ESA Climate Change Initiative
River Discharge Precursor
(RD_cci+)

D.10. Product User Guide (PUG)

Contract number: 4000139952/22/I-NB

Reference

Issue 1 –[Issue Date]

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

CHRONOLOGY ISSUES

Issue	Date	Object	Written by
1.0	30/04/24	Initial version	Laetitia Gal Sylvain Biancamaria Paolo Filippucci
1.1	19/06/2024	Modified version following ESA comments	Laetitia Gal Sylvain Biancamaria Paolo Filippucci

Checked by	S. Biancamaria - LEGOS	sylvain.biancamaria
Approved by	Alice Andral - CLS	A.Andral
Authorized by	Clément Albergel - ESA	clement.albergel

DISTRIBUTION

Company	Names	Email
ESA	Clément Albergel	clement.albergel@esa.int
ESA	Jérôme Benveniste	Jerome.Benveniste@esa.int
CLS	Alice Andral	aandral@groupcls.com
CLS	Yann Bernard	ybernard@groupcls.com
CLS	Beatriz Calmettes	bcalmettes@groupcls.com
CLS	Daya Ceccone	dceccone@groupcls.com
CLS	Nicolas Taburet	ntaburet@groupcls.com
CNRM	Simon Munier	simon.munier@meteo.fr
EOLA	Elena Zakharova	zavocado@gmail.com
Hydro Matters	Malik Boussaroque	malik.boussaroque@hydro-matters.fr
Hydro Matters	Laetitia Gal	laetitia.gal@hydro-matters.fr
Hydro Matters	Adrien Paris	adrien.paris@hydro-matters.fr
IRPI	Silvia Barbetta	silvia.barbetta@irpi.cnr.it
IRPI	Paolo Filippucci	paolo.filippucci@irpi.cnr.it
IRPI	Debi Prasad Sahoo	debi.prasad.sahoo@irpi.cnr.it
IRPI	Angelica Tarpanelli	angelica.tarpanelli@irpi.cnr.it
LEGOS-CNRS	Sylvain Biancamaria	sylvain.biancamaria@univ-tlse3.fr
LEGOS-CNRS	Benjamin Kitambo	benjamin.kitambo@univ-tlse3.fr
LEGOS-CNES	Julien Lefebve	julien.lefebve@univ-tlse3.fr
LEGOS-IRD	Fabrice Papa	fabrice.papa@ird.fr
Magellium	Gilles Larnicol	gilles.larnicol@magellium.fr
Magellium	Vanessa Pedinotti	vanessa.pedinotti@magellium.fr

LIST OF CONTENTS/SOMMAIRE

LIST OF TABLES AND FIGURES

REFERENCE DOCUMENTS

[RD-1] D.1. User Requirements Document for CCI River Discharge precursor project (CCI-Discharge-0003-URD, Issue 1.1), https://climate.esa.int/media/documents/D1_CCI-Discharge-0003-URD.pdf

[RD-2] D.2. Selection of river basins. CCI River Discharge precursor project Document (CCI-Discharge-0004-RP_WP2, Issue 1.1), https://climate.esa.int/media/documents/D2_CCI-Discharge-0004-RP_WP2_v1-1.pdf

[RD-3] D.3. Water Surface Elevation (WSE) Algorithm Theoretical Basis Document (ATBD) (CCI-Discharge-0005-ATBD-WSE, Issue 1.2), https://climate.esa.int/media/documents/CCI-Discharge-0009-ATBD-WSE_v1-2.pdf

[RD-4] D.3. River Discharge (Q) from Altimeters and Ancillary data, multispectral images and data combination - Algorithm Theoretical Basis Document (ATBD) (CCI-Discharge-0012-ATBD, Issue 1.1), https://climate.esa.int/media/documents/CCI-Discharge-0012-ATBD-Discharge.pdf

[RD-5] The ESA River Discharge CCI+ precursor project website, https://climate.esa.int/en/projects/river-discharge/about-the-river-discharge-project/

[RD-6] ESA Climate Office (2021). CCI Data Standards. Ref. CCI-PRGM-EOPS-TN-13-0009, issue 2.3, available at https://climate.esa.int/documents/1284/CCI_DataStandards_v2-3.pdf

[RD-7] D.4. Release Note on Water Level Climate Research Data Package (CCI-Discharge-0006-WSE-CRDP-ReleaseNote, Issue 1.1, https://dap.ceda.ac.uk/neodc/esacci/river_discharge/docs/WL/v1.1/CCI-Discharge-0006-WSE-CRDP-ReleaseNote_v1.1.pdf)

[RD-8] D.5. Release Note on Altimetry-Based River Discharge Climate Research Data Package (CCI-Discharge-0013-Qalti-CRDP-ReleaseNote, Issue 1.0)

LIST OF ACRONYMS

ATBD	Algorithm Theoretical Basis Document
AlTiS	Altimetry Time Series software
CCI	Climate Change Initiative
CDF	Cumulative Distribution Function
CDR	Climate Data Record
CF	Climate and Forecast
CM	Difference between a Calibration (*C*) and a Measurement (*M*) pixels
CRDP	Climate Research Data Package
CSV	Comma-Separated Values
CTOH	Centre for Topographic Studies of the Oceans
ECMWF	European Centre for Medium-Range Weather Forecasts
ECV	Essential Climate Variable(s)
Envisat	Environmental Satellite
EO	Earth Observation
ERA5	ECMWF Reanalysis v5
ERS-1	European Remote Sensing satellite 1
ERS-2	European Remote Sensing satellite 2
ESA	European Space Agency
FDC	Flow Duration Curve
GCOS	Global Climate Observing System
GDR	Geophysical Data Records
GRDC	Global Runoff Data Centre
J1	Jason-1
J2	Jason-2
J3	Jason-3
LEGOS	Laboratoire d'Etudes en Géophysique et Océanographie Spatiales
LRM	Low Resolution Mode
MCMC	Markov Chain Monte Carlo
MH	Metropolis-Hasting sampler
MODIS	Moderate resolution imaging spectroradiometer
NetCDF	Network Common Data Form
NIR	Near-InfraRed
NUTS	No-U-Turn Sampler
OCOG	Offset Centre of Gravity
PUG	Product User Guide
RC	Rating Curve
RD	River Discharge
RDC	Recession Duration Curve
RD-alti	River Discharge from altimeters and ancillary data
RD-multi	River Discharge from multispectral images and ancillary data
S3A	Sentinel-3A
S3B	Sentinel-3B
S6	Sentinel-6 Michael Freilich
Saral	Satellite with ARgos and ALtika
SWOT	Surface Water and Ocean Topography
TP	Topex-Poseidon
UTC	Coordinated Universal Time
VS	Virtual Station
WL	Water Level
WMO	World Meteorological Organization
WP	Work Package
WSE	Water Surface Elevation

1. Introduction

This document provides all the information needed by the users to successfully employ the CCI River Discharge CRDPs in their work. The PUG may reuse information that was originally provided in the ATBDs and release notes for each product.

Section 2 is dedicated to the general background of the project and its objectives. Section 3 focuses on the description of the products and the methodologies used to compute them with a part on the limitations and weaknesses associated to these products. Section 4 provides a detailed description of the data format, with guidance for their reading and display. For readers information, Calibration/Validation periods defined for in situ data used to compute satellite-data/discharge relationship for each station are provided in [RD-4, Annex]. In all sections, references to specific CCI documents are provided for readers needing more details.

2. General Background

Satellites are essential tools for observing Earth's surface and monitoring its changes. The Global Climate Observing System (GCOS) has identified numerous Essential Climate Variables (ECVs), more than half of which can benefit from Earth Observation (EO) data collected by satellites (GCOS, 2022). The European Space Agency (ESA) has recognized the significance of utilizing satellite EO data for climate monitoring and has thus initiated the Climate Change Initiative (CCI) over a decade ago. The primary objective of CCI is to leverage long-term global EO archives to realize the full potential of satellite observations in understanding and addressing climate change (ESA Climate Change Initiative).

As of early 2023, the CCI projects have focused on 27 ECVs, including some like river discharge, which are still in the development phase (ESA Climate Change Initiative). River discharge, defined by the World Meteorological Organization (WMO) as the volume of water flowing through a river or channel per unit of time, is crucial for understanding the dynamics of the water cycle (WMO, 2012). It plays a pivotal role in transporting water from land to oceans, with approximately 0.0002% of Earth's total water stored in river networks, amounting to around 36,000 km^3 per year (Gleick, 1996; Milliman and Farnsworth, 2013).

Climate change significantly impacts the water cycle, necessitating long-term monitoring of river discharges to assess its effects on continents and facilitate adaptation strategies (Trenberth, 2011). However, obtaining consistent and reliable long-term river discharge data is challenging due to various factors. These include difficulties in accessing remote gauge locations, a decreasing number of gauges worldwide, limitations imposed by national or regional agencies on gauge time series sharing, and economic constraints hindering the maintenance or expansion of gauge networks (Global Runoff Data Centre [GRDC]; Milliman and Farnsworth, 2013).

The decline in the availability of in-situ gauge measurements has been observed since the mid-20th century, exacerbating the challenges associated with monitoring river discharge (Milliman and Farnsworth, 2013). Furthermore, many in-situ gauge measurements, especially in transboundary river basins, are not shared publicly due to sensitivity concerns.

In this context, EO satellites offer a promising solution to preserve and enhance our capacity to observe and understand climate change's impact on continental freshwater resources. The CCI River Discharge precursor project aims to capitalize on EO data, including nadir radar altimeters and multispectral images, spanning over two decades [RD-5]. However, utilizing EO data poses challenges, as highlighted in previous studies (Biancamaria et al., 2017; Crétaux et al., 2023; Tarpanelli et al., 2021).

For instance, while multispectral images provide superior temporal resolution and spatial coverage compared to altimetry data, their effectiveness is impeded by cloud cover, particularly in mountainous regions. Concerning nadir altimeter, they have two main limitations: their coarse spatial sampling (i.e. measurements are done only along the satellite ground tracks) and temporal sampling (from 10 days to 35 days, depending on the mission).

Despite these challenges, the CCI precursor project endeavors to meet the stringent requirements of various scientific communities [RD-1]. For example, oceanographers require discharge data to understand freshwater inputs into oceans, necessitating precise measurements over extended periods. Similarly, hydrologists utilize discharge data to model river flows, assess water availability, and manage water resources effectively. Their requirements [RD-1] include geophysical measurements of river discharge, monthly average time series products, and a time span extending from 2002 to 2022, with a goal to cover the 1995-2022 time span. 18 rivers basins have been selected to be representative of diverse climatic zones and anthropization levels, albeit excluding mountain basins due to data limitations, as well as 54 stations to take into account the requirements to each scientific community (Figure 1) [RD-2].

Through meticulous data integration and adherence to predefined requirements tailored to each scientific community [RD-1], the CCI River Discharge precursor project aims to deliver robust discharge products [RD-3, RD-4] meeting the diverse needs of the oceanographic, water cycle, hydrology, and related communities [RD-1].

In summary, the main goal of the CCI River Discharge precursor project is to integrate multispectral and altimetry satellite data to compute long-term river discharge time series [RD-5]. By addressing user requirements defined in [RD-1], the project seeks to advance our understanding of hydrological processes and climate change.

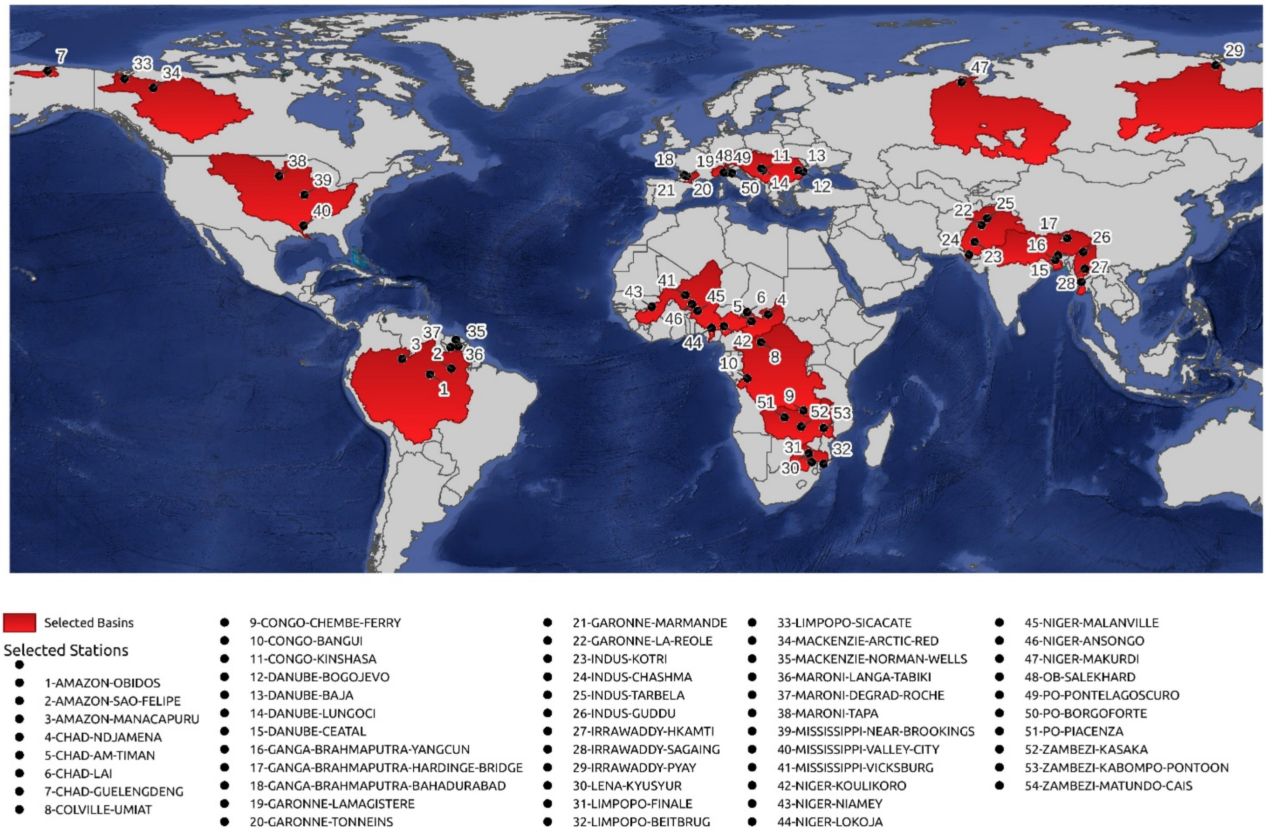


Figure 1: Selected basins (red) and selected locations (black points) near which nadir radar altimeter WSE time series have been computed

3. Products descriptions

To assess the potential for deriving long-term discharge ECV time series from remote sensing observations and ancillary data, three types of products have been provided at selected locations (WSE, Altimeters-base RD and Multispectral images-based RD) [RD-3, RD-4]. The table below summarizes the methodologies employed for each station. These methodologies will be described in the following sections.

