EO4FLOOD

WATER CYCLE HYDROLOGY SCIENCE CLUSTER - ADVANCING FLOOD FORECASTING

Algorithm Theoretical Basis Document

ESA Contract 4000145540/24/I-EB

Project reference: EO4FLOOD_ESA_ATBD_D2.2
Issue:3.0
17/02/2026

Change Record

Date	Issue	Section	Page	Comment
08/09/2025	1.0			First version - DRAFT
05/12/2025	2.0	All	All	Second version - DRAFT
17/02/2026	3.0	3	55, 56, 65, 66, 71	Third version - DRAFT

Control Document

Process	Name	Date
Written by:	A. Tarpanelli, K. Nielsen, L. Ciabatta, S. Barbetta, P. Filippucci, C. Massari, M.J. Tourian, P. Saemian, O. Elmi, C. Chewing, R. Palmitessa, P. Tamagnone, G. Schumann, D. Dettmering, D. Scherer, D. Gustafsson, Y. Hundecha, K. Nielsen, M. Sadki, V. Pedinotti, E. Cantoni, B. Revilla-Romero, K. Larnier, L. Gal, S. Köhn	08/09/2025
Checked by	E. Volden, A. Vrettou	20/10/2025
Reviewed by:	A. Tarpanelli, K. Nielsen, L. Ciabatta, S. Barbetta, P. Filippucci, C. Massari, M.J. Tourian, P. Saemian, O. Elmi, C. Chewing, R. Palmitessa, P. Tamagnone, G. Schumann, D. Dettmering, D. Scherer, D. Gustafsson, Y. Hundecha, K. Nielsen, M. Sadki, V. Pedinotti, E. Cantoni, B. Revilla-Romero, K. Larnier, L. Gal, S. Köhn	05/12/2025
Reviewed by:	A. Tarpanelli, K. Nielsen, L. Ciabatta, P. Filippucci, C. Massari, M.J. Tourian, P. Saemian, O. Elmi, C. Chewing, C. Kittel, P. Tamagnone, D. Scherer, D. Gustafsson, Y. Hundecha, M. Sadki, K. Larnier, S. Köhn	17/02/2026
Approved by:	A. Tarpanelli	17/02/2026

	Signature	Date
For EO4FLOOD team	Angelica Tarpanelli 	17/02/2026
For ESA		

Table of Content

List of acronyms used

AMSR2: Advanced Microwave Scanning Radiometer 2
ASCAT: Advanced Scatterometer
AVHRR: Advance Very High-Resolution Radiometer
CGLOPS: Copernicus Global Land Operations
CH-TESSEL: Carbon Hydrology-Tiled
CHIRPS: Climate Hazards group Infrared Precipitation with Stations ECMWF Scheme for Surface Exchanges over Land
CLMS: Copernicus Land Monitoring Service
CRDP: Copernicus Research Data Package
DEM: Digital Elevation Model
DTM: Digital Terrain Model
EASE: Equal-Area Scalable Earth
ECMWF: European Centre for Medium-Range Weather Forecasts
EEA: European Environment Agency
EO: Earth Observation
EPSG: European Petroleum Survey Group
ERA5: ECMWF Reanalysis 5
ESA CCI: European Space Agency Climate Change Initiative
FSC: Fractional Snow Cover
FSCOG: Fractional Snow Cover on-Ground
FSCTOC: Fractional Snow Cover on Top of Canopy
GCOS: Global Climate Observing System
GFSC: Ground Fraction Snow Cover
GPM: Global Precipitation Measurement
GSMAP-NRT: Global Satellite Mapping of Precipitation Near Real Time
GSMAP-MVK: Global Satellite Mapping of Precipitation Microwave-Infrared combined product using Kalman filter
H SAF: EUMETSAT Satellite Application Facility on support to Operational Hydrology and Water Management
HR-WSI: High Resolution Water, Snow and Ice
HR-S&I: High Resolution Snow and Ice
IMERG: Integrated Multi-satellitE Retrievals for GPM
JAXA: Japan Aerospace Exploration Agency
JRC: Joint Research Council
LSTM: Long short-term memory
MCP: Model Conditional Processor
MGB: Modelo de Grandes Bacias
METOPA: (MetOp-A) Meteorological Operational satellite (A)
MODIS: Moderate Resolution Imaging Spectroradiometer
NAM: Nedbørs-Afstrømning rainfall runoff model
NASA: National Aeronautics and Space Administration
NDSI: Normalized Difference Snow Index
NOAA 18-19: National Oceanic and Atmospheric weather satellites 18&19
NRT: Near Real Time
PPS: Precipitation Processing System
PU: Predictive Uncertainty
SAR: Synthetic Aperture Radar
SCDA: Simple Cloud Detection Algorithm
SCE: Snow Cover Extend
SCFG: Snow Cover Fraction Ground
SM2RAIN: Soil Moisture to Rainfall
SMAP: Soil Moisture Active Passive satellite mission
SMOS: Soil Moisture and Ocean Salinity
S-NNP VIIRS: Suomi National-Orbiting Partnership, Visual Infrared Imaging Radiometer Suite
SPP: Satellite precipitation product
SWE: Snow Water Equivalent
SWOT: Surface Water and Ocean Topography
SWS: SAR Wet Snow
UTM: Universal Transverse Mercator
VIIRS: Visible Infrared Imaging Radiometer Suite
WGS: World Geodetic System
WSE: Water Surface Elevation
WSS: Water Surface Slope

1. Introduction

The EO4FLOOD project aims at demonstrating the maturity and effectiveness of cutting-edge satellite data in enhancing flood forecasting systems. The project focuses on leveraging advanced satellite technologies and algorithms to improve the accuracy and timeliness of existing hydrological and hydraulic models, resulting in more reliable and precise flood predictions. EO4FLOOD is structured around three key pillars: 1) Development of an Advanced Open Earth Observation Dataset (EO4FLOOD dataset) that leverages the latest products from both ESA and non-ESA satellite missions, ensuring global coverage with high spatial and temporal resolutions; 2) Integration of the EO4FLOOD Dataset into Flood Forecasting Models through the combination of hydrological, hydraulic, and flood models with machine learning techniques; 3) Demonstration of the EO Data and Models for Science and Society to show how the integration of EO data and models can improve flood forecasting and risk management.

1.1. Scope of the report

This document is the Algorithm Theoretical Basis Document for the EO4FLOOD and represents D2.2 of the project. The document presents the algorithms behind the fundamental data sets and models in the project.

1.2. Applicable documents

AD-01: Water Cycle Hydrology Science Cluster - Advancing Flood Forecasting – Statement of Work, V1.

1.3. Document Organization

After this introductory section, Section 2 provides the individual product descriptions and Section 3 the individual model descriptions. In Table 1.3.1, we will provide a list of product types and the responsible organizations that supplied each product. Figure 1.3.1 shows the Product Breakdown Structure of the EOdataset.

Table 1.3.1: List of the product type and the responsible organization.

Product overview	Product type	Responsible organization
Precipitation	IMERG Early Run	CNR
	IMERG Late Run	
	IMERG Final Run	
	CHIRPS	
	SM2RAIN_ASCAT	
	ERA5-Land	
	GSMAP-NRT	
	GSMAP-MVK	
	GSMAP-Gauge	
Soil Moisture	SMOS	CNR
	SMAP	
	ASCAT	
Snow	CLMS Fractional Snow Cover (FSC)	Magellium

	CLMS daily-cumulative Gap-filled Fractional Snow Cover (GFSC)	
	CLMS SAR Wet Snow (SWS)	
	CLMS Snow Cover Extent (SCE) - 500m - Europe	
	CLMS Snow Cover Extent (SCE) - 1km - Northern Hemisphere	
	Snow_CCI SCFG MODIS V3.0	
	Snow_CCI SCFG AVHRR METOPA V3.0	
	Snow_CCI SCFG AVHRR NOAA18 V3.0	
	Snow_CCI SCFG AVHRR NOAA19 V3.0	
	Snow_CCI SWE v3.1	
Width	Landsat	GIS
	Sentinel-2	
	MODIS	
Surface reflectance	Sentinel-2	CNR
	MODIS	
Water level virtual stations	Sentinel-3, Sentinel-6, SWOT	TUM
Multi-mission water level time series and slope	Sentinel-3, Sentinel-6, SWOT, Cryosat-2, ICESat-2	DTU
River discharge	from Width	GIS
	from Reflectance	GIS
	from Water level	GIS
	From multi mission (gauged)	GIS
	From multi-mission (ungauged)	TUM
Flood Extent	VIIRS	GMV
	Sentinel-1	GMV
	Sentinel-2	RSS-Hydro
Hydrological models	MGB	HydroMatters
	DHI-GHM	DHI
	HYPE	SMHI
	Hybrid-AI	DHI
	MCP probabilistic forecast	CNR
Hydraulic models	Mike+	DHI
	HEC-RAS	RSS-Hydro
	DassFlow 1D/2D	HydroMatters
	LISFLOOD-FP (2D) / GVF (1D)	SMHI
NRT precipitation	IMERG-ER	CNR
	IMERG-LR	
	GSMAP-NRT	
NRT water level	Sentinel-3 and Sentinel-6	TUM
NRT river discharge	From Water Surface Elevation	GIS

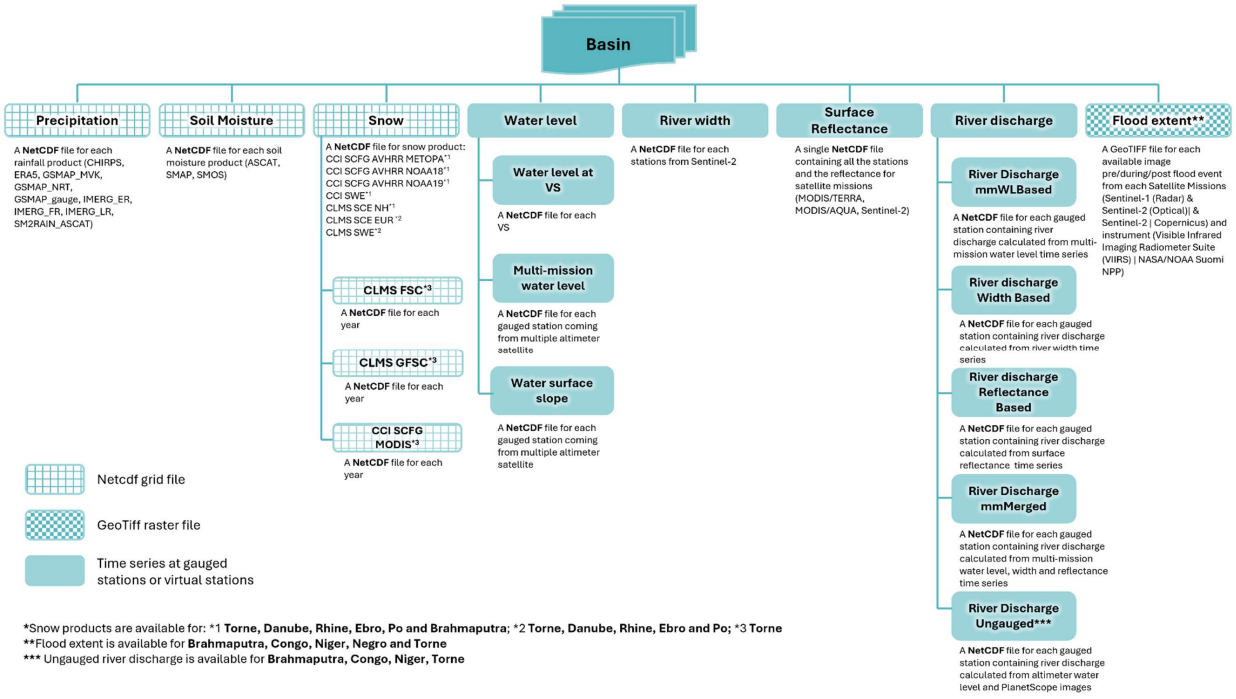

Figure 1.3.1 – Product Breakdown Structure of the EOdataset

2. Data products

2.1. Precipitation

Precipitation is the main driver for river discharge; therefore, an accurate estimation of this variable is of paramount importance for any activity related to flood monitoring and forecasting. The most accurate source of precipitation data are represented by rain gauges. These instruments provide accurate and reliable estimates of precipitation at the installation site. Anyhow, rain gauges are impacted by two main limitations that prevent the use over large areas: 1) the spatial representativeness of the estimate is referred to the area surrounding the instrument and 2) rain gauges are unevenly distributed throughout the globe. A valuable source of precipitation data could be found in remote sensing derived datasets. Satellite precipitation products (SPPs) are available now with both high spatial and temporal resolutions, with low latency and with high accuracy that allow their use for operational hydrological applications. Due to the different algorithms used for the generation of precipitation estimates used by the state-of-the-art SPPs and to perform a robust analysis, several datasets were collected and extracted over the study basins. The selected SPPs along with their main features are summarized in Table 2.1.1.

Table 2.1.1: Main features of the selected satellite precipitation products used in the EO4FLOOD project

Precipitation product	Temporal sampling	Spatial sampling	Temporal coverage	Latency
IMERG-ER	1 h	0.1°	2016-2024	4 hours
IMERG-LR	1 h	0.1°	2016-2024	18 hours
IMERG-FR	1 h	0.1°	2016-2024	3 months
CHIRPS	daily	0.05°	2016-2024	3 weeks
SM2RAIN_ASCAT	daily	0.1°	2016-2024	N/A
ERA5-Land	1 h	9 km	2016-2024	5 days
GSMAP-NRT	1 h	0.1°	2016-2024	4 hours
GSMAP-MVK	1 h	0.1°	2016-2024	3 days
GSMAP-Gauge	1 h	0.1°	2016-2024	3 days

For each of the selected study basins, the precipitation data has been extracted following the procedure outlined in Figure 2.1.1. In detail, the shapefile of each catchment is used to define the extraction area.

A buffer of 0.5° is then applied to the shapefile extent in the North, South, East and West directions in order to consider a wider rainfall field that can be used for spatial interpolation if needed. For each timestep of the parent SPP, only the pixels falling into this area have been extracted and stacked in order to create a 3D cube (longitude, latitude and time) for each basin during the analysis period, 2016-2024. It is worth mentioning that the native spatial and temporal resolutions of the original SPPs are used in order to avoid interpolation and regridding issues. The only exception is represented by the temporal resampling of the IMERG products, that have been cumulated to reach the hourly temporal resolution. An example of the data extracted over the Ebro River basin is reported in Figure 2.1.2.

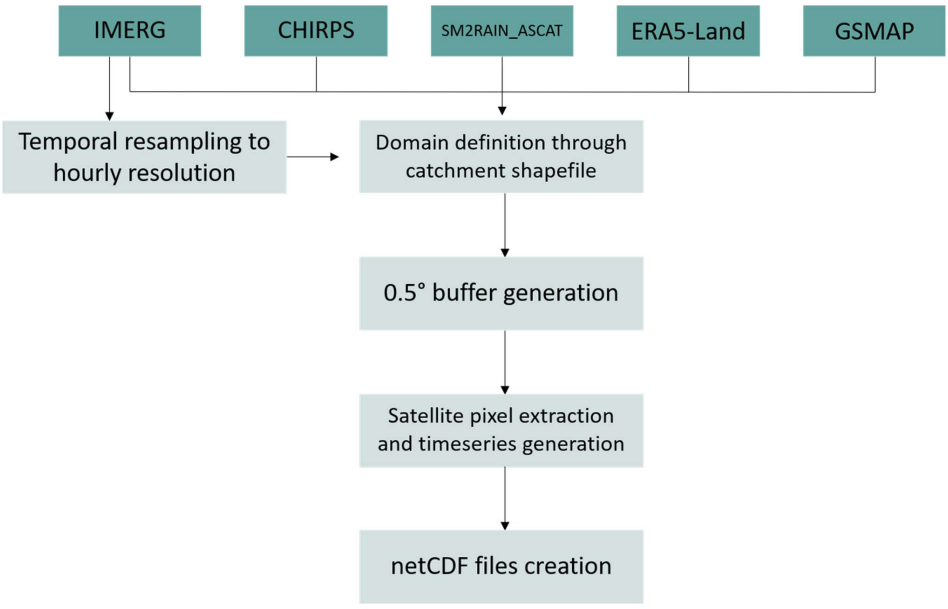

Figure 2.1.1: Flow chart of the precipitation data extraction over the analysis basins.

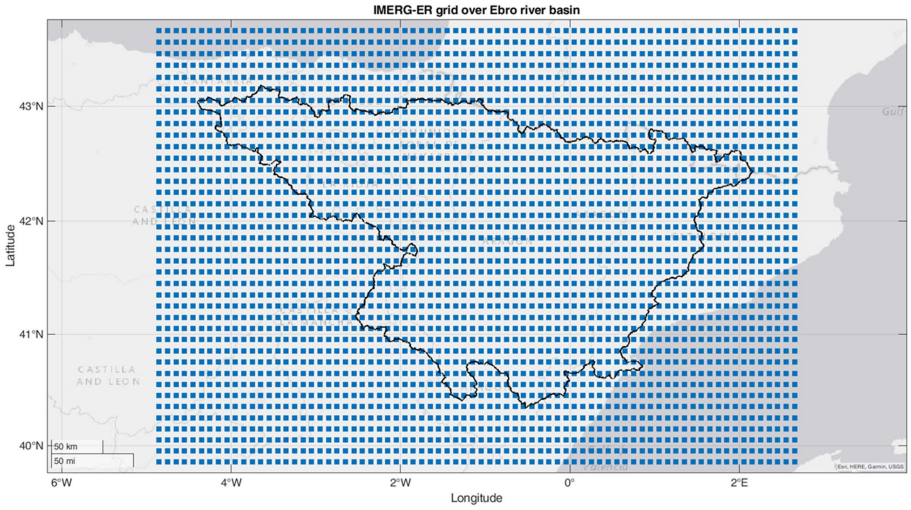

Figure 2.1.2: Example of data extraction for IMERG-ER over the Ebro River Basin. A buffer of 0.5° has been considered to extract the data over the basin boundary.

In the following, the main characteristics of the retrieval algorithms will be presented. For a more detailed description, the reader is referred to the cited literature.

- IMERG

The NASA IMERG products take advantage of a constellation of satellites working in the microwave and IR frequencies. The microwave instruments, both active and passive, provide the backscatter/emission of the hydrometeors, also by vertical profiling of precipitating clouds. The IR data are used to fill the gaps between subsequent microwave overpassed using a Kalman Filter technique. The data are then merged with a weighted combination of passive, active and IR data. The data are provided on a global scale over a 0.1° grid every 30 minutes (in EO4FLOOD the data have been resampled to 1 hour).

The algorithm releases three different products according to the data used that are characterized by different latency. The Early Run (ER) releases data as soon as possible using a forward propagation approach and it is characterized by a latency of 4 hours. The Late Run (LR) uses more satellite overpasses and takes advantage of both forward and backward propagation, allowing for interpolation. The LR also uses a climatological correction that incorporates gauge data. The product is available about 18 hours after sensing. The Final Run (FR) takes advantage of more satellite observations (not arrived on time for the release of LR) and uses a gauge-based adjustment to correct the estimates. It is available in 3 months. As already mentioned above, the data have been resampled to hourly temporal resolution to make it comparable to the other high-resolution SPP.

For further information about the IMERG products, the reader is referred to Huffman et al., (2020). The data are available upon registration through the NASA Precipitation Processing System (PPS) data servers.

- CHIRPS

The Climate Hazards group Infrared Precipitation with Stations (CHIRPS, Funk et al., 2015) product provides precipitation through interpolation and Cold Cloud Duration (CCD) observations based on IR data.

The algorithm is based on the estimation of precipitation probabilities using CCD approach (the colder clouds the higher the probability). In this phase, the product uses IR data from geostationary satellites. The data obtained in this way are then compared with a long-term climatology (CHPclim) to normalize the data and to bias correct errors. The anomalies are then used to estimate precipitation. Rain gauge stations are then used to evaluate a precipitation field through kriging and the merged with the satellite data. The product is available at 0.05° of spatial resolution on a daily basis for the 50°S-50°N latitude band. This means that CHIRPS is not available for high latitude basins like the Torne catchment. For further information about CHIRPS dataset, the reader is referred to Funk et al., (2015). The data are provided by the Climate Hazards Center of the University of Santa Barbara. The CHIRPS processing code is provided upon request.

- SM2RAIN-ASCAT

This dataset is based on the inversion algorithm named SM2RAIN (Brocca et al., 2014) that allows to estimate rainfall by knowing the soil moisture variation between two subsequent observations. The dataset used in EO4FLOOD is based on the application of the SM2RAIN algorithm to ASCAT soil moisture observations as described in Brocca et al. (2019).

The product is going to be released operationally within EUMETSAT H SAF (hsaf.meteoam.it) and it

provides rainfall observations at daily time scale over a 10 km regular grid on a global scale. Being based on soil moisture observations, the quality of the derived rainfall estimates is dependent on the quality of the raw satellite observations. This means that poor quality is expected over dense vegetated area, deserts and frozen soils, like Congo and Negro river basins. The SM2RAIN algorithm code is available on the CNR-IRPI GITHUB page (https://github.com/IRPIhydrology/sm2rain). The reader is referred to the H SAF website for release notes and product documentation.

- ERA5-Land

ERA5-Land provides hourly data of 50 land surface variables from 1950 onwards, combining the output of the modelling chain running at ECMWF with satellite and ground observations. It was produced by replaying the land component of the ECMWF ERA5 climate reanalysis and it is characterized by an improved spatial resolution (9 km), while the temporal resolution is 1 hour (Hersbach et al., 2020, Munoz Sabater et al., 2019). The CH-TESSEL (Carbon Hydrology-Tiled ECMWF Scheme for Surface Exchanges over Land) model ingests satellite observations of air temperature and humidity, wind speed and direction, ToA radiance and temperature and humidity atmospheric profiles along with temperature, humidity, pressure and wind data through ground monitoring network. The modelling framework assimilates these observations into the lowest ERA5 atmospheric model level (level 137, at 10 meters above the surface) which provides air temperature, specific humidity, wind speed, and surface pressure. The data are available upon registration through the Copernicus Climate Data Store (https://cds.climate.copernicus.eu/datasets/reanalysis-era5-land?tab=download).

