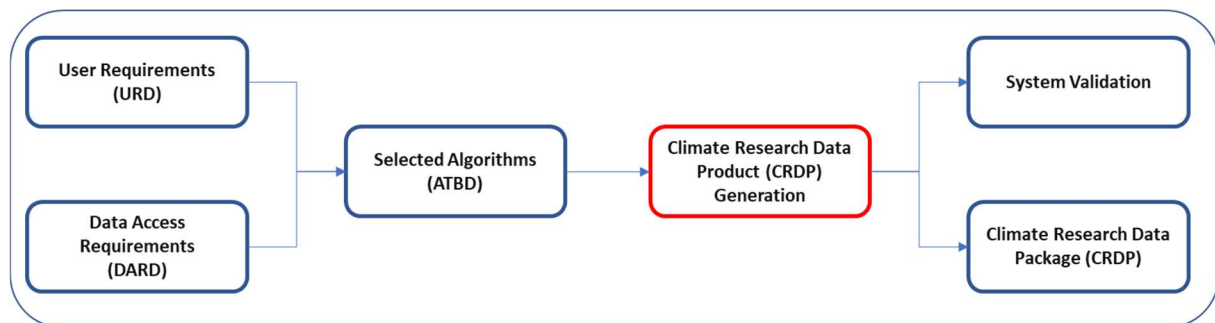

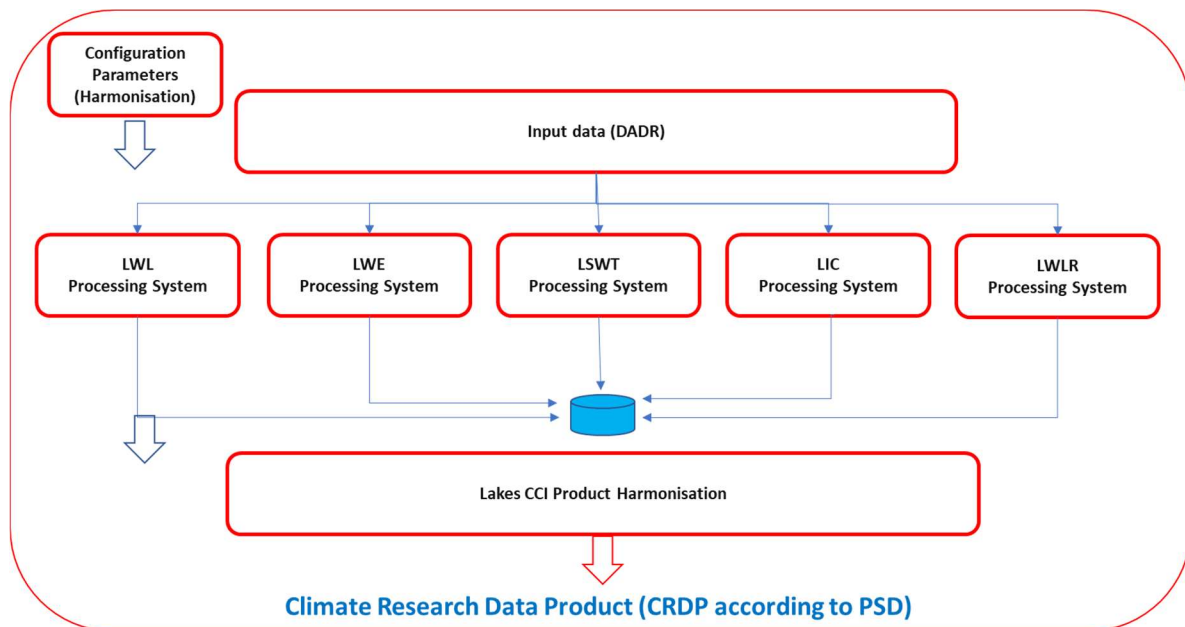**Figure 1. Structure for the product generation.**

Figure 2. General Principle for harmonisation of Lakes ECV product

Each processing system generates their products that become the inputs of the CRDP generation system. The harmonisation process starts by applying a set of harmonisation parameters and ends by merging all outputs from the different systems (see §2.2).

3. Processing Systems

3.1. Lake Water Level (LWL) processing system

3.1.1. General Description

This section provides an overview of the LWL system with its system context, its main function and processing chain, and its architecture (Figure 3).
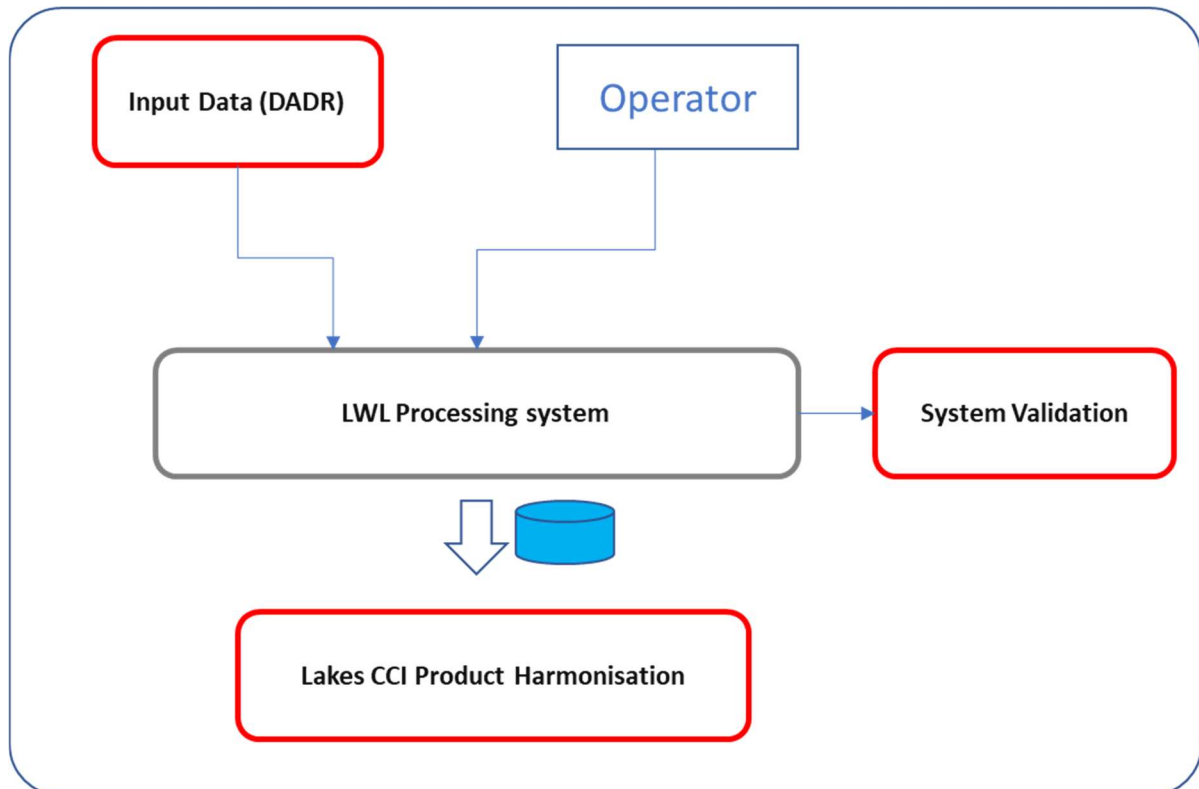
Figure 3: LWL Processing system: General Diagram

3.1.2. Main functionalities

The processing framework of Lake Water Level in the CCI Lakes project is based upon the scientific base of the THEIA/Hydroweb project (RD- 6).