BASINS	STATIONS	PRODUCTS						
		WSE		Altimeters-based RD		Multispectral-based RD		
		Single mission	*Merge mission*	*Overlap approach*	*Non-overlap approach*	*cal-BestFIT*	*uncal-CDF*	*cal-Copula*
AMAZON	OBIDOS	X	X	X		X	X	X
	SAO-FELIPE	X	X	X		X	X	X
	MANACAPURU	X	X	X		X	X	X
CHAD	NDJAMENA	X					X	
	AM-TIMAN	X	X		X		X	
	LAI	X	X		X		X	
	MAILAO	X	X		X		X	
COLVILLE	UMIAT	X	X	X		X	X	X
CONGO	CHEMBE-FERRY	X	X	X			X	
	BANGUI	X	X	X		X	X	X
	KINSHASA	X	X	X		X	X	X
DANUBE	BOGOJEVO	X	X	X			X	
	MOHACS	X	X		X		X	
	LUNGOCI	X	X	X			X	
	CEATAL	X	X	X			X	
GANGES-BRAHMAPUTRA	YANGCUN	X	X		X		X	
	HARDINGE-BRIDGE	X	X		X		X	
	BAHADURABAD	X	X		X		X	
GARONNE	LAMAGISTERE	X	X	X		X	X	X
	TONNEINS	X	X	X		X	X	
	MARMANDE	X	X	X		X	X	X
	LA-REOLE	X	X	X		X	X	X
INDUS	KOTRI	X	X		X			
	CHASMA	X	X					

	TARBELA	X	X					
	GUDDU	X	X					
IRRAWADDY	HKAMTI	X	X	X			X	
	SAGAING	X	X	X			X	
	PYAY	X	X	X			X	
LENA	KYUSUR	X	X	X		X	X	X
LIMPOPO	FINALE	X	X		X	X	X	X
	BEITBRUG	X	X	X		X	X	
	SICACATE	X	X		X		X	
MACKENZIE	ARCTIC-RED	X	X	X		X	X	X
	NORMAN-WELLS	X	X	X		X	X	X
MARONI	LANGA-TABIKI	X	X	X		X	X	X
	DEGRAD-ROCHE	X	X		X	X	X	X
	TAPA	X	X	X		X	X	X
MISSISSIPPI	NEAR-BROOKINGS	X	X	X		X	X	X
	VALLEY-CITY	X	X	X		X	X	X
	VICKSBURG	X	X	X		X	X	X
NIGER	KOULIKORO	X	X	X			X	
	NIAMEY	X	X	X		X	X	X
	LOKOJA	X	X		X		X	
	MALANVILLE	X	X		X		X	
	ANSONGO	X	X	X			X	
	IBI	X	X		X		X	
OB	SALECKHARD	X	X	X		X	X	X
PO	PONTELAGOSCURO	X	X	X		X	X	X
	BORGOFORTE	X	X	X		X	X	X
	PIACENZA	X	X	X		X	X	X
ZAMBEZI	KASAKA	X	X	X			X	
	KABOMPO-PONTOON	X	X	X			X	
	MATUNDO-CAIS	X	X		X		X	

Table 1: Summarize the computation of time series for each station and product based on the methodology employed.

3.1. Water Surface Elevation from altimetry

3.1.1. Data definition

Long time series Water Surface Elevation (WSE), which could also be referred to as Water Level (WL) in the literature or CCI products, are measured from different satellite nadir altimeter missions since 1992

[RD-3]. This data provides valuable information about river water levels dynamics, which is crucial for understanding hydrological processes, managing water resources, and assessing environmental impacts. It is also an ECV defined by GCOS and is a good proxy of river discharge, through the so-called rating curve approach (see [RD-4] for more details).

WSE is defined as the distance between the river water surface and a reference surface (ellipsoid or geoid) [RD-3]. The chosen reference surface for the CCI River Discharge project is the WGS84 ellipsoid. The geoid is more meaningful from a hydraulic point of view. However, as WSE is not used to compute river slopes in this project and as many global to national geoids are available (and multiple versions of a specific geoid might exist), it is better suited to use a mathematically defined surface, i.e. an ellipsoid.

This precursor project used WSE time series near locations identified in [RD-2] already computed and available on the Hydroweb database (https://hydroweb.theia-land.fr). This database contains mainly time series from S3A/B, J3 and S6 missions' Virtual Stations (VS). However, it has been needed to extensively extend this database for past missions and VS with current missions not available on Hydroweb [RD-3].

The corresponding DOI for this product is: 326a574b-42f9-4039-99e5-dcbd81e041e4.

3.1.2. Data characteristics

The timeline and repeat cycle of all nadir radar altimeter missions used are provided in Figure 2. This figure gathers, through the same color code, missions that were on the same orbit tracks. This means they observe the same locations with the same repetitiveness. If the TP/J1/J2/J3/S6A observe the same VS every 10 days from 1992 to now, with some time overlaps between consecutive missions, this is not the case for other orbits. The ERS-1/ERS-2/Envisat/Saral 35-days orbit tracks are not sampled since 2016, when Saral satellite began to drift and not being anymore on a repeat orbit. Another issue is the absence of time overlap between Envisat on its nominal orbit and Saral launch, leading to a few years observation gap. S3A and S3B missions are on another orbit, with a better time sampling (27 days), but have been launched quite recently (2016 and 2018, respectively). To sum up nadir radar altimeter missions used are the following: ERS-2, Envisat, Saral, Topex-Poseidon, Jason-1, Jason-2, Jason-3, Sentinel-3A/B, and Sentinel-6 [RD-3, RD-7]. Use of Cryosat-2 has been investigated to complete and correct bias in WL time series between Envisat and Saral, but because of river slopes and crude time sampling of the mission, it has been decided not to use Cryosat-2 data in this precursor project [RD-7].

The satellite orbit defines both the spatial and temporal sampling of the nadir altimeter mission. They change in opposite directions: the greater the number of tracks in an orbit, the finer its spatial sampling is, but the greater its repeat period (i.e. the time taken for the satellite to fly over the same point again), the coarser its temporal sampling is. Therefore, Jason series allows a better time sampling than other orbits (i.e. 10 days), but the counter part is the scarcity of its spatial sampling (nadir altimeter observing only within the footprint of the instrument at the nadir of the satellite and Jason series intertrack distance at the equator is equal to 315 km). It means that for some selected locations, only observations from altimeters on Envisat orbit could be used, leading to observation gaps (at least between Envisat change of orbit and launch of Saral) and a coarser time sampling of 35 days. Furthermore, due to some technical limitations, satellite ground tracks are controlled to within ± 1 km around their nominal positions for most altimeter missions (for more details, see product handbook for each mission).

It should also be noted that the oldest missions are the least accurate. The most important issue arises for J1, which was finer tuned to observe the ocean than TP, resulting in less data acquisition over continental water bodies.

The intersection of the satellite ground track with a targeted water body (e.g. a river reach) is usually referred to as "virtual station" (VS) in scientific literature. Its definition is therefore intrinsically linked with the orbit of the radar nadir altimeter mission considered. The VSs from all available missions tracks near the selected locations in [RD-2] have been processed to compute WSE time series.

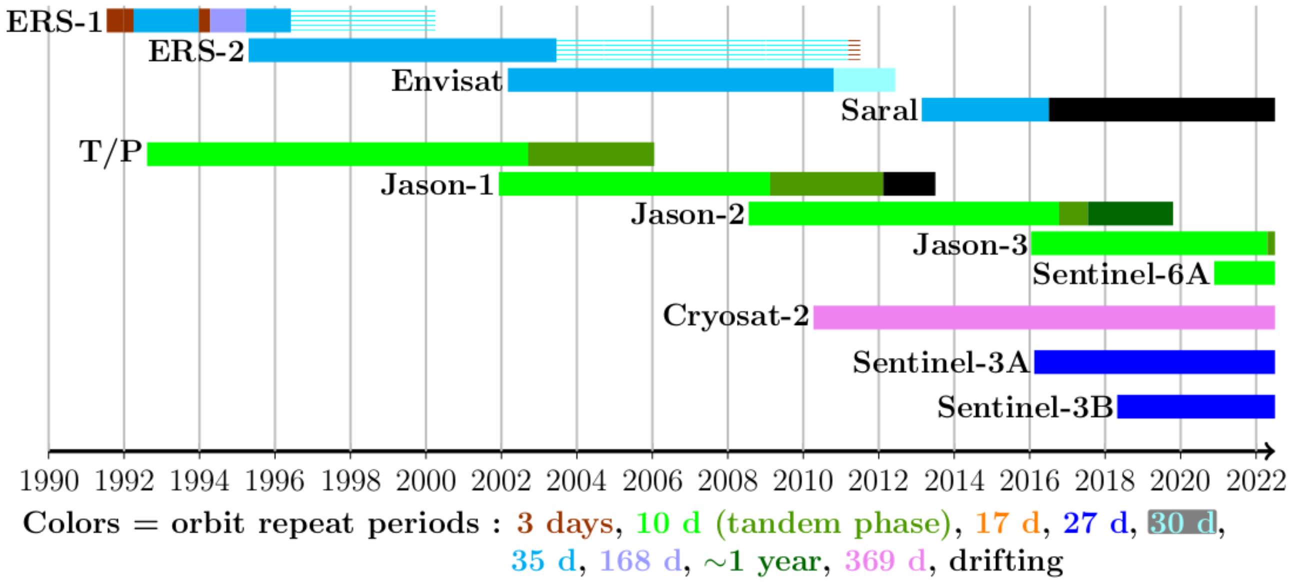


Figure 2. Timeline of the altimetry missions considered in this precursor project. Colors correspond to missions' orbits repeat periods. After June 1996, ERS-1 is in back-up mode and no measurements are recorded and from Mid-2003, altimeter onboard ERS-2 stopped working. That's why the boxplot patterns for these two missions after these dates are changed to show the absence of measurements.

3.1.2.1. Time series per mission

The methodology employed for deriving Water Surface Elevation (WSE) time series from individual radar nadir altimeter missions over rivers is quite standard now (see for example Cretaux et al. 2017, 2023). It requires various corrections and selection processes to ensure the accuracy and reliability of the generated data. These corrections encompass ionospheric correction, dry and wet atmospheric corrections, solid Earth correction, and pole tides correction. These corrections are directly provided by the space agencies in the Geophysical Data Records (GDR) products. Notably, for missions such as TP, where atmospheric corrections are absent at multiple locations, a substitute approach is adopted. Leveraging climatology derived from corrections made during the Jason-1 to Jason-3 period, this ensures continuity and consistency in the derived WSE time series.

Nadir altimeter measurements basically provide a “waveform” that is used to compute the range, i.e. the distance between the satellite center of mass and the surface of the river. CCI River Discharge WSE have been computed using ranges estimated from the waveforms using the so-called “Ice-1” or “OCOG” retracker (Wingham et al., 1986; Bamber, 1994) [RD-3]. These retracked ranges are directly available in the GDR files. This retracker is commonly used to retrieve WSE over rivers, as stated, for example, in Cretaux et al. (2017).

In cases where data is not readily available within the Hydroweb database, an intricate manual selection process is initiated, utilizing specialized software tools, such as AlTiS developed at LEGOS/CTOH (https://gitlab.com/ctoh/altis). This manual selection involves several steps: including the delineation of a polygon at the intersection of the satellite ground track and the observed river reach, visualization and analysis of WSE for all cycles and measurements within this polygon, removal of outliers based on predefined criteria, computation of the median of selected WSE for each cycle, and subsequent exportation of the WSE time series (Santos da Silva et al., 2010) [RD-3].

