ESA Climate Change Initiative
River Discharge Precursor (RD_cci+)

D.3 River Discharge (Q) from Altimeters and Ancillary data, multispectral images and data combination - Algorithm Theoretical Basis Document (ATBD)

Contract number: 4000139952/22/I-NB

Reference

Issue 1.1 – 12/01/2024

CHRONOLOGY ISSUES

Issue	Date	Object	Written by
1.0	18/12/23	Initial version	Laetitia Gal and Angelica Tarpanelli
1.1	12/01/24	New version replying to ESA comments	Laetitia Gal and Angelica Tarpanelli

Checked by	S. Biancamaria - LEGOS	Sylvain Biancamaria 16/01/2024
Approved by	Alice Andral - CLS	A. Andral 16/01/2024
Authorized by	Clément Albergel - ESA	Clement Albergel 16/01/2024

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@legos.obs-mip.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.2. Selection of river basins. CCI River Discharge precursor project Document (CCI-Discharge-0004-RP_WP2, Issue 1.0)

[RD-2] D.3.1. Water Surface Elevation (WSE) Algorithm Theoretical Basis Document (ATBD) (CCI-Discharge-0005-ATBD-WSE, Issue 1.0)

[RD-3] he ESA river discharge CCI project - https://climate.esa.int/en/projects/river-discharge/about-the-river-discharge-project/

LIST OF ACRONYMS

ATBD	Algorithm Theoretical Basis Document
CCI	Climate Change Initiative
CDF	Cumulative Distribution Function
WSE	Water Surface Elevation
Q	Discharge
WP	Work Package
GRDC	Global Runoff Data Centre
AIPo	Agenzia Interregionale del Fiume Po
SCHAPI	Service Central d'Hydrométéorologie et d'Appui à la Prévision des Inondations
ArcticGRO	The Arctic Great Rivers Observatory
SO-HYBAM	Service d'Observation - HYdro-géochimie du Bassin AMazonien
RivDIS	The Global River Discharge
USGS	The U.S. Geological Survey's
MCMC	Markov Chain Monte Carlo
RC	Rating Curve
sDT	starting datetime
eDT	ending datetime

1 Introduction

This document provides an overview of the theoretical framework underlying the algorithm employed in computing long discharge time series data (Q) using two distinct approaches. The first approach involves computing discharges time series derived from the altimeter time series defined in WP3.1 [RD-2]. The second approach uses multispectral images to derive discharge time series.

Section two of this document will focus on discussing the available ancillary data essential for implementing these two approaches. Following that, the third section will delve into defining calibration and validation periods for each station as defined in [RD-1] and updated when WSE time series have been computed [RD-2].

Subsequently, section four will elaborate on the methodology used to derive discharge from altimeters. This section will incorporate the discussion of three different methods based on data availability. Additionally, it will explain the three approaches derived for these methods: the Bayesian approach when overlap time is present between discharge data and water surface elevation, the quantile approach when it is not and specific approach especially for arctic basins.

The fifth section will describe the procedure to derive river discharge from multispectral images. First, a description on the extraction of reflectance indices is provided along with a multi-mission approach to generate a single time series. Successively, the river discharge estimation is presented based on the similar approaches used for the altimeters (rating curves and quantile approach).

Finally, in section six, the multi-sensor river discharge approach is presented. Several merging procedures at Level-2 and at Level-3 are considered. Indeed, on one hand a first attempt is to generate river discharge by the ingestion of the water level by altimetry and reflectance indices following the RIDESAT (https://eo4society.esa.int/projects/ridesat/) approach (Level 2). On the other hand, the estimation of river discharge at Level-3 includes the combination of multiple river discharge products elaborated with the different approaches (by altimetry and by multispectral images). These analyses involve three methods to merge altimetry-derived and multispectral reflectance-derived river discharge: 1) the CDF techniques to densify the altimetry-derived river discharge with the multi-spectral reflectance-derived river discharge; 2) the copula method and 3) Machine learning approach.

All these products will be released to be analysed in the round robin and validation phase (WP4).

2 Available ancillary data

An initial analysis of the available data was conducted in WP2 [RD-1] to select stations where the estimation of long-term flow would be particularly valuable and feasible. Since WP2 [RD-1], and specifically during the generation of water height time series using altimetry (WP3.1 [RD-2]), this list of stations has been revised for various reasons, as detailed in the ATBD of WP3.1 [RD-2]. For the remainder of the project, we will investigate a total of 54 stations distributed across 18 different basins (see here: https://climate.esa.int/fr/projects/river-discharge/ and [RD3]) . From this updated list, we procured combined time series of water heights dating back to at least 2002, as well as time series from associated missions. In addition to these water height data, observed flow data from various global databases were incorporated:

· **GRDC**: The Global Runoff Data Base (GRDB) maintained by the Global Runoff Data Centre (GRDC) has been the primary dataset used in large-scale hydrological studies, with more than 9000 stations available to the research community (GRDC, 2015). The GRDC is an international archive of data up to 200 years old and fosters multinational and global long-term hydrological studies. Originally established three decades ago, the aim of the GRDC is to help earth scientists analyse global climate trends and assess environmental impacts and risks. https://www.bafg.de/GRDC/

· **AIPo**: The Italian hydrological monitoring network is managed at regional level by different agencies. For the Po basin, the Agenzia Interregionale del Fiume Po (AIPo) is responsible for the coordination of the hydraulic activity, the management and improvement of river navigation infrastructures, environmental and river protection and the coordination of the flood service. For the management of extreme events, AIPo is involved in forecasting and monitoring. Specifically, the website of the agency (https://www.agenziapo.it/content/monitoraggio-idrografico-0) shows real-time and historical measurements that can be freely downloaded by any users.