The LWL product is measured using satellite radar altimetry. Radar altimeters send an electromagnetic pulse to the satellite nadir and record the propagation time to and from the emitted wave and its echo from the surface. Algorithm Theorical Basis Document (RD- 1) includes the complete description of the algorithm used to estimate the Lake Water Level

Pre-processing needed to get the input data streams (RD- 2) into the correct format to be handled by the Hysope processor. A post process is also needed to adapt the data to the Lakes_cci specifications and it's added around the Hysope processor core.

3.1.3. Architecture

In the case of the LWL, two types of processing are used to generate the dataset. The first one, used for historical data is performed by LEGOS using GDR data from past missions (Topex/Poseidon,

Jason 1 and 2, Envisat, Cryosat-2) and the second one, using NRT data from current missions (Jason 3, Sentinel 3A) in operational mode, performed by CLS in an operation model with close cooperation with LEGOS.

The Technical Platform and Operations Framework is composed of a set of hardware and software components interacting with each other (Figure 4).

The CCI lakes LWL Technical Platform is realized at CLS and CNES providing two major hardware components. These hardware components are:

- The CLS cluster: 9 batch servers / 216 cores (432 threads) 128GB memory / server
- The CNES cluster (HAL): 300Tflops, 380 batch servers / 8400cores, 128GB memory / server, 6,2 PB GPFS / 200TB burst buffer/ 100GBs bandwidth Infiniband, low latency network, 4 GPGPU Nvidia Volta V100

The HAL cluster is used to gather and enhance the historical altimetry data thanks to its significant storage capability and computing power. The historical altimetry databases (L2E-HR) on the server are maintained and operated with a CLS software and database proprietary format directly compatible with LWL system.

The CLS cluster is used to copy and store only the relevant input data from the altimetry L2E-HR databases on the HAL cluster. This data is stored temporarily on a dedicated partition (netapp4-L2P-HYDRO, 3To) for the operation team use. The partition also stores the relevant missing ancillary data and houses the operational C3S LWL Processing Framework.

The Operations Framework is a based on a source code library on Github[1] and its web-based repository manager GitLab[2], holding all code and configurations fragments needed to run the operational service. Writing access to the GitLab code repository is restricted to the members of the Lake Water Level development team. Specific version of code packages can be cloned or downloaded to the Technical Platform. This process is automated by making use of a CLS overlayer software to the Git tools. The documentation platform of all packages is hosted on a dedicated Microsoft SharePoint repository with restricted access to the Science Team, the System Development Team and the Operations Team.
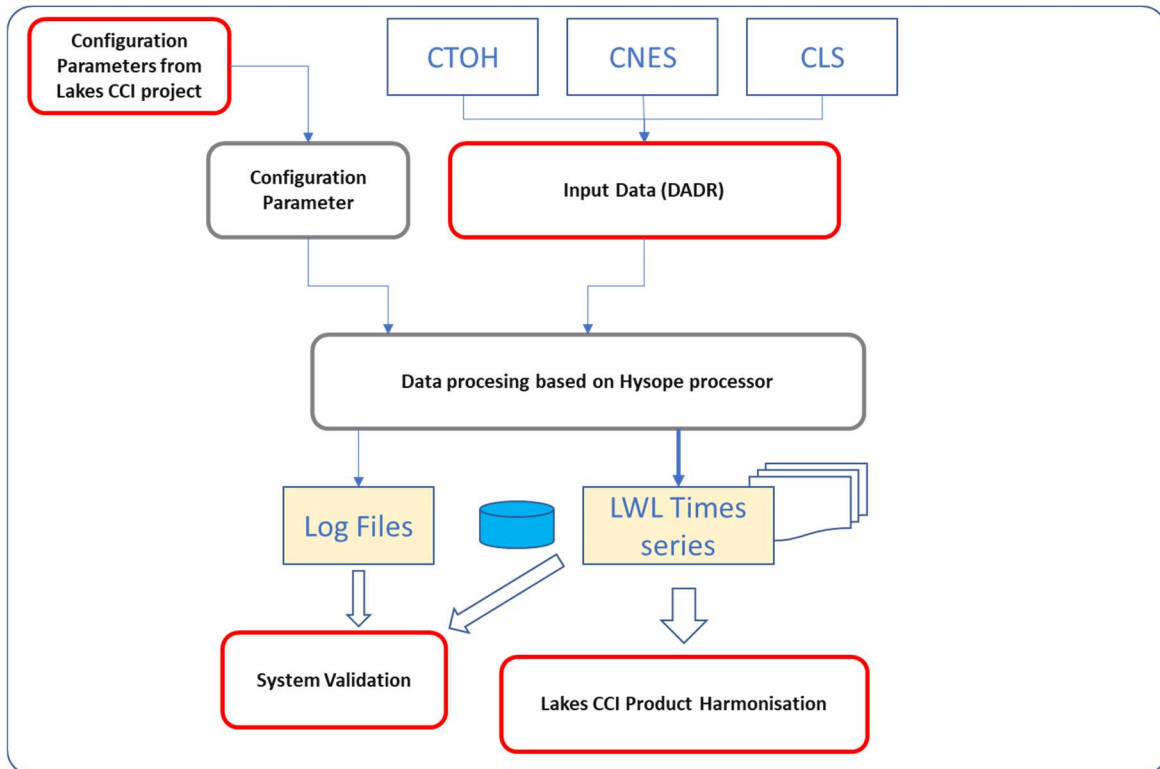
---

[1] https://github.com/

[2] https://about.gitlab.com

Figure 4. Overview of the elements and interfaces of the CCI LakesLWL Processing Framework

3.1.4. Input

The original Hysope processor was designed to ingest Level-2 altimetry data. The input format can either be the NetCDF official formats delivered by the space agency (e.g. cophub) or the CLS proprietary format. This last functionality allows the Hysope processor to ingest the historical altimetry database L2E-HR housed by and maintained on the HAL cluster. Output

Data used in the CCI lakes LWL can be discriminated into historic and NRT inputs. Historic data are level-2 altimetry products from already decommissioned satellite missions generated in one processing cycle. These datasets are reprocessed on a regular basis by space agencies to enhance them with the state-of-the art algorithms. The data archive is available on the disk storage of CLS cluster and on the CTOH (Centre for Topographic studies of the Ocean and Hydrosphere), data centre of the Science Team. The data archive and is complemented by regular backups of the data.