- GSMAP

As per IMERG, GSMAP (Global Satellite Mapping of Precipitation, Kubota et al., 2020) estimates precipitation over the globe by combining three types of sensors installed on a constellation of satellites working in the microwave and IR frequency. Each sensor has its own advantages and disadvantages in terms of observation frequency and accuracy. The basic assumptions of the NASA algorithm are valid also for the generation of this kind of product. The microwave retrievals available on a global scale every 1 hour are merged together and then the IR data are used to fill the gaps. JAXA releases several versions of GSMAP products, according to the latency and the data used to generate the estimate. Among them, in the framework of EO4FLOOD project, the GSMAP_NRT, GSMAP_MVK and the GSMAP_Gauge products will be used. The GSMAP_NRT product uses microwave and IR data providing fast estimates over the 0.1° global grid every 1 hour. It is characterized by a latency of 3 hours. GSMAP_MVK is the standard version and takes advantage of more satellite overpasses, being available 3 days after sensing. The GSMAP_Gauge product uses the same approach as the MVK version but integrates correction through the use of the WMO SYNOP product. The correction considers the spatial distribution of rain gauges, and it is available with 3 days of latency. The data are available through the JAXA data repository (https://sharaku.eorc.jaxa.jp/)

2.2. Soil Moisture

The runoff response to rainfall is significantly influenced by soil moisture, which is the determinant of flood severity (Massari et al., 2020; Ortenzi et al., 2025). Monitoring soil moisture helps maintain accurate flood model predictions via data assimilation therefore contributing to improve flood forecasting (Massari et al., 2015). Current quasi-operational satellite products providing soil moisture data include the Soil Moisture Active and Passive (SMAP) mission, the Advanced Microwave Scanning Radiometer 2 (AMSR2), and the Advanced SCATterometer (ASCAT) with resolutions and revisiting times suitable for flood forecasting. Methods for enhancing the spatial resolution of these products

involve various data types and modelling approaches, while newer missions like the Sentinel-1 satellites and the CYGNSS mission offer promising enhancements for soil moisture retrieval and flood forecasting applications (Chew and Small, 2018; Azimi et al., 2020, Chew et al., 2023, Betchold et al., 2024). Three different datasets have been extracted and gridded over the nine study basins which are summarised in Table 2.2.1. The extraction and processing are summarised in Figure 2.2.1. As interpolation technique, it was used the nearest neighbour to avoid modifying the original soil moisture signal dynamics. The selected grid was the SMOS EASE grid (Equal-Area Scalable Earth Grid) which is the standard spatial reference system used for distributing SMOS Level-3 (L3) products processed by CATDS. The choice of this grid was motivated by the intermediate spatial resolution between SMAP and ASCAT products.

For soil moisture, we selected L3 versions of the ASCAT, SMAP, and SMOS soil-moisture products. Using L3 datasets guarantees a consistent spatial and temporal representation of the observations and supports a robust analysis of their sensitivity in flood forecasting at the research level. To emulate their use in near-real time (NRT) operations, we will additionally provide an algorithm for reading the data, enabling users to reproduce NRT ingestion conditions without relying on the operational data streams.

It should be noted that storing NRT soil-moisture datasets in a centralised database is not meaningful here and not compliant for their use in NRT. For example, in the case of ASCAT, multiple NRT products exist, each characterised by different processing chains, and update frequencies. Creating a homogenised database from these sources would require significant reprocessing, producing datasets that would no longer comply with the EUMETSAT product-specification and archival policies. For this reason, we supply harmonised L3 products and provide tools to emulate their operational use for the specific experiment, rather than maintaining a dedicated NRT archive, which will be the one then used for NRT applications. However, in the following we will provide a brief overview of the NRT version.

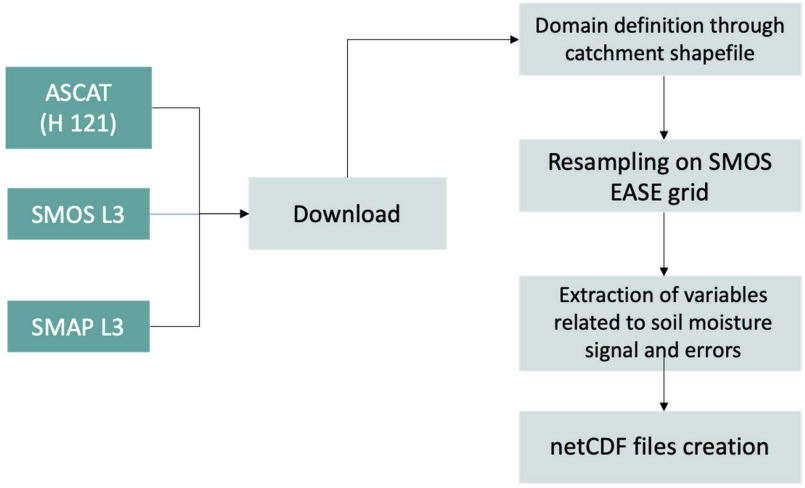

Figure 2.2.1: Flow chart of the processing of the satellite soil moisture datasets.

Table 2.2.1: Main features of the selected satellite soil moisture products used in the EO4FLOOD project

Soil moisture product	Revisit time	Spatial sampling	Temporal coverage	Latency
SMAP L3	2-3 days	36km	2016-2024	50 hours
SMOS L3	3 days	25-50km	2016-2024	18 hours
ASCAT (H 121)	Daily	12.5km	2016-2024	N/A

1) The Soil Moisture Active and Passive (SMAP) mission, launched by NASA in March 2015, provides global soil moisture observations with a spatial resolution of approximately 36 km and a temporal sampling of 1–3 days (Entekhabi et al., 2010). The mission employs an L-band radiometer to measure brightness temperature, which is sensitive to the soil's dielectric constant and therefore closely related to surface soil moisture (SM). The SMAP soil moisture retrieval algorithm is based on a radiative transfer model, which inverts the observed brightness temperatures using ancillary information on land cover, surface temperature, and vegetation. Specifically, the tau–omega model is used to account for vegetation attenuation and surface roughness effects, enabling accurate estimation of the soil dielectric constant—and thus volumetric soil moisture—in the top ~5 cm of the soil layer. NASA provides Level 3 soil moisture products on a global 36 km grid, updated every 1–3 days with a latency of about 50 hours. These products are derived from the radiometer-only configuration after the failure of the active radar component shortly after launch. The data are freely available from the National Snow and Ice Data Center (NSIDC): https://nsidc.org/data/smap/data. An example of the product is illustrated in Figure 2.2.2.

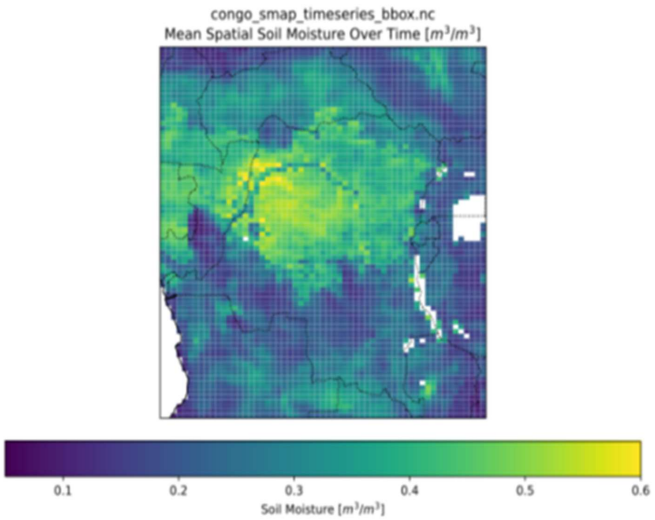

Figure 2.2.2: Example of mean (2016-2024) soil moisture over the Congo River Basin from SMAP L3 product.

2) The Soil Moisture and Ocean Salinity (SMOS) mission, launched by the European Space Agency (ESA) in November 2009, was the first satellite mission specifically dedicated to global soil moisture observation (Kerr et al., 2012). SMOS carries the Microwave Imaging Radiometer using Aperture Synthesis (MIRAS), an L-band (1.4 GHz) interferometric radiometer that measures brightness temperature across a wide field of view. These observations are used to retrieve surface soil moisture over land with a spatial resolution of 36–50 km and a temporal sampling of 1–3 days, depending on location and coverage. The soil moisture retrieval algorithm employed by SMOS is based on the L-band Microwave Emission of the Biosphere (L-MEB) model. This physically based model accounts for vegetation opacity, surface roughness, and dielectric properties of the soil to estimate surface soil moisture content in the top few centimeters. The retrieval is sensitive to land cover and atmospheric conditions and benefits from multiple-angle observations provided by MIRAS. The dataset used here is the Level 3 filtered soil moisture product, generated by the Centre Aval de Traitement des Données SMOS (CATDS). The dataset is available starting from January 2010. The data can be accessed in two ways: i) ESA FTP Server: ftps://smos-diss.eo.esa.int and ii) CATDS Data Portal: https://www.catds.fr/Products/Available-products-from-CEC-SM/CEC-SM-OS.

In addition to the standard CATDS Level-3 soil moisture product described above, we will make the SMOS Near Real Time Neural Network (SMOS NRT-NN) soil moisture product for specific case-study experiments. The SMOS NRT-NN product provides a fast-delivery soil moisture estimate derived from MIRAS L-band brightness temperatures using neural-network-based statistical coefficients trained on long-term SMOS retrievals (Rodríguez-Fernández et al., 2017). The product is available globally at a spatial resolution of approximately 36–50 km and with an effective temporal sampling of 1–3 days, consistent with the SMOS overpass frequency listed above. Its main operational feature is the very low latency of less than 3 hours, enabling near-real-time hydrological applications. Although the SMOS NRT-NN dataset is not listed in the main data table, it will employed for targeted experiments designed to emulate near-real-time conditions. A dedicated reading algorithm will be provided to support the ingestion of SMOS NRT-NN data in these simulations. The product is available from May 2015 to present and can be accessed through the ESA OADS data service (https://smos-diss.eo.esa.int/oads/access/). Users should note that, while fully operational, the NRT-NN product may be affected by residual radio-frequency interference (RFI) and currently has no follow-on mission planned.

3) For ASCAT, the dataset that will be provided is the H121 soil moisture product, which integrates observations from the ASCAT sensors onboard Metop-A, Metop-B, and Metop-C into a consistent long-term data record. The retrieval algorithm is based on the well-established change detection method developed at the Vienna University of Technology, where backscatter measurements are normalized between dry and wet reference values with an explicit correction for vegetation dynamics (Wagner et al., 2013). Unlike SMOS and SMAP, the ASCAT soil moisture variable represents a degree of saturation rather than volumetric soil moisture. The H121 dataset is available globally from January 2007, provided on a 12.5 km grid with daily temporal sampling. The product is generated by the EUMETSAT H SAF (Satellite Application Facility on Support to Operational Hydrology and Water Management) and is freely accessible via the H SAF product portal (https://hsaf.meteoam.it/Products/ProductsList?type=soil_moisture). Note that since this is a data records produced every year the latency information in Table 2.2.1 does not apply.

As mentioned before, in addition to the H121 long-term data record, here we described ASCAT Near Real Time (NRT) products for specific experiments designed to emulate real-time operational conditions. These include the individual platform-based soil-moisture products H101 (Metop-A), H16 (Metop-B), and H104 (Metop-C), which provide NRT soil moisture estimates at the same spatial sampling (12.5 km) with a latency of approximately 1.5 hours after sensing. These NRT products use the same change-detection retrieval methodology as the H121 dataset but are generated independently

for each Metop satellite. They are employed here exclusively in the NRT experiment framework and are therefore not listed in the main product table. As for SMOS an algorithm will be provided to read in NRT the different products.

2.3. Snow

Snowmelt is a major contributor to river systems, particularly in mountainous and high-latitude regions, providing a critical source of freshwater. Understanding snowpack conditions is essential for accurate flood forecasting, as rapid snowmelt can lead to sudden increases in river discharge.

Satellite observations can help to address critical data gaps by providing:

- snow cover maps derived from visible/infrared sensors allow consistent monitoring of snow extent, either provided as binary maps, or as snow fraction information.
- snow state conditions with wet snow maps from microwave-based sensors can detect presence of liquid water in the snowpack, correlated with the onset of melting.
- snow Water Equivalent (SWE) estimates — while SWE is not directly measured from space, satellite data can be combined with in situ observations and modelling to infer SWE values.

For several EO4FLOOD river basins characterized by a nival regime, different gridded snow datasets have been extracted and pre-processed. These datasets are summarized in Table 2.3.1.

Copernicus Land Monitoring Service snow products over Europe

The Copernicus Land Monitoring Service (CLMS) delivers snow information across Europe and the Northern hemisphere in operational production conditions. Among the CLMS snow portfolio, the following products were prepared for the EO4FLOOD project:

- **CLMS Fractional Snow Cover (FSC)**

The FSC product is part of the pan-European component of the CLMS portfolio, managed by the European Environment Agency (EEA) (https://doi.org/10.2909/3e2b4b7b-a460-41dd-a373-962d032795f3). Data are available over EEA38 + United Kingdom and was produced from September 2016 to mid-January 2025 in the frame of the HR-S&I project. An upgraded production is underway in the frame of HR-WSI (High Resolution Water, Snow and Ice) project.
Snow detection is performed on Sentinel-2 level-2A products generated with the atmospheric correction and cloud-screening MAJA software[1]. Snow is identified at 20m x 20m pixel spacing from the normalised difference snow index (NDSI) and the reflectance in the red band (Gascoin et al., 2019). A top-of-canopy snow fraction (FSCTOC) is then estimated from the NDSI for each snowy pixel (Hall et al., 2002, Salomonson and Appel, 2004). The on-ground snow fraction (FSCOG) corresponds to FSCTOC, except in forested areas, where the fraction under the canopy is estimated from FSCTOC. Snow presence can be derived from snow fraction values higher than 0. Note that the water areas are not masked out in the product. One might use its own water extent to mask out snow information above rivers, lakes, and seawater areas. The main limitations of the FSC product are related to the optical observation's limitations: the presence of clouds and cloud shadows and low illumination conditions (polar night) totally prevent or challenge the snow detection.

[1] https://gitlab.orfeo-toolbox.org/maja/maja accessed in July 2025.

Table 2.3.1: Key features of the selected satellite snow products used in the EO4FLOOD project (FSC: fractional snow cover, GFSC: gap-filled FSC, SCE = snow cover extent, SCFG = snow cover fraction on-ground, SWE = snow water equivalent). The temporal coverage corresponds to the availability of the data over the EO4FLOOD period study (2016-2024). Latency indicates the average time for data release after satellite acquisition.

Variable	Snow product	Revisit time	Spatial sampling	Temporal coverage	Latency
Snow cover fraction	CLMS FSC (HR-S&I)	5 days	20m	Sept. 2016-Dec. 2024	< 10 hours
Snow cover fraction	CLMS GFSC (HR-S&I)	Daily	60m	Sept. 2016-Dec. 2024	< 1 day
Wet Snow cover	CLMS SWS (HR-S&I)	6 days	60m	Sept. 2016-Dec. 2024	< 10 hours
Snow cover fraction	CLMS SCE	Daily	500m	March 2017-Dec. 2024	< 8 hours
Snow cover fraction	CLMS SCE	Daily	1km	Jan. 2018-Dec. 2024	< 1 day
Snow cover fraction	ESA CCI SCFG AVHRR ME-TOPA v3.0	Daily	0.05 degree	Jan. 2016-Dec. 2021	Not operational. Research context
Snow cover fraction	ESA CCI SCFG AVHRR NOAA18 v3.0	Daily	0.05 degree	Jan. 2016-Dec. 2022	Not operational. Research context
Snow cover fraction	ESA CCI SCFG AVHRR NOAA19 v3.0	Daily	0.05 degree	Jan. 2016-Dec. 2022	Not operational. Research context
Snow cover fraction	ESA CCI SCFG MODIS v3.0	Daily	0.01 degree	Jan. 2016-Dec. 2022	Not operational. Research context
Snow Water Equivalent	CLMS SWE v1.0	Daily	0.05 degree	Jan. 2016-Jun. 2024	< 12 hours
Snow Water Equivalent	ESA CCI SWE v4.0	Daily	0.1 degree	Jan. 2016-Dec 2023	Not operational. Research context

- **CLMS SAR Wet Snow (SWS)**

The SWS product is part of the pan-European component of the CLMS portfolio, managed by the EEA[2]. Data are available over high mountain areas of the EEA38 + United Kingdom and was produced from September 2016 to mid-January 2025 in the frame of the HR-S&I project. An upgraded production is underway in the frame of HR-WSI (High Resolution Water, Snow and Ice) project. The

[2] https://sdi.eea.europa.eu/catalogue/srv/api/records/47ac5988-e647-4481-9479-ccee7c2cd9b9, accessed in July 2025

product provides binary information on the wet snow extent and the snow free or patchy snow or dry snow extent in high mountain areas with a spatial resolution of 60 m x 60 m.
The presence of liquid water within the snowpack is detected with Sentinel-1 radar observations, whatever the meteorological and illumination conditions. Wet snow in this context means the presence of liquid water content within the snowpack, not necessarily at the snow surface. For the SWS product generation, "high mountains" means areas above a certain elevation where no human activities, such as tilling agricultural areas, affect the synthetic aperture radar (SAR) signal. As an example, the Torne River basin studied in the EO4FLOOD project is not covered by the SWS production. The retrieval of wet snow from SAR data are unreliable in forested areas, over water bodies and in urban areas, so these surface types are masked out. The SAR-based algorithm for mapping snowmelt area extent applies change detection based on the ratio of the reduced backscatter coefficient of wet snow versus the conditions when the surfaces are snow-free or covered by dry snow, the latter being transparent at the C-band from the Sentinel-1 instrument (Nagler et al., 2016).

- **CLMS daily-cumulative Gap-filled Fractional Snow Cover (GFSC)**

The GFSC product is part of the pan-European component of the CLMS portfolio, managed by the EEA[3]. Data are available over EEA38 + United Kingdom and was produced from September 2016 to mid-January 2025 in the frame of the HR-S&I project. An upgraded production is underway in the frame of HR-WSI (High Resolution Water, Snow and Ice) project. The product provides the extent of the snow cover per pixel as a percentage (0% – 100%) with a spatial resolution of 60 m x 60 m. It is based on radar satellite data from the Sentinel-1 constellation and on optical data from the Sentinel-2 constellation. The product uses Copernicus FSC (Fractional Snow Cover) and Copernicus SWS (SAR Wet Snow) products as input to form a spatially complete composite of snow conditions, to reduce observational gaps due to clouds and lack of sensor coverage on a daily basis. Gap-filling is limited to the last 7 days of input data and the snow coverage availability in the input products. Snow information from the SWS product is limited to the detection area of the SWS product in selected mountainous tiles. In practice, this means that over the Torne River basin, the GFSC data corresponds to a composite of FSCs only from the 7 past days. A conversion from the pixel snow fraction to the pixel snow presence can be done in the same way as the FSC (snow fraction > 0). Note that the water areas are not masked out in the product. One might use its own water extent to mask out snow information above rivers, lakes, seawater areas.

All the CLMS pan-European snow datasets (FSC, SWS, GFSC) from the HR-S&I project are freely available through an API after registration: https://github.com/eea/clms-hrsi-api-client-python (accessed in July 2025). The product documentation can be found online on the CLMS portal[4], with one ATBD dedicated to the FSC product[5] and another to the SWS and GFSC products[6]. FSC/GFSC/SWS data are delivered as Cloud-Optimized GeoTiff rasters on Sentinel-2 tiles (110km x 110km area) in Universal Transverse Mercator (UTM)/WGS84 or EPSG:326XX projection where XX indicates the UTM zone over which the product is set.

To reduce the volume of data to be processed within the EO4FLOOD project, the CLMS pan-European snow data variables, i.e; FSC, SWS and GFSC, are binarized into two variables: a *snow mask* and a *no-data mask*. These two masks summarize the relevant information for these products. Through them, the user can determine whether snow is present or not on a given pixel, , or if there is

[3] https://sdi.eea.europa.eu/catalogue/srv/fre/catalog.search#/metadata/befca22b-af37-4cd2-8e99-e09b3330fdc1, accessed in July 2025
[4] https://land.copernicus.eu/en/products/snow, accessed in July 2025
[5] https://land.copernicus.eu/en/technical-library/hrsi-snow-atbd/@@download/file.
[6] https://land.copernicus.eu/en/technical-library/hrsi-snow-s1-atbd/@@download/file

no information available. The user is referred to the PSD for more information on the binarization variables and their use.

In brief, for the EO4FLOOD application, data are binarized, reprojected to EPSG:4326 (WGS 84), merged and delivered with a 0.5° buffer around the river basin of interest. Figure 2.3.1 presents a flow chart summarizing the main steps involved in the preparation of the CLMS high resolution snow product (FSC, GFSC and SWS) for the EO4FLOOD project.

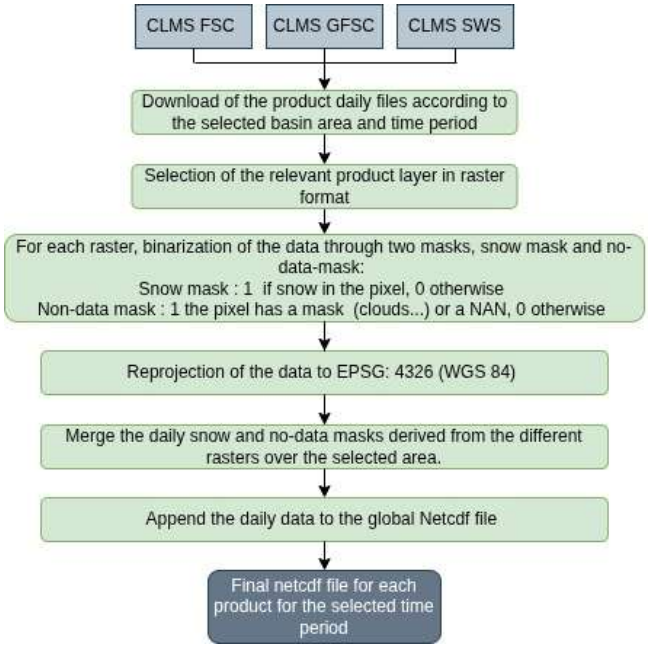

Figure 2.3.1: Flow chart illustrating the sequential steps undertaken to deliver the CLMS over Europe

Copernicus Land Monitoring Service snow products over the Northern Hemisphere

- **CLMS Snow Cover Extent (SCE) - 500m - Europe**

The SCE 500m product is part of the global component of the CLMS portfolio, managed by the Joint Research Council (JRC)[7]. Data in version 1 is available over continental Europe and is produced from March 2017 to present in the frame of the Copernicus Global Land Operations (CGLOPS-2) project. It provides daily updates of the fraction of snow cover on the ground (also in forested areas) per pixel in percentage (0% – 100%) in the spatial resolution of 500 m (grid resolution is 1°/200). Note that

[7] https://doi.org/10.2909/e2dd658f-8835-4b17-bcd5-eeb921a79a61, accessed in July 2025

archived products are available since 1st November 2000.
Snow classification is applied on top-of-atmosphere reflectance of visible and mid-infrared bands, and brightness temperatures of thermal infrared bands from Terra MODIS (Moderate resolution Imaging Spectroradiometer). It is based on a two-step approach: a pre-classification is performed using the NDSI which is based on the different reflectivity of snow and other surfaces in the visible and mid-infrared wavelengths (Hall et al. 2002), combined with a threshold approach on the brightness temperature from a thermal infrared band. The NDSI is also used as a basic map for cloud screening from the MODIS data since February 2016. On all pixels identified as potentially snow covered from the pre-classification, the SCAmod (Metsämäki et al., 2012) algorithm is applied to generate fractional snow cover maps. A canopy correction is applied to provide snow cover extent on the ground in forested areas. Clouds, water bodies, polar night and missing data pixels are flagged (see D2.3 Product Specific Documentation).

- **CLMS Snow Cover Extent (SCE) - 1km - Northern hemisphere**

The SCE 1km product is part of the global component of the CLMS portfolio, managed by the Joint Research Council (JRC)[8]. Data in version 1 is available over Northern hemisphere land areas, including Greenland, and is produced from January 2018 to present in the frame of the Copernicus Global Land Operations (CGLOPS-2) project. It provides daily maps of the fraction of snow cover on ground (also in forested areas) per pixel in percentage (0% – 100%) with a pixel spacing of about 1 km (grid resolution is 0.01° x 0.01°). The snow fraction is derived from medium resolution optical satellite data with the same approach as for the SCE 500m product, from the S-NPP Visible Infrared Imaging Radiometer Suite (VIIRS) Level 1B calibrated radiances. Clouds, water bodies, polar night and missing data pixels are flagged (see D2.3 Product Specific Documentation).