· **SCHAPI**: In France, in situ gages operated by regional public agencies (i.e. DREALs, Directions Régionales de l'Environnement, de l'Aménagement et du Logement) are collected by the SCHAPI (Service Central d'Hydrométéorologie et d'Appui à la Prévision des Inondations). SCHAPI releases these data publicly via the online “HydroPortail” national database (https://hydro.eaufrance.fr). These public agencies are responsible to observe and forecast floods, and to alert the population in case of dangerous events (https://www.vigicrues.gouv.fr/).

· **SO-HYBAM**: The HYBAM observatory is a unique facility that has been in operation since 2003, with a specialized focus on the monitoring of rivers and water resources in the Amazon region. HYBAM serves as a research support service, conducting extensive and long-term hydrological, sedimentary, and geochemical measurements to gain insights into the origin and evolution of water and transported materials (such as sediments, organic matter, nutrients, etc.) in Amazonian rivers, spanning from the Andes to the Atlantic Ocean. HYBAM collaborates with partners from all countries within the basin and extends its network by including four additional. stations located along rivers that also flow into the tropical Atlantic Ocean: the Orinoco, Congo, Maroni, and Oyapock rivers (https://hybam.obs-mip.fr/fr/donnees/).

· **HYDAT**: Hydrometric data are collected and compiled by Water Survey of Canada’s eight regional offices. The information is housed in two centrally managed databases: HYDEX and HYDAT. HYDEX is the relational database that contains inventory information on the various streamflow, water level, and sediment stations (both active and discontinued) in Canada. This database contains information about the stations themselves such as location, equipment, and type(s) of data collected. HYDAT is a relational database that contains the actual computed data for the stations listed in HYDEX. These data include daily and monthly means of flow, water levels and sediment concentrations (for sediment sites). For some sites, peaks and extremes are also recorded. The historical discharge data were extracted from the Environment and Climate Change Canada Historical Hydrometric Data web site https://wateroffice.ec.gc.ca/mainmenu/historical_data_index_e.html

· **ArcitcGRO:** The Arctic Great Rivers Observatory (ArcticGRO) is a collaborative research initiative dedicated to studying and monitoring the hydrology and discharge of some of the largest rivers in the Arctic region. By collecting and analyzing comprehensive data from these rivers, ArcticGRO contributes essential insights into the complex processes and changes occurring in the Arctic's freshwater systems, which are of significant importance in understanding climate change and its effects on the polar environment. The historical discharge data from the ArcticGRO database are available on their official website: https://arcticgreatrivers.org/discharge/.

· **RivDIS:** The Global River Discharge (RivDIS) data set contains monthly discharge measurements for 1018 stations located throughout the world. The period of record varies widely from station to station, with a mean of 21.5 years. These data were digitized from published UNESCO archives by Charles Vörösmarty, Balaze Fekete, and B.A. Tucker of the Complex Systems Research Center (CSRC) at the University of New Hampshire (Vörösmarty, 1998). River discharge is typically measured using a rating curve that relates local water level height to discharge. This rating curve is used to estimate discharge from the observed water level. The rating curves are periodically rechecked and recalibrated through on-site measurement of discharge and river stage.

· <u>USGS:</u> The U.S. Geological Survey's (USGS) supports the acquisition, processing, and long-term storage of water data. Water Data for the Nation serves as the publicly available portal to a geographically seamless set of much of the water data maintained within NWIS. Nationally, USGS surface-water data includes more than 850,000 station years of time-series data that describe stream levels, streamflow (discharge), reservoir and lake levels, surface-water quality, and rainfall. The data are collected by automatic recorders and manual field measurements at installations across the Nation. Data are collected by field personnel or relayed through telephones or satellites to offices where it is stored and processed. The data are processed automatically in near real time, and in many cases, current data are available online within minutes. Streamflow data can be download in the National Water Information System (NWIS) web interface at the following link: http://waterdata.usgs.gov/nwis.

All available data for each station and from all available sources have been merged to constitute observed in-situ discharge (Figure 1).