Active input data streams are based on observations from a series of altimeters on-board historic and current international satellite missions. Algorithms to derive lake water level from altimetry have been developed and are further research by LEGOS (Crétaux et al 2006, Crétaux et al 2011, Crétaux et al 2016).
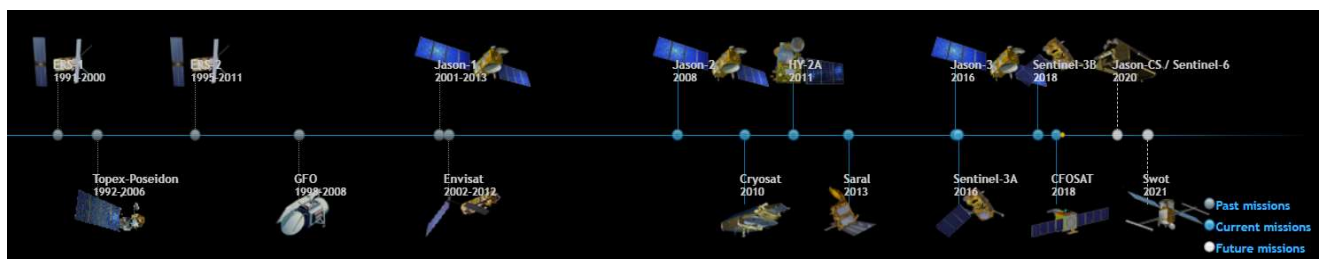

Figure 5: Altimetry constellation (source: https://www.aviso.altimetry.fr)

The Data Access Requirement Document (DARD) contains the characteristics of the input data used to estimates the LWL from altimetry (multiples missions), ancillary data (land mask) or in-situ data (for validation purposes)

3.1.5. Outputs

The Hysope processor can produce for each lake either a CSV or GeoJSON file in both modes. Consequently, the CCI lakes post processing block converts the format into the required NetCDF4 described in CCI Data Standards document and also compliant to the Product Specification Documents (PSD), mainly in a gridded output / daily basis. Given that the LWL is not a grid product, it needs to be adapted following the common defined lake mask used for all the products in the Lakes_cci project.

3.1.6. Verification tests

Verification is the process to demonstrate that the system meets the specified requirements (URD) and provides data to users. Verification methods used are tests, inspection, and monitoring:

- Unit testing: it is carried out during development of the system software and before its deployment. The expected result is defined before the code execution and the actual result is compared to the expected result.
- Regression testing: it verifies that previous results are the same after a software change
- Completeness check: it ensures that the system has calculated a complete output dataset by inspecting report files and testing that the expected output files have been generated
- Visual inspection: it is the last stage of verification. It consists of opening, with appropriate software, and a looking at the results.

3.2. Lake Water Extent (LWE) processing system

3.2.1. General Description

The processing system for LWE will drastically evolve during the project thanks to the WP6 devoted to the new methodologies for LWE. In its initial version, LWE is assessed thanks to hypsometric curve methodology which need the LWL. Therefore, the LWE processing system is included in the LWL processing system which has as inputs the hypsometric curves for a list of lakes.

The chapter below gives a first description of the processing system that are currently used in the frame of the WP6.

3.2.2. Future Processing system

This chapter gives a first description of the processing system dedicated to LWE measurement and currently defined, developed and tested in the frame of the WP6.

3.2.2.1. NORCE-SAR LWE processing system

General description

This section provides an overview of the LWE system with its system context, its main function and processing chain, and its architecture. The LWE processor classifies SAR radar backscatter images

over individual lakes into water/land classes, and estimates the overall Lake Water Extent based on the classification. The overview of the LWE processing system is shown in Figure 6.
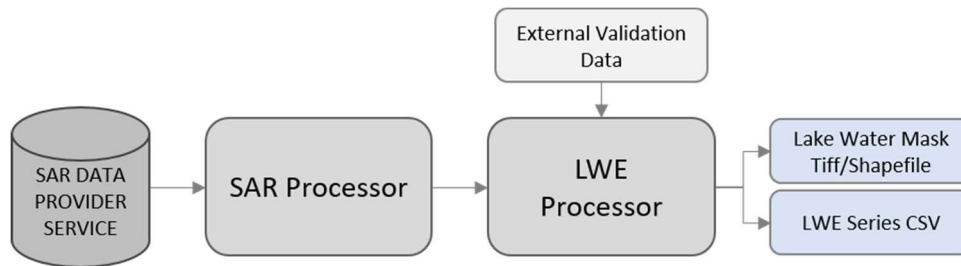
Figure 6. NORCE SAR LWE Overview System

3.2.2.1.1. Main functionalities

The Norce LWE processor applies the following processing steps:

- Geocoding of SAR GRD products into backscatter images for the VV and VH channels, including calibration, terrain correction, and multi-looking to the desired spatial resolution (20m for small/medium lakes)
- Classification of backscatter images into water masks, applying maximum and minimum extent masks
- Estimates of LWE, applying multitemporal filters to remove outlaiers
- Generation of lake masks
- Uncertainty estimates

3.2.2.1.2. Architecture

The NORCE processing architecture is based on in-house processing software GDAR implemented in Python. The software supports most historical and current SAR sensors (ERS, Radarsat-1 &2, Envisat ASAR, TerraSAR X, ALOS Palsar, Cosmo-Skymed and Sentinel-1).

3.2.2.1.3. Input

SAR GRD images from external source (e.g. SciHub) using both polarizations preferably.

Digital elevation model

Lake masks (maximum and minimum extent)

Lake Water Level

3.2.2.1.4. Output

The LWE processor generates a tiff and a shapefile containing the corresponding lake water mask and a csv file containing the lake water extent for each of the processed dates.

3.2.2.1.5. Verification tests

Several verification steps are taken in order to assure the quality of the product:

-Geocoding accuracy can be manually inspected

-Classification of water/land can be checked by comparisons with backscatter imagery.

-Multi-temporal filtering can be checked against individual LWE estimates, or by studying the RMSE.

NORCE LWE processing can also be independently verified using in situ or satellite derived LWL-measurements via the hypsometric curve or by comparisons with other satellite or airborne data (e.g. near simultaneous S2 cloud free images over the same lake) to quantify the accuracy.

3.2.2.2. TRE-Altamira's SAR General description

This section provides an overview of the LWE system with its system context, its main function and processing chain, and its architecture. The LWE processor takes advantage of the SAR processor to generate the set of SAR amplitude images to which LWE processor is applied. The overview of the LWE processing system is shown in Figure 7.
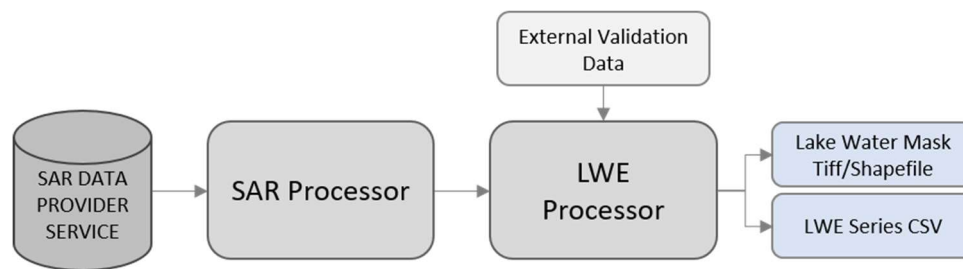
Figure 7. TRE-Altamira's SAR LWE Overview System

3.2.2.2.1. TRE's Altamira's SAR Main functionalities

The TRE-Altamira SAR and LWE processors undertake the following functions:

- SAR Satellite RAW and SLC data read of nearly all existent and past missions. Within the framework of the present Project Sentinel-1 IW SLC images are considered by default.
- Image coregistration.
- Images amplitude calibration.
- Speckle reduction.
- Generation of the Lake Water Mask.
- Calculation of the Lake Water Extent.
- As part of the outcome of this Project it will provide uncertainty figures to the Lake Water Mask pixels.