- **CLMS Snow Water Equivalent (SWE) - 5km - Northern Hemisphere - v1**

The Copernicus Global Land Operations (CGLOPS-2) foresees the delivery of Snow Water Equivalent (SWE) data in near real-time[9]. The SWE retrieval algorithm produces daily SWE data at a hemispherical scale by combining satellite-based microwave radiometer and optical spectrometer observations with ground-based weather station snow depth measurements. This approach applies the methodology developed by Pulliainen (2006) and further refined by Takala et al. (2011). Satellite data are processed through a snow emission model alongside weather station observations to generate initial SWE estimates. These satellite-derived SWE values are then combined with interpolated weather station-based SWE estimates, using a Bayesian assimilation, to produce the maximum likelihood SWE for each coarse-resolution grid cell across the Northern Hemisphere.
This daily product is typically available from late September to early June, depending on the availability of snow depth data from weather stations. If snow is observed outside this period—such as during the summer—the product may still be generated for those specific days when snow is recorded. The daily product becomes available approximately 12 hours after the necessary global satellite data has been acquired. The product covers the Northern Hemisphere between latitudes 35°N and 85°N and spans the full longitudinal range from –180° to +180°. It is provided on a regular latitude/longitude grid, using the WGS 1984 ellipsoid, with a spatial resolution of 0.05° (approximately 5 km). SWE data are not generated for mountainous regions, glaciers, or ice sheets.

[8] https://doi.org/10.2909/7f2eb891-cadb-4aa0-8b43-2c18e2a442e9 accessed in July 2025

[9] https://doi.org/10.2909/58ca9f01-a526-418d-8862-b4be43ef4738 accessed in July 2025

All the CLMS global snow datasets (SCE 500m, SCE 1km, SWE) from the CGLOPS-2 project are freely available through the Copernicus Data Space Ecosystem (CDSE) interface after registration: https://browser.dataspace.copernicus.eu/ (accessed in July 2025). The product documentation can be found online on the CLMS portal[10], with an ATBD for each product[11]. SCE 500m/SCE 1km/SWE data are delivered as multi-band netCDF4 files in a regular latitude/longitude grid (plate carrée) with the ellipsoid WGS 1984 (Terrestrial radius=6378km). For EO4FLOOD application, data are delivered with a 0.5° buffer around the river basin of interest. Figure 2.3.2 illustrates the sequence of steps required to deliver the prepared data for the CLMS SCE NH, CLMS SCE EU, and CLMS SWE products.

ESA Climate Change Initiative Snow project

The Snow CCI project is part of the ESA **C**limate **C**hange **I**nitiative programme (Solberg et al., 2025). Launched by ESA in 2008, the CCI is a research and development programme aimed at producing long-term, global climate data records from satellite observations. In the context of this programme, the snow CCI project aims to produce consistent, high-quality long-term datasets on snow that meet the standards of the Global Climate Observing System (GCOS). Its objective is to generate homogenized time series of daily global snow extent maps using optical satellite data and daily global snow water equivalent products derived from passive microwave satellite observations. The project was launched to address major differences found in existing snow datasets, which showed inconsistencies in snow cover trends and patterns.

The snow_cci climate research data package (CRDP) version 3, released in Q4 2024, contains the following time series of satellite based daily global snow products:

- Snow Cover Fraction Viewable or SCFV (snow cover fraction viewable at the surface in open areas, and in forested regions, it represents the snow visible on top of the canopy)
  - Terra MODIS data (2000–2022)
  - AVHRR data (1979–2022)
- Snow Cover Fraction Ground or SCFG (where the snow in the canopy is corrected to map snow in the ground in the forested areas).
  - Terra MODIS data (2000–2022)
  - AVHRR data (1979–2022)
- Snow Water Equivalent or SWE (1979 - 2022)

For the EO4FLOOD project, SCFG and SWE daily data from the aforementioned package are extracted, covering the period from 2016 to 2022 (see Table 2.3.1).

- **Snow_CCI SCFG MODIS V3.0**

The Snow CCI SCFG MODIS product, version 3.0 (Solberg et al., 2025; Schwaizer et al., 2024), is based on data from the MODIS (Moderate Resolution Imaging Spectroradiometer) sensor on the Terra satellite, launched in 1999. The dataset provides daily snow cover information from February 2000 to December 2022.
The SCFG MODIS product has a spatial resolution of approximately 1 km (0.01 degrees) and covers all land areas except Antarctica, the Greenland ice sheet, and regions with permanent snow and ice.

[10] https://land.copernicus.eu/en/products/snow, accessed in July 2025
[11] https://land.copernicus.eu/en/technical-library/algorithm-theoretical-basis-document-for-snow-cover-extent-europe-500-m-raster/@@download/file, https://land.copernicus.eu/en/technical-library/algorithm-theoretical-basis-document-for-snow-cover-extent-norther-hemisphere-1-km-raster/@@download/file, https://land.copernicus.eu/en/technical-library/algorithm-theoretical-basis-document-for-snow-water-extent-norther-hemisphere-5-km-raster/@@download/file

Greenland's coastal zones are included. The data are available at a daily temporal resolution.
Each dataset includes the snow cover fraction on ground (expressed as percentage per pixel), uncertainty estimates (provided as root mean square error per pixel), satellite viewing zenith angle (in degrees), and image acquisition time (in fractional hours of the day). For the EO4FLOOD project, only the snow cover fraction on ground variable, at the native spatial resolution, is extracted.
For SCFG MODIS product generation, cloud-covered pixels are first removed (masking of clouds) using an adapted Simple Cloud Detection Algorithm (SCDA2.0) (Metsämäki et al., 2015). Then, all cloud-free pixels are used for snow extent mapping. The Snow_cci snow cover mapping algorithm applies a two-step approach: first, all cloud-free pixels that are certainly snow-free are identified; secondly, the snow_cci SCFG retrieval method is applied to all remaining pixels (snow cover fraction classification). This retrieval method has been further developed and improved based on the ESA GlobSnow approach described by Metsämäki et al. (2015). Permanent snow and ice areas, along with water bodies, are masked in the SCFG products using the relevant classes from the Land Cover CCI map of the year 2000 as auxiliary data layers. There are some days with missing or incomplete data due to gaps in MODIS acquisitions, mainly between 2000 and 2006 and on a few occasions in later years.

- **Snow_CCI SCFG AVHRR V3.0**

Released in May 2023, the Snow CCI SCFG AVHRR product, version 3.0 (Xiao et al., 2024), is based on medium resolution optical satellite data from the Advanced Very High-Resolution Radiometer (AVHRR). The SCFG AVHRR product has a spatial resolution of approximately 5 km (0.05 degrees) and, similar to MODIS, covers all land areas except Antarctica, the Greenland ice sheet, and regions with permanent snow and ice. Greenland's coastal zones are included. The data are available at a daily temporal resolution, providing snow cover information from 1979 to 2022.
Different snow CCI AVHRR datasets are available depending on the specific AVHRR instrument and the satellite platform on which it is deployed. For the EO4FLOOD project, datasets from the AVHRR/3 instrument are selected because they offer the longest time coverage for the study period. The AVHRR/3 instrument scans the Earth's surface in six spectral bands and provides day and night imaging of land, water, and clouds. It also provides measures of the sea surface temperature, ice, snow, and vegetation cover. The selected datasets are derived from the following satellites:

- MetOp-A: It is the first in a series of three satellites forming the space segment of the EUMETSAT Polar System (EPS). It is also Europe's first polar-orbiting meteorological satellite, launched on 19 October 2006.
- NOAA-18: A polar-orbiting satellite operated by the National Oceanic and Atmospheric Administration (NOAA), launched on 20 May 2005.
- NOAA-19: Another NOAA polar-orbiting satellite, launched on 6 February 2009.

As a result, three distinct Snow CCI SCFG AVHRR datasets have been extracted for the EO4FLOOD project, each named after the satellite from which the data originate: AVHRR_METOPA (data selected from 2016 to 2021), AVHRR_NOAA18 and AVHRR_NOAA19 (data selected from 2016 to 2022).
The snow_cci SCFG AHVRR product uses an improved snow detection method based on the ESA GlobSnow approach (Metsämäki et al., 2015), similar to the MODIS product. It includes a pre-classification step to remove cloud-covered areas. Then, snow extent is then mapped using specific spectral bands. The snow cover mapping process has three steps: pre-classification to filter out cloud-free pixels that are clearly snow-free, snow cover estimates for the remaining pixels, a post-processing step to remove errors caused by cloud misclassification or unreliable data. ESA CCI Land Cover (2000) is used to mask out water bodies and permanent snow/ice areas if they cover more than 50% of a pixel.

- **Snow_CCI SWE v4.0**

The snow_cci SWE product version 4.0 was released in September 2025 (Solberg et al., 2025; Luojus et al., 2024). The product is derived using a Bayesian non-linear iterative data assimilation approach that integrates spaceborne passive microwave radiometer observations (SMMR, SSM/I, and SSMIS) with ground-based synoptic weather station data (snow depth measurements). The methodology builds on the work of Pulliainen (2006), Takala et al. (2011), and Venäläinen et al. (2023). The snow_cci SWE data record spans from January 1979 to December 2023, covering the Northern Hemisphere winter season from October to May, with occasional data available for June and September. SWE retrievals are produced every second day from January 1979 to May 1987, and daily from October 1987 onward. No SWE retrievals are generated during the Northern Hemisphere summer.

The dataset provides SWE information at a spatial resolution of 0.10° (approximately 10 km). It covers terrestrial, non-mountainous land areas of the Northern Hemisphere, excluding glaciers, the Greenland ice sheet, large lakes, complex terrain, and coastal regions of Greenland.

Download of the ESA Snow CCI products

All the ESA Snow CCI products and their documentation are stored in the Centre for Environmental Data Analysis (CEDA) Archive funded by the UK Natural Environment Research Council (NERC)[12]. The ESA Snow CCI products data are open access, and it can download for free through the following link: https://data.ceda.ac.uk/neodc/esacci/snow/data/ (accessed in July 2025).

Snow CCI data are delivered as netCDF4 files in a regular latitude/longitude grid with the ellipsoid WGS 1984 (Terrestrial radius=6378km). For EO4FLOOD application, daily data are downloaded, cropped to the river basin area with a 0.5° buffer, and appended to a single netCDF file. If a daily file is missing, the data for that date and area is replaced with NaNs. Figure 2.3.2 illustrates the data preparation workflow for the Snow CCI products.

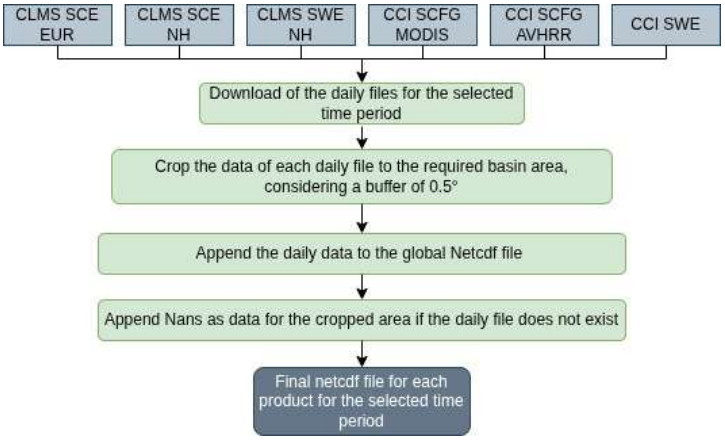

Figure 2.3.2: Flowchart representing the main steps in the data preparation process for the CLMS (SCE NH SCE EU and SWE) and Snow CCI products

[12] https://www.ukri.org/councils/nerc/, accessed in July 2025

2.4. Width

2.4.1. Short description

This algorithm estimates river width from optical satellite imagery (Landsat 8/9 and Sentinel-2) by combining water indices (Normalized Difference Water Index (NDWI), Automated Normalized Difference Water Index (ANDWI), Normalized Difference Moisture Index (NDMI), Automated Water Extraction Index - Shadow Version (AWEI-sh), and Automated Water Extraction Index - Non-Shadow Version (AWEI-nsh)), elevation data, and long-term water occurrence with a graph-based segmentation method. It is designed to extract reliable time series of river widths at selected reaches from the SWOT River Database (SWORD). The approach accounts for both spatial smoothness and temporal consistency, ensuring robust detection of the river boundary across varying hydrological conditions.

The method is tailored to operate within a predefined buffer around each reach, leveraging topographic and historical water occurrence constraints. The core of the algorithm is a Max-Flow/Min-Cut graph cut method applied to a water probability map derived from the input data, which yields a binary water/non-water classification used to compute river width over time.

2.4.2. Mathematical background

Let $\mathcal{L}$ denote the set of pixels in the buffered river region and assign a binary label $fp \in \{0,1\}$ to each pixel $p \in \mathcal{L}$, where fp=1 represents water and fp=0 represents non-water. The goal is to find a labeling that minimizes the energy:

$$E(f) = \sum_{p \in \mathcal{P}} D_p(f_p) + \lambda \sum_{(p,q) \in \mathcal{N}} V_{p,q}(f_p, f_q) \quad (2.4.1)$$

where $Dp\ (fp\ )$ is the **data term** (unary potential) representing the cost of labeling pixel *p* as water or non-water. It is derived from a sigmoid-scaled combination of the NDWI (or other water index) and the water occurrence probability from the GSW dataset. $Vp,q\ (fp\ ,fq\ )$ is the **smoothness term** (pairwise potential) promoting spatial continuity and penalizing label differences based on water index similarity.

- $\lambda > 0$ controls the trade-off between data fidelity and spatial regularization.
- N represents the set of 4-connected neighboring pixels.

The term $D_p(f_p)$ is the cost function defined in the temporal domain, which can be obtained from the monthly or long-term water occurrence map from GSW that delivers the probability of water in a certain pixel P_p. So:

(2.4.1)

$$D_p(f_p) = \begin{cases} P_p & \text{if } f_p = 1 \quad \text{(label: water)} \\ 1 - P_p & \text{if } f_p = 0 \quad \text{(label: non-water)} \end{cases} \quad (2.4.2)$$

On the other hand, the smoothness term $V_{p,q}(f_p, f_q)$ penalizes discontinuities between neighboring pixels based on water index similarity:

$$V_{p,q}(f_p, f_q) = 0 if f_p = f_q, \exp(-|I_p - I_q|/\sigma)\ if f_p \neq f_q \quad (2.4.3)$$

To solve the energy minimization problem defined by the binary labeling function f_p, we employ the

Max-Flow/Min-Cut algorithm. The formulation is cast as a graph cut problem, where each pixel corresponds to a node in a directed graph, and two additional nodes represent the source (water) and sink (non-water). The edge weights are derived from the data and smoothness terms: terminal links (T-links) encode the data cost $D_p(f_p)$ while neighborhood links (N-links) encode the pairwise smoothness cost $V_{p,q}(f_p, f_q)$. The Max-Flow algorithm (implemented using the PyMaxflow library) computes the minimum s–t cut that separates the graph into water and non-water regions, yielding the optimal binary segmentation f* that minimizes the total energy. This approach guarantees a globally optimal solution for binary segmentation under submodular energy terms.

2.4.3. Flowchart

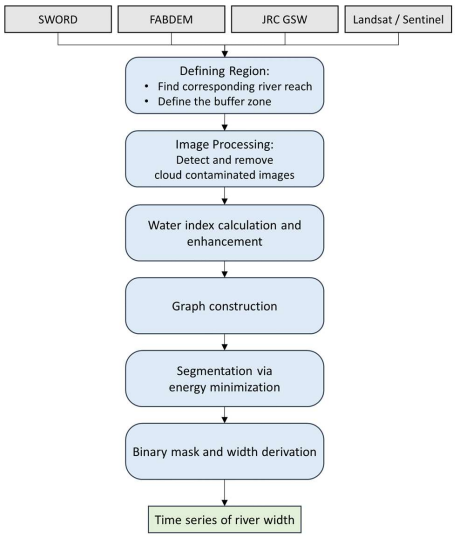

Figure 2.4.1: Flowchart of the methodology for estimating river width from optical satellite imagery.

2.4.4. Algorithm description

1. Region Definition: The algorithm starts by selecting a target river reach from the SWOT River Database (SWORD). A buffer is applied around the reach based on its average width and variability, defining the area of interest for width extraction.

2. Image Preprocessing: Time series of cloud-free optical satellite imagery (Sentinel-2 or Landsat 8/9) are filtered to retain only those pixels that satisfy terrain slope, elevation, and water occurrence criteria, using the FABDEM and the Global Surface Water Explorer (GSW) data sets.

3. Water Index Calculation and Enhancement: For each image including NDWI, ANDWI, NDMI, AWEI-sh, and AWEI-nsh are computed. The index time series is normalized and smoothed to reduce noise and enhance temporal consistency. To improve reliability, the water index is merged with a long-term water occurrence probability map, resulting in a probabilistic estimate of water presence per pixel.

4. Graph Construction: Each pixel in the buffered region is treated as a node in a graph structure.

Two additional nodes—source and sink—represent the water and non-water classes. Edges connecting pixels to the source/sink encode how likely the pixel belongs to each class. Connections between neighboring pixels promote spatial consistency and penalize abrupt transitions.

5. Segmentation via Energy Minimization: The segmentation problem is formulated as an energy minimization task. The algorithm seeks to assign a binary water/non-water label to each pixel in a way that balances the match with observed data and spatial smoothness. This is solved using a Max-Flow/Min-Cut algorithm, which guarantees an optimal binary segmentation under specific conditions.

6. Binary Mask and Width Derivation: The output is a binary mask showing water and non-water areas for each date in the image time series. The mean river width for the selected reach is calculated by dividing the water mask area by the reach length.

7. Postprocessing: include filtering out noise, correcting small, isolated regions, or applying geometric constraints to refine the final width estimates.

2.4.5. Output product

The main output is a time series of river width estimates at each selected reach. This product serves as input for river discharge estimation. The output is in CSV and netCDF format containing columns of date, river width [m] together with metadata as is described in PSD document.

2.5. Surface reflectance

2.5.1. Short description

Surface reflectance can be used to obtain a proxy of river discharge, as demonstrated by Tarpanelli et al. (2013; 2017; 2020) and Filippucci et al. (2022; 2025a; 2025b). Specifically, the different passive responses of vegetation, soil and water in the Near InfraRed (NIR) spectral band is used to infer information related to river discharge. Indeed, the NIR signal of water is very low in comparison to the one of soil and vegetation. Leveraging on this, a proxy of river discharge can be obtained analyzing the NIR signal of a periodically wet area (Measurement pixels, M), which will decrease when the area is flooded due to an increase of river discharge and increase in the opposite conditions. Moreover, including the signal of permanently wet pixels (Water pixels, W), vegetated areas (Vegetation pixels, V) and stable areas (roads or bare soil, Calibration pixels, C) near the river allows to compensate variation of NIR signal caused by other sources, such as atmospheric and climatological conditions, river sediment load within M pixels. Tarpanelli et al. (2013) laid the foundation for this analysis by applying the ratio between C and M (CM approach) to NIR data from MODIS satellite sensors. Further studies analyzed data from MERIS and Sentinel-3 multispectral sensors, with spatial characteristics consistent with MODIS (Tarpanelli et al., 2017; 2020). The work of Filippucci et al. (2022) allowed to move towards high resolution sensors such as Sentinel-2, including the sediment load correction (CMW approach) and developing an uncalibrated procedure for finding M pixels positions. The vegetation correction was then implemented in Filippucci et al. (2025a), while in Filippucci et al. (2025b) the NIR proxy from multiple sensors (MODIS, Sentinel-2, Landsat-5, -7, -8 and –9) were merged to maximize both temporal resolution and performance of the estimated river discharge across 54 sites worldwide.

2.5.2. Mathematical background

Different versions of the CM/CMW approach are used to estimate river discharge, according to the characteristics of the river reach analyzed. Specifically, following Filippucci et al. (2022) and (2025a), four formulations are proposed, according to the possibility to correct the sediment load and the presence of vegetation within the periodically wet areas M:

1. For areas with scarce/no vegetation around the embankments and where the sediment transportation is negligible, the original CM index (Tarpanelli et al., 2013) proportional to the river discharge (Q) is obtained through:

$$Q \propto CM = \frac{C}{M} \tag{2.5.1}$$

2. For areas with significant sediment transportation and scarce/no vegetation around the embankments, the CMW index is obtained through:

$$Q \propto CMW = \frac{C}{(M - wa * W + z)} \tag{2.5.2}$$

where wa is an index varying between [0 1], sensitive to the fraction of M pixels that are actually flooded, wa is given by:

$$wa = \frac{(C - M)}{(C - W)} \tag{2.5.3}$$

and it is equal to 0 when the pixels in M are completely dry (M=C) and equal to 1 when the pixels in M are completely flooded (M=W), while z is designed to fix the minimum of the denominator to be equal to the minimum of M, according to:

$$z = \max(wa * W - M) + \min(M) \tag{2.5.4}$$

3. For areas where the sediment transportation is negligible and with flourishing vegetation over the embankments, the CVM index is obtained through:

$$Q \propto CVM = \frac{(C + V)}{2M} \tag{2.5.5}$$

4. For areas with significant sediment transportation and with flourishing vegetation over the embankments, the $CVMW$ index is obtained through:

$$Q \propto CVMW = \frac{(C + V)/2}{(M - wa_v * W + z_v)} \tag{2.5.6}$$

with

$$wa_v = \frac{(C+V)/2-M}{(C+V)/2-W} \quad (2.5.7)$$

and

$$z_v = max(wa_v * W - M) + min(M) \quad (2.5.8)$$

When sediment transport is negligible for a river, the W signal effectively only adds noise to the index calculation and therefore should be avoided (cases 1 and 3). Instead, when flourishing vegetation is present over the embankments, the pixels M (that are usually found over the riverside) generally contain a portion of the vegetation signal of V: in other words, in dry condition M will not tend to approach the value of C alone but rather a combination of C and V. Therefore, C should be corrected accordingly (cases 3 and 4).

M positions can be obtained with calibration, when river discharge observation concurrent to satellite overpass are available. In such cases, when the river width variation is considerably larger than the satellite spatial resolution, a spatial kernel is applied to satellite pixels in order to match the desired width variation. When observed data are not available, uncalibrated procedures are used to determine M pixels.