Furthermore, the differences between the CCI Lake WSE and precursor CCI River Discharge WSE processing are the following: river WSE are referenced to an ellipsoid (contrarily to CCI Lake WSE referenced to a geoid) and located on a river reach, the geoid slope correction, crucial for lakes, is not applied for rivers WSE, bias between missions (see [RD-4] for more details) is computed differently as there are usually less time series overlaps.

Additionally, it is emphasized that no slope values are utilized to correct ± 1 km satellite drifts between each revisit time, because of the lack of a globally accurate river slope product. The potential consideration of such corrections, contingent upon the availability of validated SWOT river slope products, is not yet available globally [RD-3].

Lastly, dates in the time series are provided as UTC time to ensure standardization and facilitate compatibility across different datasets and analyses.

3.1.2.2. Merge time series

As there is not a single nadir altimeter mission covering the whole period of interest (2002-2022), the precursor project computed merged multi-mission WSE (WSE-merge). Users are also interested in getting merged WSE time series [RD-1]. This methodology briefly described below (see [RD-3] for more details) aims to address this need and evaluate the potential advantages of merged WSE compared to using multiple WSE time series.

The proposed approach involves computing a merged WSE at a specific Virtual Station (VS). This VS is intrinsically tied to a specific mission and its ground tracks. The selection of the reference VS prioritizes missions with the longest duration, highest time sampling, and recent launch dates to ensure accuracy. Jason-3 is suggested as the preferred reference mission due to its extended time span and high repeat period.

- **Merging Time Series on the Same Ground Track**: Biases between consecutive missions on the same ground track are computed and corrected. This involves calculating the mean bias over the common period between consecutive time series and adjusting the WSE accordingly. When multiple observations are available for a single day, priority is given to the most recent data.
- **Merging Time Series from Different Tracks**: After merging time series from missions on the same track, the WSE time series from different surrounding VSs need to be merged. A reference VS is identified, and a linear relationship between WSE time series from different VSs and the reference VS is computed, to consider bathymetry differences between VSs. A time lag is also applied, if needed, to take into account the flow propagation time.
- **Correction of Biases for Non-Overlapping Time Series**: a WSE climatology is computed for both time series and the bias is computed between these climatology. Extreme events or short high flow periods are addressed by removing the highest data points before computing the bias.
- **Monthly Discharge Time Series**: While monthly discharge time series are a goal, no similar requirement exists for WSE products. As a result, monthly WSE time series are not computed.

This methodology ensures the creation of a comprehensive and accurate merged WSE dataset with one measurement per day. It leverages data from multiple missions and employs various correction techniques to enhance accuracy, particularly for river discharge computations.

3.1.2.3. Uncertainties

Altimetry-based WSE uncertainties are quite difficult to estimate a priori and depend on any factors (sensor characteristics, orbit track orientation with reference to the river, complexity of the observation, previous measurements or data stored onboard…). We compared altimetry-based WSE to in situ WSE at some locations where such in situ data are available, to estimate a global WSE uncertainty per altimeter mission. We assumed that in situ WSE uncertainties are at least one order magnitude lower than altimetry-based WSE, therefore in situ WSE are considered as "perfect".

Given the potential for significant differences in uncertainty between Arctic rivers and other rivers, separate standard deviations for WSE per mission were computed. One standard deviation was calculated for the Arctic basin, while another one was computed for other regions.

For each of the 28 locations where in situ WSE data were available (5 locations for arctic region and 23 for lower latitudes), a comparison was made between the anomalies of these in situ WSE and the anomalies of WSE derived from altimetry merged time series over a common period. The closest dates were used as common dates if the lag time between the datasets did not exceed 24 hours. It should be noted that the absolute difference between in situ and altimetry WSE cannot be computed because in situ measurements are not made at the exact location of the VS. This also means that the difference between in situ and altimetry WSE could be due to difference in river bathymetry at the in situ and at the VS.

Following this comparison, the standard deviation of these errors was calculated for each mission (Eq.1). The decision to compute the standard deviation across all locations, rather than calculating the mean standard deviation for each location, aimed to address the challenge of assigning equal weight to each location regardless of the number of common dates per station.

$$\sigma_m = \sqrt{\frac{\sum_{i=1}^{n} (x_i - \mu)^2}{n-1}} \qquad \text{Eq. 1}$$

Where, σ_m is the standard deviation for the satellite mission m ; x_i represents each individual difference between in-situ water high anomaly and altimetry-based WSE anomaly; μ is the mean of the differences; n is the total number of observations (dates and stations).

The following table summarizes the standard deviation for each satellite mission for Arctic rivers and other rivers (lower latitudes) where data are available over the same period. Therefore, values for T/P, Jason-1, and Sentinel-6 have not been provided for Arctic rivers due to the absence of WSE data from these missions over the considered rivers. These standard deviations were calculated mostly based on a small number of points (fewer than 200 dates on average) and should therefore be considered with caution. However, they provide an initial indication of the errors related to altimetry.

	TP	*ers2*	*Envisat*	*J1*	*J2*	*J3*	*Saral*	*S3A*	*S3B*	*S6*
Arctic	-	0.50	1.13	-	0.86	1.02	1.12	1.02	0.87	-
Other	1.01	0.91	0.83	0.91	0.65	0.62	0.71	1.02	0.32	0.51

Table 2: Summary of the standard deviation of WSE anomaly difference to in situ WSE anomaly for each satellite mission, categorized into two groups: Arctic rivers and rivers from other regions. In parentheses, the number of dates used to calculate the statistics is indicated. The number of dates used to calculate the statistics is indicated in parentheses.

3.1.3. Data limitation

The main limitation is the heterogeneity of time sampling in time within each time series. This is due to the fact there is no satellite altimeter mission that lasts 20 years. As time series from multiple missions are merged, which could have different orbit characteristics, some merged time series could have a 35 days' time sampling during the Envisat period and then ten days' time sampling if some data from Jason-2 or its successor can be used. There are some important time gaps in time series, due to the non-overlapping period between Envisat and Saral, or the issue of data loss for the missions in "closed loop tracking mode" (see RD-3). This last issue is particularly important at Finale and Beitburg locations in the Limpopo basin, for which there is no data before 2017.

There are also some sensor characteristics changes between the oldest and the newest missions, which explains why the accuracy can change in time. The "Low Resolution Mode" (LRM) missions (ERS-2, Envisat, Saral, Topex-Poseidon, Jason-1, Jason-2, and Jason-3) have large footprints (8km, 18km and 30km diameter for Saral, Envisat and Jason series, respectively), which means that radar altimeter waveforms can record information from any water bodies (and even other targets) in the footprint. Therefore, waveforms can have multiple peaks, and only one corresponds to the targeted water body. This means that the OCOG retracked range and the corresponding WSE might not correspond to the river WSE. Such cases can be easily removed when the retracked data is some order of magnitude higher than the expected river WSE. But in some cases, it is not possible to filter them out. It explains why some time series are noisy and some WSE are erroneous. It usually concerns only few measurements, but, in some cases, it could concern the whole time series, like at Chembe-Ferry on the Congo basin.

Some sensors could also have some limitations that could result in higher WSE uncertainties than other missions. For example, there is the specific case of Envisat that could have a tracking window with adaptive size (64m, 256m, 1024m), but with same number of bins. For the largest tracking windows, bins will be wider and therefore WSE will be more uncertain. Jason series altimeters (i.e. Poseidon series) could saturate and lead to important uncertainties on WSE. Nevertheless, it appears that the oldest altimeter missions (e.g. ERS-2, ENVISAT and Topex/Poseidon) have the largest errors and data loss.

For the merged time series, even if the intermission bias issue has been addressed, there could still be some residual bias between different mission time periods.

3.2. Altimeters-based River Discharge (RD-alti)

3.2.1. Data definition

Altimetry-based River Discharge refers to the measurement and estimation of river discharge using data obtained from altimetry satellites. These data can serve as an alternative means of estimating river discharge when in-situ river discharge (Q) is not available [RD-4]. This data is instrumental in understanding river dynamics, monitoring hydrological processes, managing water resources, and assessing environmental impacts [RD-1].

The measurement is based on the relationship between water surface elevation (WSE) and river discharge, established through a power-law physical approach, implemented via a rating curve. This rating curve is a mathematical relationship derived from field measurements, defining the relationship between river stage (or WSE) and discharge. It enables the conversion of observed water levels into corresponding discharge values, enhancing the accuracy and reliability of discharge estimations [RD-4].

This product aims to use long time series of WSE and in-situ discharge ensuring comprehensive coverage for accurate river discharge assessments.

The corresponding DOI for this product is: 10.5285/44c930e1388f40728884fbdf7e28c109

3.2.2. Data characteristics

Just as in-situ water heigh measurements can be used to gauge river discharge, altimetry-derived water surface elevation (WSE) can serve as an alternative means of estimating river discharge when discharge time series data (Q) is not available. Several methodologies have been documented for deriving discharge time series from altimetry observations and supplementary data.

3.2.2.1. Methodology used

Two approaches have been used [see RD-4 for more details], depending on the available temporal overlap between discharge and altimetry water surface elevation (WSE) time series:

Method 1 – temporal overlap data: The preferred approach relies on the altimetry water surface elevation time series and in situ or simulated discharge time series to create a rating curve (RC) characterized by a power relationship between these two variables following a Bayesian approach (Rantz, 1982). This method has already been applied to major river basins, including the Amazon, the Niger, the Ganges-Brahmaputra, the Mekong, and the Ob, by the institutions and organizations involved in this project (Paris et al., 2022; Bogning et al., 2021; Zakharova et al., 2020). However, this method necessitates a significant overlap period between discharge data and radar altimetry measurements (e.g., Kouraev et al., 2004; Biancamaria et al., 2011; Papa et al., 2012), or it requires the assumption that the rating curve remains valid and consistent when discharge data is only available prior to the altimetry observation period (Tourian et al., 2013, 2017; Frappart et al., 2015; Bogning et al., 2018).

The rating curve is conducted over the calibration period defined as the last 2/3 of the period extending from the first to the last date of overlap (closest value, less than 24 hours) between altimetric data and in-situ discharge data (see Appendix 1). To ensure the robustness of this method, we have established that it can only be applied if the number of overlap points exceeds 15 otherwise, the Method 2 (described below) has been applied.

Tree cases has been identified to create rating curves:

- <u>Case 1:</u> General cases where we can directly compute the rating curve between the available WSE and Q.
- <u>Case 2:</u> In cases where ice cover appears intermittently over certain years and months due to local climate conditions, an alternative method involves excluding these frozen dates from the rating curve dataset. This can be achieved by using the monthly mean temperature to filter out these specific data points. For instance, at the Near-Brookings station, observations reveal the presence of a frozen river during some years between December and March, creating outliers in the rating curve (refer to the figure below). By examining the temperatures recorded during these months, based on the ECMWF fifth reanalysis for the global climate and weather (ERA5 database), we can discard data points where the monthly temperature falls below 0°C. Implementing this approach allows us to generate an alternative rating curve devoid of these outlier points.
- <u>Case 3:</u> In several specific cases especially in arctic, the relation between water surface elevation and discharge may be not uniform. This occurs, for example, near nodes of rivers confluence or during ice cover and ice breakup periods. For these cases the set of the rating curves, specific for a particular condition could be developed. A prior knowledge about these particular conditions and range of applicability of each rating curve is compulsory. The modification of flow hydraulics due to the river ice can be mapped using remote sensing techniques and even altimetry observations simultaneous with the water surface elevation retrievals (Zakharova et al., 2021). For the Arctic rivers an application of the set of the rating curves, specific for recession, ice period and for flood rise demonstrated better accuracy in several previous studies (Zakharova et al., 2020). The rating curves for recession and flood rise are approximated by the classical power equation, while for the ice period a polynomial equation of low degree may produce better accuracy. For the Ob River for example, an automated method of ice setup and breakup detection based on altimetry measurements were tested and the WSE timeseries subset for 2008-2019 ice periods based on this ice product is isolated and used for calibration/validation of ice rating curve. For other years, the Landsat and MODIS images are used for ice on/off detection. Optical

imagery is used for other Arctic test sites as well. The spring flood rise subset is extracted directly from the WSE timeseries. For four Arctic test sites located on the Colville, Mackenzie and Lena Rivers two rating curves (flood recession and merged winter - flood rise RCs) were built using modified Bayesian method (section 4.2.1), while for the Ob River the flood rise was fitted with its own RC. The modification of Bayesian method consisted in probabilistic estimate of only two RC parameters (“a” and “z0”), while b parameter was fixed at each approximation step allowing expert correction of the RC shape for winter and flood rise periods.