3.2.2.2.2. TRE's Altamira's SAR Architecture

TRE-ALTAMIRA has a dedicated in-house data processing centre, custom built to cope with the high demands of processing Synthetic Aperture Radar (SAR) data sets where the LWE is hosted. The architecture consists of a Linux-based parallel processing system, made up of approximately 30 servers and 204 cores for data processing and 20 CPUs for interactive session management. The system has a storage capacity of 350 TB accessed via a high-performance parallel file system. The SAR processor consists of some core modules in C and python. LWE processor is written in Matlab and Python.

3.2.2.2.3. TRE's Altamira's SAR Input

The SAR and LWE processors allow working with the following mission's RAW and SLC data:

- ERS
- ENVISAT-ASAR
- PALSAR, PALSAR-2
- RADARSAT-1,RADARSAT-2
- TerraSAR-X, Tandem-X, PAZ
- Cosmo-SkyMed
- Sentinel-1

3.2.2.2.4. TRE's Altamira's SAR Output

The LWE processor generates a tiff and a shapefile containing the corresponding lake water mask and a csv file containing the lake water extent for each of the processed dates.

3.2.2.2.5. TRE's Altamira's SAR Verification Tests

Several quality checks are generated to ensure data process correctness:

- Input data anomalies: correctness of input data directory, Readability, Completeness of coverage of the project area by the input data.
- Best master selection: Baseline/Doppler distribution
- Images coregistration: TOPSAR indicators, missing lines, missing data, focusing errors.
- Visual Inspection on the Contrast and entropy distributions for mask selection.
- Visual Inspection on the K-means classification.
- Visual inspection on the Lake Water Mask over the amplitude image.
- Some other as comparison against ground-truth data or the corresponding lake hypsometry curve are being currently evaluated.

3.3. Lake Surface Water Temperature (LSWT) processing system

3.3.1. General description

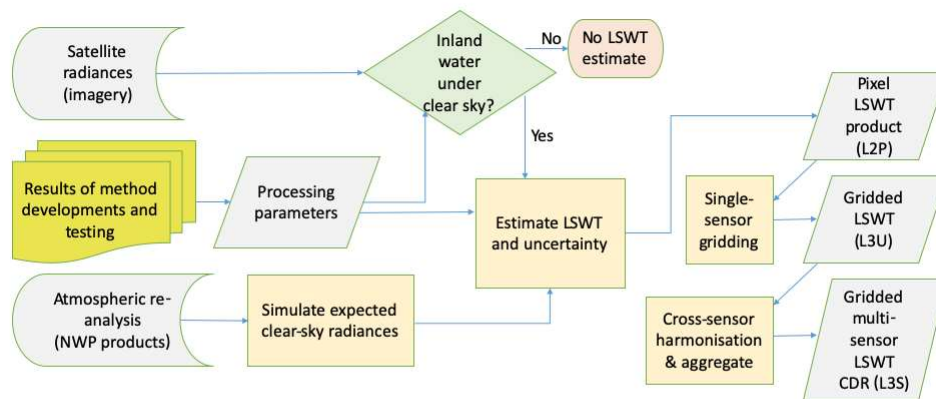The overview of the LSWT processing system is shown in Figure 8

Figure 8. LSWT Overview System

3.3.2. Main functionalities

LSWT processor undertakes the following functions:

- Satellite L1 data read (all required input data streams)
- Extraction of matching prior information from numerical weather prediction (NWP) fields
- Identification of candidate satellite pixels filled with inland water (static mask)
- Dynamic water detection on candidate satellite pixels and cloud screening
- Radiative transfer modelling of satellite radiances and Jacobians given NWP (calling external software)
- Optimal estimation of LSWT, total column water vapour (discarded) and uncertainties
- Output of L2P (full-resolution swath LSWT) products
- Averaging/regridding to regular grids creating single-sensor gridded uncollated (L3U) products
- Harmonisation of mean LSWT per lake using overlapping data across sensors
- Final output of multi-sensor gridded ("super-collated") climate data records (L3S)

3.3.3. Architecture

The LSWT processor is hosted at the Centre for Environmental Data Analysis (CEDA), in a Unix environment enabled to see multiple TB (Terabytes) of input files on "spinning disk" access. The processor consists of some core modules in Fortran augmented by Python 3. Trivially parallel processing is available, as each input file may be processed independently of all others to the point of generation of L3U. Once all L3U are generated, a separate process is initiated to combine to the L3S climate data record.

3.3.4. Input

The Data Access Requirement Document (RD- 2) contains the characteristics and full references of the input data used to estimate the LSWT from the missions listed below, the ancillary data (inland water mask and distance to land raster file, plus NWP fields) and in-situ data (for validation purposes).

The missions presently processed are:

- Along-Track Scanning Radiometer 2
- Advanced Along-Track Scanning Radiometer

- Advanced Very High Resolution Radiometer MetOp-A
- Advanced Very High Resolution Radiometer MetOp-B

The ancillary NWP have been ERA-Interim analysis.

The system development for switching to ERA-5 is complete and available for future reprocessing, and the future system will also add a data stream from the Moderate-resolution Imager Spectroradiometer (MODIS) Adding a further sensor with similar channels is a direct evolution from the system point of view: the elements that need to be modified are: the reader for ingesting satellite radiance imagery; the processing parameters (from results of scientific algorithm development); and the simulation of expected clear-sky radiances (configuration of simulation for new sensor). Other elements in the processing chain operate as for other sensors.

3.3.5. Output

Outputs are netCDF files compliant with the recommendations of the CCI Data Standards Working Groups.

The distributed outputs are the L3S outputs which are generated on two grids: the legacy grid of 0.05 degrees; and the Lakes CCI common grid (1/120th degree). The latter data are distributed only within the all-variables Lakes CCI product.

3.3.6. Verification Tests

Verification comprises:

- Visual inspection of plots from random samples of each run
- Validation of extracted data against matched in situ LSWT observations, using the quality levels and water detection generated by the processor

The results of validation of extracted data are described in detail in the Product Validation and Inter-Comparison Report (RD- 5).

3.4. Lake Ice Cover (LIC) processing system

3.4.1. General description

An overview of the LIC processing system is given in Figure 9. The goal of the LIC processing system is to determine the state of lake surfaces. The state can be assigned as either ice or water, or cloud in which case no observation is able to effectively observe the lake surface. The surface of the Earth (and therefore the lakes) is gridded by latitude and longitude into "squares" whose edges subtend 1/120th of a degree (approx. 1 km at the equator.) The cells of this grid which are of interest conform to the specifications of ESA.