2.5.3. Flowchart

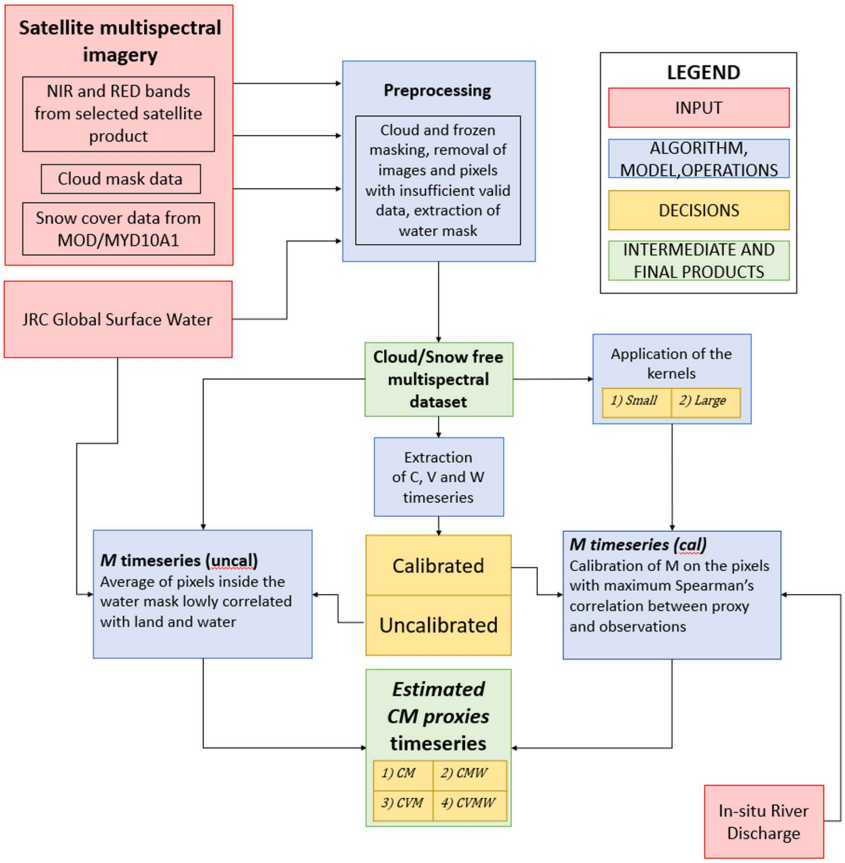

Figure 2.5.1: Flowchart regarding the surface reflectance methods

2.5.4. Algorithm description

- Inputs collection: NIR and RED bands from the selected satellite products (MODIS AQUA: MYD09GQ.061, MODIS TERRA: MOD09GQ.061 and Sentinel-2: Harmonized Sentinel-2 L1c) were retrieved from Google Earth Engine (GEE) platform using the Python API and stored locally. Corresponding cloud-masking datasets were also obtained (MODIS AQUA: MYD09GA.061, MODIS TERRA: MOD09GA.061 and Sentinel-2: Cloud Score+ S2_HARMONIZED V1 and Sentinel-2: Cloud probability). Snow-cover data (MOD10A1.061 and MYD10A1.061) were collected from GEE to exclude frozen and snow-covered periods from the analysis. In addition, JRC Global Surface Water data were downloaded from GEE to derive water masks for the MODIS products, to estimate the kernels for resampling M pixels, and for use in one of the uncalibrated procedures.

- Preprocessing: the satellite data from MODIS AQUA, MODIS TERRA and Sentinel-2 are masked for clouds and frozen periods. Images with less than 20% valid data are excluded from the analysis, as well as pixels with less than 5% valid observations. Water masks are derived from JRC Global Surface Water data for the coarse-resolution sensors (MODIS) and from a threshold applied to the 5th percentile of the NDVI timeseries for Sentinel-2. Images with less than 20% valid data within the water-mask area are also excluded from the analysis.

- C, V and W signals: The C signal is obtained by averaging the NIR reflectance of all pixels outside the water mask that have a coefficient of variation (standard deviation divided by the mean) below the 2nd percentile. The V signal is obtained by averaging the NIR reflectance of all pixels outside the water mask with an average NDVI value above the 95th percentile. The W signal is obtained by averaging the NIR reflectance of all pixels inside the water mask where the product of the mean NIR value and its standard deviation is below the 5th percentile.

- Calibrated M: the JRC water-mask occurrence information is used to obtain a first estimate of the river width variation. This estimate, combined with the spatial resolution of the sensor, is used to determine the number of satellite pixels required to cover the target area. A large and a small kernel are then derived by rounding the estimated pixel number up and down, respectively. These kernels are applied to the cloud- and snow-free dataset, and the CM proxy formulations are calculated. The position of the M pixels is identified by those with the highest Spearman correlation between the CM proxies and the in-situ observations. Surface reflectance proxies are then obtained by applying the corresponding formulations.

- Uncalibrated M: In case observed data are not available in the same period of the satellite overpasses, uncalibrated procedures are used to derive M pixels. According to the data availability two different procedures are applied. The first one is applied when altimetry data are available and consists in calibrating M pixel position using interpolated altimetry measurements. The second one, based on Filippucci et al. (2025a), is applied in the other cases and relies in selecting as M all the pixels in the water mask area which are lowly correlated with C, V and W. Surface reflectance proxies are hence obtained by applying the relative formulations.

2.5.5. Output product

The main output is a time series of river discharge proxies for each selected reach. This product is used as input for river discharge. The results are provided as netCDF files, one for each basin, containing the date, the coordinates of each river reach, the river name, the station name (or code), the satellite products used, and the time series of surface reflectance proxy estimates derived from all products, together with the relevant metadata.

2.6. Water level virtual stations

2.6.1. Short description

Water level timeseries from satellite altimetry are provided by DGFI-TUM through the Database for Hydrological Time Series of Inland Waters (DAHITI) at so-called virtual stations where the satellites' nadir ground tracks intersect the river centrelines. For EO4FLOOD, water level data from seven active altimeter satellites are provided through DAHITI: Sentinel-3A and -3B (S3A/B), Sentinel-6A (S6A), and the Surface Water and Ocean Topography (SWOT) mission, with a 10 to 27-day repeat interval, and Jason-3, Cryosat-2, and Saral, which are currently (07/2025) on longer repeat orbits with little use for monitoring at virtual stations. S3A/B and S6A use nadir SAR altimeters, limited to observations

along the nadir ground track, and SWOT uses a wide-swath altimeter capable of measuring water levels within a 120 km swath.

DAHITI water level time series data usually have a latency of more than a week, caused by the latency of the satellite data (NTC; non-time-critical) and of the models used to derive the measurement corrections. For EO4FLOOD, S3 NRT was implemented to reduce the latency to one day and thus improve flood monitoring, but using mission-specific corrections, which can lead to slightly lower accuracy. In general, the accuracy of the DAHITI data in rivers based on the S3A/B, S6A, and SWOT missions ranges from 5 to 20 cm RMSE compared to in situ gauges.

2.6.2. Flowchart

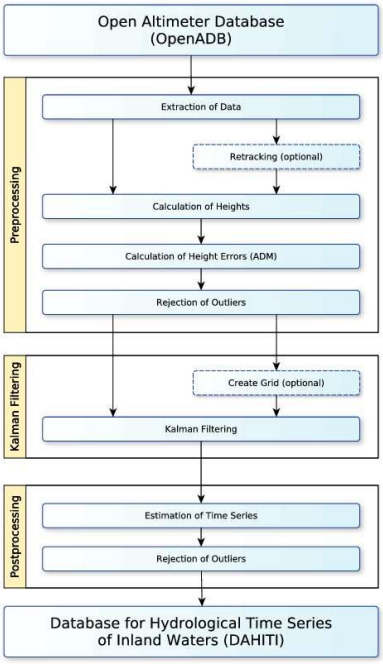

Figure 2.6.1: Processing strategy for the computation of water level time series for inland waters in DAHITI from Schwatke et al. (2015).

2.6.3. Algorithm description

DAHITI extracts altimeter data from the Open Altimeter Database (OpenADB, Schwatke et al., 2023) and applies the improved threshold retracker, an advanced outlier rejection, and Kalman filtering (Schwatke et al., 2015). Additionally, DAHITI applies consistent corrections for atmospheric delays and geophysical effects across the different satellites and, additionally, corrections for systematic orbit errors estimated with the Multi-Mission Crossover Analysis (MMXO) by Bosch et al. (2014). In this way, observations from different satellites can be combined.

2.6.4. Output product

DAHITI provides water level timeseries data in ASCII, JSON, CSV, and netCDF format containing metadata such as target name, coordinates, and applied geoid, alongside columns of date, time, water level [m], and error [m].

DAHITI provides an API to retrieve virtual station time series and metadata. The API is documented at https://dahiti.dgfi.tum.de/en/api/doc/v2/ with examples and can also be accessed via the python3-dahitiapi package (https://pypi.org/project/dahitiapi/). Using the API requires an API key available after free registration for DAHITI. The most significant functions for EO4FLOOD are the *list-targets, download-water-level*, and the *get-raw-water-level* calls:

list-targets lists all DAHITI virtual stations, optionally filtered by basin name, country code, continent name, or coordinate boundaries.

download-water-level downloads the water level time series of a virtual station specified by the requested DAHITI ID in ASCII, JSON, or netCDF format.

get-raw-water-level was specifically created for EO4FLOOD. Instead of the resulting water level time series, it returns the coordinates, mission, cycle and pass number, WSE, and WSE uncertainty of each altimetry measurement used to derive the water level time series, which are usually multiple observations per cycle. This data is used to create the multi-mission water level time series described in the next section.

2.7. Multi-mission Water level time series

2.7.1. Short description

The multi-mission approach is a way to obtain temporally densified water level time series by merging observations of water level, located along a river reach of up to ~120 km, from several altimetry missions. Here we apply observations from the missions; CryoSat-2, SARAL/AltiKa, Sentinal-3A/S3B, ICESat-2, Sentinel-6, and SWOT. The data merging is done via s state-space model (described below). The model output is grid of predicted water levels and associated standard deviations as a function of space (chainage) and time. Additionally, a water level time series is reconstructed at the location of the provided gauge station.

2.7.2. Mathematical background

In this section, we describe the mathematical formulation to derive the gridded multi-mission water level. Here, we assume that a water level observation is a function of 1D space (chainage along the river center line) and time. The observation equation is presented in Eqn. (1) below. Here ω is a Gaussian Markov random field accounting for the spatial and temporal correlation (with the dimension N space steps x M time steps) defined as an AR1 process in both the space and time dimensions (see Eqn. (2)), τ is a cubic spline (forced to be increasing in the upstream direction) representing the average river water level profile, β is a bias to account for the potential elevation difference between the satellite mission, and ε is the noise term assumed Gaussian distributed. The model is implemented via the R-package RTMB (Kristensen et al., 2016) which applies the Laplace-approximation to integrate out the unobserved random effects.

$$h_{s,t} = \omega_{s,t} + \tau(s_x) + \beta(sat) + \epsilon_{s,t} \quad \epsilon_{s,t} \sim N(0, \sigma^2_{obs}) \tag{2.7.1}$$

$$\omega \sim N(0, \sigma^2_p \Sigma) \quad cor(\omega_{s,t}, \omega_{s',t'}) = \rho_s^{\Delta s} \rho_t^{\Delta t} \tag{2.7.2}$$

2.7.3. Flowchart

Figures 2.7.1 and 2.7.2 present the workflow of the multi-mission water level time series approach. The left-hand side describes the integration of the altimetry water level observations with the SWOT river database (SWORD) (Altenau et al., 2021). The individual steps are described in more detail in the next subsection below.

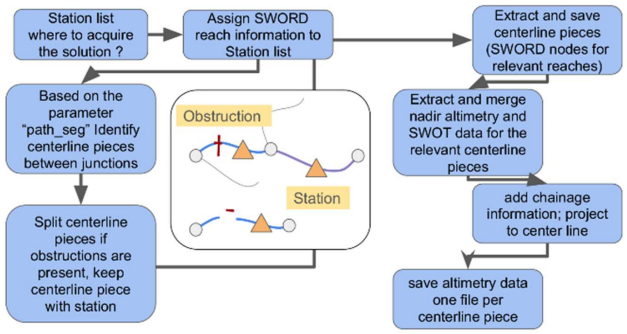

Figure 2.7.1: Multi-mission altimetry water level integration with SWORD

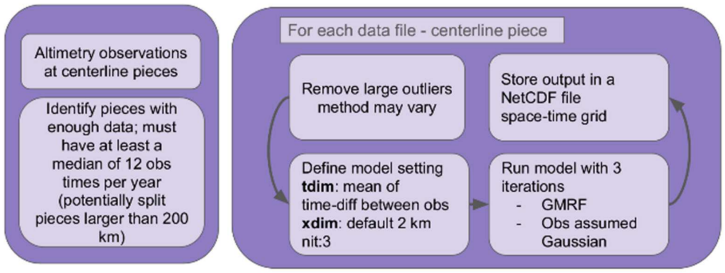

Figure 2.7.2: Filtering and model workflow

2.7.4. Algorithm description

Integration with SWORD

1. The starting point is a station list with coordinates of where the model should be predicted. This would typically be at locations where there is in-situ data of e.g. river discharge or water level for comparison.

2. As the altimetry is organized in a database indexed based on the river reach ID of SWORDv17

it is necessary to identify the SWORD reach ID where the station is located.

3. As the model (see Eqn. (1) and (2)) assumes a positive correlation in the spatial dimension (and temporal dimension), a continuous water surface elevation profile is preferable. To identify connected reaches, we use the SWORD parameters "*path_segs*" or "*path_order*". Where "*path_segs*" is an ID for reaches between river junctions and *"path_order"* is an ID for reaches on a given river branch (see the small illustration in the flowchart).

4. Next, we use the SWORD parameter *"obstr_type"* to identify reaches with an obstruction i.e. a dam or a waterfall. All the reaches may no longer be connected, and we only keep the fraction of the connected reaches where the station of interest is located (see the small illustration in the flowchart).

5. From the SWORD nodes related to the reaches of interest, we extract the node coordinates and selected attributes and save these as the centerline.

6. From the identified reaches of interest, we can extract the nadir altimetry and the SWOT RiverSP data and merge it to a common altimetry product.

7. Lastly, the altimetry data are projected to the local centerline. For each station of interest, we now have a local centerline file and a data file with water surface elevations from satellite altimetry.

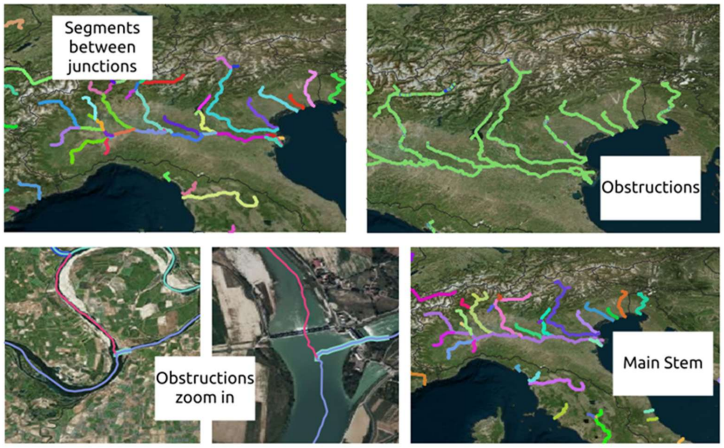

Figure 2.7.3: An example of the SWORD parameters "path_segs", "path_order", and "obstr_type" illustrated over the Po River.

Filtering and model workflow

1. Not all connected reaches are suitable for the multi-mission approach. Some reasons are that the length of the connected reaches is too short (< 10 km), only few nadir mission data besides SWOT are available, or the river is narrow (< 100 m) resulting in only few measurements. Hence, we require that there must be at least a median of 12 observation times pr year (2016-2025) to apply the muti-mission approach.

2. Before running the model, outlier filtering is performed.

    - DEM filter – remove crude outliers based on the SWORD DEM
    - Occurrence filter – The Occurrence percentages from the Global Surface Water Product (Pekel et al., 2016) is used to remove observations potentially over land. In general, the Occurrence threshold is set to 90%, but for narrow rivers or meandering rivers it is necessary to set a lower threshold to ensure enough data.
    - SWOT RiverSP data filter – can be noisy or erroneous (phase unwrapping errors or if islands are present). The SWOT RiverSP data are first filtered based on various parameters in the product (dark_frac < 0.5, xovr_cal_q < 2, wse_u < 0.2, node_q_b <= 2097152, and ice_clim_f!=2). To eliminate remaining outliers et water surface profile is generated for each time stamp and observations outside a fixed distance w.r.t the profile is discarded. All observations to a given time stamp are discarded if too few observations are available or if the profile has a negative slope.
    - Final filter - To remove obvious persistent outliers. A fixed upper and lower boundary, e.g., 5 m (this is case specific) around the mean river profile. Observations outside are discarded.
    - Seasonal filter (sometimes necessary) - often many outliers at low flow – knowledge about seasonality is necessary.

3. The number of space and time steps in the output grid must be defined. Here we aim at a space step size of 2 km. The time step size will depend on the available data. We use the mean time difference between the observations as the time step size.

4. The model is now run iteratively three times where observations related to the lower tail of the model residuals, are down weighted in the next iteration to reduce the influence of remaining outliers.

2.7.5. Output product

The model output is a grid of reconstructed water surface elevations as a function of chainage (distance along the centerline) and time. To make the solution usable outside the local coordinate system the longitude and latitude of the chainage points are also provided. As an addition to the grid a water level time series at the station location is also provided.

2.8. Water surface slope based on the multi-mission water level time series

2.8.1. Short description

Time series of surface water slope is generated directly from the grid of water level time series

2.8.2. Mathematical background

The water surface slopes are calculated directly for the multi-mission water level time series (the model is described in Section 2.7.2). The water level timeseries are reconstructed on a grid with time and chainage in each direction. Hence, the slope time series can be reconstructed by subtracting to columns in the grid, two water level time series separated by a defined distance and then divide by the distance.

2.8.3. Flowchart

Please see the flow chart for the water level time series (Figure 2.7.1 and Figure 2.7.2)

2.8.4. Algorithm description

Please see Section 2.7.4.

2.8.5. Output product

Time series of water level slope at equidistance locations along the river centerline.

2.9. River discharge from water level, width, and reflectance

2.9.1. Short description

Many river gauge stations worldwide are no longer active, and for many of them, the last discharge measurements date back several decades. Satellite observations, however, enable the reconstruction and extension of discharge records at these sites. Variables such as river water height (Section 2.6), river width (Section 2.4), and surface reflectance (Section 2.5) can be used for this purpose and are hereafter collectively referred to as predictors.

In this project, river discharge (Q) is estimated from satellite-based predictors using the stochastic quantile mapping approach of Elmi et al. (2021). Unlike traditional rating curves, which require simultaneous in-situ and satellite observations, this method relates discharge and predictors through their statistical distributions, allowing the use of historical discharge records together with satellite-era data. Being nonparametric, it assumes no predefined model structure, making it robust against modeling errors. Furthermore, by employing Monte Carlo simulations, the approach provides uncertainty estimates and yields a flexible, data-driven relationship between discharge and predictor variables.

2.9.2. Mathematical background

The quantile function $q(\cdot)$ provides a statistical representation of a dataset by relating each data value to its cumulative probability. For a given predictor variable X (e.g., river height, width, or reflectance) and discharge Q, their quantile functions are defined as:

$$q_X(p) = \inf\{X \in \mathbb{R}: p \leq F_X(X)\} \quad (2.9.1)$$

$$q_Q(p) = \inf\{Q \in \mathbb{R}: p \leq F_Q(Q)\} \quad (2.9.2)$$

where $F_X(.)$ and $F_Q(.)$ are the cumulative distribution functions (CDFs) of the predictor and discharge, respectively, and p ∈ [0,1] denotes the cumulative probability. In the stochastic quantile mapping framework, discharge and predictor variables are linked through a monotonic transformation T(·):

$$Q = T(X) \quad (2.9.3)$$

Rather than mapping the raw data directly, the relationship is established between their quantile functions:

$$q_Q(p) = T(q_X(p)) \quad (2.9.4)$$

This quantile-based formulation preserves the rank structure between discharge and predictor variables and does not assume any predefined functional form, making it nonparametric and robust against modeling errors. To account for measurement uncertainty, multiple Monte Carlo realizations of *Q* and *X* are generated. Quantile functions are computed for each realization, forming an ensemble of mapping functions $T_i(\cdot)$. The final discharge estimation model and its uncertainty are obtained taking the mean and standard deviation of the ensemble of mapping functions. The flowchart in Figure 2.9.1 illustrates the workflow for estimating T(·) using the nonparametric stochastic quantile mapping approach.

2.9.3. Flowchart

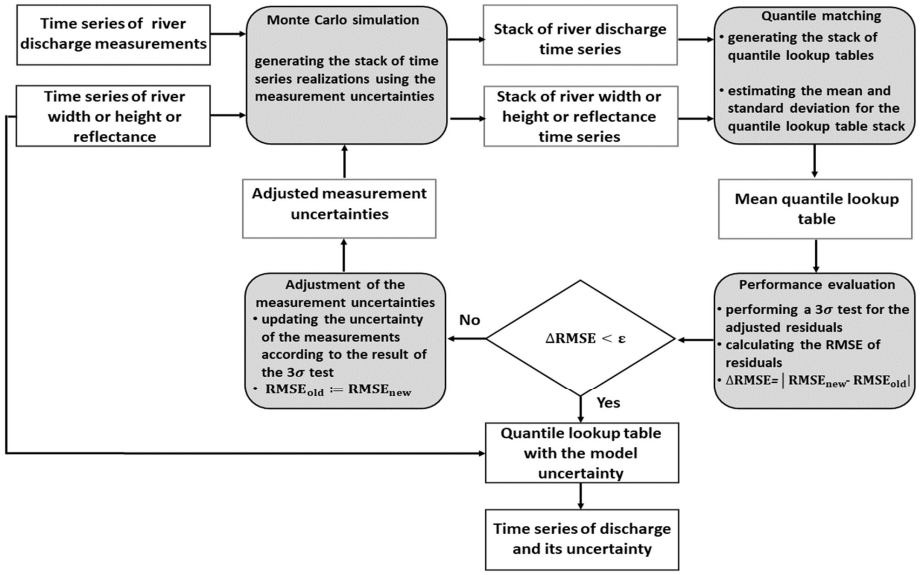

Figure 2.9.1: Flowchart of the nonparametric quantile matching approach for estimating discharge and uncertainty

2.9.4. Algorithm description

The algorithm performs the following steps to obtain the stochastic quantile mapping function (Figure 2.9.1):

- Monte Carlo sampling: Generate multiple Monte Carlo realizations of both the discharge and predictor time series according to their respective measurement uncertainties.
- Quantile mapping function generation: For each realization, derive quantile functions of the discharge and predictor, and construct a collection of quantile mapping functions by matching all possible permutations of their quantiles.
- Mean and uncertainty estimation: Compute the mean quantile mapping function and estimate the uncertainty for each percentile based on the spread of the derived functions.
- Model evaluation: Assess the performance of the derived mapping by comparing the estimated and observed discharge values using a three-sigma test.
- Convergence check: Terminate the algorithm if the variation in RMSE between consecutive iterations is smaller than a predefined threshold, otherwise proceed to the next step.
- Uncertainty update: Update the measurement uncertainties based on the 3σ test results and return to Step 1 for the next iteration.

Afterward, the RS-based discharge estimates are generated using the developed discharge estimation model and satellite-derived predictors (river height, width, or reflectance), along with their corresponding uncertainty estimates. To ensure the reliability of the estimated discharge, a two-step quality control procedure is applied. In the first step, the Kling–Gupta Efficiency (KGE) is calculated to assess model performance. When simultaneously observed and estimated discharge values are available, KGE is computed using the coincident discharge values in both datasets. Otherwise, it is calculated from the mean monthly values of the observed and estimated discharge. Discharge time series with a KGE lower than 0.2 are discarded. In the second step, the remaining time series are visually inspected to identify and remove unrealistic fluctuations or inconsistencies. This process ensures that only high-quality discharge estimates are retained for further analysis.

2.9.5. Output product

The output product includes river discharge estimates with corresponding uncertainties at the temporal sampling of the predictor time series (river height, width, or reflectance).

2.10. River Discharge from the combination of different sensors data

2.10.1. Short description

This algorithm (Scherer et al., 2024) estimates reach-scale river discharge using remote sensing data from multiple satellite missions. It is developed for application in regions without in-situ measurements and combines long-term satellite altimetry data from DAHITI with PlanetScope high-resolution satellite imagery and slope estimates from the ICESat-2 mission (Scherer et al., 2023). Discharge is computed based on reconstructed channel geometry and slope, with the roughness parameter optimized to satisfy mass conservation between adjacent cross-sections. The method is intended for rivers with uniform or gradually varied flow conditions and quantifies uncertainty in the discharge estimates.