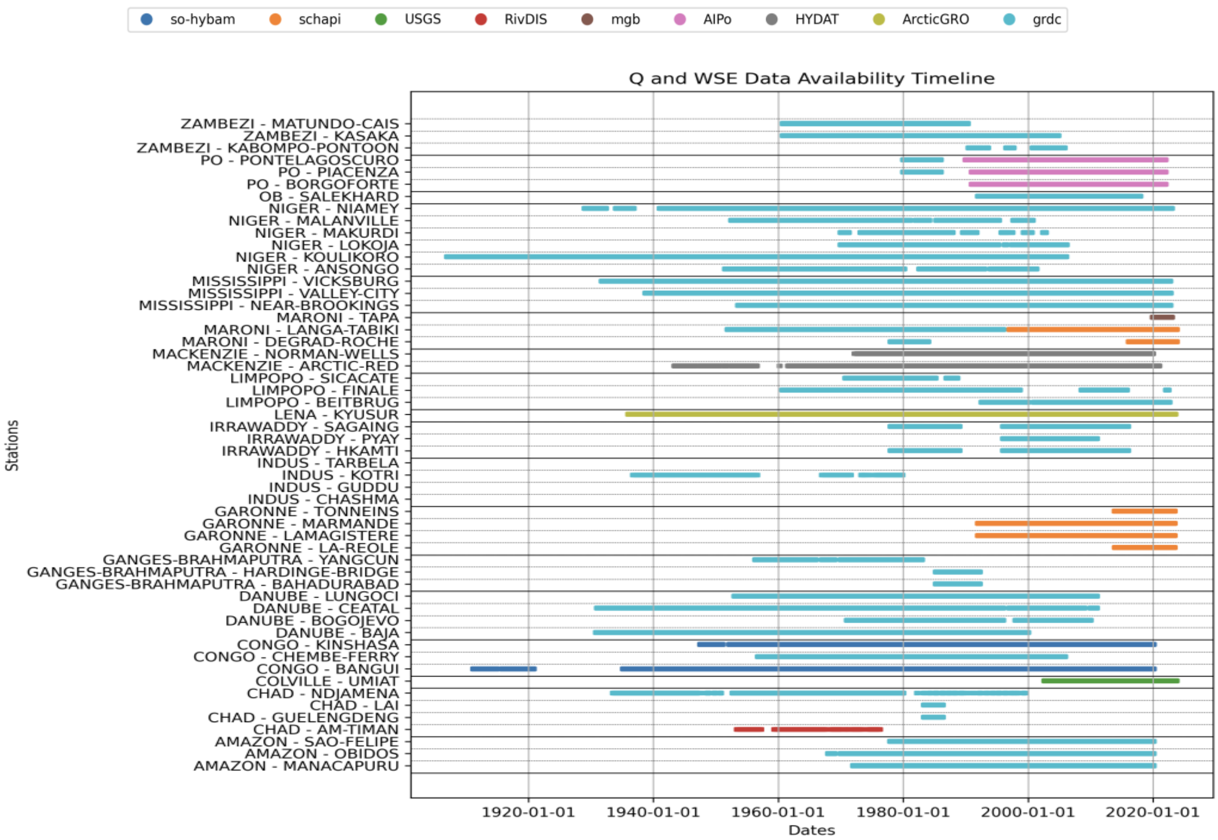

Figure 1: Available discharge data for all stations from different sources

3 Calibration – Validation periods

The methodology used to derive discharge from altimetry WSE will highly depend on ancillary data available and especially discharge time series used for calibration.

The selection of the calibration and validation period for altimeter calibration curves, correlating discharge and water surface elevation, is critically important for result precision. In our approach, we have chosen, in agreement with WP3.1 and WP3.1.2 teams, a method involving the use of the period from the earliest date where simultaneous in situ discharge and merged water surface elevation from the altimeter data (from WP3.1) are available to the latest date when these two data sets overlap (Figure 2). This period is divided into three sections. The first section, in chronological order, is allocated for the validation period, during which altimeter data is compared to reference data to assess model performance. The subsequent two sections are dedicated to the calibration period (Table 1). Using data from these latter two-thirds enables us to take advantage of the most recent satellite constellations and the least biased a priori data, thus ensuring optimal altimeter calibration. This methodological choice is designed to secure the reliability and precision of altimeter measurements across diverse hydrological conditions.