The input data is provided at multiple resolutions with subpixels that are centered at locations that are perturbed from an ideal grid. Pixel correlation is done by nearest neighbor resampling. Data from each pixel is then fed to the LIC Retrieval Algorithm which produces a label for the pixel.
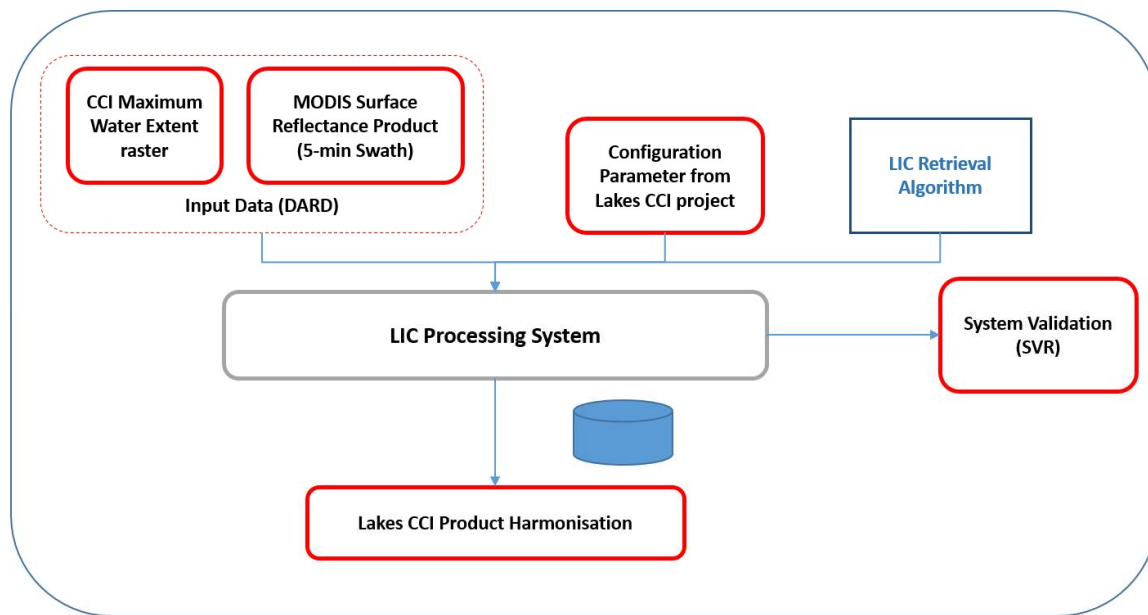
Figure 9. LIC Processing System: Context Diagram

3.4.2. Main functionalities

The LIC product is generated from a threshold-based algorithm using MODIS surface reflectance data, which records the percentage of light reflected by the surface of the earth.

The retrieval algorithm using a set of thresholds turns the satellite observations into labels as either lake ice, open water or cloud cover. Moreover, a global water mask is employed as auxiliary data to filter land and ocean pixels. The ATBD (RD- 1) provides the complete details of the algorithm to produce LIC in v1.0.

Prior to generating the LIC product, the required MODIS product is downloaded from NASA's server and loaded into the processing chain. Subsequent to algorithm retrieval the LIC product is written output in the Lakes_cci specifications.

3.4.3. Architecture

The LIC processing system is written in C++11 with support modules written in Python 3.6. The software is mostly hardware independent, though single precision floating point calculations are used. It should be the case that all IEEE 754 compliant hardware and compilers produce effectively identical results. The software requires a POSIX environment. Figure 10 shows the external interfaces of the data processing software.

The program requires approximately 7 GB of RAM to run one instance (this depends on the total area of the lakes of interest). Instances are fully independent. Each instance requires one process with two threads. The threads are only for pipelining – close to optimal throughput can be achieved with only one physical core.

One day of input (one instance) is approximately 100 GB. Total input size is currently 598 TB for 20 years of MODIS surface reflectance level 2 data (Terra MOD09: 2000-2019, Aqua MYD09: 2002-2019).

The software is deployed on the Graham heterogeneous cluster which is operated by Sharcnet in Ontario, Canada (Table 1). The cluster is attached to multiple storage systems which total to more than 95 PB of storage. A low-latency high-bandwidth InfiniBand fabric connects all nodes and scratch storage. Nodes each have multiple CPUs. The CPUs vary among the Xeon E5 v4, Xeon E7 v4, and Xeon Gold product lines.

Table 1. Overview of Graham HPC cluster

Operating System	CentOS 7
Number of nodes	1185
Number of cores per node	32 ~ 64
Memory per server	128 ~ 3070 GB
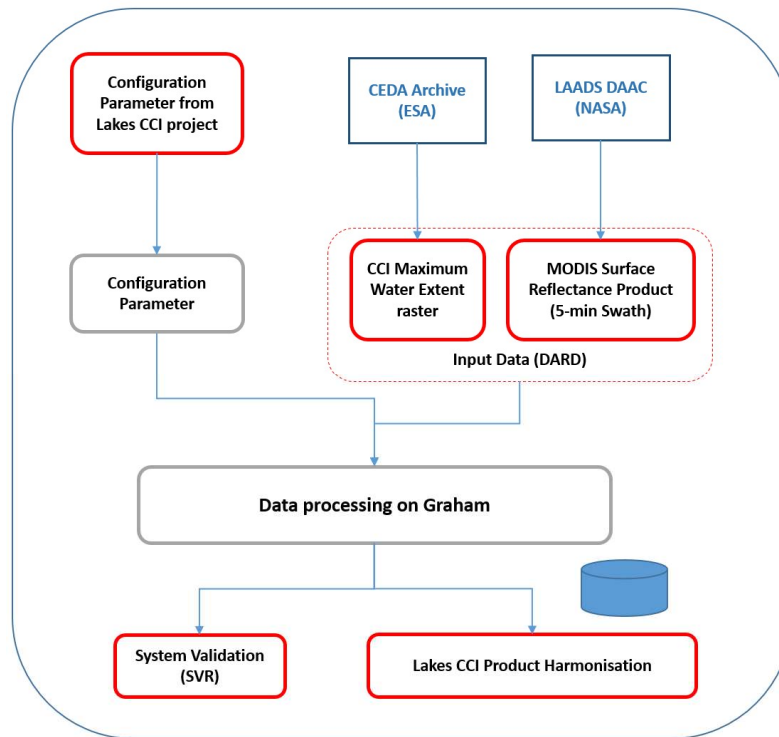
Figure 10. Overview of the elements and interfaces of the CCI lakes

LAADA DAAC (NASA) refers to the Level-1 and Atmosphere Archive & Distribution System Distributed Active Archive Center (DAAC) serving the global NASA Terra, and Aqua MODIS products. The CEDA Archive used by ESA refers to the UK Natural Environment Research Council's Data Repository hosting the ESA CCI Water Extent product.

3.4.4. Input

The LIC producing program requires two input data sources, MODIS surface reflectance (MOD09/MYD09) product and water extent raster.

The MODIS instruments onboard Terra and Aqua have been delivering data since 2000 and 2002, respectively. The MODIS surface reflectance product records Earth observations in 5-min orbital swath format without projection. The generation of the MODIS surface reflectance product requires an atmospheric correction algorithm to correct the effects of atmospheric gases and aerosols from top-of-the-atmosphere reflectance level 1 product. The MODIS product format is the Hierarchical Data Format (HDF5), which contains multidimensional arrays to store data from multiple channels and associated metadata. In addition, each HDF file provides longitude and latitude bands in WGS 84 as georeferenced data.

The maximum water extent provided in ESA CCI Land Cover (v4.0) at 150-m resolution is employed as lake mask to filter land and ocean pixels from MODIS.

3.4.5. Output

The output data is produced in the harmonised grid format. The edge of each grid cell subtends 1/120th degrees latitude/longitude. The output variables in the LIC product include label assigned to grid cell (band 1) and uncertainty of the label (%) (band 2). The ATBD (RD- 1) describes the complete details of the output variables of LIC product v1.0.