2.10.2. Mathematical background

The discharge *Q(t,x)* at time *t* and chainage *x* is computed by summing the partial discharges from n=30 vertical segments of a river cross-section:

$$Q(t,x)=\sum_{i=1}^{n}\overline{v}_i(t,x)\cdot A_i(t,x) \tag{2.10.1}$$

where $v_i(t,x)$ is the depth-averaged flow velocity in segment *i* and $A_i(t,x)$ is the cross-sectional wetted area of the segment. The depth-averaged velocity is estimated using the Gauckler-Manning-Strickler equation:

$$\overline{v}_i(t,x)=k_i(t,x)\cdot R_i(t,x)^{2/3}\cdot I(t,x)^{1/2} \tag{2.10.2}$$

where $k_i(t,x)$ is the Strickler roughness coefficient, $R_i(t,x)=A_i(t,x)/P_i(t,x)$ is the hydraulic radius of the segment with the wetted parameter $P_i(t,x)$, and $I(t,x)$ is the water surface slope. To account for flow depth $d_i(t,x)$ and channel sinuosity $s(x)$, the roughness coefficient is defined as:

$$k_i(t,x)=\frac{k_B(x)}{d\cdot s(x)} \text{ with } d=\begin{cases}\left(\frac{d_i(t,x)}{d_0}\right)^{-\epsilon}, \wedge d_i(t,x)>d_0\\ 1, \wedge \text{otherwise}\end{cases} \tag{2.10.3}$$

The parameters $k_B(x)$, d_0, and ε are determined through optimization by minimizing the discharge differences between all cross-sections:

$$\min_k \frac{max(k_B)}{min(k_B)}\sum_{a=1}^{m}\sum_{b=1,a\neq b}^{m}\left(\frac{1}{\overline{Q}(a)}\sqrt{\frac{1}{l}\sum_{t=1}^{l}(Q(t,a)-Q(t,b))^2}\right)^2 \tag{2.10.4}$$

where is *l* the number of all timestamps *t* in the water level time series, *m* is the number of cross-sections at which the discharge *Q* is estimated, and *a* and *b* are pairs of cross-sections, with the parameter vector:

$$\dot{k}=(k_B(a),\ldots,k_B(m),d_0,\epsilon). \tag{2.10.5}$$

To combine the unsynchronous observations of water levels from satellite altimetry and surface area from optical sensors, the following hypsometric relationship (Schwatke et al., 2020) is fitted to the data:

$$y=\left[\frac{x_{min}-x}{x_{min}-x_{ip}}\cdot\frac{x_{max}-x_{ip}}{x_{max}-x}\right]^{z}\cdot y_{scale}+y_{min} \tag{2.10.6}$$

where y is the water level, x_{min} and x_{max} define the minimum and maximum surface area, y_{min} is the bed elevation, and y_{scale} the water level variation. The parameters describing the shape of the curve (z) and the abscissa of its inflection point (x_{ip}) must be fitted to the data.

2.10.3. Flowchart

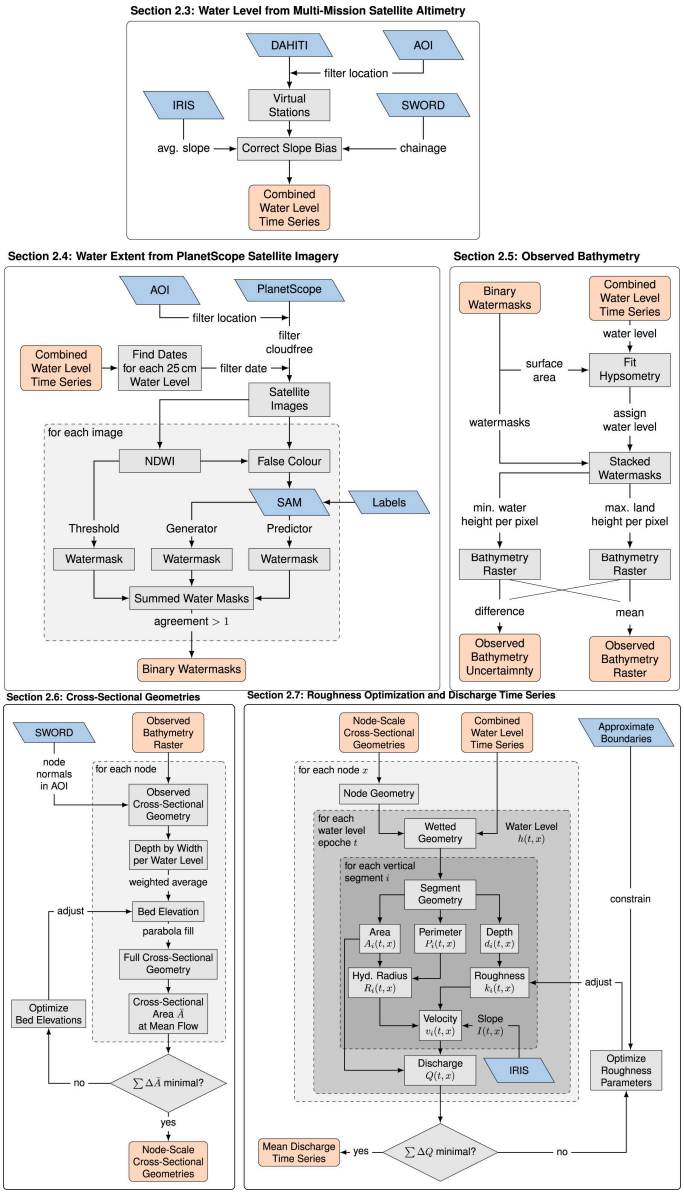

Figure 2.10.1: Flowchart from Scherer et al. (2024) summarizing the steps to obtain the river bathymetry and the optimization of the cross-sectional geometries and roughness parameters to get discharge.

2.10.4. Algorithm description

The approach involves the following main components:

1. Cross-section selection: Geometry is extracted at multiple locations along a river reach using the SWORD database.
2. Water level data: Water level time series are derived from multiple altimetry missions and corrected for slope-induced biases.
3. Bathymetry estimation: A hypsometric relationship is fitted between observed surface areas and water levels to infer riverbed topography.
4. Submerged geometry: Below the observed minimum water level, the geometry is extrapolated using an empirical width–depth relationship.
5. Velocity and discharge: Segment-wise velocities are estimated using the Manning–Strickler formulation. Discharge is computed accordingly.
6. Optimization: An L-BFGS-B algorithm is used to find roughness parameters that minimize discharge inconsistencies across cross-sections.
7. Uncertainty analysis: Variability in input parameters is propagated through the model to define a 90% uncertainty envelope for each discharge estimate.

2.10.5. Output product

The output products are time series of river discharge at the temporal resolution of the input water level data with a 90% confidence interval.

2.11. Multi-mission river discharge

2.11.1. Short description

For each target river reach, river discharge is estimated using all three derived predictors: river height, river width, and reflectance. The temporal resolution of each discharge time series depends on the availability of its corresponding estimates. This section aims to describe the procedure of a multi-mission river discharge time series by merging the discharge estimates derived from all three predictors.

2.11.2. Flowchart

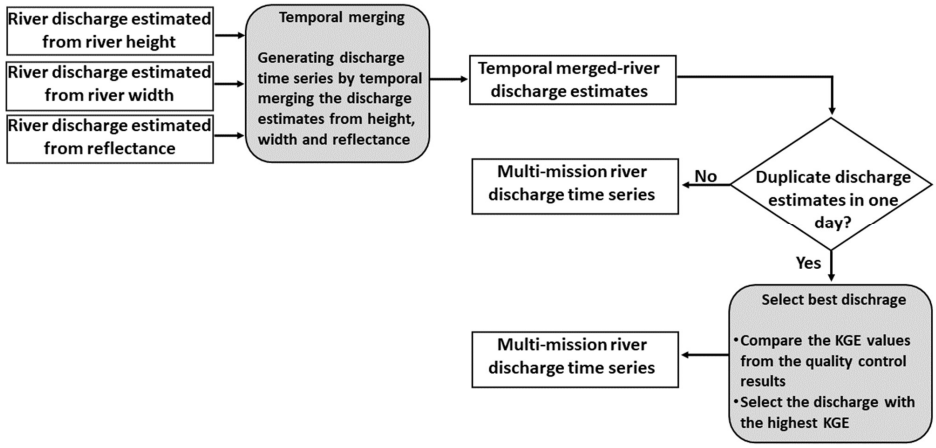

Figure 2.11.1: Flowchart of the procedure for merging discharge estimates derived from river height, width, and reflectance into a multi-mission river discharge time series.

2.11.3. Algorithm description

The algorithm performs the following steps to generate a multi-mission river discharge time series:

- On a daily basis, temporally merge the discharge estimates derived from river height, width, and reflectance to form a unified time coordinate.
- For each date, check for duplicate discharge estimates. If multiple values exist, select the one associated with the highest Kling–Gupta Efficiency (KGE) from the quality control results.
- Generate the final multi-mission discharge time series and its associated uncertainty based on the selected estimates.

2.11.4. Output product

The output is a multi-mission river discharge time series with its associated uncertainty. The discharge values are provided at the temporal resolution of the satellite-derived predictors (river height, width, and reflectance).

2.12. Flood extent from Sentinel-1

2.12.1. Short description

This product provides flood extent mapping for selected flood events using SAR based on Sentinel-1. The flood detection algorithm runs in the Google Earth Engine python API.

2.12.2. Algorithm description

The method implemented is a modified version of the UN-SPIDER recommended practice, in which flood delineation is based on the ratio between two backscatter images (σ^0_E): a reference image (or mosaic) representing dry conditions, and the image to be evaluated. A pixel is considered flooded if the ratio exceeds a pre-defined threshold. The original UN-SPIDER uses a default threshold value that, when exceeded, indicates that water has covered the elements depolarizing the signal and reducing the cross-polarized backscatter. The methodology applied here leverages the same concepts but incorporates the VV channel, improves the flood detection threshold estimation based on soil moisture data for an enhanced flood extent delineation, and defines references images as a mosaic of dry conditions rather than relying on a single image.

It is vital to specify both the reference, and the evaluated images have been acquired with the same scanning mode and the same observation geometry (same orbit track/relative orbit) in order to prevent data artefacts that may originate from differences in the tracks.

Input data

Sentinel-1 mission was designed with two polar-orbiting satellites launched in 2014 (Sentinel-1A) and 2016 (Sentinel-1B). Both satellites carry a C-band synthetic aperture radar (SAR) imager, capable of acquiring images regardless of cloud cover or night conditions. Together, they enabled a repeat image acquisition every 6 days along the same orbital track. However, Sentinel-1B experienced a technical anomaly in December 2021, reducing the revisit time to 12 days. In August 2022 was officially declared non-operational. To restore full mission capacity, Sentinel-1C was launched in December 2024, but it is still not operational.

Flood delineation analysis was conducted using images acquired in interferometric wide swath mode (IW, 10-m resolution), considering all images captured during the period of interest (i.e., flooding events to be analyzed) for each basin.

Additionally, ancillary data layers were used to support the analysis. For instance, monthly average soil moisture from ERA5-Land was used to determine the driest period within the Sentinel-1 temporal range in order to identify the optimal reference period for each basin. The Harmonized Landsat Sentinel (HLS) dataset supported the determination of flood detection thresholds. Furthermore, the GHS built-up surface layer was used in post-processing to mask urban areas and remove false positives, along with the Height Above Nearest Drainage (HAND) layer derived from the Fathom DEM (30 m).

Methodology

The flood delineation methodology is based on the analysis of backscatter signal ratios derived from Sentinel -1 imagery. Specifically, the method compares a reference image, representing typical dry conditions, with an image acquired during or shortly after a suspected flood event. The reference is constructed as a by-track composite (median) of multiple pre-flood acquisitions, reducing reliance on a single date, which may not uniformly represent dry conditions across the area of interest (AOI). This

is particularly important in large heterogeneous basins where some regions may already be flooded while others remain dry during the same period. To ensure geometric consistency and avoid artefacts, both the reference composite and the evaluated image are required to be acquired in the same SAR scanning mode (interferometric wide swath, IW) and from the same orbit track and relative orbit.

Flood detection is based on the ratio between the backscatter values of the evaluated image and the reference composite. A pixel is classified as flooded if this ratio exceeds an empirically defined threshold, which reflects a significant departure from typical surface backscatter. The original UN-SPIDER method relied solely on VH polarization, under the assumption that cross-polarized backscatter is reduced when water covers surface elements that would otherwise depolarize the signal. In this modified version, both VH and VV channels are used to improve floods detection under the canopy.

The methodology distinguishes between two flood types: open-water flooding (smooth water surface) and flooding beneath vegetation canopies. Although their radar backscatter signatures differ, both are considered indicative of flooding. Detection thresholds for each flood type are empirically defined and should be re-evaluated and calibrated for each new Area of Interest (AOI).

Reference mosaic selection

The selection of a reference mosaic representing dry conditions is a critical step in ensuring reliable flood detection. Here, the reference period is determined dynamically for each study area based on climatic and temporal variability. More specifically, the reference period is identified using the minimum average monthly soil moisture data from the ERA5-Land dataset, within the full Sentinel-1 data availability range. This ensures the selected period is most representative of local dry conditions. The length of the reference period varies by catchment, typically ranging from several weeks to a few months, depending on the temporal distribution of dryness.

Table 2.12.1: Sentinel-1 reference periods per basin, identified based on dry conditions using the ERA5-Land soil moisture dataset.

Basin	Reference period
Niger	November (2022) - May (2023)
Congo	June (2024) - August (2024)
Brahmaputra	December (2020) - April (2021)
Torne	September (2022) - October (2022)
Negro	September (2023) - October (2023)

To construct the reference mosaic, a median composite is generated using all Sentinel-1 images acquired within the identified dry period, ensuring consistency in orbit track and acquisition geometry. The median is preferred over a single-date image to reduce noise and mitigate the risk of biased representation caused by anomalous acquisitions. While longer periods offer more observations, they also increase the risk of incorporating flood events, which would result in underestimation (false negatives). The soil moisture-based selection strategy addresses this trade-off by selecting the driest plausible period while minimizing the likelihood of including flood-affected images.

Flood threshold estimation

Threshold selection for flood detection plays a critical role in ensuring the accuracy and reliability of

the derived flood extent. In this methodology, the thresholds for both VV and VH polarization channels are empirically determined for each basin by integrating quantitative analysis with expert visual interpretation.

To support this process, Harmonized Landsat Sentinel (HLS) images, including both Sentinel-2 and Landsat data, were selected for dates close to the identified flooding events. These optical images were visually interpreted to identify flooded and non-flooded areas. The resulting interpreted flood masks were compared with the corresponding areas in Sentinel-1 imagery where pixel-level backscatter ratios (reference vs. evaluated image) were computed.

This comparative analysis was used to assess the distribution of VV and VH ratio values across flooded and non-flooded areas. A percentile-based analysis was then applied to explore a range of potential thresholds, and their performance was evaluated using F1-score metrics to quantify the balance between precision and recall.

The threshold values resulting from this statistical analysis were used as baseline and were further refined using expert judgment. This step ensured that the final thresholds effectively captured flood signatures while minimizing false positives and negatives, particularly in complex land cover types such as urban areas and vegetated floodplains.

*Table 2.12.2: Thresholds used for S-1 flood maps to classify **flood**.*

Basin	Threshold
Torne	Band VH: 0.7 Band VV: 1.29
Negro	Band VH: 0.59 Band VV: 1.41
Niger	Band VH: 0.70 Band VV: 1.36
Congo	Band VH: 0.68 Band VV: 1.29
Brahmaputra	Band VH: 0.60 Band VV: 1.32

Postprocessing

The flood extent maps obtained from the flood detection approach described above were postprocessed to reduce noise and minimize false positives associated with the SAR-based flood detection. The following techniques were applied:

1. **Slope-based filtering**: A slope mask was derived from the DEM (Fathom, 30m) to exclude flood detection in areas with terrain slopes greater than 5%, as they are generally inconsistent with floodwater accumulation.
2. **Hydrologic filtering with HAND:** The Height Above Nearest Drainage (HAND) layer, was derived from the DEM (Fathom, 30 m) and was used to filter out false positive in hydrologically implausible areas. More specifically, flood detections located more than 15 m above the main river channel were removed, as these are unlikely to be hydraulically connected to the flooding event.

3. **Urban area masking:** Built environments are known to produce high backscatter variability in SAR imagery, which often results in false positives during flood classification. For this reason, the Global Human Settlement Layer (GHS-Built-up Surface, 2018, 10 m resolution) was used to mask out urban areas.
4. **Noise reduction:** A rolling window–based local density filter was applied to reduce the salt-and-pepper effect in the flood extent maps. This filtering step smooths isolated pixels and enforces spatial coherence, ensuring that flood detections are consistent with the expected spatial continuity of inundation patterns.

Strengths and Limitations

Strengths

- The workflow can be easily applied to different areas.
- Fully automatic after specifying the area of interest, time periods and thresholds.
- Near-real time monitoring of flood events over large spatial scales
- Cloud independent flood monitoring
- Cloud-processing allows the use of auxiliary datasets to delineate the flood extent (e.g., slope)
- Comparisons are performed by relative orbit, preventing commission errors.
- Suited for most cases. Weak under vegetation and wetland areas
- It is not dependent on a single reference image.

Limitations

- Highly dependent on the threshold selection.
- Some false positives can appear over cropping areas.
- Difficulties of detecting flood in urban or densely vegetated areas
- Difficulties capturing the flood maximum extent due to the Sentinel-1 frequency of acquisition

2.12.3. Flowchart

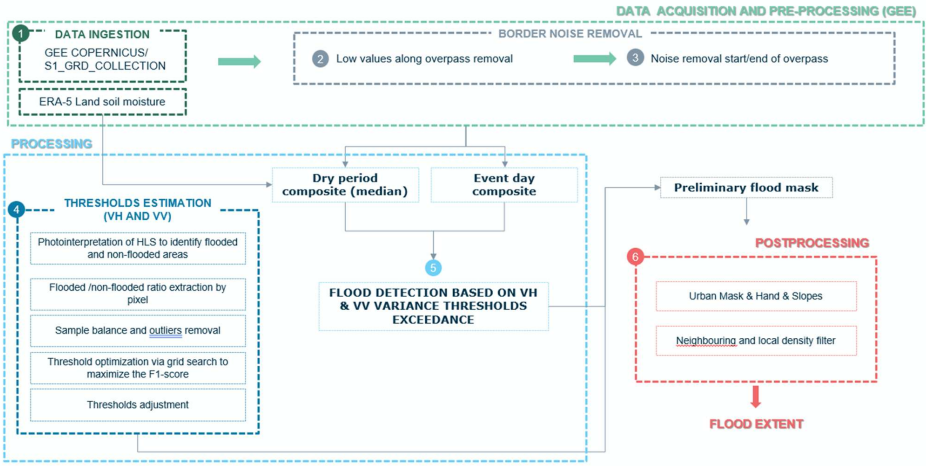

Figure 2.12.1: Sentinel-1 flood data processing workflow (deviation) to derive flood maps.

2.12.4. Output product

The final output of the flood mapping process consists of a series of flood extent masks generated for each available Sentinel-1 acquisition during the flood event considered, within each basin of interest. These products are delivered in GeoTIFF format, containing pixel-based classifications indicating the presence or absence of flooding, based on the results of the VV and VH backscatter ratio analyses. The pixel values in the GeoTIFF files represent the following categories: a value of 0 indicates no flood detected, 1 corresponds to flood detected by either the VV or VH polarization, and 2 indicates flood detected by both VV and VH polarizations.

This classification scheme allows users to distinguish between varying levels of confidence in the flood detection: pixels flagged as 2 represent stronger flood signals detected consistently across both SAR channels, while those marked as 1 may represent more uncertain detections, such as those occurring under vegetation or partial surface coverage.

2.13. Flood extent from Sentinel-2

2.13.1. Short description

The flood extent from Sentinel-2 imagery has been derived by processing the satellite imagery using FloodSENS, the ML-based flood detection algorithm developed by RSS-Hydro.

The MultiSpectral Instrument (MSI) on board the Sentinel-2 satellites acquire imagery across a wide 290 km swath. This broad coverage enables rapid acquisition of extensive areas.

Thanks to the two-satellite constellation, Sentinel-2 offers a high revisit frequency. The Earth is revisited every 5 days at mid-latitudes, with even more frequent revisits at higher latitudes due to orbital overlaps. Individually, each Sentinel-2 satellite completes a revisit cycle every 10 days. Rather than distributing entire swaths, Sentinel-2 products are organised into fixed-size "tiles," also known as granules. Each tile measures 110 km x 110 km. To ensure continuous coverage and simplify the processing of adjacent data, these tiles incorporate a 10 km overlap with neighbouring tiles. This means that while the effective usable area within a tile is 100 km x 100 km, the tile itself is larger to accommodate this overlap. This tiling system standardises data distribution and facilitates efficient data management and processing for users.

Consequently, the flood extent within each of these 110 km x 110 km tiles is precisely analysed by FloodSENS, with the results then symbolised and visualised for interpretation.

2.13.2. Algorithm description

FloodSENS is a machine learning algorithm developed by RSS-Hydro for robust flood detection. It utilises a U-Net architecture enhanced with a Squeeze-and-Excitation (SE) block. This integration significantly improves the model's ability to learn more discriminative features by adaptively recalibrating channel-wise feature responses, which is crucial for accurate flood mapping.

FloodSENS is specifically designed for the semantic segmentation of floodwaters from satellite imagery. Its U-Net foundation enables efficient end-to-end learning, capturing both fine-grained details and broader contextual information. The inclusion of the SE mechanism further refines the feature maps, allowing the model to focus on the most relevant information for flood detection while suppressing less important features.

The algorithm is trained on a combination of EO data, specifically Sentinel-2 imagery, and hydrological information derived from Digital Elevation Models (DEMs). This combined approach of remote sensing and field/hydrological data allows the model to better understand the probability of a pixel being flooded.

Currently, FloodSENS performs inference on Sentinel-2 products, generating a three-class probability GeoTIFF that indicates the likelihood of flood presence.

For both training and inference, FloodSENS requires local execution via a dedicated launcher on a Linux machine, exclusively managed by the RSS-Hydro team.

FloodSENS pipeline follows 3 main steps:

1. Acquisition and preprocessing

Data acquisition for Sentinel-2 products is directly facilitated by the Copernicus API, allowing the download of any available products from its catalogue. From the Sentinel-2 product's extent, the corresponding SRTM DEM is downloaded from the Google Earth Engine API.

Both Sentinel-2 products and DEMs undergo necessary preprocessing steps, including georeferencing, alignment, and a scaling of the resolution, before being utilized. From the DEM, four different hydrological indices are computed:

- Slope
- Height Above Nearest Drainage (HAND)
- Topographic Wetness Index (TWI)
- Flow accumulation

All of these features and bands are then stacked to form the input data for both training and inference.

2. Training

Each FloodSENS model is obtained by training on a proprietary dataset of flood events, which varies in size. RSS-Hydro's dataset of flood events is meticulously classified by expert remote sensing and disaster analysts. These human-labelled samples ensure the highest possible quality for training, leading to robust and accurate models.

Consequently, each model is unique and can be custom-tailored and updated over time to improve performance or adapt to specific requirements.