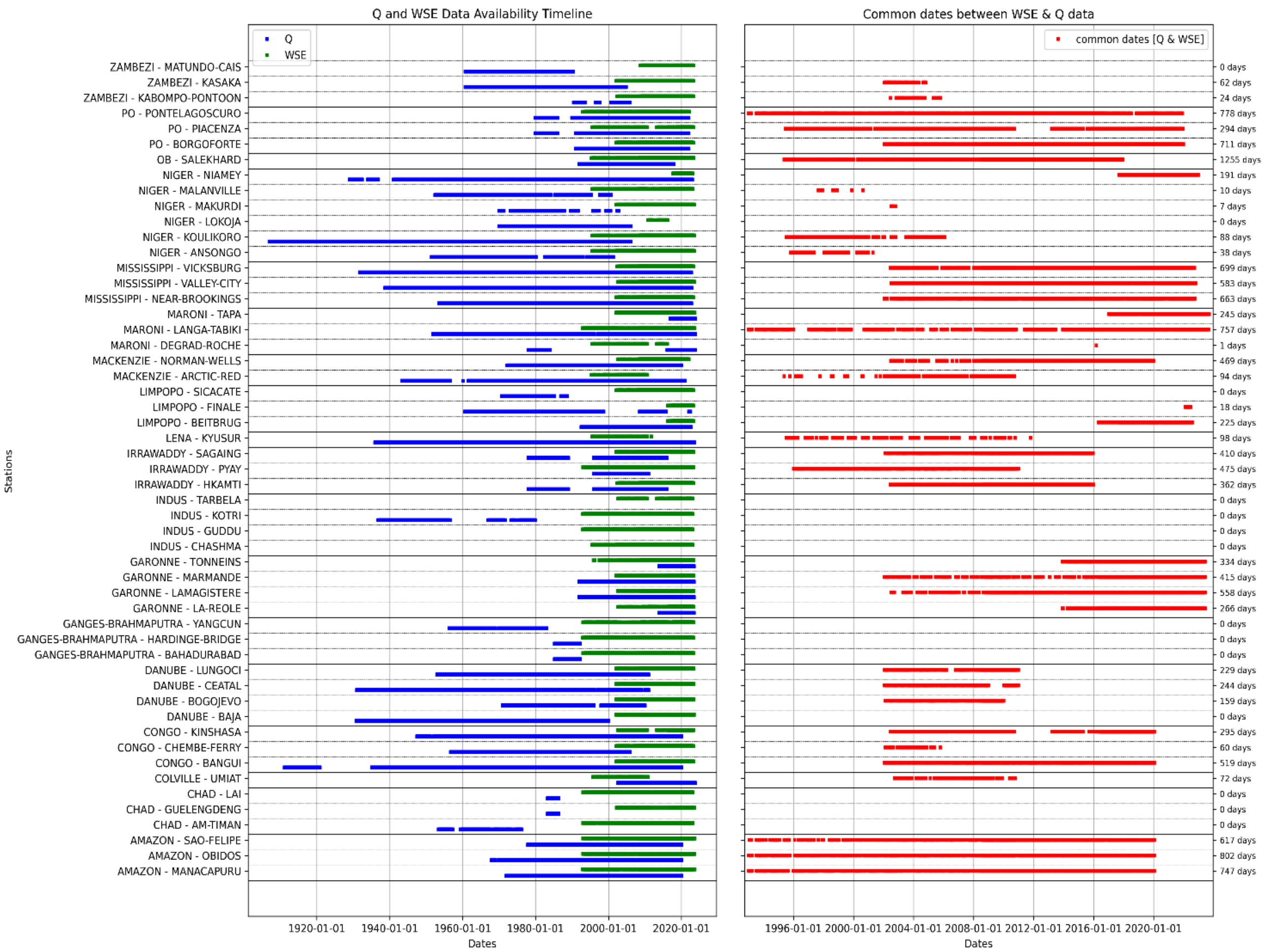

Figure 2: Available data for in situ discharge from various sources (blue) and merged water surface elevation from altimeter (green) from each station. The period where the two data sets overlap is represented in red with the number of overlap's days.

Table 1: Summary of the calibration and validation periods for each station, along with the number of days available for both calibration and validation. Stations marked with an asterisk (*) indicate cases where there is too little overlap between discharge and water surface elevation data to separate into calibration and validation periods. In these instances, all available data will be used for calibration or another method to derive discharge from altimetry will be use.

Basins	Stations	Calib sDT	Calib eDT	days	Valid sDT	Valid eDT	days
AMAZON	MANACAPURU	2002-01-05	2020-01-28	548	1992-12-24	2002-01-04	199
AMAZON	OBIDOS	2002-01-08	2020-01-24	599	1992-12-30	2002-01-07	203
AMAZON	SAO-FELIPE	2002-01-19	2020-01-28	529	1993-01-13	2002-01-18	88
CHAD	AM-TIMAN	-	-	-	-	-	-
CHAD	GUELENGDENG	-	-	-	-	-	-
CHAD	LAI	-	-	-	-	-	-
CHAD	NDJAMENA	-	-	-	-	-	-
COLVILLE	UMIAT	2005-05-30	2010-10-10	50	2002-09-22	2005-05-29	22
CONGO	BANGUI	2008-01-29	2020-01-28	377	2002-01-28	2008-01-28	142
CONGO	CHEMBE-FERRY	2003-05-05	2005-10-20	42	2002-02-09	2003-05-04	18
CONGO	KINSHASA	2008-05-05	2020-01-28	235	2002-06-22	2008-05-04	60
DANUBE	BAJA	-	-	-	-	-	-
DANUBE	BOGOJEVO	2004-09-26	2009-12-30	124	2002-02-07	2004-09-25	35
DANUBE	CEATAL	2005-01-08	2010-12-22	161	2002-01-16	2005-01-07	83
DANUBE	LUNGOCI	2005-01-16	2010-12-24	164	2002-01-27	2005-01-15	65
GANGES-BRAHMAPUTRA	BAHADURABAD	-	-	-	-	-	-
GANGES-BRAHMAPUTRA	HARDINGE-BRIDGE	-	-	-	-	-	-
GANGES-BRAHMAPUTRA	YANGCUN	-	-	-	-	-	-
GARONNE	LA-REOLE	2017-02-07	2023-06-09	214	2013-12-07	2017-02-06	52
GARONNE	LAMAGISTERE	2009-06-30	2023-06-10	495	2002-07-10	2009-06-29	63
GARONNE	MARMANDE	2009-03-06	2023-06-10	347	2002-01-17	2009-03-05	68
GARONNE	TONNEINS	2017-02-02	2023-06-10	230	2013-11-29	2017-02-01	104
INDUS	CHASHMA	-	-	-	-	-	-
INDUS	GUDDU	-	-	-	-	-	-
INDUS	KOTRI	-	-	-	-	-	-
INDUS	TARBELA	-	-	-	-	-	-
IRRAWADDY	HKAMTI	2006-12-18	2015-12-24	302	2002-06-14	2006-12-17	60
IRRAWADDY	PYAY	2001-01-02	2010-12-29	320	1996-01-04	2001-01-01	155
IRRAWADDY	SAGAING	2006-09-20	2015-12-19	315	2002-02-03	2006-09-19	95
LENA	KYUSUR	2000-12-11	2011-10-18	63	1995-07-09	2000-12-10	35
LIMPOPO	BEITBRUG	2018-06-04	2022-08-07	175	2016-05-01	2018-06-03	50
LIMPOPO	FINALE	2022-03-25	2022-06-18	11	2022-02-09	2022-03-24	7
LIMPOPO	SICACATE	-	-	-	-	-	-
MACKENZIE	ARCTIC-RED	2000-06-25	2010-09-14	83	1995-05-16	2000-06-24	11
MACKENZIE	NORMAN-WELLS	2008-04-24	2019-12-24	432	2002-06-24	2008-04-23	37
MARONI	DEGRAD-ROCHE *	2016-03-10	2016-03-09	0	2016-03-09	2016-03-09	1
MARONI	LANGA-TABIKI	2003-03-23	2023-09-10	563	1992-12-26	2003-03-22	194
MARONI	TAPA	2021-01-02	2022-12-23	73	2020-01-07	2021-01-01	36
MISSISSIPPI	NEAR-BROOKINGS	2008-12-17	2022-10-02	522	2002-01-24	2008-12-16	141
MISSISSIPPI	VALLEY-CITY	2009-04-12	2022-10-26	492	2002-07-05	2009-04-11	91
MISSISSIPPI	VICKSBURG	2009-03-19	2022-09-20	564	2002-06-17	2009-03-18	135
NIGER	ANSONGO	1997-08-22	2001-04-12	24	1995-10-26	1997-08-21	14
NIGER	KOULIKORO	1999-01-21	2006-01-30	54	1995-07-17	1999-01-20	34
NIGER	LOKOJA	-	-	-	-	-	-
NIGER	MAKURDI *	2002-08-06	2002-10-14	3	2002-07-01	2002-08-05	4
NIGER	MALANVILLE *	1998-08-28	2000-08-20	5	1997-08-31	1998-08-27	5
NIGER	NIAMEY	2019-06-13	2022-12-23	128	2017-09-06	2019-06-12	63
OB	SALEKHARD	2002-11-24	2017-12-10	1194	1995-05-16	2002-11-23	124
PO	BORGOFORTE	2008-09-11	2021-12-23	541	2002-01-19	2008-09-10	170
PO	PIACENZA	2004-04-23	2021-12-19	210	1995-06-24	2004-04-22	84
PO	PONTELAGOSCURO	2002-08-18	2021-12-02	571	1992-12-24	2002-08-17	207
ZAMBEZI	KABOMPO-PONTOON *	2003-07-27	2005-10-20	16	2002-06-13	2003-07-26	8
ZAMBEZI	KASAKA	2002-12-21	2004-10-25	35	2002-01-17	2002-12-20	27
ZAMBEZI	MATUNDO-CAIS	-	-	-	-	-	-