3.4.6. Verification Tests

In addition to the main processing chain (Map Producer), another branch of the LIC mapping system has been designed for data sampling (collect, train/optimization, test) and product validation (Figure 11). This branch can produce RGB colour composites and a graphical user interface (GUI) developed in Python allows users to access and display RGB images to manually sample pixels with labels.

The output of labelled samples from the sampling GUI and LIC maps can be read by the validation program to conduct accuracy assessment. Moreover, the validation program also can export misclassified observations and samples with spectral data for further testing and research. The ATBD describes the complete details of accuracy assessment of the LIC product v1.0.
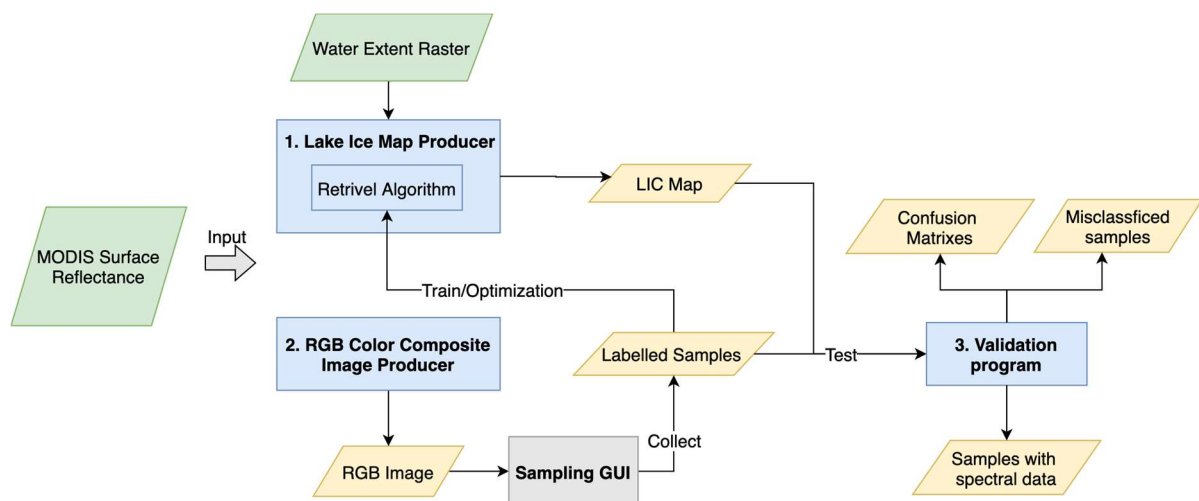
Figure 11. Overview of the LIC mapping system

The verification of the LIC product involves quantitative and qualitative assessments derived from the computation of confusion matrices and manual inspection, respectively. Manual inspection of the product is done via loading and display of a sample of images from various seasons, regions, and ice years in either image analysis or GIS software packages (ESA SNAP and ArcMap). Verification involves examination of metadata as well as visual inspection of pixel values for ice map classes (ice/water/cloud/bad data) against RGB colour composites for quality assurance and identification of errors to consider for future algorithm improvements leading to CDRP V2.0.

3.5. Lake Water-Leaving Reflectance (LWLR) processing system

3.5.1. General description

The *Calimnos* processing chain combines data discovery, subsetting by target area (individual water bodies), radiometric and atmospheric corrections, pixel identification (land/cloud/water/ice), optical water type classification, individual algorithms (per parameter and water type), algorithm blending, conversion and aggregation into a single processing chain.

A schematic overview of *Calimnos* is given in Figure 11. The main processing stages and their corresponding algorithms are listed in the next section and are detailed in the ATBD (RD- 1).
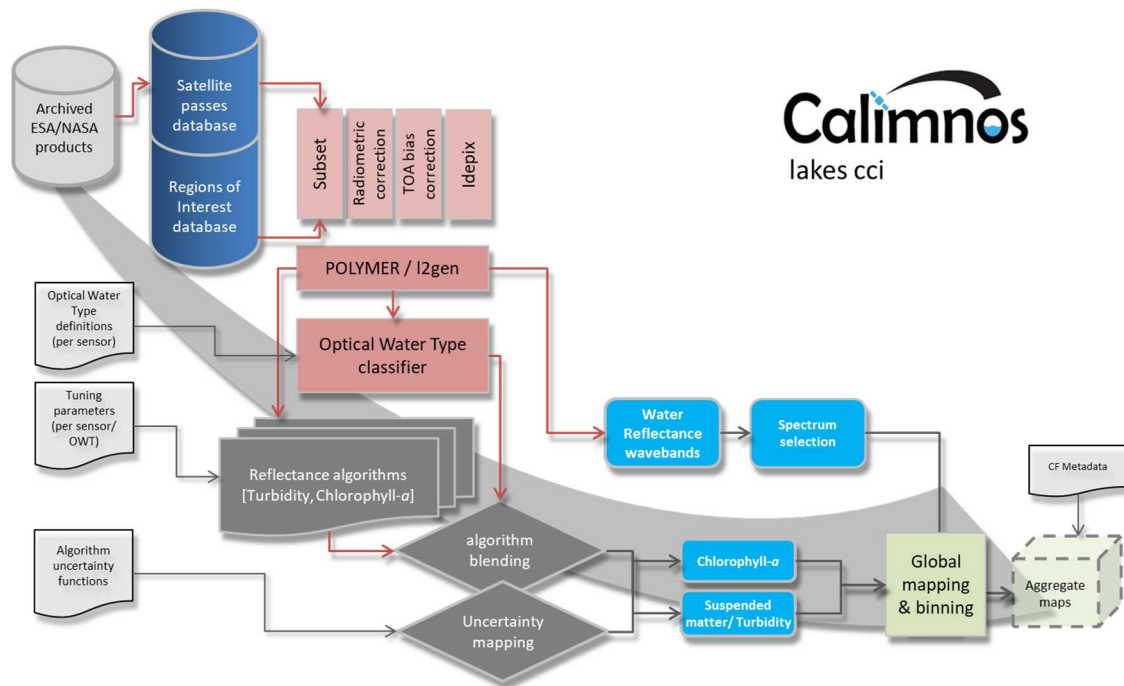
Figure 11: Schematic overview of the Calimnos processing chain for LWLR, Chlorophyll-*a* and turbidity or suspended matter.

3.5.2. Main functionalities

L2 processing steps

To produce Lake Water-Leaving Reflectance:

- Data discovery
- Subsetting around the lake areas of interest
- Radiometric and/or sensor Bias corrections
- Pixel identification as water/land/ice/cloud/cloud-shadow
- Atmospheric correction

To produce derived water-column properties:

- Optical water type classification
- Algorithm mapping and blending

For uncertainty characterization:

- Uncertainty mapping per algorithm and per optical water type

L3 processing steps

For merged lakes ECV product format consistency:

- Aggregation to 1-day intervals and Reprojection to a common planetary lat/lon grid
- Mosaicking into a single global product

3.5.3. Architecture

The processing system is built around the concept of workspaces which are individually submitted to a high performance computing environment.

L2 workspaces are created to contain symbolic links to all required input data for each given combination of target area and sensor overpass. L3 processing workspaces similarly combine all L2 inputs for a given aggregation period. Workspace creation and processing is done using in-house software written for Python 2.7, submitting each workspace to a set of processing stages (described in 3.5.2) which are individually monitored and timed. Log files are created for bulk workspace creation and individual workspace monitoring. For operational processing, a job monitoring database (postgreSQL) is used to follow the timing of completion of L2 jobs, re-start jobs that are in error, and launch L3 processing when all upstream jobs are completed. This process is fully automated.