3. Inference

Model Execution: The stacked input tensor is fed into the FloodSENS U-Net model. The forward pass leverages the learned features through its encoder-decoder structure. The SE block embedded within the U-Net adaptively recalibrates channel-wise feature responses at various depths, enhancing the model's focus on discriminative flood-related patterns and suppressing irrelevant noise. This ensures that features like water absorption in SWIR bands or elevation gradients are optimally weighted.

Output: The output of the U-Net's final layer is a raw logit tensor. This tensor is passed through a softmax activation function to generate a per-pixel probability distribution across the defined classes.

For FloodSENS, this results in a three-channel output, where each channel represents the pixel-wise probability.

2.13.3. Flowchart

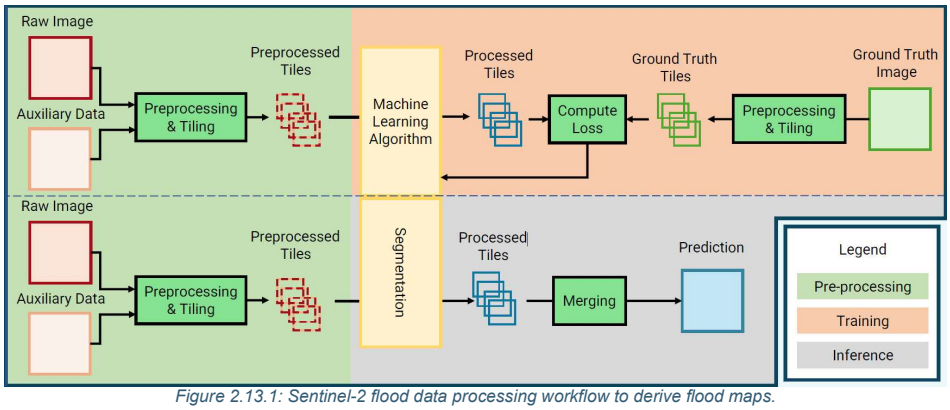

Figure 2.13.1: Sentinel-2 flood data processing workflow to derive flood maps.

2.13.4. Output product

When FloodSENS processes a Sentinel-2 product, the primary output is a GeoTIFF file containing three distinct bands:

- Band 1: Other - This band encompasses all other land cover types that are neither cloud nor water, such as vegetation, bare soil, and urban areas.
- Band 2: Water - This band represents the probability of a pixel being water
- Band 3: Cloud - This band indicates the probability of a pixel being clouds.

These bands are designed to represent the classification results from a machine learning model, categorising each pixel into one of three predefined classes, which were established during the training phase.

For each pixel in the GeoTIFF output, the combined probabilities across these three bands will sum to 100%. This probabilistic distribution reflects the machine learning model's confidence in classifying the nature of that specific pixel.

Example: Consider Pixel A. It might exhibit a probability of 0.03 (3%) in Band 1 (cloud), a high probability of 0.70 (70%) in Band 2 (water), and a probability of 0.27 (27%) in Band 3 (other). This example suggests that the model is highly confident that Pixel A represents water, with minor uncertainties regarding cloud cover or other land types.

This probabilistic output allows users to assess not just the most likely class, but also the model's certainty for each classification.

For data storage optimization the FloodSENS output has been transformed in a single band GeoTIFF format by classifying each pixel into the following categories: value of 1 indicates “other land cover”, pixel with highest % in Band 1, 2 indicates water, pixel with highest % in Band 2, and 3 indicates

clouds, pixel with highest % in Band 3. Lastly the maps have been converted and compressed in Cloud Optimized GeoTIFF (GOG) format.

2.14. Flood extent from VIIRS

2.14.1. Short description

This product provides flood extent delineation based on imagery from the Visible Infrared Imaging Radiometer Suite (VIIRS) sensor.

2.14.2. Algorithm description

The method leverages VIIRS's high temporal resolution to map areas affected by flooding by using cloud-filtered composite products. To improve spatial detail, the native resolution of the VIIRS flood product (375 meters) is downscaled to 30 meters using a spatial disaggregation approach. The resulting high-resolution flood extent masks enable more detailed flood mapping over broad geographic areas.

Input data

The VIIRS sensor is mounted on five platforms of the NOAA Joint Polar Satellite System and provides daily images with 375 or 750 m spatial resolution, depending on the specific band used. The analysis conducted relies on the 5-day composite VIIRS flood products (375 m resolution) provided by the Cooperative Institute for Meteorological Satellite Studies (CIMSS). This 5-day product is generated by aggregating the daily (1-day) flood masks over a rolling five-day window, applying a maximum water-fraction composition approach. This process filters out cloud contamination and captures the maximum observed water extent during the period, making it particularly suitable for use in regions with frequent cloud cover. While the 1-day product is available for real-time applications, the 5-day composite is generally more robust for accurate flood extent delineation.

A key advantage of using VIIRS for flood monitoring is its high temporal frequency, which enables the consistent capture of maximum flood extent thanks to its daily revisit capability. This makes it particularly valuable in dynamic or rapidly evolving flood situations. However, the relatively coarse native spatial resolution of 375 meters can limit its applicability for detailed analyses. To address this limitation, a downscaling process is applied, using the Fathom DEM at 30 m. This results in an improved spatial resolution to the DEM resolution (30 meters), thereby improving the product's utility for local- to regional-scale flood mapping.

Data specifications:

- **Spatial coverage**: Global
- **Spatial resolution**: 375 m (Input) →Downscaled by GMV to 30m (Output)
- **Temporal Coverage**: 2012-present
- **Temporal resolution**: Daily (1-day or 5-days composites)
- **Latency**: A few hours, an operational flood monitoring product can be set-up

Main uses:

The resulting flood product can be used to:

1. **Delineate** specific **flood events** or time intervals. (Processing of specific images, as done in EO4FLOOD project)

2. Create **flood hazard maps**: In this case, the entire time series needs to be processed and calculated the flooding frequency.

3. Set-up an operational **flood monitoring** tool leveraging the short latency of the inputs.

Methodology

NOAA Joint Polar Satellite System and provides daily images with 375 or 750 m spatial resolution, depending on the specific band used. The analysis conducted relies on the 5-day composite VIIRS products (375 m resolution). The VIIRS 5-day composited flood products are recommended for use given that they filter out the cloud cover through a maximal water-fraction composition process, accurately reflecting the maximal flood extent during the last five days.

The VIIRS flood product distinguishes between permanent water and flood water, estimating flood presence through flood fraction values, reported in 1% intervals. Additional classifications include dry land, snow, and ice. The full list of pixel classifications is provided in the accompanying reference table.

ID	Meaning
1	Fill value
15	Floodwater without water fraction
16/17	Clear-sky dry land
38	snow/ice water or mixed water
20	Snow
25	Ice
30	Cloud
50	Cloud shadow
99	Normal open water

ID	Floodwater fraction
100-120	1-20%
121-130	21-30%
131-140	31-40%
141-150	41-50%
151-160	51-60%
161-170	61-70%
171-180	71-80%
181-190	81-90%
191-200	91-100%

Figure 2.14.1: VIIRS raw pixel IDs (left) and Floodwater fraction values (%)(right)

The flood fraction pixel values are used to perform the downscaling in order to obtain a flood mapping product with an improved spatial resolution. The downscaling technique used leverages the concept of height above nearest drainage (HAND), one of the possible derivatives of digital elevation models (DEM). For this purpose, the Fathom 30 m DEM is used to compute HAND, which represents the vertical distance from each pixel to its nearest drainage point (e.g., stream or riverbed). This allows identification of low-lying areas where floodwaters are most likely to accumulate.

To ensure accurate downscaling, areas where more than 1% of VIIRS flood fraction pixels are classified as permanent water are also included in the drainage network. The downscaling is applied to areas where HAND is below 30 meters and the VIIRS pixel is marked as flooded.

For each coarse-resolution (375 m) VIIRS 'parent' pixel, the corresponding 30 m 'child' pixels from the DEM are ranked by HAND value (from lowest to highest). The number of child pixels to be classified as flooded (n) is calculated by multiplying the VIIRS flood fraction (expressed as a value between 0 and 1) by the total number of child pixels within the parent pixel. The n lowest HAND pixels are then labeled as flooded, simulating the concentration of water in topographically lower areas.

Finally, morphological operations are applied to the downscaled flood mask to enhance spatial coherence and continuity, reducing isolated noise and ensuring more realistic flood extent patterns.

Strengths & Limitations:

Strengths

- Peer-reviewed methodology which outputs are used over the whole world.
- The Flood fraction can be used to increase flood mask granularity when combined with a DEM (downscaling).
- Near-real time monitoring of flood events over large spatial scales.
- Daily temporal sampling increases likelihood of capturing maximum flood extent.

Limitations

- It is necessary to download the data before usage or to combine it with auxiliary datasets
- Images are affected by cloud cover that may be partially retained in the 5-day product.
- Low spatial resolution even after the downscaling process.
- Downscaling of permanent water bodies is not possible (no water fraction data), retaining the original resolution.

2.14.3. Flowchart

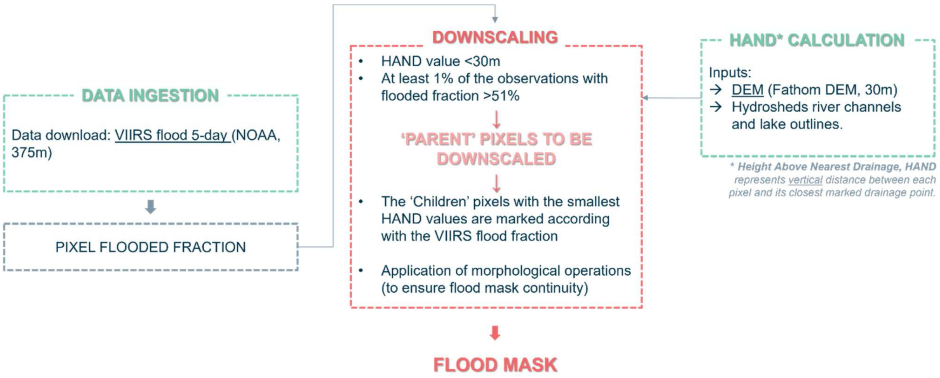

Figure 2.14.2: VIIRS flood 5-day data processing workflow to derive flood maps.

2.14.4. Output product

The final output of the VIIRS-based flood mapping process consists of flood extent masks generated for each flood event and area (defined as the hydraulic model area within a basin), provided in GeoTIFF format. The GeoTIFF outputs preserve the original classification scheme of the native VIIRS product, which includes values representing dry land, snow, ice, permanent water, and flooded areas. However, due to the application of the downscaling process, flood fraction values, originally expressed in percentage intervals (100–200), are replaced by a single standardized value for all flood detections.

In the downscaled product:

- All flooded pixels are assigned a value of 200, regardless of the original flood fraction, to indicate that these pixels have been classified as flooded based on the redistribution of water within the parent 375 m VIIRS pixel.

- All other pixel values, including those representing permanent water, dry land, snow, or ice, are retained as defined in the native VIIRS product. The full classification scheme is provided on the accompanying reference table.

3. Models

3.1. MGB

3.1.1. Short description

The Modelo de Grandes Bacias (MGB) is a physical-based, semi-distributed hydrological model designed to simulate the transformation of rainfall into runoff across various spatial and temporal scales. Developed initially in Brazil by the Institute of Hydraulic Research at the Federal University of Rio Grande do Sul (IPH-UFRGS), MGB is recognized for its adaptability and computational efficiency. It is compatible with both Linux and Windows operating systems, making it accessible for a wide range of users and applications (Collischonn et al., 2007).

MGB has been extensively applied in diverse watersheds across South America, Africa and Asia, including basins as diverse as the Amazon River basin and the Maroni River basin in South America, the Niger River basin, the Congo River basin, the Ogooué River basin and the Chad River basin in Africa, and the Mahanadi and Cauvery river basins in Asia, among several others. Its design is particularly advantageous for regions with limited gauging infrastructure, as it can operate effectively using almost exclusively Earth Observation (EO) datasets. This capability makes MGB an excellent tool for local to regional flood modelling in data-scarce regions (Wongchuig et al., 2023).

The model's structure is built to accommodate a wide range of applications, from local flood simulations to large-scale continental hydrological modelling. Its flexibility and efficiency are key attributes that enhance its suitability in water resource management and flood forecasting. MGB's ability to integrate with near real-time satellite data for monitoring and prediction further underscores its relevance in modern hydrological practices (Paris et al., 2022). The official MGB website (https://www.ufrgs.br/lsh/mgb/what-is-mgb-iph/) provides extensive resources and guidance for further details or accessing the model and its comprehensive documentation.

Recent initiatives have been undertaken to develop real-time hydrological monitoring systems based on MGB fed by EO datasets, and first proofs of concept were deployed on the Amazon (https://www.ufrgs.br/sardim/), on the Niger (https://mgb-hyfaa.pigeo.fr/site/) and on the Maroni (https://sagui.hydro-matters.fr/) river basins.

3.1.2. Mathematical background

The MGB model écombines physical and conceptual equations to simulate hydrological processes on the land surface. It works by dividing a basin into smaller unit catchments, usually named “Mini basins” (following the Maidment method, 2002), which are then further broken down into Hydrological Response Units (HRUs). These HRUs are defined by specific combinations of soil type and land use, allowing the model to account for spatial variability in factors like agricultural practices, reservoirs, and soil properties.

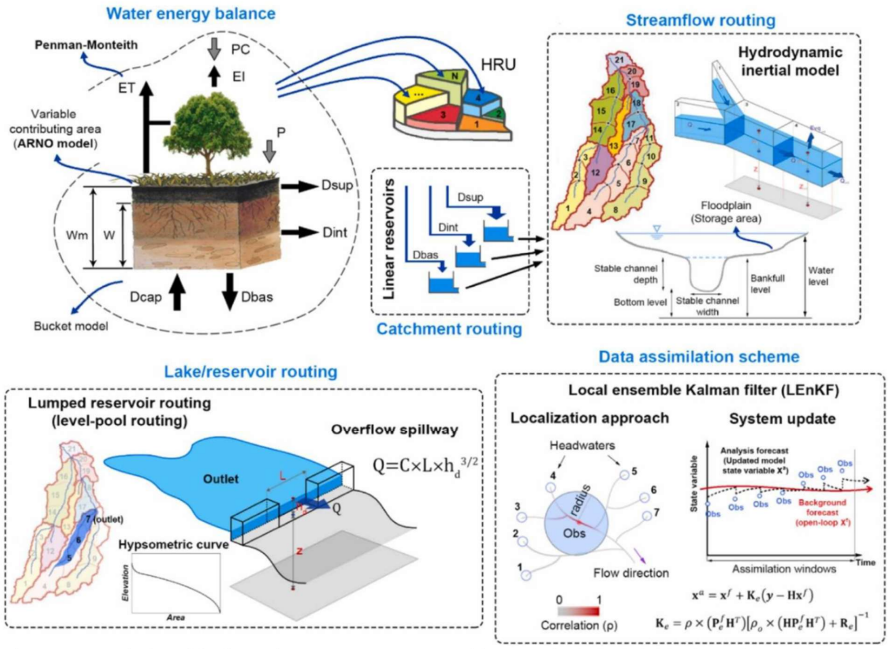

Figure 3.1.1: Model flowchart

Land Surface Model / Vertical processes

The MGB model captures several key vertical hydrological processes:

- Soil Water Budget: is modelled using a bucket approach, which simplifies the soil into a single layer that stores and releases water based on precipitation and evapotranspiration. It's a straightforward way to understand how water is held and moved within the soil.

- Energy Budget and Evapotranspiration use the Penman-Monteith equation to calculate evapotranspiration. This equation considers various meteorological factors to estimate how much water is lost from the soil and vegetation. It's a crucial part of understanding the energy dynamics in the hydrological cycle.

- Interception: converts precipitation into effective precipitation due to plant & canopy interception. It's an important part of the water cycle, as it affects how much water actually reaches the ground.

- Infiltration and Runoff: are calculated based on the idea of variable contributing areas, which considers how different parts of a watershed contribute to runoff under varying conditions. This helps the model simulate how water moves across the landscape and into streams.

Flow routing / Horizontal processes

Flow routing in the MGB model is a critical aspect that determines how water moves through river networks. The model uses methods like Muskingum-Cunge or a simplified version of the 1D Saint-Venant equations, known as the local inertial method. This simplification makes the calculations more efficient while still providing accurate results. The key equations used are:

$$\frac{\partial A}{\partial t} + \frac{\partial Q}{\partial x} = 0 \quad (3.1.1)$$

$$\frac{\partial Q}{\partial t} + gA\frac{\partial h}{\partial x} = gAS_0 - gAS_f \quad (3.1.2)$$

In these equations, Q is the flow rate, A is the cross-sectional area, h is the flow depth, S_0 is the bed slope, and S_f is the friction slope. The t variables and x represent time and space, respectively.

3.1.3. Input

The MGB model is a comprehensive hydrological tool that requires various types of input data to simulate water dynamics effectively within a basin. Each type of data plays a crucial role in ensuring the accuracy and reliability of the model's outputs. Below is a detailed description of the input data needed for the MGB model, including their properties and descriptions. Data are of two kind: dynamic (precipitation, climate, hydrology) and static (land use, soil, topography). The dynamic datasets of precipitation and hydrological variables are provided by WP200, while other variables com from globally available and free of use datasets.

Precipitation Data

Precipitation data are crucial for hydrological modelling. This data needs to have high spatial and temporal resolution to capture the variability of precipitation events accurately. It is fundamental for simulating processes such as runoff and infiltration, providing the primary input of water into the hydrological system. The dataset should include detailed temporal resolution, such as hourly or daily measurements, and spatial coverage that matches the study area, ensuring accurate representation of precipitation patterns and their impacts on the hydrological cycle. MGB can work with in situ datasets (that are, in such cases, interpolated toward the mini-basin centroid to have one precipitation value for each mini-basin and timestep) or EO datasets (in that case, the value for each mini-basin corresponds to the value of the pixel for each timestep). In Africa and South America, we use mostly the GPM products IMERG (available at https://gpm.nasa.gov/resources/documents/imerg-v07-release-notes) and GSMAP (available at https://sharaku.eorc.jaxa.jp/GSMaP/index.htm), that provide NRT precipitated volumes at adequate spatiotemporal scales.

Specifically, in the Niger River Basin we use daily resampling of IMERG-LR, SM2RAIN_ASCAT and GSMAP-Gauge. In the Congo River Basin, we use CHIRPS and IMERG-FR; In the Negro River Basin we use CHIRPS.

Meteorological Variables

Meteorological variables are essential for calculating evapotranspiration. These variables need to have a high temporal resolution to effectively capture diurnal and seasonal variations. Key parameters include temperature, wind speed, humidity, solar radiation, and atmospheric pressure, which are vital for understanding the energy budget and water loss from soil and vegetation. Accurate representation of these variables is crucial for modelling hydrological processes. As of today, hydro-meteorological variables can be taken whether from long-term means (CRU,

https://crudata.uea.ac.uk/cru/data/hrg/) or from atmospheric models outputs such as ERA5.

In the considered basins, daily temperature (variable “2m_temperature”), wind speed (combination of “10m_u_component_of_wind” and “10m_v_component_of_wind” variables), humidity (variable “relative_humidity”), solar radiation (variable “surface_solar_radiation_downwards”) and atmospheric pressure (variable “surface_pressure”) are taken from ERA5 database. A long term mean climatology is also built from CRU dataset (variables relative humidity, sunlight (%), temperature, elevation, wind speed) in order to fill possible gaps or no-data.

Land Use and Soil Type Data

Data on land use and soil types should have a high spatial resolution to accurately capture the variability in land cover and soil types. It is essential for defining Hydrological Response Units (HRUs) and understanding how different land covers and soil types influence water infiltration and runoff. Detailed classifications of land use and soil types are crucial for accurate hydrological modelling.

We use the FAO Harmonized World Soil Database (https://www.fao.org/soils-portal/data-hub/soil-maps-and-databases/harmonized-world-soil-database-v12/en/) for soil types, while land use land cover is taken from Hansen et al. (2022, https://glad.umd.edu/dataset/global-land-cover-land-use-v1) in the Congo River basin and ESA world cover (https://esa-worldcover.org/en/release-worldcover-map-2021) in the Niger and Negro river basins. Both are resampled to native resolution of topographic data.

Topographic Data

Topographic data, such as Digital Elevation Models (DEMs), should have a high spatial resolution to effectively capture the topographic features of the basin (around 30m). These data are crucial for understanding water flow and accumulation, and for delineating watersheds. A combination of high accuracy and adequate spatial resolution in DEMs ensure that topographic features, such as floodplains, are captured effectively, providing a detailed representation of the terrain for hydrological modelling.

MERIT-Hydro (Yamazaki et al., 2014, http://hydro.iis.u-tokyo.ac.jp/~yamadai/ MERIT_ Hydro/) is used for the Congo and Niger river basins, while bare-earth SRTM (O’loughlin et al., 2016) is used in the Negro River basin.

Hydrological Data

Hydrological data, obtained from historical records at gauging stations and satellite observations, should have a high temporal resolution to capture variations in river flow and water levels. This data are crucial for calibrating and validating the model, helping to set up initial conditions and validate model outputs. Time series data for streamflow and water levels are essential for accurate simulation and validation of hydrological processes. In poorly gauged situations, in situ hydrological data can be substituted by EO datasets. Here, WL from altimetry plays a crucial role, since they provide a distributed view of surface water dynamics.

3.1.4. Output product

MGB outputs are linked to the automatically extracted discretization of each basin (based on DEM), and are associated to linear elements (e.g. streamflow on a reach) or to polygon elements (e.g. evaporation in a mini-basin). We provide a detailed look at the various outputs of the MGB model. There are two kind of products, which can be associated to two different specifications: point-wise products

(streamflow and water level) and spatialized products which are gridded-like (water budget variables). They are provided for each mini basin and for each date (day), so the spatial resolution depends on model's discretization. Mini basins usually are of few hundreds of squared kilometers.

The outputs of MGB include:

- River discharge timeseries in m^3/s at basin outlets along river network (.csv &/or .nc). Product type (PT): "q"
- Base Flow in m^3/s for each elementary catchment (.nc). Product type (PT): "bf"
- Evapotranspiration in mm/timestep for each elementary catchment (.nc). Product type (PT): "et"
- Total water storage in mm for each elementary catchment (.nc). Product type (PT): "tw"

Simulated streamflow

The MGB model produces simulated streamflow data, which is crucial for understanding water dynamics within a watershed. This data represents the calculated water flow at the outlet of each of the smaller unit catchments which corresponds to different points in the river network. Simulated streamflow data can be generated at different time intervals depending on the specific needs of the application and on the precipitation temporal resolution in input.

Water level estimations

Hydrological data, obtained from historical records at gauging stations and satellite observations, should have a high temporal resolution to capture variations in river flow and water levels. This data are crucial for calibrating and validating the model, helping to set up initial conditions and validate model outputs. Time series data for streamflow and water levels are essential for accurate simulation and validation of hydrological processes.

Water budget variables

The MGB model can provide valuable information on several key hydrological variables, each computed at the scale of smaller catchment areas, referred to as mini-basins. These variables include:

- Soil Moisture: This variable represents the water content present in the soil across various points within each "Mini" catchment. Soil moisture is crucial for understanding processes such as infiltration and runoff. It is computed as a mean value for each "Mini" catchment, with the timestep for these calculations depending on the temporal resolution of the input precipitation data.