Method 2 – no temporal overlap data: The second approach was employed when there is no temporal overlap between in-situ or simulated discharge and water surface elevation data (or not enough – less than 15 overlap dates), it assumes that the validity and stability of the rating curve persist across the various time periods covered by the two datasets. Both time periods should be sufficiently long to encompass a wide range of events. With this assumption, Tourian et al. (2013, 2017) introduced a method for calculating the rating curve, not based on the time series of discharge and water surface elevation, but on the distribution of their quantiles. This method has been adopted by a limited number of recent studies (e.g., Belloni et al., 2021). However, it’s important to note that this methodology naturally introduces higher errors when compared to the preferred approach. For this reason, this methodology will be validated over some stations with various hydrological dynamics and satisfying previous methods (overlap period exists between WSE and Q).

For both methodologies, the Bayesian approach has been used to compute the rating curve, except for specific cases of arctic basin as explained before. This method is a robust statistical approach used for constructing a rating curve, frequently applied in the field of hydrology when the goal is to estimate unknown parameters from observed data, while taking into consideration the associated uncertainty in these estimates (Gelman et al., 2013). This method is grounded in the computation of posterior probabilities for the model parameters, employing Bayes' theorem (Eq.2):

$$P(\theta/D) = P(D)\cdot P(D/\theta)\cdot P(\theta) \qquad \text{Eq. 2}$$

Here, $P(\theta/D)$ denotes the posterior probability of the model's parameters θ, which is the value we are attempting to estimate. $P(D/\theta)$ signifies the likelihood of the data D given a specific set of parameters θ, typically based on the chosen probabilistic model. $P(\theta)$ represents the prior probability of the parameters, derived from our prior knowledge or assumptions, while $P(D)$ is the marginal probability of the data, serving as a normalization factor.

According to this, the estimation of the rating curve using the Bayesian method involves several steps:

- The initial step entails defining a probabilistic model that describes the relationship between observed data and the parameters we aim to estimate. In many hydrological applications, the relationship between discharge data (Q) and water surface elevation data (WSE) is often expressed as a power function (Eq.3):

  $$Q = a\cdot(WSE-z0)^{b} \qquad \text{Eq. 3}$$

  Here, *a, z0* and *b* are the parameters of the rating curve. *a,* is a scaling coefficient governing the magnitude of the Q-WSE relationship, *b,* characterizes the nature of this relationship, and *z0*, represents the height of the free surface above the reference point, corresponding to the river bottom's altitude.
  The power relationship is especially pertinent due to its consistency with numerous hydrodynamic phenomena. The exponent b within the equation allows for the representation of distinctive flow characteristics, including factors like roughness and channel geometry. Moreover, it offers adaptability in modelling to accommodate variations in flow characteristics, whether they are

turbulent or laminar. This relationship, despite its mathematical simplicity, facilitates the fine-tuning of model adjustments in accordance with observed data (Chow, 1959).

- The second step involves the use of prior normal distributions, reflecting our prior knowledge about these parameters. These distributions can either be informative or uninformative, depending on our level of knowledge.
  The limits and ranges for a, z0 and b can vary depending on the specific context of the study, the dataset used, and the characteristics of the river or channel being analysed.
  Coefficient “a”:
  “a" is an adjustment parameter for the rating curve representing the scaling factor for discharge. Its value can significantly fluctuate based on various factors such as the characteristics of the river or channel, hydraulic conditions, and other influencing factors. Consequently, "a" must be non-negative and constrained within a sensible range specific to the system under study.
  Following the Manning equation, “a” must be equal to $W/n*S^{1/2}$ (Chow et al., 1988) where W is the river’s width (m), n the Manning’s roughness coefficient and S the slope (m/m). Given the considerable variability in river width and slope across different stations, a feasible range for this coefficient can be considered as:
  $$a \in [0; 3000]$$
  Coefficient “b”:
  “b” is also an adjustment parameter representing the exponent of the rating curve and indicating the hydraulic condition of the study site. Like "a," this value must comply with physical constraints and cannot be negative.
  Following the Manning equation, “b” must be equal to 5/3 for reference hydraulic condition (Rantz, 1982). To accommodate the variability in system characteristics across sites, the following range values can be considered for this coefficient:
  $$b \in [0; 5]$$
  Coefficient “z0”:
  “z0" represents an offset or the elevation at which discharge begins. It should be within the range of elevations relevant to your study. For this *reason, the value* cannot exceed the minimum value of water surface elevation (WSE) and the range value need to consider of the variability in term of water depth over the sites. A feasible range for this coefficient can be considered as:
  $$z0 \in [\min(WSE)-50; \min(WSE)]$$
- The final step involves parameter estimation. The posterior distribution of the parameters yields probabilistic estimates of the rating curve parameters in the form of mean values (optimal values) and credibility intervals (95th percentiles). This accounts for the uncertainty associated with these parameters and is achieved through Markov Chain Monte Carlo (MCMC) sampling from the posterior distribution (Robert and Casella, 2004). Two commonly employed MCMC algorithms are "NUTS" (No-U-Turn Sampler) and "Metropolis-Hastings". The Metropolis-Hasting sampler "MH" algorithm, which is relatively simple and efficient where a balance between exploration and exploitation is desired. This algorithm can be adapted to sample from discrete state spaces (Geyer, 2011).

3.2.2.2. Uncertainties

Uncertainty propagation through mathematical models plays a crucial role in estimating the reliability of derived results in various scientific fields. In the context of hydrology and discharge estimations, the propagation of uncertainties in parameter estimation, such as those in the parameters of the discharge equation, becomes essential for assessing the reliability of the calculated discharge values. Utilizing a Gaussian error propagation method provides a systematic approach to quantify the uncertainties associated with parameters a, WSE, b, and z0 from the power law function to express the relation between Q and WSE.

This method involves employing statistical principles to propagate uncertainties through the mathematical relationships between the parameters and the discharge equation. By considering the Gaussian distribution of errors in these parameters, this approach enables a more comprehensive evaluation of the overall uncertainty in discharge estimations (e.g., McMahon and Peel, 2019, Tourian et al., 2017).

Given the mean values and standard deviations (σ) for each parameter, the uncertainty in discharge (δ(Q)) due to uncertainties in these parameters can be computed as follows (Eq.4):

$$\partial(Q)=\sqrt{\left(\frac{\partial Q}{\partial a}\cdot\partial a\right)^2+\left(\frac{\partial Q}{\partial WSE}\cdot\partial WSE\right)^2+\left(\frac{\partial Q}{\partial b}\cdot\partial b\right)^2+\left(\frac{\partial Q}{\partial z0}\cdot\partial z0\right)^2}$$

Eq. 4

$$\partial(Q)=\sqrt{\left((WSE-z0)^b\cdot\sigma a\right)^2+\left(a\cdot b\cdot(WSE-z0)^{b-1}\cdot\sigma WSE\right)^2+\left(a\cdot(WSE-z0)^b\cdot\ln(WSE-z0)\cdot\sigma b\right)^2+\left(-a\cdot b\cdot(WSE-z0)^{b-1}\cdot\sigma z0\right)^2}$$

Where, σa, σb, σz0 and σWSE correspond to the standard deviations of parameters a, b, z0 and WSE respectively. The standard deviations for a, b, and z0 will be determined using the Bayesian approach through the MCMC algorithms. Due to modification in RC fitting method, the uncertainties in b parameter were not evaluated for the Arctic sites. The term σb for the Arctic was set up to the global mean equal 0.1.

This formula uses the standard deviations as measures of uncertainty in each parameter and calculates the overall uncertainty in discharge considering the propagation of these uncertainties through the power law equation relating discharge and the parameters a, b, z0 and WSE.

It is important to notice in one hand, that this equation assume that the uncertainties in the parameters (a, b, z0) and WSE are independent, and in another hand, that the propagation of uncertainties provides an estimate based on the assumption of linearization around the mean values of the parameters.

3.2.3. Data limitation

Creating accurate streamflow time series based on altimetry data is a complex endeavor due to several inherent challenges. Firstly, the reliability on rating curves, which establish the empirical relationship between WSE and discharge, poses a significant limitation. These curves, established based on a specific period or over two different periods for each variable, can be irrelevant due to changes in hydraulic dynamics within the river system over time. Factors such as dam construction, river morphology alterations through natural processes or human intervention, and land use changes can all impact the river's flow characteristics. Consequently, the established rating curves may no longer accurately represent the true relationship between WSE and discharge, leading to potential inaccuracies in streamflow estimations.

Secondly, the spatial disparity between virtual stations and in-situ discharge stations introduces additional challenges. Ideally, an in-situ station and the associated virtual station should be close to each other to facilitate the computation of reliable rating curves. However, achieving this proximity can be logistically challenging, particularly mostly due to the disparity in space and time of the altimetry and discharge data. The distance between the altimetry track and the discharge station can introduce errors in flow estimation, especially when hydraulic conditions vary significantly between the two locations (e.g.

Mohacs station over the Danube basin or Ibi station over the Niger basin both with a distance between SV and in-situ discharge data of more than 200 km). Moreover, errors may arise from variations in slope between these two stations, further complicating the accuracy of streamflow estimations.

Furthermore, the general limitations of altimetry data also play a crucial role in the accuracy of river discharge estimations. Altimetry data may be affected by various factors, as described in the previous section on altimeter data limitations (section 3.1.2). These uncertainties propagate into altimetry-based streamflow estimations, potentially compromising the overall accuracy of the derived time series.

Addressing these challenges requires a comprehensive approach that integrates regular updates of rating curves to reflect changes in hydraulic dynamics, strategic placement of monitoring stations to minimize spatial disparities, and the utilization of advanced remote sensing technologies to mitigate the limitations of altimetry data. Additionally, ongoing monitoring, validation, and calibration efforts are essential to improve the reliability of altimetry-based RD estimations and ensure accurate representation of river discharge dynamics over time.

3.3. Multispectral images-based River Discharge (RD-multi)

3.3.1. Data definition

Multispectral-based RD refers to estimating river discharge by exploring the spectral behaviour of pixels with the presence and absence of water in the near-infrared (NIR) band of the electromagnetic spectrum. Specifically, adhering the spectral variability of the different land uses (viz., soil, water, vegetation) along the river and near river environment, numerous spectral indices are developed that can be used to estimate river discharge. To develop those spectral indices, the multispectral images are processed to retrieve the signals (hereafter, denoted as CM and based on the different behaviour in the NIR region between a Calibration (*C*) and a Measurement (*M*) pixels) and further utilized for discharge estimation along the sparsely gauged rivers.

The corresponding DOI for this product is: 10.5285/a8422dd3766c447d8b5fa80920649f31

3.3.2. Data characteristics

Like the in-situ observations of the river hydraulic variables, the temporal dynamics of CM signals can be used as the proxy variables to study the discharge dynamics along the river. To derive the long-term time series of river discharge, all the available multi-spectral images from the available sensors (i.e., Aqua and Terra MODIS, Landsat Series, Sentinel 2 and 3 Series) are merged for the analysis. The detailed formulation of the algorithms by considering these pixels can be found in the ATBD report [RD-4]. To formulate the river discharge algorithm, we need to calibrate the CM signals against the contemporary in situ river discharge for any typical river sites. However, along the selected river sites, the in-situ observations are often unavailable during the period in which satellite data are available (2006-2005 for Landsat 5, 2021-2022 for Landsat-9; 2016-2022 for the remnants). Therefore, two different analyses were carried out depending on the availability of the in-situ data: 1) calibrated approach (when coincident observation of in situ Q and CM signals are available) and 2) uncalibrated approach (when only in situ observation non-contemporary to satellite data are available). Among the selected 54 gauging sites, only 22 sites have the facility to test the calibrated approach as presented in Figure 3. The detailed framework of calibrated and uncalibrated approaches is explained in the ATDB [RD-4].

Figure 3: Observed data availability for the 54 stations. In blue, the selected calibration period, In red the stations in which the observed data overlap with the calibration period.

3.3.2.1. Calibrated Approach

With the coincident data availability of CM and Q, the calibrated approaches are categorized into two types of formulation: 1) Empirical Formulation and 2) Probabilistic Formulation. Notably, the calibrated procedure is performed following the pre-fixed calibration and validation period.

- In the case of empirical formulation (**cal-BestFIT**), four potential distributions (linear, quadratic, power, and exponential) are selected as potential laws between Q and CM data. CM and Q time series are, therefore, trained with the aforementioned formulations. To check the best-fit solution from the selected distribution, a model evaluation criterion has been set considering Akaike Information Criteria (AIC; Eq.5a), Bayesian Information Criteria (BIC, Eq.5b), and Pearson correlation coefficient (*r*). The best fit of any site has been obtained by lower values of AIC and BIC with a higher value of *r*; thus, a composite index (CI; Eq.5c) is formulated to evaluate the overall model scores to determine the best-fit model for the selected site.

$$AIC = 2k - 2\ln L \quad \text{Eq.5a}$$

$$BIC = k*\ln n - 2\ln L' \quad \text{Eq.5b}$$

$$CI = r + (1\text{-}AIC) + (1\text{-}BIC) \quad \text{Eq.5c}$$

  where k equals to the number of parameters used in the model; n equals to the sample size, and L' is the maximum value of the likelihood function for the model.

- In the case of probabilistic formulation (**cal-Copula**), the widely used Copula function is selected for the analysis. Here, the framework proposed by Sahoo et al. (2020) is being adopted. First, the CM and corresponding Q values are considered as pairs to compute Kendall's tau (τ) and the dependence parameter (θ). Second, the five Archimedean family copulas are being formulated, and the best copula fit is obtained by performing Kolmogorov-Smirov (KS) statistic test. In all the 22 analyzed sites, the Frank copula is found to be the best-fit copula for the analysis, and subsequently, the formulation is performed to derive the discharge. For more information about the copula approach, the reader is referred to Sahoo et al (2020).