For a typical lake area, the number of L2 processing workspaces for MERIS and OLCI observations is in the order of 3000-5000 observations per lake or approximately 1 Million individual processing jobs for the 250 lakes included in CDRP V1.0.

Workspaces are submitted to a Sun Grid Engine processing queue and controlled using environmental resource allocations. Memory (RAM) and temporary local storage criteria are set at the individual job level depending on lake size to ensure that processing nodes with different capabilities are optimally used. Each gridnode has access to a shared software repository and storage media containing the input data, through a Gigabit network interface. The processing grid is normally configured to provide up to 850 individual processing slots, which are used in parallel. A containerised implementation of the processing chain is under development, which will reduce any risk of software version deviations going unnoticed on older or newly introduced hardware.

When an instance of a workspace is started on a gridnode, it is copied to local storage and a supervisor script then executes each stage of the processing chain, logging both standard and error output streams to disk. The processing stages are coded into an xml file where each stage can be switched on or off depending on requirements. A separate configuration file stores common algorithm parameterisation settings and instructions for individual processing stages, which allows a common processing stage configuration to be used with project-specific parameterisation.

Upon completion of a workspace, any files that need to be archived are copied to their archiving destination. Workspace logs are normally kept up to 1 month following production. A copy of the configuration settings is kept with the output data.

External factors such as server disk space and internet connections are monitored for the processing site using Nagios, with personnel on call to resolve any issues.

3.5.4. Input

Input satellite data consist of MERIS Reduced Resolution L1B from the 3rd reprocessing (the fourth reprocessing became available in 2020 and will be used in future) and OLCI full resolution L1B data from the Sentinel-3A and 3B platforms. All downloaded products (a full archive is kept) are referenced in a postgreSQL database.

A polygonised and manually corrected version of the ESA CCI Land Cover (v4.0) maximum water extent at 150-m resolution is employed as lake mask. The polygons are available at

https://github.com/pmlrsg/lake-polygons-PML. They are accessible to Calimnos from a postgreSQL database.

Configuration files are used as input to each L2 or L3 workspace and comprise:

- Processing stages in xml format
- Common algorithm parameterization in ascii format compatible with Python configparser.
- Optical water type spectra in comma separated ascii format
- Processing environment requirements in ascii format, generated from a common template at workspace creation for each lake/sensor combination.

3.5.5. Output

The outputs comprise:

- L2 products, at the native sensor resolution, 1 file per combination of satellite overpass and region of interest. These contain most intermediary products and can be used for match-up validation and further algorithm development. They are not disseminated.
- L3 products, at a common grid, 1 file per combination of region of interest and aggregation time frame.
- Mosaicked L3 product, 1 file per time step containing all regions for which input data was available.
- Log files of each processing level, separated by processing stage.

3.5.6. Verification Tests

The Calimnos codebase is versioned using Git (private Gitlab repository) and tagged by software version number. The software version number for the Lakes_cci CDRP v1.0 is 1.4.

Integration tests are completed as part of code review, for each compatible sensor, against a template configuration which includes all processing stages.

Selected lakes and time periods are produced and inspected prior to large-scale processing. Log files are screened for common errors (such as network glitches) while any remaining errors are manually inspected and resolved. Consistency in file size is checked and selected files are visually inspected.

Implementation of new versions of software dependencies normally results in a minor version increase and will be subject to the above tests.

4. Requirement coverage

This part indicates if the parameter/product is compliant with the Cardinal requirements (CR) and Technical requirements (TR) indicated in the SoW as well as the requirements from the users.

From	Target	LWL	LWE	LSWT	LIC	LWLR
CR-1	Develop and validate algorithms to approach the GCOS ECV and meet the wider requirements of the Climate Community (i.e. long term, consistent, stable, uncertainty-characterized) global satellite data products from multi-sensor data archives.	Yes	Yes	Yes	Yes	Partial
CR-2	Produce, validate and deliver consistent time series of multi-sensor global satellite ECV data products for climate science	Yes	Yes	Yes	Yes	Yes
CR-3	Maximise the impact of European EO mission data on climate data records	Use all altimetric missions	Yes	Yes	No for v1.0. Yes for v2.0.	Yes
CR-4	Generate and fully document a production system capable of processing and reprocessing the data in CR-2, with the aim of supporting transfer to operational activities outside CCI (such as C3S).	Yes	Yes	Yes	Yes	Yes
R-3	Each CCI project (the contractor) shall make significant progress towards meeting the corresponding GCOS requirements for their ECV	Yes	Yes	Yes	Yes	Yes
R-7	Each CCI project (the contractor) shall take into account the legacy of the CCI in their projects. This involves: • adopting the community consensus concept • learning from, understanding and building upon the success achieved in CCI • espousing the implicit request to contribute actively to the core elements of the CCI Programme, in particular Colocation and Integration meetings, working groups, cross-programme activities (Open Data Portal, Toolbox, Knowledge Exchange) and cross-project initiatives.	Yes	Yes	Yes	Yes	Yes

From	Target	LWL	LWE	LSWT	LIC	LWLR
R-9	Each CCI project team (the contractor) shall take full account of the following key technical constraints when planning and implementing the CCI project: During Phase 1 the project consortium shall need to respond to the following technical constraints: • Need for scientific consensus on detailed ECV product and performance specifications • Availability and quality input data from EO Archives (ESA and non-ESA) • Availability and quality of associated metadata, cal/val data, and documentation • Compatibility of data from different missions and sensors • Trade-offs between cost, complexity and impact of new algorithms to be developed and validated during the project • Advance planning for data from new missions to be integrated during the project • End-to-end throughput of ECV production systems • Re-use of existing capabilities within Europe • Compliance to applicable standards • Availability of external validation data • No duplication of activities covered by other projects or programmes (e.g.H2020, Copernicus, national funding)	Yes	Yes	Yes	Yes	Yes
R-16	Each CCI project team (the contractor) shall integrate data from the Copernicus Sentinels and other key satellite missions within the relevant CCI processing systems and ECV data products	Yes	Yes	Yes (Sentinels will be V2.0)	Yes	Yes

From	Target	LWL	LWE	LSWT	LIC	LWLR
R-17	Each CCI project team (the contractor) shall ensure that the system is adequately dimensioned to accommodate the growing volumes of input and output data, and the increasing computational loads needed to process, reprocess, quality control, validate, and disseminate multi-decadal, global, ECV data products, of the required climate quality, in a timely manner.	Yes	Yes	Yes	Yes	Yes
TR-3	The Contractor shall confirm and demonstrate that all work undertaken in Lakes_cci is complementary to other Lake ECV development and delivery activities being conducted by ESA and other agencies e.g. GloboLakes, HydroWeb, Copernicus Climate Change Service and Copernicus Global Land Services. It shall coordinate its activities with these efforts to ensure the information on lakes is effectively and efficiently produced for the investment from these different projects.	Yes	Yes	Yes	Yes	Yes
TR-4	The project shall be consistent and compatible with previous CCI projects, in particular, LandCover_cci in terms of its Permanent Water Bodies product	Yes	Yes	Yes	Yes	Yes
TR-7	The project shall ensure consistency across the different lake products and deliver all the products to established lakes databases (GTN-H - HYDROLARE, HYDROWEB, NSIDC Global Lake and River Ice Phenology, GLTC).	Yes	Yes	Yes	Yes	Yes
TR-8	The project shall ensure consistency in production across multiple satellites, in particular focusing on different spatial and temporal resolutions offered consistent with the range of lake sizes that require monitoring	Yes	Yes	Yes	No for v1.0. Yes for v2.0.	Yes
TR-10	The contractor shall aim to generate the most accurate, stable and uncertainty characterised products by continuously evaluating and developing the selected algorithms and keeping abreast of new developments in the field	Yes	Yes	Yes	Yes	Yes