- Evapotranspiration: This variable accounts for the total water transfer from the land surface to the atmosphere, encompassing both soil evaporation and plant transpiration. Evapotranspiration is vital for assessing the water budget within a watershed. Like soil moisture, it is also calculated as a mean value for each "Mini" catchment, with the timestep aligned with that of the precipitation data input.

- Groundwater Storage: This variable indicates the volume of water stored in underground aquifers within each "Mini" catchment. It plays a significant role in influencing processes such as aquifer recharge and discharge. Groundwater storage is also computed as a mean value for each "Mini" catchment, with the timestep matching that of the precipitation input data.

Each of these variables is essential for a comprehensive understanding and analysis of the hydrological conditions and dynamics within a watershed, aiding in effective water resource management and

environmental assessments.

3.2. DHI-GHM

3.2.1. Short description

The DHI-GHM produces historical, operational and forecasted predictions of runoff and discharge globally and at catchment and grid level. The distributed rainfall-runoff models include basin-scale hydrological processes and produce important hydrological outputs at the grid and catchment level for the requested time frequency. The model has been benchmarked against other GHMs and provides a detailed representation of the physical catchment, with a runtime of less than one hour for the entire globe at comparable spatial and temporal resolution. Additionally, DHI-GHM can provide forecasts at 10 km2 resolution, updated hourly, which is highly relevant for extreme event warning systems. The model domain covers 80°N to 60°S and is divided into 0.1° x 0.1° grid cells, with over 1.5 million grid cells covering the earth's terrestrial surface. The benchmarked model is forced and parametrized with global datasets for catchment characteristics (topography, land use, soil type) and climate datasets (rainfall, temperature and potential evaporation). Rainfall-runoff processes are simulated using distributed Nedbørs-Afstrømning Models (NAM, DHI A/S, 2022) at grid level, which are then connected at catchment level with a fast Kinematic River Router. Routing is performed at grid level (removing runoff from grid edge), basin level (moving runoff to the hydrological catchment outlet) and along river networks (routing between catchments to the sea). The model structure is conceptually illustrated below and the full model structure, validation and comparison to alternative GHMs is detailed in Murray et al. (2023).

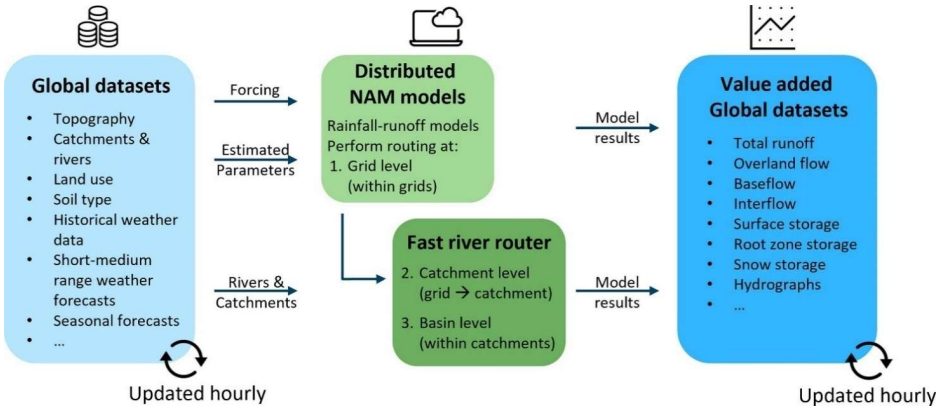

Figure 3.2.1: DHI Global Hydrological Model structure from Murray et al. (2023)

3.2.2. Mathematical background

Nedbørs-Afstrømning (NAM) rainfall runoff model

The NAM model is a well-proven deterministic, lumped, conceptual model that has been applied to many catchments with different hydrological regimes and climatic conditions around the world. NAM simulates the rainfall-runoff process by continuously accounting for the water content in four different and mutually interrelated storages that represent different physical elements of the catchment. Based on the meteorological input data NAM produces catchment runoff as well as information about other

elements of the land phase of the hydrological cycle, such as the temporal variation of the evapotranspiration, soil moisture content, groundwater recharge, and groundwater levels. The resulting catchment runoff is split conceptually into overland flow, interflow and baseflow components. A brief description of how water content is accounted for in each storage is presented below and a more detailed description can be found within the Mike1D reference guide:

https://manuals.mikepoweredbydhi.help/latest/Water_Resources/MIKE_1D_reference.pdf

- Snow storage
    - Controlled by the temperature conditions and the current amount of snow on the catchment surface. Snowmelt is transferred to surface storage. In the DHI-GHM, elevation zones are established that distribute the temperature vertically and distribute the snow. This is for mountainous areas with large elevation variation within grid cells.
- Surface storage
    - Represents moisture interception on the vegetation, as well as water trapped in depressions and in the upper-most cultivated part of the ground. The amount of water in the surface storage is continuously diminished by evapo-transpirative consumption as well as by horizontal leakage (interflow) in the upper ground level. In situations when there is a maximum surface storage achieved, some of the excess water will enter the streams as overland flow, whereas the remainder infiltrates into the lower root zone, which can in turn continues as percolation into the groundwater storage.
- Root zone storage
    - Defined as a ground layer below ground surface from where the vegetation can draw water for transpiration. Moisture in this zone is subject to consumptive loss from evapotranspiration. This process starts when the surface storage is completely exhausted, and in that situation the actual evaporation depends on the relative contents of the moisture in the lower zone. Additionally, the amount of moisture in the lower zone influences the process of overland flow and amount of water that enters the groundwater storage as recharge.
- Groundwater storage:
    - Filled from root zone storage and contributes to baseflow.

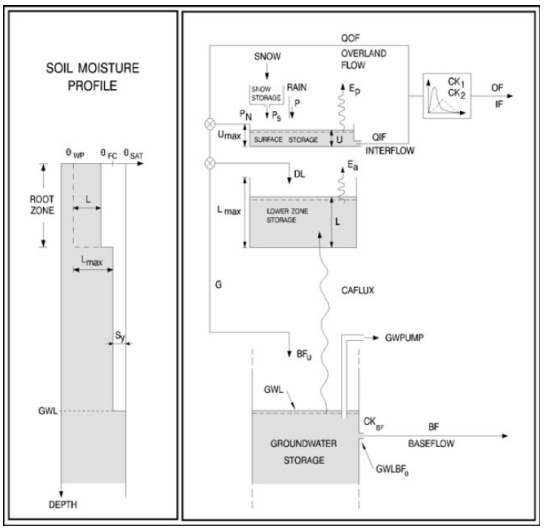

Figure 3.2.2: Conceptual representation of NAM model from Mike1D Reference Guide

Kinematic fast river router

The kinematic router within the GHM is designed to route water from individual grid cells, each with their distributed NAM model, to its representative sub-basin outlet and then downstream between sub-basins for sub-daily, real-time simulation and short-range forecasting. The router leverages the Manning formula for open channel flow assuming a triangular cross section in the river and calculating the river slope from the DEM. A relationship is established between the NAM parameters describing overland flow in the basin and the manning coefficient. This relationship allows the manning coefficient to be adjusted individually for each basin and further refinement of the model. Withdrawal for irrigation is also included within the router.

3.2.3. Input

The DHI-GHM requires various global spatial/landmass datasets, meteorological data, and observed discharge data used for model creation, forcing and performance evaluation. The model can handle any resolution of input data, though data are gridded to 0.1-degree resolution, matching the resolution of each distributed NAM model. A more detailed description of how each dataset is used can be found in (Murray et al., 2023).

Spatial/landmass Data

DHI-GHM requires that catchment characteristics are defined across several topographical, hydrologic and geographical datasets. These types of datasets and their sources are briefly outlined in the Table 3.2.1.

Table 3.2.1: Required datasets for the DHU-GHM model

Purpose	Dataset

Catchment delineation and characteristics	HydroBASINS (Level 12 basins used) (Lehner and Grill, 2013)
Reservoir/Lake delineation and characteristics	HydroLAKES (Includes lakes greater than 10km2 and on the main river channel) (Messager et al., 2016)
River delineation and characteristics	Global Rivers Classification (GloRiC) (Dallaire et al., 2019)
Land cover types and mask	Moderate Resolution Imaging Spectroradiometer (MODIS) land cover products MCD12Q1.006 and MOD44W.006 (Carroll et al., 2017, Friedl et al., 2019) Annual International Geosphere-Biosphere Programme (IGBP) land cover classification from 2017 Global Forest Cover Change Water Cover 2000 (Townshend, 2016)
Elevation data	Void-filled 90 m STRM Digital Elevation Database v4.1 (Jarvis et al., 2008) MERIT Hydro 3 arc-second resolution raster river network hydrography (Yamazaki et al., 2019)
Soil data	Depth to bedrock dataset (Shangguan et al., 2017) HYSOGs250m soil groups dataset (Ross et al., 2018) HiHydroSoil v2.0 data layers (Simons et al., 2020)
Irrigation data	Water withdrawal time series from Huang et al. (2018)

Meteorological Data

DHI-GHM requires meteorological data providing precipitation, temperature and potential evaporation. For hindcast forcings and model evaluation data from the Copernicus European Centre for Medium-Range Weather Forecasts Reanalysis 5th Generation (ERA5) dataset (Hersbach et al., 2020) is typically used. Precipitation and temperature forecasts for operational short term discharge forecasts are sourced from the IBM Weather Company database (IBM, 2018a) while long term ensemble forecasts are sourced from NCEP/NCAR Reanalysis 1 dataset for seasonal forecast simulations (Kalnay et al., 1996).

Observed river discharge data

Observed river discharge data from various sources (for example from the Global Runoff Data Centre (GRDC)) can be injected into the DHI-GHM for calibration and validation of the results.

3.2.4. Output product

DHI-GHM produces gridded hydrological data including runoff, overland flow, interflow, baseflow, snow coverage, actual evaporation, and soil moisture at 0.1°×0.1° resolution for the globe from 60°S to 80°N. River discharge outputs are also produced for each HydroBASINS level 12 sub-basin outlet point. Data can be produced as a historical simulation at an hourly timestep, short to medium-range forecast at hourly timesteps or seasonal forecasts at daily timesteps.

3.3. HYPE

3.3.1. Short description

The Hydrological Predictions for the Environment (HYPE) is a continuous process-based model, which simulates components of the catchment water cycle and water quality (Lindström et al. 2010). The model is a semi-distributed conceptual model, in which a river basin may be subdivided into multiple subcatchments, which are further subdivided into homogeneous hydrological response units (HRUs) based on combined soil type and land use classes. It can run at daily or hourly time steps. Model outputs are normally computed at the outlet of the sub-catchments.

HYPE has been implemented operationally for several applications from national to global scales. A Swedish national setup of the model is employed as a national operational tool for hydrological applications, including flood warning. A pan-European model setup of HYPE is also employed as an operational tool for seasonal hydrological forecasting of European rivers, and a global setup of the model is employed as an operational global flood forecasting and seasonal hydrological forecasting tool. More details on the operational use of HYPE can be obtained at https://hypeweb.smhi.se/

3.3.2. Mathematical background

HYPE is a process-based model and it employs conceptual routines for most of the major land surface and subsurface processes (snow/ice accumulation and melting, evapotranspiration, surface and macropore flow, soil moisture, discharge generation, groundwater fluctuation, aquifer recharge/discharge, irrigation, abstractions and routing through rivers, lakes and reservoirs, solid matter, and dissolved nutrient pools). It also offers options for selecting different model routines for some of the processes, such as evapotranspiration and snowmelt, depending on availability of data and local conditions. The processes are controlled by a number of parameters that generally need to be estimated through calibration. Most of the parameters are linked to catchment physiography, which allows parameter regionalization over a large domain or in ungauged catchments by calibrating the model for selected representative gauged basins (RGBs).

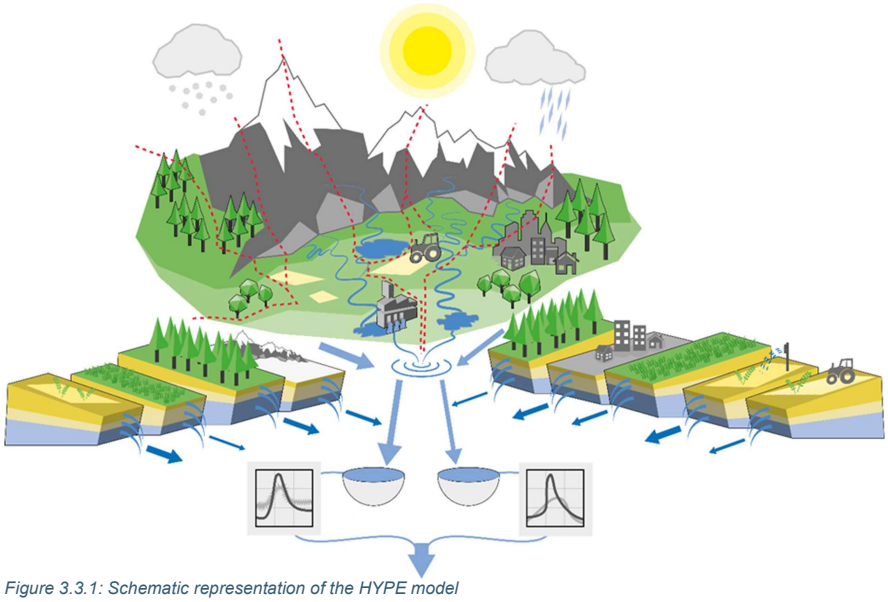

Figure 3.3.1: Schematic representation of the HYPE model

Snow accumulation and melt

HYPE offers different options for the snow accumulation and melting process. The simplest option is implementation of the degree-day method with land use dependent parameters. Precipitation accumulates as snow when the temperature is below the threshold, and snowmelt ensues when the temperature is above the threshold. The other option is a temperature and radiation index model, where snowmelt is additionally driven by radiation. The potential snowmelt due to radiation is computed from shortwave radiation and albedo of the snow. In this later method, snow cover scaling may also be applied to adjust the snowmelt.

Land processes

The soil zone can be subdivided into up to three layers and a fraction of the rainfall or snowmelt infiltrates into the topsoil layer, which is limited by a soil type dependent maximum rate. If the soil moisture in the upper soil layer exceeds a threshold for macropore flow, part of the remaining water forms macropore flow. Another part of the remaining water is transformed into surface runoff. The remaining water forms a surface pool and overland flow. Water percolates from upper to lower soil layers when the soil moisture in the upper layer exceeds the field capacity, and the rate is determined using a soil type dependent percolation rate. The ground water level is estimated based on the level in the soil zone where the pore space is filled.

Evapotranspiration

Evapotranspiration from the soil is assumed to occur from the upper two layers, and the potential evaporation is assumed to decrease exponentially with depth. The actual evaporation from the soil is limited by the availability of water in the soil above the wilting point. Evaporation is at potential rate only if the actual soil moisture exceeds a large proportion (lp) of the soil field capacity and for soil moisture below this limit, the actual evapotranspiration decreases linearly to zero at wilting point. HYPE offers different options for the computation of potential evapotranspiration, which includes: a simple temperature index method, a modified Jensen-Haise method, a modified Hargreaves-Samani method, the Priestly-Taylor method, and the FAO Penman-Monteith model.

Flow routing

The discharge generated in a subcatchment is routed through each subcatchment and between subcatchments using a river routing routine which simulates flow attenuation and delay using a flow wave and delay parameters, respectively. If lakes and reservoirs are present within a subcatchment, the flow is routed in the lake or reservoir using a rating curve with two parameters.

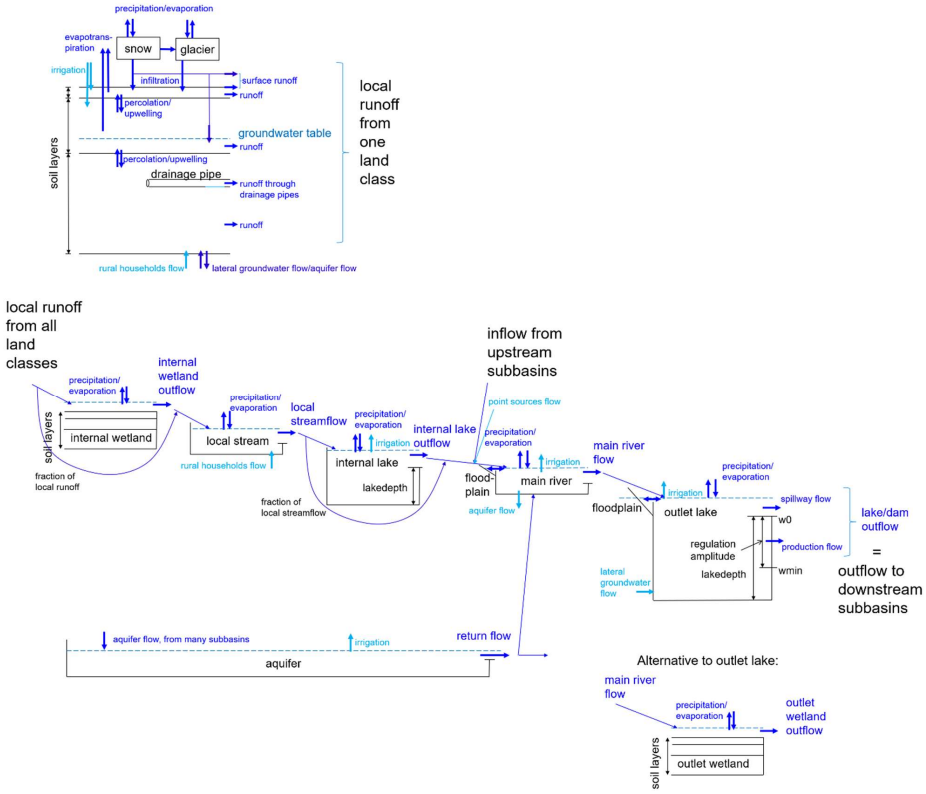

Figure 3.3.2: Structure of the HYPE model

3.3.3. Inputs

HYPE requires a number of physiographic data, which are employed for model setup and parameter estimation; meteorological forcing, and river flow or other catchment response data for model calibration and validation. There is no fixed requirement on the temporal and spatial resolution of input data or the size of the sub-catchments in the model, but both temporal and spatial resolution are important for the model's ability to predict the hydrological phenomena of interest and the objectives of the simulation. Typically, daily data are used both for meteorological input and other data such as discharge and water level. For the EO4FLOOD use cases, the aim for the hydrological modelling with HYPE is first to resolve the locations of the selected stations along the river with EO based discharge data - this means that model sub-catchment outlets need to be located at these points. Secondly, the flow direction and sub-catchment delineations should preferably be aligned with the DEM used to set up the hydraulic model (Lisflood-FP). The EO4Flood models are mainly based on the WW-HYPE model, which is consistent with the MERIT Hydro data (~90x90m), while the Lisflood-FP models will be set up using the FABDEM 30x30m data. Thus, special care will be taken to obtain spatial consistency when using discharge predicted by HYPE as input to the Lisflood-FP models, as well as when using EO based water level data to generate stage discharge relationships for HYPE sub-

catchment outlets.

Meteorological data

Precipitation and temperature are the main meteorological inputs required for HYPE. Typically, daily sums of precipitation and mean daily temperature are required, but daily maximum and minimum temperatures can also be used depending on the routines selected for different processes. Hourly precipitation and temperature may be required for modeling the impact of high intensity short duration precipitation. Representative values of precipitation and temperature for each sub-catchment are required, and they are typically estimated through interpolation of station observations or other gridded products at the centroid of each sub-catchment. Daily precipitation and temperature of the global HydroGFD data set (Berg et al. 2021), which is based on ERA5 reanalysis bias adjusted against gridded global observations (GCPP, CRU), will be used. In addition, gridded daily precipitation and temperature data synthesized from SMHI ground observations across Sweden will be used for the Torne catchment.

Physiographic data

Digital elevation models are required for delineation of sub-catchments and derivation of slope and elevation distributions within sub-catchments. DEM from National laser scanning was used for the Torne river S-HYPE setup, while the MERIT Hydro DEM was used for the WWHYPE setup that is used for the Niger and Negro River models.

Spatial soil and land use information are required for derivation of hydrological response units (HRUs) within sub-catchments. Land use data from the Land Cover Project of the ESA Climate Change Initiative (CII) as well as soil data from the FAO Harmonized World Soil Database are used in the derivation of the HRUs in the WWHYPE model, while data from the Swedish Land survey is used in the S-HYPE model.

Data sources on lake and reservoir and as well as irrigated areas can also be used as input. The HydroLAKES database is used for the global WWHYPE model setup. For Sweden, the SVAR database, developed by SMHI and the Swedish Landsurvey is used.

Hydrological data

Hydrological response data are used for model calibration and validation. In-situ discharge data at a temporal resolution corresponding to the model's temporal resolution at the outlets of one or several subcatchments is typically used. Other hydrological data, such as lake water levels as well as snow water equivalent, snow cover and evapotranspiration, have been used in the calibration of the WWHYPE model as well as the S-HYPE models. To a large extent, the in-situ river discharge data collected for the EO4FLOOD project have also been used in the previous calibrations of the S-HYPE and WWHYPE models. Nevertheless, it will be interesting to investigate the impact of the new satellite-based data when used in model calibrations and assimilation experiments, compared to the previously used in-situ data. In contrast, the satellite- based water level data in lakes and rivers, and the Copernicus snow data products have not been used as input to the HYPE model in earlier studies.

3.3.4. Output product

HYPE produces a number of output variables for each subcatchment. These include time series of simulations and observations (for each time step or averaged/summed over a longer period) as well as model performance results, where there are observations against which model performance can

be assessed. The output variables include, among many others: local runoff, river discharge, river and lake levels, snow depth, and snow water equivalent. More descriptions of the HYPE model can be found at http://hype.smhi.net/wiki/doku.php.

In general, the output can have the same or lower temporal resolution as the input data, and the same or lower spatial resolution as the model sub-catchments. However, for outputs calculated at the hydrological response unit scale, such as snow depth, soil moisture, evaporation, it is also possible to downscale the model output to the scale of the underlying data used to delineate the hydrological response units – which in most cases mean the resolution of the land cover data. For the Torne river models, the resolution of the land cover data are 25x25m and for the models used in the Niger it is 300x300m. With regard to the EO4Flood use cases, snow cover, snow water equivalent, and soil moisture will be downscaled using this technique before comparison with the EO based data. On the other hand, the in-stream variables, river discharge and water level are defined in the model at the sub-catchment outlets, and not along the river reaches as in the hydraulic models.

3.4. Hybrid-AI

3.4.1. Short description

The AI framework leverages physics-informed deep learning to integrate large-scale Earth observation (EO) data while maintaining physical consistency. The hydrologic and hydraulic models are trained against SWOT pixel cloud data, using the output of the hydrologic model as the input to the hydraulic model. Joint training allows both models to benefit from the information contained in the SWOT data (WSE, WSS), and, potentially, satellite earth observations of additional state variables (e.g., soil moisture, evapotranspiration, terrestrial water storage). A conceptual flowchart is presented below.