3.3.2.2. Uncalibrated procedure

In the absence of coincident observations of Q and CM time series, the uncalibrated procedure (**uncal-CDF**) uses the same framework proposed by Tourian et al. (2013). Here, the available discharge and retrieved CM signal time series are sorted independently in descending order. Subsequently, the corresponding exceeding probability of each value in the time series is computed for both Q and CM time series individually by considering their percentage of the observation periods. Despite the Flow Duration Curves are obtained with the same method, it is better to not merge too much information. For this reason, the paper by Tourian et al. (2013) is referenced, and the method for Reflectance indices is termed "CDF matching. For each site, the basic assumption is made that the in-situ Q and CM signals have the same exceedance probability. Developing the joint probability distribution between these two curves can be a suitable solution to estimate river discharge from the CM signals when the in-situ observation is not available. Following the aforementioned steps, the uncalibrated procedure has been designed for each site to derive the long-term discharge time series from the CM signals (see Tarpanelli and Domeneghetti; 2021 for more details).

3.3.3. Data limitation

For deriving the long-term discharge time series from CM signals, the major limitation is the cloud cover, as the signals are retrieved from the optical images. As the optical images are frequently contaminated with clouds, the signals received from the images with unmasked clouds, unmasked cloud shadows, and unmasked ice can add a source of noise to the analysis. Moreover, the majority of the flood events occurred during the cloudy period; therefore, capturing this dynamic using the CM signals is still a challenging task with the unavailability of images during this period. Using the coarser resolution pixels (e.g., 250 m resolution of MODIS), the retrieval of CM signals along narrow or small rivers is prone to noise due to the interference of adjacent pixels. Additionally, the selection of W pixel along the narrow river stretch is still a challenging task for which there is limited application of uncalibrated approach along the small rivers. Retrieval of CM signals along the braided rivers can also be affected by the menders while the river changes its courses over a certain period due to the change in flow dynamics.

The merging procedure to generate the long-term CM signals is based on the contemporary availability of data from different sensors. Without having the contemporary data, one can successfully obtain the merged time series, but those signals may not be a suitable choice for capturing the flow dynamics. Thus, for the implementation of the merging procedure, contemporary observations are needed.

For generating long-term discharge time series using the CM signals, the calibrated procedure is based on the coincident observations of in situ Q and CM. The absence of CM signals during flood events due to the presence of cloud cover in the images may affect the model parameterization to capture the high flow dynamics both in cal-BestFIT and cal-Copula solutions. Although the uncal-CDF procedure is independent of the coincident observations of in situ Q and CM, the availability of the in-situ Q data period is still a key concern. ~~For~~ instance, the hydrograph generated from a short event may not represent the long-term period. Consequently, the derived Flow Duration Curve (FDC) and Recession Duration Curve (RDC) may not yield accurate results when determining the joint distribution. This can introduce significant uncertainties in deriving long-term discharge time series. Additionally, in both calibrated and uncalibrated procedures, there is a risk of losing flood information due to the absence of CM signals caused by frequent cloud cover.

4. RD dataset

4.1. Main characteristics

The CDR (Climate Data Record) River Discharge dataset is a merged product composed of the thematic product described in the previous section and summarized in the Figure 4. There are three CRDPs (Climate Research Data Packages) provided within the CCI River Discharge precursor project:

- WSE CRDP (doi:10.5285/c5f0aa806ec444b4a4209b49efc4bb65),
- RD-alti CRDP (doi:10.5285/44c930e1388f40728884fbdf7e28c109),
- and RD-multi CRDP (doi:10.5285/a8422dd3766c447d8b5fa80920649f31).

All products are provided with two different formats. The first one corresponds one NetCDF4 file per station following the CCI data standard [RD-6]. The second one corresponds to one CSV file per station. Next sections describe in detail the CRDP directory structure, and the CSV and NetCDF file formats computed for all CRDPs. If more information is needed, users can read [RD-7] and [RD-8], the release notes for WSE and RD-alti CRDPs, respectively.

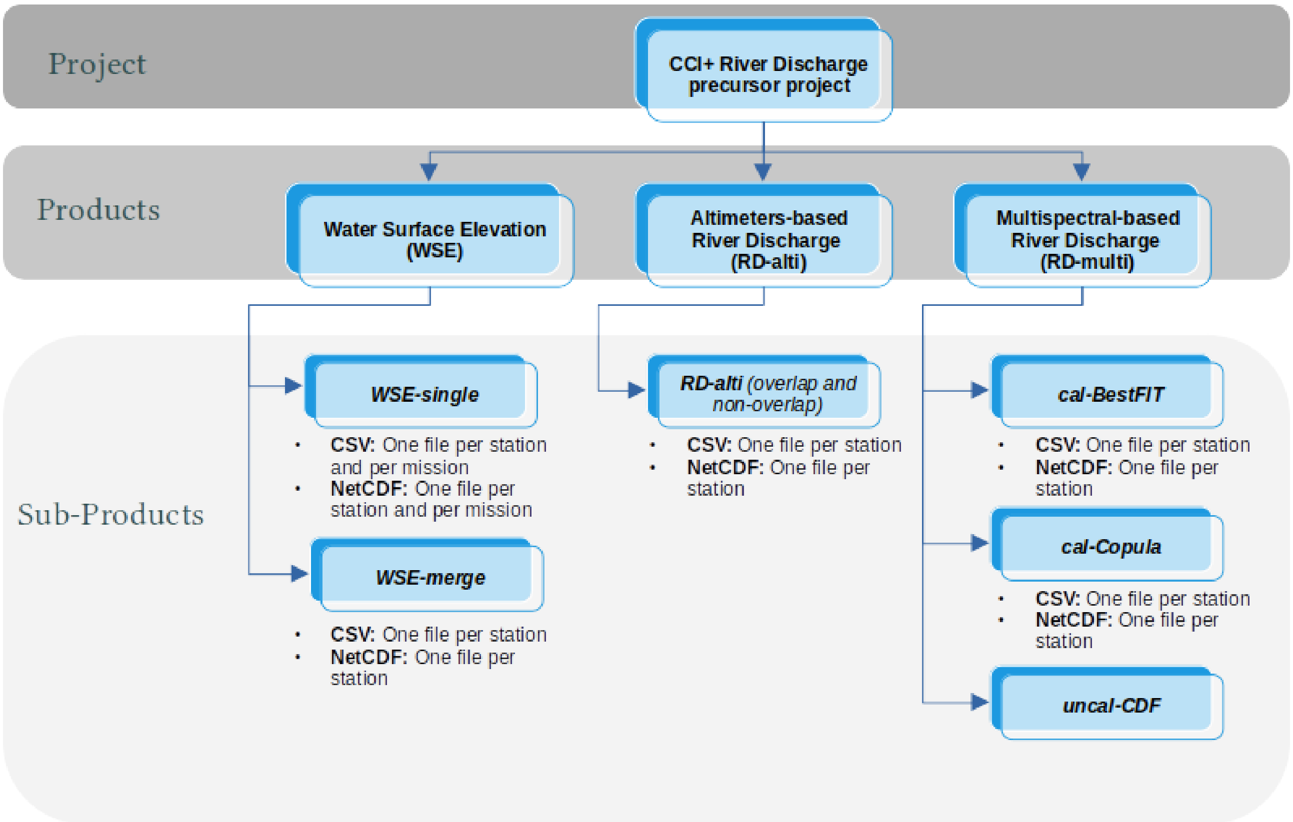


Figure 4. Organigram of the products and sub-products of the CCI+ River Discharge Precursor Project.

Products generated in this precursor CCI+ project are derived from data acquired by multiple sensors and satellites (for details see [RD-3] and [RD-4]). Therefore, they have different temporal and spatial resolutions, but also different accuracies.

Both single mission WSE time series and merged WSE time series are provided for VS near locations defined in the [RD-2] at the satellite observation times. However, in the merged time series, only one measurement per day is kept if multiple observations are available for a given day (the earliest one being retained [RD-7]).

Altimetry-based RD time series are provided at the in-situ discharge station locations defined in [RD-2]. In situ data are used over the calibration period [RD-4], when there is time overlap with altimetry data, to compute the rating curve, using merged WSE time series.

The multispectral images-based RD time series are also provided in all the stations defined in [RD-2]. As discussed in 3.3.2, three different river discharge products are provided, according to the availability of observed data that overlaps with the multispectral indices: in case of observed data availability, the RD obtained through the copula and best-fit approach are calculated, together with the RD from the uncalibrated CDF matching. In case no overlapping observed data are available, just the RD from the uncalibrated CDF matching is provided.

Valid uncertainty estimates are provided only for RD-alti and correspond to a first estimate of a partial end-to-end uncertainty budget (see [RD-4]).

4.2. CRDPs directory structure

4.2.1. WSE CRDP directory structure

Concerning WSE CRDP directory structure, the highest-level directory corresponds to the current CRDP version number. Then, there are two subdirectories, corresponding to the file formats. They are labelled

“CSV” and “NetCDF”. They contain only files with format corresponding to their name subdirectory. Subdirectories structure for these two directories are the same. Single mission time series are provided in one directory called “single_mission_timeseries”, whereas merged time series are provided in a directory called “merged_timeseries”. In both directories, there are 18 directories, one per selected basin in [RD-2], labeled with the name of the basin in capital letters. Then, in each “basin” directory, there is one directory per location defined in [RD-2] for this basin. These “location” directories are labeled with the name of the location in capital letters (see Appendix B for the name of these locations). Altimetry time series files for all VS associated to these locations are located within these "location” directories (rather in CSV format or in netCDF format, depending on the level 2 directory name). For merged time series, there is one file in each location directory. A diagram of the CRDP directory structure is shown in Figure 5.

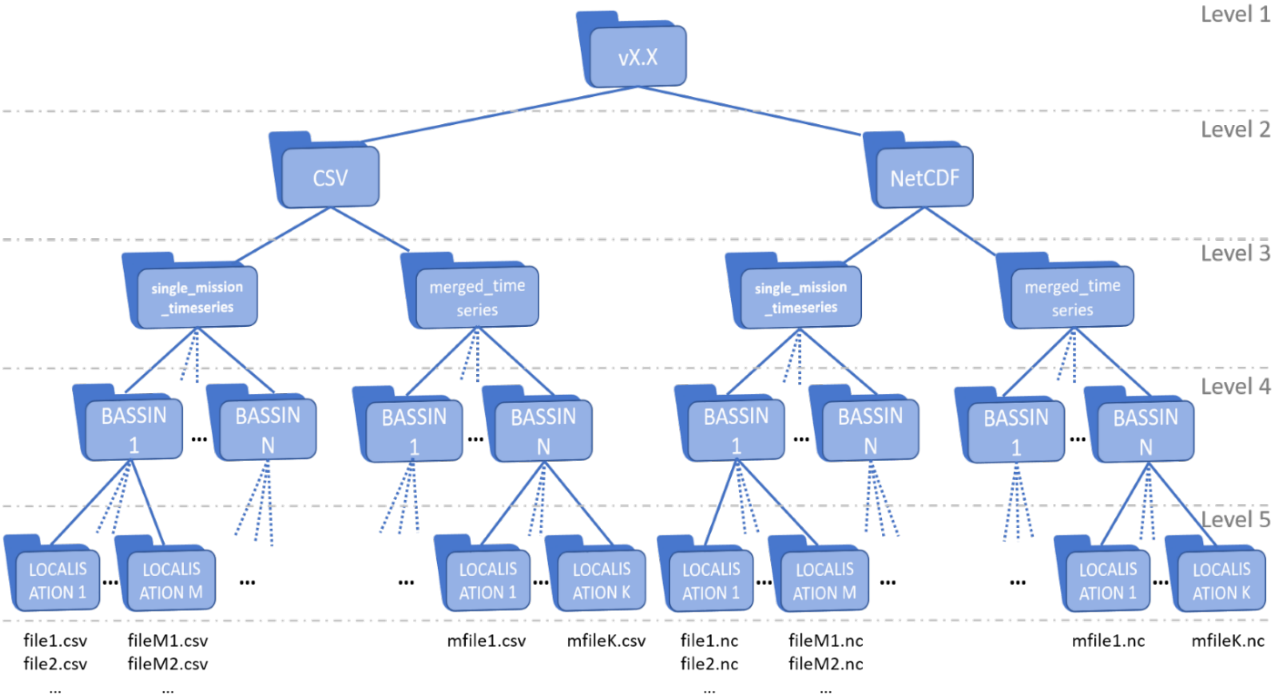


Figure 5. WSE CRDP directory structure

4.2.2. RD-alti CRDP directory structure

Like WSE CRDP, the highest-level directory for RD-alti CRDP corresponds to the current CRDP version number. Then, there are two subdirectories, corresponding to the file formats. They are labelled “CSV” and “NetCDF” and contain only file format corresponding to their name subdirectory. RD-alti time series files for all locations are provided in these two subdirectories. As the RD-alti time series have been computed using merged WSE time series, there is just one file per location. A diagram of the RD-alti CRDP directory structure is shown in Figure 6.