From	Target	LWL	LWE	LSWT	LIC	LWLR
TR-11	The contractor shall ensure that all products are consistent across product streams and thus the system delivers all products for the lakes under evaluation. This shall include consideration of baseline lake identifiers, sensor geolocation, algorithm stability and accuracy, required preprocessing/sensor corrections,	Yes	Yes	Yes	Yes	Yes
TR-14	The Contractor shall review algorithm performance and their implementation in the light of developments in the scientific literature, application to multiple sensors, robustness and accuracy on a regular basis and evaluate the need to conduct reprocessing to improve lake ECV products on a regular basis	Yes	Yes	Yes	Yes	Yes
TR-17	The contractor shall develop a prototype processing system (Lake-System) to generate all products for the Lake ECV	Yes	Yes	Yes	Yes	Supported
TR-18	The Lake-System shall provide the scientists a configurable, flexible, agile and open Lake CCI workflow for managing the evolution of the Lake ECV. It shall have the following capabilities: • Multi-sensor full mission (re-)processing to all Lake ECV products • Systematic, data-driven processing, allowing rapid ingestion of new data • Configuration management (processor versions, auxiliary data, input and output data, etc …) • Capability to integrate (existing) tools developed in different programming languages • Capability for algorithm developers to easily trial plug in, without recompilation where practical, and execute new algorithms, new versions of algorithms or new parameterisation, for testing, intercomparison and evaluation purposes • Support efficiently all the needs of the scientific algorithms and the new development	Yes (spanning LWL–LWLR)				

From	Target	LWL	LWE	LSWT	LIC	LWLR
TR-19	The Lake-System shall include tools for full mission time series analysis, round robin algorithm intercomparison, quality control and other comprehensive massive product analysis (feature extraction, dependencies (e.g. from sensor, detector, geometry, latitude/longitude, etc.). The development of these tools shall be coordinated with, and transferred to, the CCI Toolbox and, where possible, compatible with existing web services (e.g. Global Surface Water Explorer).	No plans for such tools				
TR-20	The Lake ECV products shall be made available to the users through appropriate distribution mechanisms with full user support and analysis tools	No plans for distributions. tools				
TR-21	The Lake-System shall include data access, ingestion, product conversion tools and distribution functionality, and shall address long-term archiving of both input and output products. Both requirements shall meet the generic needs outlined in the main body of the CCI SoW.	Yes				
TR-22	The design of the Lake-system shall be based on experience where relevant from previous CCI projects and/or external processing systems	Yes				
TR-23	The Lake-System shall have data access interfaces to handle efficiently, and in a standardised fashion, the massive primary and auxiliary data streams from Sentinel 1, 2 and 3. The interfaces shall support the possibility of cross-ECV synergies.	Yes				
TR-24	The contractor shall link with the CCI portal and data analysis/visualisation tools available through the CCI Toolbox	To be done by those projects				
TR-25	The Contractor shall generate global maps of lakes at multiple resolutions derived from multi-sensor Lake FCDRs. The products shall address all GCOS Lake ECV requirements, pushing products as far as scientifically and technically possible.			All GCOS addressed		

From	Target	LWL	LWE	LSWT	LIC	LWLR
TR-26	The project shall generate prototype products of: • Lake level • Lake area • Lake surface water temperature • Lake colour (and its derived properties: turbidity, chlorophyll and coloured dissolved organic matter) • Lake ice coverage (where relevant)	Yes	Yes	Yes	Yes	Partial: CDOM not yet possible
TR-27	The project shall generate products for at least 2000 lakes covering the range of lake sizes and types with priority given to those lakes of climate importance as a demonstration of system capability	250 selected for the first version of the dataset				
TR-28	The project shall focus on addressing lakes in the priority regions identified by the IPCC. Namely: • Lake surface warming, biogeochemistry and water column stratification increases in the African Great Lakes, Lake Kariba. • Increased lake water temperatures, biogeochemistry and ice phenology in the Arctic. • Associated change in lake patterns from permafrost degradation in Siberia, Central Asia, Tibetan Plateau, Arctic (both thermakarst lake loss and new lake generation in frozen peat). • Shrinking mountain glaciers across most of Asia, Andes, western North America, Arctic and associated development and expansion on foreglacier lakes. • Lake changes in response to changes in snowpack in Australasia and western North America. • Lakes associated with changes in river discharge patterns in circumpolar rivers, Amazon river, western Andes and La Plata river.	Yes				
TR-29	The project shall aim to provide the longest time series possible based on available data products with 10 years being the minimum requirement	Yes	Yes	Yes	Yes	Yes

From	Target	LWL	LWE	LSWT	LIC	LWLR
TR-30	The Lake ECV products shall: • Cover the time frame from 1992-end of this contract. • Be derived from all available and suitable satellite instruments listed in Tables 4-8 of the SoW • Include RMSE and bias uncertainty estimates on a per pixel basis following guidelines expressed in the main Statement of Work (see also [RD-8]). The confidence in these uncertainty estimates shall be stated	V1: 1992 - 2019	V1: 1992- 2019	Yes	Yes	Yes V1.0: 2002-2012 and 2016- 2019 MERIS & OLCI currently suitable (v2.0 to expand)
TR-31	The Lake ECV products shall comply with the actual version of the CCI Guidelines for Data Producers [RD-7]. Note: The CCI project shall adapt to changes in these Guidelines as these will be further developed jointly by the CCI projects.	Yes				
TR-33	The Contractor shall generate at least 2 versions of ECV products: one following the round-robin, then one incorporating the results from user feedback.	Yes	Yes	Yes	Yes	V1.0 based on prior round-robins
TR-34	Full ECV product validation shall be achieved by a combination of: a. Activities of the CCI EO Science Team b. Activities of the Climate Research Group engaged in the Lake CCI project c. Involvement in the CMUG process d. Interaction with key ecosystem modelling groups e. Involvement of key stakeholders of other international climate research projects.	Yes				

Annex A.   Project Acronyms

This is a generic list containing all the acronyms used in the project

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

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

SPONGE	SPaceborne Observations to Nourish the GEMS
SRD	System Requirements Document
SSD	System Specification Document
SST	Sea Surface Temperature
STSE	Support To Science Element
SWOT	Surface Water and Ocean Topography
TAPAS	Tools for Assessment and Planning of Aquaculture Sustainability
TB	Brightness Temperature
TM	Thematic Mapper
TOA	Top Of Atmosphere
TR	Technical Requirement
UNEP	United Nations Environment Programme
UoR	University of Reading
US	United States
VIIRS	Visible Infrared Imaging Radiometer Suite
WCRP	World Climate Research Program
WHYCOS	World Hydrological Cycle Observing Systems
WMO	World Meteorological Organization
WP	Work Package