In instances where there are fewer than 20 common dates between discharge and water surface elevation data, all available data will be utilized for the calibration process, and validation will be conducted by comparing it with alternative data sources, such as in-situ water surface elevation measurements or simulated discharge data generated from modelling.

4 Derive discharge from altimeters

4.1 Context

Just as in-situ stage 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 is available. Several methodologies have been documented for deriving discharge time series from altimetry observations and supplementary data. At least three approaches will be used, depending on the available in situ discharge and altimetry water surface elevation (WSE) time series:

- **Method 1:** The preferred approach relies on the altimetry water surface elevation time series and in situ discharge time series to create a rating curve (RC) characterized by a power relationship between these two variables following a Bayesian approach (Rantz et al., 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. 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).

- **Method 2:** An alternative approach if in situ discharge data is no available, is to use simulated river discharges from hydrological model in combination with radar altimetry water elevation data to establish rating curves. This methodology has been successfully employed in basins such as the Amazon (Paris et al., 2016), Congo (Paris et al., 2022), and Ogooué (Bogning et al., 2021), as well as in smaller watersheds like the Tsiribihina River Basin in Madagascar (Andriambeloson et al., 2020). Some recent studies (e.g., Zakharova et al., 2020) have used a physically based Manning equation along with ancillary data, such as the dynamics of river width obtained from optical or SAR sensors, to derive river discharge.

- **Method 3:** The final option, in cases where there is no temporal overlap between in-situ or simulated discharge and water surface elevation data, assumes that the validity and stability of the rating curve persist across the various time periods covered by the two datasets. Both of these 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).

4.2 Approaches to derive Rating Curve (RC)

Rating curve derived from these different approaches will be used to create a long discharge timeseries according to long water surface elevation timeseries computed in the WP3.1 [RD-2].

4.2.1 Bayesian approach

The Bayesian 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:

$$P(\theta/D) = P(D)\cdot P(D/\theta)\cdot P(\theta)$$

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:

  $$Q = a\cdot(WSE - z0)^b$$

  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. However, generally accepted ranges might be:

  <u>Coefficient "a"</u>:

  - *a* must be non-negative since it represents the scaling factor for discharge.
  - *a* should stay within a range that makes sense for the specific system being studied. However, it can vary significantly based on the characteristics of the river or channel, hydraulic conditions, and other influencing factors. Through empirical observations, an average value of 800 and a standard deviation of 300 have been determined, offering a suitable representation for the flow coefficient across various rivers [0-1700]

Coefficient "b":

- *b* should not exceed physical constraints and so it should not be negative
- *b* should stay with a typically range between 0 and 1 for most hydraulic conditions even for some cases, it can be greater than 1 (Chow et al., 1988) to align with the underlying physical processes. To represent the characteristics of the system, a mean value of 1.5 and a standard deviation of 0.5 enables the inclusion of all values within a reasonable and physically plausible range [0 - 3]

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.
- *z0* cannot exceed the minimum value of water surface elevation (WSE). A mean value equal to the minimum value of water surface elevation minus 5 meters and a standard deviation of 5 should be able to consider all studying rivers with max depth around 30m.

- 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 et al., 2004). Two commonly employed MCMC algorithms are "NUTS" (No-U-Turn Sampler) and "Metropolis-Hastings." The advanced "NUTS" algorithm, which automatically adapts the optimal sampling step, was selected for this study (Hoffman et al., 2014).