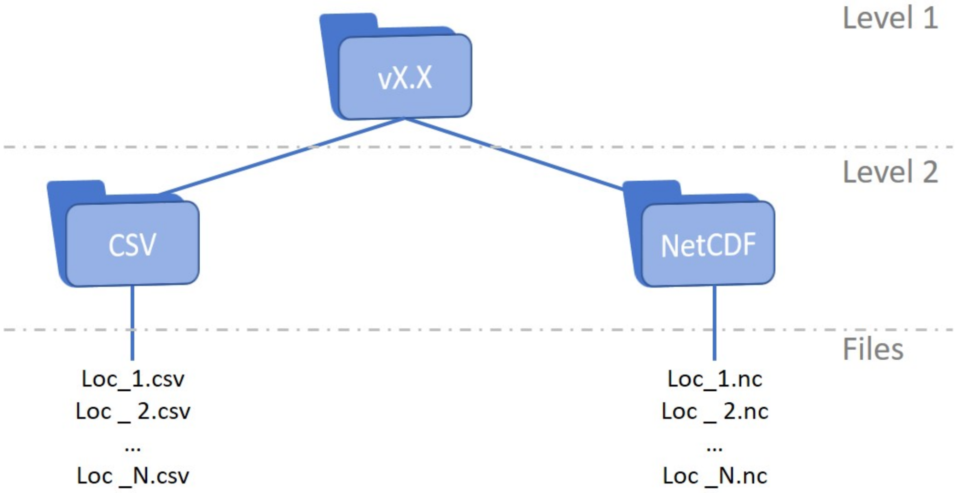


Figure 6. RD-alti CRDP directory structure

4.2.3. RD-multi CRDP directory structure

RD-multi CRDP directory structure has three levels. Level 1 and level 2 are the same as RD-alti. Level 3 corresponds to the three approaches used to compute river discharge for RD-multi (**cal-BestFIT, cal-Copula, and uncal-CDF**). Finally, RD-multi time series files for the computed locations are provided in these three subdirectories. A diagram of the CRDP directory structure is shown in Figure 7.

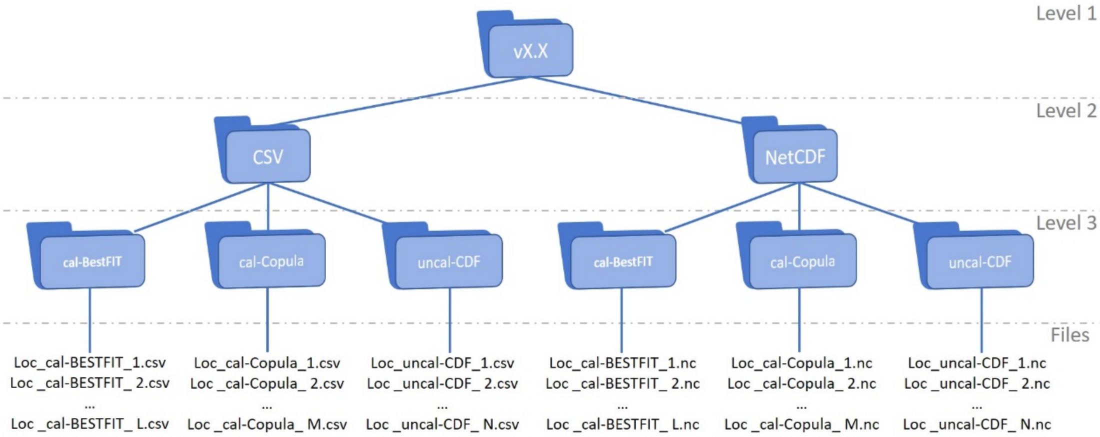


Figure 7. RD-multi CRDP directory structure

4.3. NetCDF files

The first type of file format in CCI River Discharge CRDPs corresponds to NetCDF4 files, compliant with most of the CCI Data Standards [RD-6]. There is one NetCDF file per time series. Sections below describe first the file naming convention used and the formatting.

4.3.1. File naming

NetCDF file names in the CCI+ River Discharge precursor project, are compliant with the CCI data standards [see RD-6 for more details] and follow the pattern:

ESACCI-RD-<Processing Level>-<Data Type>-<Product String>-<Additional Segregator>-<Indicative Dates>-fv<File version>.nc

Where:

<Processing Level> equal to “L3C” for single mission WSE time series or “L3S” for merged WSE time series whereas, is equal to “L4” (i.e. level 4) for RD time series created from the level 3 (L3S) satellite data or from multispectral ratio.

<Data Type> equal to “RD” for River Discharge time series, or equal to “WL” for WSE time series, to follow the CCI data standards [RD-6]

<Product String> equal to “SINGLE_nobiascorrection” (for single mission WSE time series), “MERGED” (for merged WSE time series), “ALTIBASED” for RD-alti, “CMCALBESTFIT” for RD-multi/Cal-BestFit product, “CMCALCOPULA” for RD-multi/Cal-Copula product, or “CMRATIO” for RD-multi/uncal_CDF product.

<Additional Segregator> equal to “BASIN_RIVER_STATION_MISSIONNAME_TRACKNUMBER_LATID” for WSE time series and equal to “BASIN_RIVER_STATION” for RD time series.

With: BASIN = Basin name in capital letters

RIVER = River name in capital letters

STATION = Location name defined in [RD-2, RD-7] in capital letters

MISSIONNAME = VS mission name, in lower case. It corresponds to the mission’s name of the time series in the file among the following values: ers2, envisat, saral, topex, jason1, jason2, jason3, sentinel3a, sentinel3b, and sentinel6. For merged time series, it corresponds to the mission’s name of the reference VS (see [RD-3]), preceded with the word “merged”

TRACKNUMBER = Mission orbit track number associated to the VS, coded on 4 digits

LATID = It is an ID defining the mean latitude of the VS. It begins with “N” if the mean latitude of the VS is in the Northern Hemisphere or with “S” if it is in the South Hemisphere. Then, it is followed by the mean latitude value with two decimals and without any point for decimal separator, coded on four digits (for example, if the mean latitude of the VS is 44.24 °N, then LATID = N4424)

<IndicativeDates> corresponds to the first date and the last date in the time series, separated with “_”. Dates are provided in the form YYYYMMDD, where YYYY is the four digits year, MM is the two digits month, and DD is the two digits day of the month.

<File version> is the file version.

4.3.2. Format

The River Discharge dataset is stored in the NetCDF4 classic format (Network Command Data Form) using the CF (Climate and Forecast) metadata convention (v1.8) and CCI Data Standards (v2.1), as requested in [RD-6].

The following sections describe the content of NetCDF files for each CRDP.

4.3.2.1. Global attribute

The 41 global attributes correspond to the ones required in [RD-6]. They are self-explanatory and only the main ones are described in the Table 3.

Attribute Name	Attribute description
Source	sources of the data used to compute the time series. It is set to the name of the space agencies that conceived the satellite altimetry mission.

time_coverage_resolution	time resolution of the dataset. As measurement varies in time, depending of the satellite mission(s) available, it is set to "satellite_orbit_frequency"
spatial_resolution	spatial resolution of the dataset. As measurements are provided at some specific location and not on a grid, it is set to "Point-based measurement at the satellite nadir" for WSE time series, "Point-based measurement of the insitu discharge data" for RD-alti time series, or "Point-based from pixel reflectance index over an area ranging from 0.04degx0.04deg to 0.15degx0.15deg" for all RD-multi products.
platform and sensor	name of the satellites/platforms and the associated sensor, respectively, of which data has been used to produce the time series.

Table 3. Global attributes

In addition to these attributes, there are also a few global attributes specific to this CRDP and depending on the dataset (WSE or RD time series).

Attribute Name	Attribute description
basin_name	the name (in upper case) of the basin river (see Table 1)
river_name	the name (in upper case) of the river
Location	the name (in upper case) of the station (see Table 1)
reference_virtual_station	the name/ID given by this CCI project to the VS - only for the WSE dataset
Methodology	methodology used to compute the rating curve (see [RD-4] for more details) - only for RD dataset
Doi	the Digital Object Identifier (DOI) associated to this CRDP and should be used when citing this CRDP

Table 4. Additional Global attributes

4.3.2.2. Dimensions

Following the CCI data standards, the products have three dimensions: time, latitude and longitude. All the included variables share the same dimensions.

4.3.2.3. Variables

The attributes of the variables in the NetCDF files follow the CCI data standards guidelines [RD-6] and consequently, the CF recommendations. All variables have only one dimension, named *time*, which has unlimited dimension and corresponds to time dimension of the time series. This *time* dimension differs between products. Only the variable "plateform", corresponding to the platform name (character array), has another dimension, labeled *strlen* to follow CF Metadata Conventions (see [RD-6]). This *strlen* dimension corresponds to the longest platform name.

Table 5 and 6 present the variables in the NetCDF files for WSE and RD datasets respectively.

Variable Name	Variable description
time(time)	corresponds to satellite measurement times in the time series for the VS. It is provided as seconds from 1970-01-01.

lat(time) and lon(time)	correspond to WGS84 latitude and longitude, respectively, of each VS for WSE time series, or locations defined in [RD-2] for RD-alti and RD-multi products. No data (or fill value) are set to 9.969209968386869e+36.
water_surface_height_above_reference_datum(time)	key variable and corresponds to the WL (in m) time series at the VS. No data (or fill value) are set to 9.969209968386869e+36
water_surface_height_uncertainty(time)	for each measurement time, it is the median absolute deviation (in m) of the selected along-track measurements at the measurement time to compute *water_surface_height_above_reference_datum* (see [RD-3]). No data (or fill value) are set to 9.969209968386869e+36.
orbit_track_number(time)	corresponds to the satellite orbit track number for the platform that did the measurement. No data (or fill value) are set to -2147483647.
mission_cycle_number(time)	corresponds to the satellite orbit cycle number for the platform that did the measurement. No data (or fill value) are set to -2147483647.
platform(time, strlen)	is only defined for merged WSE time series and corresponds to the name of the platform that made the measurement for each time step. The platform names are the ones defined in the CCI ontology table (see [RD-6]). No data (or fill value) are set to '\x00'.

Table 5. List of global variables in the NetCDF file for the WSE dataset

Variable Name	Variable description
time(time)	corresponds to satellite measurement times in the time series for the VS. It is provided as seconds from 1970-01-01.
lat(time) and lon(time)	correspond to WGS84 latitude and longitude, respectively, of the location where discharge is provided
float_water_volume_transport_in_river_channel(time)	key variable and corresponds to the satellite-based RD (in m3/s) time series at the location (i.e. in-situ station). No data (or fill value) are set to NaNf.
float_water_volume_transport_in_river_channel_uncertainty(time)	for each measurement time, it is the uncertainty associated to the satellite-based RD (in m3/s) using methodology defined in [RD-4]. No data (or fill value) are set to NaNf.
platform(time, strlen)	corresponds to the name of the satellite mission/platform that made the measurement used to derived discharge for each time step. No data (or fill value) are set to '\x00'.

Table 6. List of variables in the NetCDF file for the RD dataset

4.4. CSV files

The second type of file format in CCI River Discharge CRDPs corresponds to CSV files. They are generated in addition to the NetCDF files, and they contain the same time series and metadata. There is one CSV file per time series. Sections below describe first the file naming convention and the formatting.

4.4.1. File naming for WSE CRDP

The same naming convention than the Hydroweb time series is used and expanded for the WSE time series. It follows the following pattern:

R_BASIN_RIVER_KMXXXX_MISSIONNAME-TRACKNUMBER_LATID.csv

with: BASIN = Basin name in capital letters

RIVER = River name in capital letters

XXXX = Distance from river mouth (curvilinear abscissa). If not known, it is set to “XXXX”.

MISSIONNAME = VS mission name. It is in lower case and corresponds to the mission’s name of the time series in the file among the following values: ers2, envisat, saral, topex, jason1, jason2, jason3, sentinel3a, sentinel3b, and sentinel6. For merged time series, it corresponds to the mission’s name of the reference VS, preceded with the word “merged”

TRACKNUMBER = Mission orbit track number associated to the VS, coded on 4 digits

LATID = It is an ID defining the mean latitude of the VS. It begins with “N” if the mean latitude of the VS is in the Northern Hemisphere or with “S” if it is in the South Hemisphere. Then, it is followed by the mean latitude value with two decimals and without any point for decimal separator, coded on four digits (for example, if the mean latitude of the VS is 44.24°N, then LATID = N4424)

4.4.2. File naming for RD-alti

File naming used for RD-alti is derived from the GRDC one, slightly expanded and follows the pattern:

BASIN_STATION_Q_Day.Cmd.csv

with: BASIN = Basin name in capital letters

RIVER = River name in capital letters

4.4.3. File naming for RD-multi

File naming used for RD-multi is similar to RD-alti, with information concerning follows the pattern:

BASIN_STATION_Q_Day.METHOD.csv

with: BASIN = Basin name in capital letters

RIVER = River name in capital letters

METHOD = Method used to compute RD-multi products. It is equal to “CM_cal_BestFit” for RD-multi/Cal-BestFit product, “CM_cal_Copula” for RD-multi/Cal-Copula product, or “CM_uncal_CDF” for RD-multi/uncal_CDF product.