4.2.2 Quantile approach

The Quantile approach employs statistical modelling using quantile functions to create a rating curve, eliminating the necessity for overlapping measurements. This algorithmic method enables the estimation of river discharge using satellite altimetry, even in instances where there are no in situ measurements within the altimeter's timeframe. This approach has undergone application and validation in diverse river basins spanning different climatic zones, such as the Amazon, Brahmaputra, Danube, Niger, and Ob (Tourian et al., 2013).

Assuming a stationary flow behaviour and no modification in the river bathymetry both at the altimetry virtual station and at the in-situ gage, this approach ensures the utilization of historical in situ data in current applications. This method computes the quantile functions of the altimetry water surface elevation on one hand and of the discharge time series on the other hand. Then a scatter plot of these in-situ discharge quantiles versus altimetry water surface elevation quantiles is computed to establish the rating curve. The quantile function, denoted as Q(p), is the statistical description of a dataset (Gilchrist, 2000). The quantile functions for altimeter water surface elevation, $Q_W(p)$, and in situ discharge, $Q_R(p)$, datasets result in the following equations:

$$Q_R(p) = \inf \{X_R \in R : p \leq F(X_R)\}$$

$$Q_W(p) = \inf \{X_W \in R : p \leq F(X_W)\}$$

Where, X_R and X_W refer to discharge and water surface elevation values, and $F(\cdot)$ represents the Cumulative Distribution Function (CDF). The quantile function specifies, for a given probability $0 < p < 1$, the maximum value that X_R or X_W can reach with this probability. To derive discharge from satellite altimetry, we need to model a function $T(\cdot)$:

$$X_R = T(X_W)$$

The function $T(\cdot)$ is proven to be non-decreasing, according to the demonstration in Gilchrist (2000), the transformation rule Q is valid. This rule suggests that a non-decreasing function $T(\cdot)$ applied to a decreasing function (quantile function) remains non-decreasing. Therefore, we propose to obtain the functional relationship between water surface elevation and discharge, T(·), through their quantile functions, rather than relying on the raw data:

$$Q_R = T(Q_W)$$

Since the temporal coordinate is not explicitly involved (not overlap period between WSE and Q), we eliminate the need for synchronized datasets. To empirically obtain the quantile functions for water surface elevation and discharge, data are sorted in ascending order. The rank of each dataset is normalized:

$$p_i = \frac{k_i}{N+1}$$

Where k_i is the rank of sorted values, N is the number of measurements, and p_i is the probability of water surface elevation or discharge data.

Due to the limitations of satellite altimetry, which offers data at intervals of either 10, 27 or 35 days, discrepancies may arise when compared to discharge data, which are frequently recorded at daily time scale. To address this issue, Tourian et al (2013) considered using monthly time-step data as a temporal resolution to build the rating curve using quantile function. The mean monthly discharge and altimeter water surface elevation are computed respectively from historical daily river discharge and merged multi-missions water surface elevation at satellites observations times (from WP3.1 [2]).

However, due to the inherent limitations of monthly mean values in capturing non-stationary river flow dynamics over time, we opted to explore an alternative approach to compute rating curve. Recognizing that river flow can exhibit variability and fluctuations that are not adequately captured by monthly averages, we compared the rating curve derived from previous approach to the rating curve derived from water surface elevation values at observed times and daily discharge values. By this revised approach, we aimed to create a rating curve that accounts for the temporal variability in river flow (extreme events for example). This approach allowed for a more detailed and time-sensitive analysis, offering insights into the relationship between water surface elevation and discharge that might better accommodate the non-stationary nature of river flow dynamics. Comparing these two rating curves derived from different data representations enables a more comprehensive understanding of how temporal variability impacts the relationship between water level and discharge in the river system

The two different approaches will be used to compute rating curve for non-overlap period between WSE and Q and will be compare in the Work Package 4.2. For both methodologies, the scatter plot of discharge quantiles versus merged multi-missions water surface elevation quantiles is constructed with a step of an equidistant quantile of 5 % for both quantile functions. From this scatter plot of quantile, the Bayesian method to compute the rating curve (as describe in section 4.2.1) is applied as for the Method 1 (see section 4.1) to derive river discharge from altimetry water surface elevation. The same approach is also used for each single altimetry mission water surface elevation. Recomputing the quantile function for each single time series might improve discharge estimation for some missions which have less uncertainty than others (past altimetry missions).

4.2.3 Specific approach

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.

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, 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 (as depicted in the figure below).

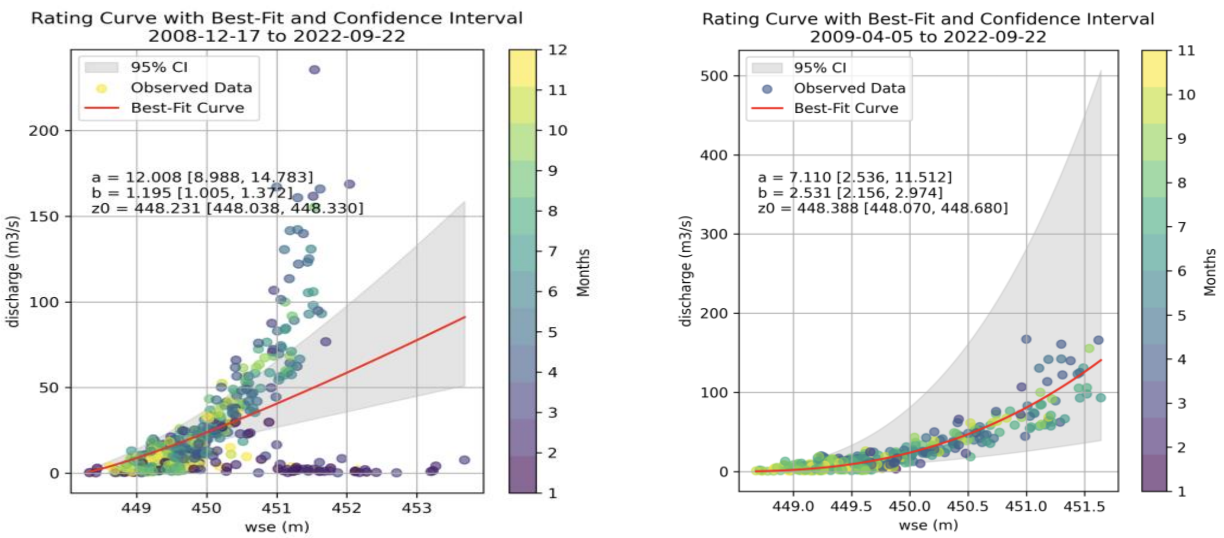

Figure 3: Rating curve for the Near-Brookings station (Mississippi basin) before (left) and after (right) excluding data points associated with temperature variations (below 0°C).

4.3 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 defined in section 4.2.1 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:

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

$$\partial(Q)=\sqrt{\left(a\cdot(WSE-z0)^b\cdot\left(\frac{\sigma a}{100}\right)\right)^2+\left(a\cdot b\cdot(WSE-z0)^{b-1}\cdot\left(\frac{\sigma WSE}{100}\right)\right)^2}$$
$$\overline{+\left(a\cdot(WSE-z0)^b\cdot\ln(WSE-z0)\cdot\left(\frac{\sigma b}{100}\right)\right)^2+\left(-a\cdot b\cdot(WSE-z0)^{b-1}\cdot\left(\frac{\sigma z0}{100}\right)\right)^2}$$

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 (see section 4.2.1). The standard deviation for the WSE will be, initially, consistent across all stations and for each mission. This value will later be adjusted through updates during the validation process.

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.

5 Derive discharge from multispectral images

5.1 Context

The method for estimating river discharge from multispectral images is based on the studies of Tarpanelli et al. (2013) and Filippucci et al. (2022), in which the differences between the passive response of the reflectance signal from the soil and that from the water are used to identify a change in the land area near the river channel that is shown to be strongly correlated with river discharge. An increase in river discharge produces an increase in wetted area, and the area near the river changes its reflectance response, which decreases. For an area near the river that is not affected by water, reflectance remains almost constant (except for changes in vegetation cover). Its relationship with the reflectance of the wetted area is used to more accurately determine the estimate of changes in hydrologic forcing, compared to the wetted area alone. Consequently, in the case of flooding, the reflectance ratio between the dry pixel (called the calibration pixel, C) and the wet pixel (called the measurement pixel, M) is sensitive to the increase of water in the wet pixel and, therefore, is directly related to the increase of river discharge. With respect to the first study by Tarpanelli et al. (2013) where the reflectance ratio C/M has been extracted from a temporal series of seven years of almost daily images of MODIS over four stations along the Po River, the recent analysis by Filippucci et al. (2022) demonstrated that the role of sediments and vegetation in the formulation was important to correct the reflectance ratio C/M during flood events. This was possible with the use of finer resolution images from Sentinel-2 and the new approach was tested over two Italian rivers, Po and Tiber.

5.2 Reflectance indices definition

In the project, we tested several algorithms for the estimation of the reflectance index. Such algorithms come from the combination of pixels: C for calibration, W for the sediments, V for the vegetation, M for measurements of the variation. The main steps for the application of the procedure are listed below:

- The study area is predetermined as a square of fixed side (e.g. 0.04 degree for Sentinel-2 data, 0.05 degree for Landsat data and 0.15 degree for MODIS data) around the selected station.
- The collection of the desired data in the chosen area is then obtained for the available period.
- Cloud products are considered to mask the cloud presence in the single images. After this masking, the total number of valid pixels in each image is computed. If the fraction of valid pixel was less than 0.2, the full image was discarded. Similarly, the fraction of valid reflectance value was calculated for each pixel during the study period: if this value was below the 5%, the pixel was deemed invalid and removed by the analysis, to avoid its selection in the mask calculation.
- Snow or ice presence is masked through other products available in the collection of data. The fraction of snow pixels is calculated for each available image and the resulting time series is averaged for each day of the year, to obtain a standard year probability of snow presence. A windowed moving average filter of 14 days was then applied to reduce the noises.
- The different categories of pixels (water area, vegetation, bare soil and field) are classified through masks derived by NDVI index, coefficient of variation (NIR standard deviation divided by NIR average), mean and standard deviation.
- The final step consists in the calculation of four algorithms of the reflectance indices to be compared with the in situ river discharge.

For some classes we used different formulations (e.g. for vegetation there is a distinction between forest and fields; for water we considered the mask with the mean or with the product mean times standard deviation) and for large rivers we use to aggregate the pixels resampling the images (especially when we use Landsat and Sentinel-2).

Moreover, because the analysis involved long temporal period, we need to consider the natural evolution of the river morphology. This means that in some areas a static temporal analysis that considers the same pixels for the entire period is not sufficient to describe the dynamic of the river. In such cases it is necessary to perform the analysis for brief periods. We considered two years in two years maintaining one year in common between a period and another in order to avoid big changes in the transition period. The procedure, called "multi-year", is used along with the "full period" for evaluating the benefit to separate the period with respect to the static investigation.