4.4.4. Format

The WSE time series file format for this precursor project is the same one as the Hydroweb expert river data format. The RD time series file format is inspired by the GRDC discharge data format and is the same one for RD-alti and RD-multi. These data formats have been chosen as they are quite well known and used in the satellite hydrology science community.

4.4.4.1. Header

Every file starts with a fixed header, containing information on the contents of the file. The lines of the header are preceded by the hash character (i.e. #). This character may only be used in the header of the file. Header data is not required but will always be exported to make the data files more intelligible for humans.

The Tables 7 and 8 present the header information in the CSV files for WSE and RD CRDP, respectively.

Lines 1 to 7 correspond to hydrological metadata:
#BASIN: Basin Name
#RIVER: River name
#ID: VS ID (unique identifier), it is of the form mission-tracknumber_LXXXX, where mission is the mission name, tracknumber is the mission orbit track number associated to this VS, L is equal to N if the mean latitude of the VS is in the Northern Hemisphere or S if it is in the South Hemisphere, XXXX is the mean latitude value with two decimals and without any point for decimal separator on four digits (i.e.if the latitude is in between 10°S and 10°N, the first digit is zero, 0). For merged time series, the mission's name corresponds to the one from the reference VS, preceded with "merged". Tracknumber corresponds to the reference VS mission orbit track number.
#TRIBUTARY OF:: Upstream river name (equals to "NA" if unavailable)
#APPROX. WIDTH OF REACH (m):: River width estimation (equals to "NA" if unavailable)
#SURFACE OF UPSTREAM WATERSHED (km2): Estimated upstream watershed surface (equals to "NA" if unavailable)
#RATING CURVE PARAMETERS A,b,Zo such that Q(m3/s) = A[H(m)-Zo]^b:: as only WL time series are provided, they are equals to "NA NA NA"
Lines 8 to 18 correspond to geographical metadata:
#REFERENCE ELLIPSOID:: Ellipsoid of reference
#REFERENCE LONGITUDE:: VS longitude
#REFERENCE LATITUDE:: VS latitude
#REFERENCE DISTANCE (km):: Distance from river mouth (curvilinear abscissa; equals to "NA" if unavailable)
#GEOID MODEL:: Reference geoid (geoid model version), set to NA as WL are referenced to WGS84 ellipsoid
#GEOID ONDULATION AT REF POSITION(M.mm):: Geoid value at the VS location, set to NA
#MISSION(S)-TRACK(S):: List of the used mission(s)-track(s) to build the timeseries
#STATUS: could be "Operational (daily processed)" or "RESEEEEARCH" (set to "RESEARCH" in this CRDP)
#VALIDATION CRITERIA:: "EXPERT", or "AUTOMATIC" (i.e. VS validated from hydrological expert, or automatically from statistical criteria)
#MEAN ALTITUDE (M.mm):: Mean WSE
#MEAN SLOPE (mm/km):: Mean slope over the river section (equals to "NA" if unavailable)
Lines 19 to 28 correspond to product metadata:
#NUMBER OF MEASUREMENTS IN DATASET:: Measurements number
#FIRST DATE IN DATASET:: First date in the timeseries
#LAST DATE IN DATASET:: last date in the timeseries
#DISTANCE MIN IN DATASET (km):: Minimal distance from river mouth (equals to "NA" if unavailable)

#DISTANCE MAX IN DATASET (km):: Maximal distance from river mouth (equals to "NA" if unavailable)
#PRODUCTION DATE:: Production date
#PRODUCT VERSION:: Product version
#PRODUCT CITATION:: Product citation (equal to "ESA CCI+ River Discharge precursor project, DOI:10.5285/c5f0aa806ec444b4a4209b49efc4bb65")
#SOURCES:: Product sources (empty value for the moment)
#PRODUCT CONTENT:: Description of the data content (empty value for the moment)
Lines 29 to 45 describes each column of the file:
#COL 1: DATE(YYYY-MM-DD)
#COL 2: TIME(HH:MM:SS)
#COL 3: ORTHOMETRIC HEIGHT (M) OF WATER SURFACE AT REFERENCE POSITION
#COL 4: ASSOCIATED UNCERTAINTY(M)
#FIELD SEPARATOR (":")
#COL 5: LONGITUDE OF ALTIMETRY MEASUREMENT (deg)
#COL 6: LATITUDE OF ALTIMETRY MEASUREMENT (deg)
#COL 7: ELLIPSOIDAL HEIGHT OF ALTIMETRY MEASUREMENT (M)
#COL 8: GEOIDAL ONDULATION (M) at location [5,6]
#COL 9: DISTANCE OF ALTIMETRY MEASUREMENT TO REFERENCE POSITION (KM)
#COL 10: SATELLITE (for merged time series, it corresponds to the one from the reference VS, preceded with "merged")
#COL 11: ORBIT / MISSION
#COL 12: GROUND-TRACK NUMBER
#COL 13: CYCLE NUMBER
#COL 14: RETRACKING ALGORITHM
#COL 15: GDR VERSION
Note: the field separator between columns 4 and 5 is introduced by Hydroweb team to separate data that are provided in the "basic" files (only the four first columns) and the "expert" files (all columns).
- The last header line separates the header to the data and corresponds to:
##

Table 7. List of variables in the CSV header file for the WSE dataset

Lines 1 to 11 correspond to hydrological metadata:	
# Title:	Title of the data with project name and type of data ---------------
# Format:	Data's format (CSV)
# Field delimiter:	Field delimiter (;)
# missing values	Value for no data (nan)
# file generation date:	Date of data generation (%Y-%m-%d)
#	
# Basin:	Basin Name in capital letters. Spaces have been replaced by "-"
# River:	River Name in capital letters. Spaces have been replaced by "-"
# Station:	Station Name in capital letters. Spaces have been replaced by "-"

# Country:	Country Name in capital letters. Spaces have been replaced by "-"
Lines 12 to 16 correspond to geographical metadata:	
# Latitude (DD):	Latitude in decimal degrees [-90,90] with 4 decimals – corresponding to the insitu discharge station [see Appendix A and Appendix B]. If multiple sources, longitude and latitude from GRDC is privileged.
# Longitude (DD):	Longitude in decimal degrees [-180,180] with 4 decimals - corresponding to the insitu discharge station [see Appendix A and Appendix B]. If multiple sources, longitude and latitude from GRDC is privileged.
# Catchment area (km^2):	Catchment area in km^2. Equals to "nan" if unavailable
# Altitude (m ASL):	The altitude of the station is expressed in meters above sea level. Equals to "nan" if unavailable
# Next downstream station:	Next downstream station. Equals to "nan" if unavailable
Lines 17 to 29 correspond to product metadata:	
# Institution:	Institutions responsible for producing the data.
# Owner and License:	Owner of the data and associated license.
# doi:	Doi of the dataset (same for all altimetry-based RD time series)
#	**
#	
# Data Set Content:	Type of data in this file - RIVER DISCHARGE (RD)
#	------------------
# Unit of measure:	Unit of measure - m^3/s
# Time series:	First and last date of available data. Both are notified in this format: %Y-%m-%d and separate by a "-" as %Y-%m-%d - %Y-%m-%d
# Last update:	Date of the last update in %Y-%m-%d
# Methodology:	Methodology used to compute the rating curve. The first part describes the approach used to compute the RC and the second part, separated by "_", describes the algorithm used. To avoid any issue for the reader the spaces have been replaced by "-". For RD-alti, two approaches ("Overlap-approach", where there is time overlap between in situ and altimeter data, or "Quantile-approach" where there is no or not enough time overlap) and two algorithms ("Bayesian-algorithm", or "Multiple-algorithms" used for Arctic rivers experiencing frozen periods) are available. For RD-multi, it could be "Calibrated CM approach - best fit regression" for RD-multi/Cal-BestFit product, "Calibrated CM approach - copula regression" for RD-multi/Cal-Copula product, or "Uncalibrated CM approach - CDF regression" for RD-multi/uncal_CDF product.
# Insitu discharge:	Insitu discharge database used to compute the RC. If multiple sources have been used to compute the RC, they are separated by "/". 9 sources haves been identified: - Global Runoff Data Centre (GRDC) - 56068 Koblenz – Germany - Global River Discharge (RivDIS) data set - U.S. Geological Survey (USGS) - Amazon basin water ressources observation service (SO-HYBAM) - Service Central d'Hydrometeorologie et d'Appui a la Prevision des Inondations (SCHAPI) HydroPortail - Arctic Great Rivers Observatory (ArcticGRO) - Environment and Climate Change Canada Historical Hydrometric Data web site (HYDAT) - Simulation from MGB hydrological model - Research Group (IPH - UFRGS) - Porto Alegre

# Calibration period:	- Agenzia Interregionale del Fiume Po (AIPo) Calibration period. Period uses to compute the RC with first date (%Y-%m-%d) and last date (%Y-%m-%d) separated by “-” as %Y-%m-%d - %Y-%m-%d
Lines 30 to 39 describes each column of the file and information on data lines	
# Table Header: # Date # Time # Value # Uncertainty # Satellite # # # Data lines: # DATA	 - Date’s format - YYYY-MM-DD - Time’s format - hh:mm:ss - original (provided) data – here discharge data in m^3/s - Value's uncertainty (same unit than “Value”) - Altimetry mission source (envisat, topex, ers2, saral, jason1, jason2, jason3, sentinel3a, sentinel3b, sentinel6) ** Number of data – integer value

Table 8. List of variables in the CSV header file for the RD dataset

4.4.4.2. Data

The data are provided in the CSV file just after the header. Each line corresponds to a different measurement time. Columns are separated by a space character for WSE data and by a semicolon character for the RD data.

Measurement data are provided within columns. The content of each column is described in the header (see previous section).

5. Guidance for reading and visualizing

The CCI River Discharge data are stored both in CSV and NetCDF formats. The same content is provided in these two types of formats.

Unidata web site (https://www.unidata.ucar.edu/software/netcdf/software.html) proposes a wide range of software for manipulating and displaying NetCDF data. They can be used to explore and process CCI River Discharge NetCDF products.

Concerning CSV products, they could be opened with any text editor. Typically, spreadsheet programs can open them and plot time series. Alternatively, scientific software or programming languages (like Matlab, R or python) can easily load, process and/or plot data stored in this format.

Furthermore, the CSV file format used for the RD and WSE time series follows the standard formats commonly used by many scientists. The CSV file format for the WSE corresponds to the expert Hydroweb CSV files, while the CSV file format for the RD is inspired by the file format of the GRDC.

6. References

Belloni, R., Camici, S., & Tarpanelli, A. (2021). Towards the continuous monitoring of the extreme events through satellite radar altimetry observations. *Journal of Hydrology*, *603*, 126870.

Biancamaria, S., Durand, M., Andreadis, K. M., Bates, P. D., Boone, A., Mognard, N. M., ... & Clark, E. A. (2011). Assimilation of virtual wide swath altimetry to improve Arctic river modeling. *Remote Sensing of Environment*, *115*(2), 373-381.

Biancamaria, S., Frappart, F., Leleu, A.-S., Marieu, V., Blumstein, D., Desjonquères, J.-D., Boy, F., Sottolichio, A., & Valle-Levinson, A. (2017). Satellite radar altimetry water elevations performance over a 200 m wide river: evaluation over the Garonne River. *Advances in Space Research*, 59, 128-146, doi:10.1016/j.asr.2016.10.008.

Bogning, S., Frappart, F., Paris, A., Blarel, F., Niño, F., Picart, S. S., ... & Braun, J. J. (2021). Hydro-climatology study of the Ogooué River basin using hydrological modeling and satellite altimetry. *Advances in Space Research*, *68*(2), 672-690.

Bogning, S., Frappart, F., Blarel, F., Niño, F., Mahé, G., Bricquet, J. P., ... & Braun, J. J. (2018). Monitoring water levels and discharges using radar altimetry in an ungauged river basin: The case of the Ogooué. *Remote Sensing*, *10*(2), 350.

Chow, V. T. (1959). *Open-Channel Hydraulics.* McGraw-Hill Education.

Chow, V. T., Maidment, D. R., & Mays, L. W. (1988). Applied Hydrology. McGraw-Hill Education. [Chapter 10: Open-Channel Flow, specifically sections discussing rating curves and power-law relationships]

Cretaux, J.-F., Nielsen, K., Frappart, F., Papa, F., Calmant S., & Benveniste, J. (2017). Hydrological applications of satellite altimetry: rivers, lakes, man-made reservoirs, inundated areas in Satellite Altimetry Over Oceans and Land Surfaces, Earth Observation of Global Changes (éd. Stammer, D. & A. Cazenave) 459-504 (CRC Press, Boca Raton, USA). isbn: 978-1-4987-4345-7.

Cretaux, J.-F., Calmant, S., Papa, F., Frappart, F., Paris, A., & Berge-Nguyen, M. (2023). Inland Surface Waters Quantity Monitored from Remote Sensing. *Surveys in Geophysics*, 44(5), 1519-1552, doi:10.1007/s10712-023-09803-x

Frappart, F., Fatras, C., Mougin, E., Marieu, V., Diepkilé, A. T., Blarel, F., & Borderies, P. (2015). Radar altimetry backscattering signatures at Ka, Ku, C, and S bands over West Africa. *Physics and Chemistry of the Earth, Parts A/b/c*, *83*, 96-110.