Based on what we describe above, the total number of algorithms to be considered in the analysis is 24 to be applied to the 11 satellite products.

5.3 Multi-mission reflectance time series

Once identified the 24 reflectance indices a first analysis is necessary to identify the best algorithm for the definition of river discharge. This discrimination is carried out based on the maximization of the Spearman correlation coefficient, to consider the non-linearity between the reflectance index and the river discharge. Because the procedure includes several combinations, we proceed to distinguish for each satellite products the following sub-cycles:

1- the best temporal method between full period and multi-year;
2- the best spatial resampling of the images with the aggregation;
3- the best formulation between the simple C/M or the ingestion of the sediments with CMW;
4- the best formulation with the inclusion of the vegetation

Once identified the best algorithm to define the reflectance, this is applied to the single satellite product. Because the analysis involves temporally many satellite products, it is fundamental to maintain a consistent approach for the same site. This is almost always guaranteed even if sometimes the different resolution of the satellite products induces to changes in the algorithms especially if the river is narrow (here, it is difficult to discriminate between W and M pixels).

The resulting temporal series appears quite noisy due to the different sensors involved and the algorithm that could be different. In order to have a single consistent time series for each site, a multi-mission approach to merge the different satellite products needs to be applied. Several attempts are tested to merge the temporal series:

- a simple average of the reflectance value at daily scale coupled with a moving window filter to remove the jumps between the successively values;
- weighted interpolation based on the temporal distance from the satellite passages. This includes a pre-processing for removing of the bias between the missions with mean and standard deviation (based on the master of the MODIS Terra product), and a post-processing for removing the outliers based on the daily minimum and maximum;
- a machine learning approach based on random forest technique.

The three procedures are applied to some test sites to identify the best solution to be implemented over the different areas. Once the best approach has been determined, itis applied to all the sites.

5.4 River discharge estimation from reflectance indices

The estimation of river discharge from reflectance indices is very similar to the rating curve approach applied for the water levels by altimetry. Indeed, here the relationship is based on the evaluation of a non-linear regression relationship between the multi-mission time series and the observed river discharge values. The relationship is built for the calibration period and successively it is applied for the rest of the period to generate the simulated river discharge.

When the in-situ data are not available, an approach similar to the one from Tourian et al. (2013) is applied. The CDF curves are calculated and compared to generate the percentiles associated to the discharges. With the relative correspondences between percentiles, it is possible to generate river discharge form the reflectance time series.

6 Derive discharge from multi-mission products

The traditional process to estimate river discharge that uses data from altimetry is here advanced with the contribution of the multispectral images to overcome the altimeter limits related to the temporal frequency. The merging procedure is carried out through four different methods:

- RIDESAT approach: the river discharge is based on the merging of data from two sensors altimeter and multispectral. Based on the traditional definition of river discharge as the product of river flow area and velocity, the two satellite sensors are used to define the two quantities, respectively. Indeed, the flow area is calculated as a function of the water level derived by satellite altimetry, whereas the flow velocity, traditionally measured through specific instruments installed in-situ (current meter, Acoustic doppler current profiler, velocimeter), is here a proxy coming from the reflectance measured by the Near Infrared signal of the multispectral sensor (Tarpanelli et al., 2015). The approach here is considered at Level-2 of the products.
- CDF method: the river discharge derived by multi-spectral reflectance is used to densify the river discharge obtained from altimetry by an approach based on the Cumulative Density Function. By linking the percentiles of the two river discharge time series through a specific relationship, it is possible to apply the same relationship to the multi-spectral reflectance indices and derive simulated discharge altimetry values used to fill the original altimetry-derived discharge time series.
- Copula method: Copula is a bivariate cumulative distribution function which is applied between any two random variables to get their joint probability distribution. The copula approach is based on the marginal distribution of the individual variables which can be coupled to derive the joint distribution of target variables characterized by nonlinearity and randomness. In the project the Frank copula based nonlinear model is used to model the high-frequent (almost daily) river discharge monitoring at a river section.
- Machine learning: some attempts are carried out to generate merged time series of river discharge with machine learning techniques. Preliminary analyses are here introduced to evaluate the potential of the common techniques (Artificial Neural Network, Random Forest, …) to estimate river discharge. The big challenge in the use of these approaches is related to the managing of discontinuous time series.

7 References

Andriambeloson, J. A., Paris, A., Calmant, S., & Rakotondraompiana, S. (2020). Re-initiating depth-discharge monitoring in small-sized ungauged watersheds by combining remote sensing and hydrological modelling: a case study in Madagascar. *Hydrological Sciences Journal*, *65*(16), 2709-2728.

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.

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]

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.

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

Gilchrist, W. (2000). *Statistical modelling with quantile functions*. CRC Press.

Hoffman, M. D., & Gelman, A. (2014). "The No-U-Turn Sampler: Adaptively Setting Path Lengths in Hamiltonian Monte Carlo." Journal of Machine Learning Research, 15(1), 1593-1623

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.

Liu, J., Bauer-Gottwein, P., Frias, M. C., Musaeus, A. F., Christoffersen, L., & Jiang, L. (2023). Stage-slopedischarge relationships upstream of river confluences revealed by satellite altimetry. Geophysical Research Letters, 50, e2023GL106394. https://doi. Org/10.1029/2023GL106394

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.

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.

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.

Vörösmarty, CJ et al. (1998): Discarge compilation from The Global River Discharge (RivDIS) Project. Distributed Active Archive Center, Oak Ridge National Laboratory, PANGAEA, https://doi.org/10.1594/PANGAEA.859439

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