GCOS (2022). The 2022 GCOS ECVs Requirements. GCOS-No. 245, Geneva, World Meteorological Organization, https://library.wmo.int/index.php?lvl=notice_display&id=22135#.Y-Tiv3bMKbg (last accessed 24 April 2024).

Gelman, A., Carlin, J. B., Stern, H. S., Dunson, D. B., Vehtari, A., & Rubin, D. B. (2013). "Bayesian Data Analysis." Chapman and Hall/CRC.

Geyer, C. J. (2011). Introduction to Markov Chain Monte Carlo. In S. Brooks, A. Gelman, G. Jones, & X.-L. Meng (Eds.), Handbook of Markov Chain Monte Carlo (pp. 3–48). Chapman and Hall/CRC.

Gleick P. (1996). Water resources in Encyclopedia of Climate and Weather (éd. Schneider, S. H.) 817-823 (Oxford University Press, New York, USA).

Kouraev, A. V., Zakharova, E. A., Samain, O., Mognard, N. M., & Cazenave, A. (2004). Ob'river discharge from TOPEX/Poseidon satellite altimetry (1992–2002). *Remote sensing of environment*, *93*(1-2), 238-245.

McMahon, T. A., & Peel, M. C. (2019). Uncertainty in stage–discharge rating curves: application to Australian Hydrologic Reference Stations data. *Hydrological Sciences Journal*, *64*(3), 255-275.

Milliman J. and K. Farnsworth (2013). River discharge to the coastal ocean - A global synthesis. ISBN: 978-0-521-87987-3 (Cambridge University Press, Cambridge, United Kingdom).

Papa, F., Bala, S. K., Pandey, R. K., Durand, F., Gopalakrishna, V. V., Rahman, A., & Rossow, W. B. (2012). Ganga-Brahmaputra river discharge from Jason-2 radar altimetry: an update to the long-term satellite-derived estimates of continental freshwater forcing flux into the Bay of Bengal. *Journal of Geophysical Research: Oceans*, *117*(C11).

Paris, A., Calmant, S., Gosset, M., Fleischmann, A. S., Conchy, T. S. X., Garambois, P. A., ... & Laraque, A. (2022). Monitoring Hydrological Variables from Remote Sensing and Modeling in the Congo River Basin. *Congo Basin Hydrology, Climate, and Biogeochemistry: A Foundation for the Future*, 339-366.

Paris, A., Dias de Paiva, R., Santos da Silva, J., Medeiros Moreira, D., Calmant, S., Garambois, P. A., ... & Seyler, F. (2016). Stage-discharge rating curves based on satellite altimetry and modeled discharge in the Amazon basin. *Water Resources Research*, *52*(5), 3787-3814.

Rantz, S. E. (1982). *Measurement and computation of streamflow* (Vol. 2175). US Department of the Interior, Geological Survey.

Robert, C. P., & Casella, G. (2004). "Monte Carlo Statistical Methods." Springer.

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

Tarpanelli, A., Camici, S., Nielsen, K., Brocca, L., Moramarco, T., Benveniste, J. (2021) Potentials and limitations of Sentinel-3 for river discharge assessment. *Advances in Space Research*, 68(2), 593-606. https://doi.org/10.1016/j.asr.2019.08.005

Tourian, M. J., Schwatke, C., & Sneeuw, N. (2017). River discharge estimation at daily resolution from satellite altimetry over an entire river basin. *Journal of Hydrology*, *546*, 230-247.

Tourian, M. J., Sneeuw, N., & Bárdossy, A. (2013). A quantile function approach to discharge estimation from satellite altimetry (ENVISAT). *Water Resources Research*, *49*(7), 4174-4186.

Trenberth K. (2011). Changes in precipitation with climate change. Climate Research, 47, 123-138, doi:10.3354/cr00953.

Wingham, D.J., Rapley, C.G., Griffiths, H., 1986. New techniques in satellite altimeter tracking systems. In: ESA (Ed.), Proceedings of IGARSS’86 Symposium, Zürich, 8–11 September 1986, SP-254, pp. 1339-1344

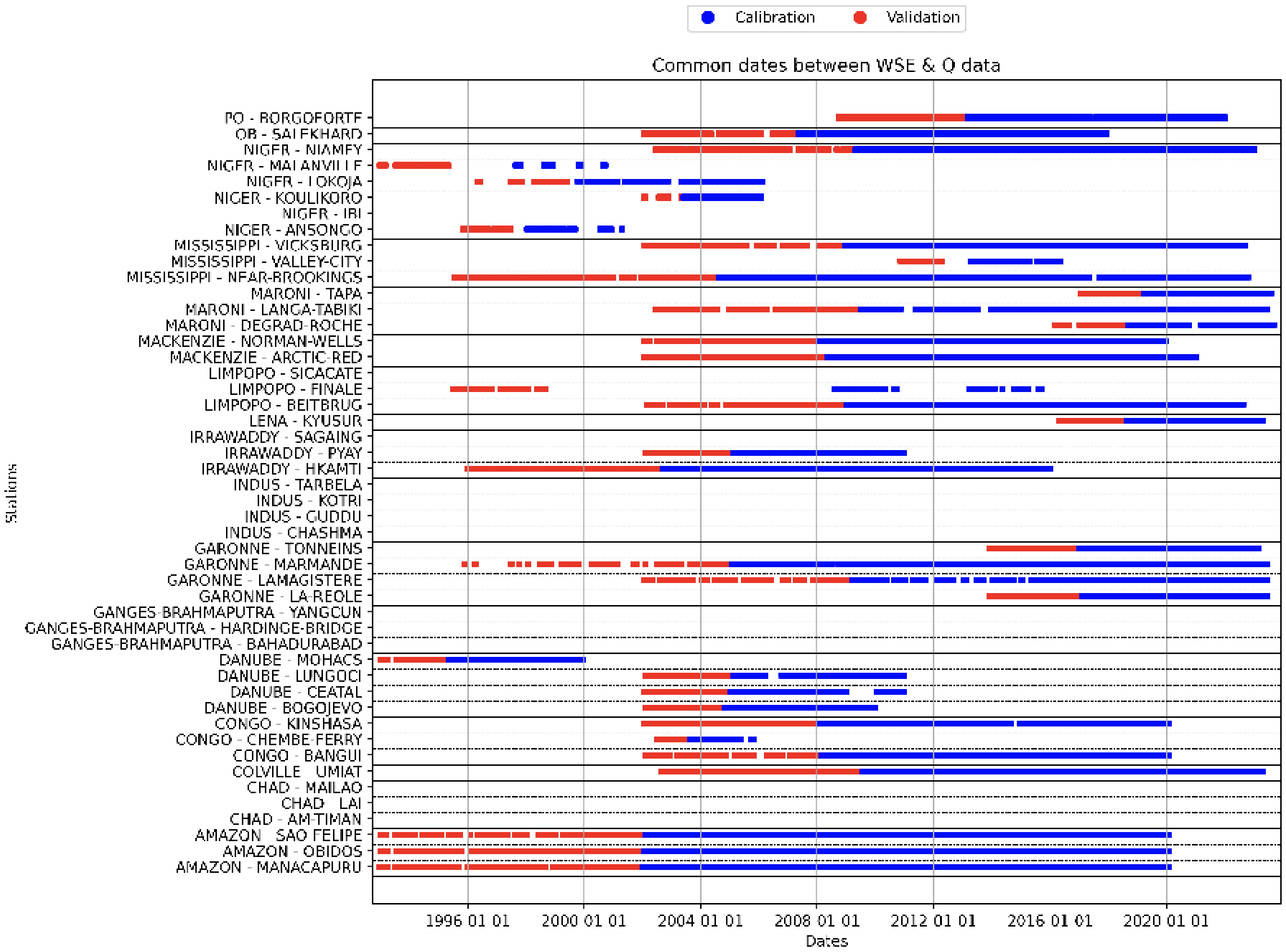
WMO (2012). International Glossary of Hydrology. WMO-No. 385, Geneva, World Meteorological Organization, ISBN 978-92-3-001154-3, https://library.wmo.int/records/item/35589-international-glossary-of-hydrology (last accessed 24 April 2024).

Zakharova, E., Nielsen, K., Kamenev, G., & Kouraev, A. (2020). River discharge estimation from radar altimetry: Assessment of satellite performance, river scales and methods. *Journal of Hydrology*, *583*, 124561.

Zakharova E., S Agafonova, C Duguay, N Frolova, A Kouraev. (2021). River ice phenology and thickness from satellite altimetry: potential for ice bridge road operation and climate studies. The Cryosphere, 15 (12), 5387-5407

Appendix A - Cal/val periods for each in situ station

Table A1. Calibration/Validation periods according to data availability for each station



Appendix B - Name and coordinates of each in situ station used for cal/val

Table A2. Location (with longitude and latitude) for each selected station [RD-2].

id	Basins	Station	lon	lat	source
1	AMAZON	MANACAPURU	-60.6094	-3.3106	GRDC
2	AMAZON	OBIDOS	-55.5131	-1.9192	GRDC
3	AMAZON	SAO-FELIPE	-67.3128	0.3717	GRDC
4	CHAD	AM-TIMAN	20.28	11.03	RivDIS
5	CHAD	LAI	16.3	9.4	GRDC
6	CHAD	MAILAO	15.28	11.58	GRDC
7	CHAD	NDJAMENA	15.03	12.12	GRDC
8	COLVILLE	UMIAT	-152.1227	69.3605	USGS
9	CONGO	BANGUI	18.5833	4.3667	So-hybam
10	CONGO	CHEMBE-FERRY	28.75	-11.9666	GRDC
11	CONGO	KINSHASA	15.3008	-4.2823	So-hybam
12	DANUBE	BOGOJEVO	19.08	45.53	GRDC
13	DANUBE	CEATAL	28.7167	45.2167	GRDC
14	DANUBE	LUNGOCI	27.5122	45.5559	GRDC
15	DANUBE	MOHACS	18.67	46	GRDC
16	GANGES-BRAHMAPUTRA	BAHADURABAD	89.67	25.18	GRDC
17	GANGES-BRAHMAPUTRA	HARDINGE-BRIDGE	89.03	24.08	GRDC
18	GANGES-BRAHMAPUTRA	YANGCUN	91.88	29.28	GRDC
19	GARONNE	LA-REOLE	-0.036	44.5776	schapi
20	GARONNE	LAMAGISTERE	0.831	44.121	schapi
21	GARONNE	MARMANDE	0.156	44.5	schapi
22	GARONNE	TONNEINS	0.301	44.389	schapi
23	INDUS	CHASHMA	71.38	32.43	nan
24	INDUS	GUDDU	69.713	28.419	nan
25	INDUS	KOTRI	68.317	25.442	nan
26	INDUS	TARBELA	72.698	34.09	nan
27	IRRAWADDY	HKAMTI	95.7	26	GRDC
28	IRRAWADDY	PYAY	95.22	18.8	GRDC

29	IRRAWADDY	SAGAING	96.1	21.98	GRDC
30	LENA	KYUSUR	127.39	70.68	ArcticGRO
31	LIMPOPO	BEITBRUG	29.9903	-22.2261	GRDC
32	LIMPOPO	FINALE	30.7414	-24.3311	GRDC
33	LIMPOPO	SICACATE	33.5431	-24.7444	GRDC
34	MACKENZIE	ARCTIC-RED	-133.745	67.458	HYDAT
35	MACKENZIE	NORMAN-WELLS	-126.85	65.27	HYDAT
36	MARONI	DEGRAD-ROCHE	-53.87	3.42	GRDC /schapi
37	MARONI	LANGA-TABIKI	-54.43	4.98	GRDC /schapi
38	MARONI	TAPA	-55.697	3.18	MGB
39	MISSISSIPPI	NEAR-BROOKINGS	-96.7489	44.18	GRDC
40	MISSISSIPPI	VALLEY-CITY	-90.6454	39.7034	GRDC
41	MISSISSIPPI	VICKSBURG	-90.9058	32.315	GRDC
42	NIGER	ANSONGO	0.5	15.6667	GRDC
43	NIGER	IBI	9.7333	8.2	GRDC
44	NIGER	KOULIKORO	-7.55	12.8667	GRDC
45	NIGER	LOKOJA	6.7667	7.8	GRDC
46	NIGER	MALANVILLE	3.4	11.88	GRDC
47	NIGER	NIAMEY	2.09	13.52	GRDC
48	OB	SALEKHARD	66.6	66.63	GRDC
49	PO	BORGOFORTE	10.7554	45.0449	AIPo
50	PO	PIACENZA	9.6667	45.0167	GRDC /AIPo
51	PO	PONTELAGOSCURO	11.6	44.8833	GRDC /AIPo
52	ZAMBEZI	KABOMPO-PONTOON	24.2166	-13.6	GRDC
53	ZAMBEZI	KASAKA	28.2166	-15.8166	GRDC
54	ZAMBEZI	MATUNDO-CAIS	33.5917	-16.15	GRDC