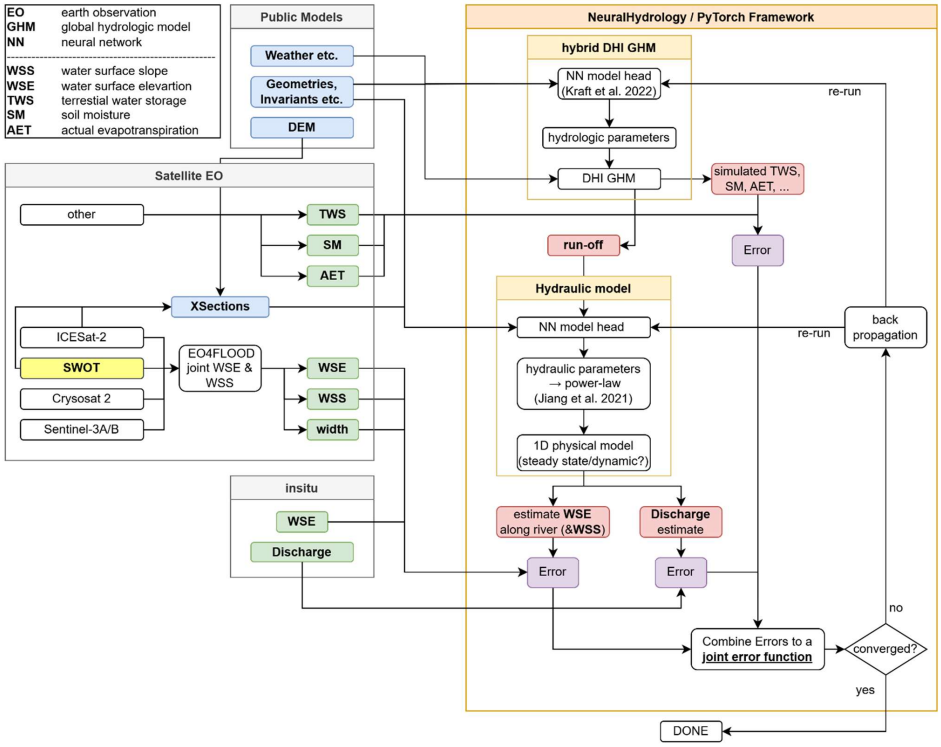

Figure 3.4.1: Model flowchart

3.4.2. Mathematical background

The model structure is twofold. Both hydraulic and hydrologic model consist of a LSTM model head estimating the physical parameters required by the attached physical models – the NAM on the hydrologic side (see above for more information on the NAM) and a steady state solver on the hydraulic side. To allow for joint training, all model parts are implemented in a differentiable manner (e.g. in PyTorch) to utilize backpropagation. The training loss is gathered at different points in the model and combined to a joint loss function to be minimized.

The hydraulic parameters are estimated via a power-law relationship as shown in Jiang et al. (2021), while the estimation of the hydrologic parameters follows the concept presented in Kraft et al. (2022).

3.4.3. Input

Both respective models require different inputs. The hydrologic model is forced with all available weather data as well as terrain information, basin extends, and other invariants. The model expects them to be handed over as .nc files that are converted to tensors to be ingested by the model. The hydraulic model requires static information about the cross section, which is to be taken from ICESat-2 or topographic LIDAR depending on user choice and local availability.. Additionally, the hydraulic model is to ingest the hydraulic model's runoff estimate. Here, no user input is required.

3.4.4. Output product

The joint model outputs different variables (red in flowchart).

- WSE timeseries (.csv) at river node resolution
- WSS for a given time stamp (.csv) at river node resolution
- Discharge timeseries (.csv) at river node resolution
- Run-off (.csv) as a 0.1° grid (analogous to DHI-GHM)
- Terrestrial water storage, soil moisture, actual evaporation (.csv) as a 0.1° grid (analogous to DHI-GHM)

3.5. MCP probabilistic forecast

3.5.1. Short description

The Model Conditional Processor (MCP) is a Bayesian approach used for estimating the predictive uncertainty (PU) defined as the probability density of a discharge outcome conditional on the available information provided by hydrological-hydraulic models. The main characteristics of MCP are presented in what follows, while for additional details the reader is referred to Barbetta et al. (2017). MCP was originally developed for predictive uncertainty estimate to support the flood forecasting activities in real-time and it is essentially based on the definition of the multivariate conditional distribution, i.e., the density of the predictand (water level, discharge, etc.) conditional on multiple deterministic model predictions (predictors).

3.5.2. Mathematical background

The mathematical basis of the multi-temporal approach of the Model Conditional Processor (MCP-MT) is here presented. MCP was first derived in single-temporal approach and is essentially based on four steps (Figure 3.5.1):

1. the observations, y, and the model estimates, $\hat{y}_k$, (k=1…M, M=number of estimates of M different models), are converted into the normal space using the Normal Quantile Transform (NQT). The variables y and $\hat{y}_k$, whose empirical cumulative distribution functions are computed using the Weibull plotting position, are converted to their transformed values, η and $\hat{\eta}_k$, respectively, which are normally distributed with zero mean and unit variance. The probability of each element of η and $\hat{\eta}_k$ is the same as its original corresponding y and $\hat{y}_k$.
2. In the normal space, the joint probability distribution of observed and estimated variables, $f(\eta,\hat{\eta}_k)$ is assumed to be a Normal Multivariate Distribution.
3. The predictive density is obtained by applying the Bayes Theorem: $f(\eta|\hat{\eta}_k) = \frac{f(\eta,\hat{\eta}_1,\ldots\ldots,\hat{\eta}_M)}{f(\hat{\eta}_1,\ldots\ldots,\hat{\eta}_M)}$

   where the numerator functions are the marginal probability distributions of the predictors. The predictive density is normally distributed.
4. The PU in the normal space is finally reconverted to the real space through the Inverse NQT.

In the above-described process the use of the Weibull plotting position implies the need of using an additional model to fit to the tails of all the variables, i.e. the observations and the model forecasts, in

the real space. The best tail model is identified considering the actual distribution of the data (*Coccia and Todini, 2011*).

This approach is applicable for single and multi-model cases, i.e. *M*=1 or *M*>1, respectively, and provides the probability of exceeding fixed water levels, a fundamental information within Flood Forecasting and Warning Systems based on hydrometric thresholds. Specifically, the probability to exceed a threshold, *a*, can be directly derived from the knowledge of the PU as the integral above the threshold in the normal space (Figure 3.5.1, left side).

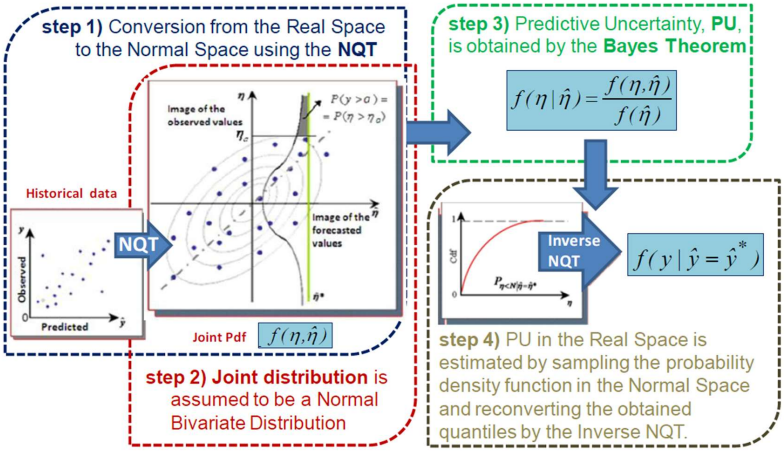

Figure 3.5.1: Scheme of the four main steps of the Model Conditional Processor

3.5.3. Input

MCP application requires as input the forecasted discharge values provided by different forecasting models. The forecast time series must be characterized by the same lead-times, i.e. 1 day, 2 days forecasting horizons, etc.) and time resolution (i.e. daily, hourly, etc.). Moreover, a long enough observed discharge time series is required for calibration and validation purposes. If flood forecasting is of interest, the calibration and validation series must consist of data from flood events periods. The outputs of the hydrological-hydraulic models applied in EO4FLOOD (i.e. MGB, DHI-GHM, HYPE, AI, MIKE, HEC-RAS, DassFlow, LISFLOOD-FP) are used as input for MCP. The different forecasted discharge values provided by the models, applied for each case study, at specific river cross-sections represent the forcing for the statistical investigation. The analysis is developed at selected river cross sections where the water level/discharge is monitored and where recorded time series are available for MCP calibration. Obviously, the forecasts of the hydrological/hydraulic models have to be provided for the same monitored periods.

3.5.4. Output product

MCP application provides as output the density of the predictand conditional on the deterministic predictions. For each lead-time and for each time step, MCP provides as output the expected value of the identified multi-variate conditional distribution, the cumulated 0.05 probability quantiles (5%, 10%, . . . , 95%) and the probability of discharge thresholds exceedance within the considered lead-

time. Specifically, MCP provides probabilistic forecasts at gauged cross-sections by exploiting the models' future estimates at those sites and the observations (historical discharge time series). In order to develop a multi-model approach, all the used forecasted discharges (coming from different models) must be characterized by the same time resolution (e.g., hourly data, daily data, etc.) and the same lead times (e.g., the forecasts have to be provided with the same advance time period, such as 1 day, 2 days, 12 hours). The outputs of MCP consist in probabilistic forecasts (mainly the expected value and the upper and lower limits of the 90% uncertainty band are provided) and are computed at selected river cross sections where the calibration of MCP can be carried out considering the historical time series availability.

3.6. Mike+

3.6.1. Short description

The MIKE+ software portfolio is widely applied in research and commercial projects for detailed operational river modelling and flood forecasting with high accuracy. It includes comprehensive functionalities to model large complex river and channel networks, lakes, and reservoirs, as well as river structures, such as gates, sluices, and weirs. DHI's MIKE 1D hydrodynamic model engine (also known as MIKE 11 or MIKE HYDRO River) is used in numerous real-time flood forecasting systems and detailed river operations and planning studies. The MIKE HYDRO River model can be set up to describe all flow paths of the river network sufficiently accurate to enable detailed mapping of floods at and near the rivers and is able to solve the fully dynamic Saint-Venant equations. 1D channels/rivers in MIKE+ can be coupled to a fixed grid or a flexible mesh 2D component for the flood plains. Couplings are possible for natural channels and flood plains as well as structures and river upstream or downstream ends. The 2D Overland Module in MIKE+ is used for modelling 2D free-surface flows – previously known as MIKE FLOOD. It allows simulation of hydraulic phenomena on land surfaces, overland water bodies, as well as coastal areas. The 2D Overland Hydrodynamic (HD) module simulates water level variations and flows in response to a variety of forcing functions. The combined 1D-2D modelling system is used for the flood forecasting modelling framework.

3.6.2. Mathematical background

1D Hydrodynamic Engine

The 1D engine solves the problem of water flowing through a network of reaches, nodes, and structures and is fully specified by boundary conditions at the network boundaries and initial conditions. The engine solves the one-dimensional Saint-Venant equations which can be derived from the Navier-Stokes equations under the assumptions that water is incompressible and homogeneous, the bottom slope of pipes and canals is small, wave lengths are large compared to the water depth and there is no acceleration of water perpendicular to the flow direction of a reach. The Saint-Venant equations are only valid for free surface flows. Since pressurized flow in closed reaches is an important part of the 1D problem space, an approximate treatment of pressurized flow has been adopted. To solve the Saint-Venant equations, a finite difference scheme is implemented. This solution scheme is only accurate for sub-critical flows (Froude number less than 1). For supercritical flows (Froude number larger than 1), a simpler approximation where non-linear terms are suppressed provides an accurate flow description. Finally, it should be noted that the solution scheme does not include a detailed description of hydraulic jumps. However, flow conditions both upstream and downstream of a hydraulic jump are accurately described. More information can be found within the Mike 1D Scientific manual:

https://manuals.mikepoweredbydhi.help/latest/Water_Resources/MIKE_1D_reference.pdf .

2D Overland Flow Model

The model is based on the numerical solution of the two-dimensional shallow water equations. Thus, the model consists of continuity and momentum equations. The spatial discretization of the governing equations in conserved form is performed using a cell centered finite volume method. The time integration is performed using an explicit scheme. Here either a first-order Euler method or a second-order explicit Runge-Kutta scheme can be applied. The interface convective fluxes are calculated using an approximate Riemann solver. This shock-capturing scheme enables robust and stable simulation of flows involving shocks or discontinuities such as bores and hydraulic jumps. More information can be found within the Mike 21 FM Scientific Manual:

https://manuals.mikepoweredbydhi.help/latest/Coast_and_Sea/MIKE_21_Flow_FM_Scientific_Doc.pdf .

3.6.3. Input

Static inputs defining the physical characteristics of the model domain are necessary for a proper result. The specific input data used, varies on a case-by-case basis depending on the availability of local, regional and global datasets. These include definitions of the 1D and 2D domains and information on how to couple these domains, cross-sections defining the river geometry within the 1D domain, information on hydraulic structures affecting 1D and 2D flows, a digital terrain model to define the topography of the 2D domain. If available information on surface cover type, infiltration properties etc. can also be included. For the EO4FLOOD project, Land Cover data will be used from the freely available 2020 global dataset provided by the Copernicus land monitoring service at 10 m resolution (https://doi.org/10.2909/602507b2-96c7-47bb-b79d-7ba25e97d0a9) to provide information on surface cover type and to estimate infiltration properties. Boundary conditions should be set for inflow and outflow of the model domain based on observed or simulated discharge timeseries. Finally, information on simulation parameters and settings must be provided to establish the temporal and spatial resolution as well as computational efficiency of the model. Most vector and raster data formats can be ingested into Mike+.

3.6.4. Output product

Mike+ can produce several standard product types for flood analysis. These include simulated water levels at all points and timesteps within the model's domain, maps of flow rates along the river, flow velocities in both the river and the floodplain and inundation maps. For a more detailed description of outputs, refer to the Mike+ user guides. Products will be produced at the spatial resolution of the available DEM for the model domain, likely 30 meters, and at an hourly temporal resolution.

3.7. HEC-RAS

3.7.1. Short description

HEC-RAS is a simulation software designed for computational fluid dynamics, specifically tailored to model the hydraulics of water flow through natural rivers and other open channels. HEC-RAS is engineered to simulate various hydraulic phenomena, including the development of water surface profiles, the evaluation of flow depth within stream channels, and the assessment of flow velocities. The software can effectively model a full network of stream channels, including dendritic systems or single river reaches, providing flexibility for diverse geographical contexts. It supports the simulation of subcritical, supercritical, and mixed flow regimes, and is capable of analyzing the hydraulic effects of

various structures such as bridges, culverts, weirs, and stream junctions on flow dynamics and floodplains. Among others, HEC-RAS finds extensive application across various facets of water resources engineering and management such as: a. floodplain Management, b. flood Insurance Studies, c. flood risk assessment, d. hydraulic structure design and testing, e. levee and topography modification studies, f. dam breach analysis.

3.7.2. Mathematical background

The simulation capabilities of HEC-RAS are underpinned by a rigorous set of governing equations and numerical methods. The choice of equations and solution algorithms varies depending on whether the flow is one-dimensional (1D) or two-dimensional (2D), and whether it is steady or unsteady. Considering the nature of the flood forecasting system based on producing the evolution of the flood propagation in a pre-defined temporal horizon, only the mathematical framework concerning the unsteady simulations is discussed.

Governing Equations for Water Flow

- *1D Flow Equations*
  HEC-RAS solves the full, dynamic one-dimensional Saint-Venant Equations. These equations constitute a system of hyperbolic partial differential equations derived from the fundamental principles of conservation of mass (the continuity equation) and conservation of linear momentum (the momentum equation). The 1D Saint-Venant equations are designed to incorporate the main characteristics of the channel cross-sectional shape and include terms that account for bed slope, friction slope (which is related to flow velocity through empirical formulas like Manning's, Darcy-Weisbach, or Chézy's), and pressure forces.

- *2D Flow Equations*
  the 2D flow routing is based on two primary equation sets: the Shallow Water Equations (SWE) and the Diffusion Wave Equations (DWE). These equations are derived by depth-integrating the Navier-Stokes equations under the critical assumption that the horizontal length scale of the flow is significantly greater than its vertical length scale. While the SWE considers all terms of the equations, leading to a high hydrodynamic fidelity, the DWE offer the computational advantage of being faster and more stable in their execution by omitting the local acceleration and convective acceleration terms.

Numerical Methods and Solution Algorithms

- *1D Flow Scheme*
  HEC-RAS utilizes an implicit, finite difference method. Specifically, it implements the four-point implicit scheme.

- *2D Flow Scheme*
  HEC-RAS utilizes an implicit finite volume algorithm. For combined 1D/2D modelling scenarios, the respective solution algorithms are tightly coupled on a time step by time step basis.

3.7.3. Input

As every numerical model, the accuracy and reliability of HEC-RAS simulations are directly contingent upon the quality and comprehensiveness of the input data provided. The input data can be classified in 3 main blocks:

Geometric Data

Geometric data defines the physical characteristics of the river system and its associated floodplain, and it is composed of: river cross-sections, digital terrain model, roughness and infiltration layers, and hydraulic structures.

Geometric data are determined per each model. However, to maintain coherence between the models, each model will be built on a common DTM and land cover source, meaning FATHOM DEM 30m and ESA WorldCover 2021 10m resolution.

Unsteady Flow Data

Unsteady flow data are necessary for performing 2D unsteady flow simulations and it is composed of: boundary conditions (e.g., upstream inflows, downstream outflows), initial conditions (e.g., flow and water surface elevation at the commencement of the simulation).

Unsteady flow data ingested in the models is derived from the outcomes of the hydrological models and EO data products (water level and discharge) over the period identified for the analyses.

Computational Parameters

Computational parameters are crucial for controlling model stability and accuracy during unsteady flow computations. These include the computational time step, time frame, the maximum number of iterations, Theta implicit weighting factor and the equation sets to be used.

Computational parameters are calibrated per each model to optimize the numerical stability.

3.7.4. Output product

HEC-RAS can generate a diverse array of outputs, ranging from fundamental hydraulic parameters to advanced geospatial post-processed products, which are crucial for comprehensive hydraulic analysis.

The primary simulation outputs used for understanding water flow behavior within a modeled system are: water surface profiles, depth of flow, velocity distributions, flow and stage hydrographs at the cross-sections of interest.

Beyond the basic hydraulic results, HEC-RAS can produce advanced geospatial and analytical products such as: inundation maps, shear stress, floodway encroachment data.
The simulation results can be exported either as derived geospatial data, such as in the form of flood inundation maps, or as tabular and graphical reports.
The outputs produced by each model, and their specification, are listed in the document called EO4FLOOD_D2.3_PSD in section 3.15.

3.8. DassFlow 1D/2D

3.8.1. Short description

DassFlow (Data Assimilation for Free Surface Flows) 1D and 2D are two computational codes built for hydraulic modelling of free surface river flows. The kernel codes are written in Fortran and wrapped in Python, therefore enabling the easy use of other Python libraries and is under development by the MathHydroNum team. Both DassFlow 1D and 2D have demonstrated capabilities of assimilation of

large heterogeneous datasets to estimate large spatio-temporal arrays of flow model parameters. In their studies, Pujol et al. (2020), Malou et al. (2021), and Larnier et al. (2025) successfully coupled DassFlow with MGB on respectively the Negro River and the Niger River. MGB was used as boundary condition for upstream discharge and lateral tributaries. Stage discharge rating curve proposed from Paris et al. (2016) were used to estimate river bathymetry. This existing coupling framework between MGB and DassFlow ensure that, in the frame of this activity, we will manage to improve the existing experiment and deploy the necessary tools in order to achieve project's objectives and provide an efficient framework for flood modelling and forecasting.

3.8.2. Mathematical background

DassFlow 1D and 2D embed variational data assimilation methods. DassFlow 1D has been specially developed to exploit SWOT and other EO data with variational data assimilation technique as showcased on large river networks portions (Malou et al., 2021). Moreover, it is the core of the HiVDI discharge algorithm, one of the reference SWOT discharge estimation algorithms (Larnier et al., 2021). DassFlow 2D enables fine 2D hydraulic modelling over structured or unstructured meshes of floodplains, also in multi-D at river network scale with zooms on floodplains, as well as variational data assimilation. DassFlow 2D is HPC ready and thus enables effective scalability in forward and backward mode.

3.8.3. Inputs

DassFlow 1D and 2D models require static information such as topographic data, land use and soil type data as well as boundary conditions for flood modelling and forecasting. The topographic data should be high spatial resolution Digital Elevation Models (DEMs) improved with river bathymetry when available. Boundary conditions are required such as upstream discharge and lateral tributaries and/or precipitation data. Computational parameters including time step or model mesh size are finally crucial to control the simulations.

DassFlow 1D model inputs

- geometry: river network (GIS linestrings dataset), cross-sections locations (GIS points dataset), ideally at 50m to 500m resolution,
- Upstream boundary conditions: timeseries of discharge (CSV file) for each upstream boundary condition, hourly to daily timestep.
- Downstream boundary condition(s): timseries of water level (CSV file), hourly to daily timesteps.

DassFlow 2D model inputs

- geometry: topography, i.e. DEM with riverbed breaching (GeoTIFF file), ideally at resolution ranging from 10m to 100m,
- Buildings/structures: shapefiles of polygons (buildings) and linestrings (structures such as levees, roads, etc.),
- Upstream boundary conditions: timeseries of discharge (CSV file) for each upstream boundary condition, hourly to daily timestep.

- Downstream boundary condition(s): timeseries of water level (CSV file), hourly to daily timesteps. This is optional as a free outlet boundary condition is implemented.

3.8.4. Output product

DassFlow 1D model generates one output file that contains all the hydraulic variables (depth, wse, width, discharge) for each cross-section and for each output timestep (typically 3h). This output file can either be generated in CSV or netCDF format.

DassFlow 2D model generates multiband rasters in GeoTIFF format. The bands in this rasters correspond to the hydraulic variables: height (also sometimes refered as depth), velocity vedctor components, norm of the velocity vector). The resolution of the output grid is by default equal to the DEM resolution or can be coarser if needed for computational cost reasons.

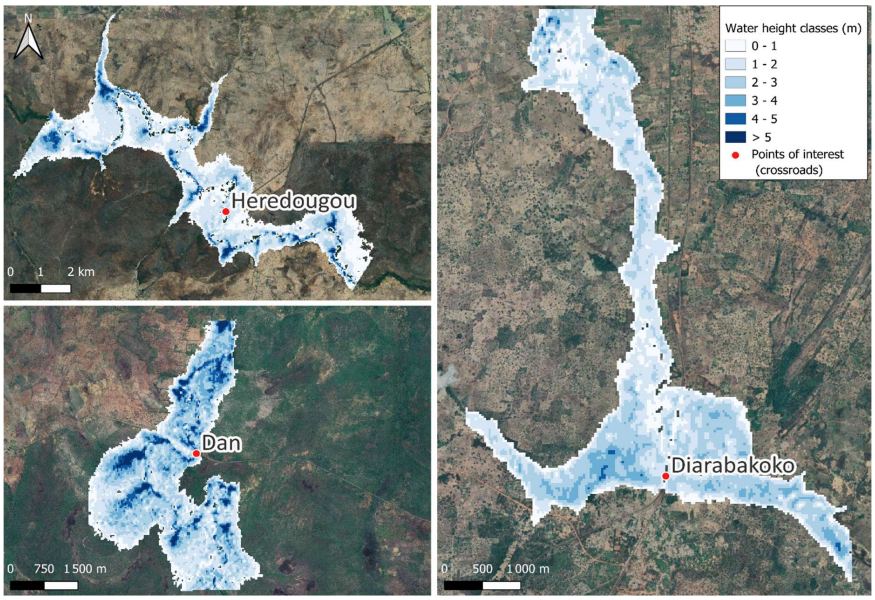

Figure 3.8.1: River flood simulations obtained with DassFlow 2D at three points of interest (intersections of national roads with rivers) in Burkina Faso.

3.9. LISFLOOD-FP

3.9.1. Short description

LISFLOOD-FP (Bates and De Roo, 2000) is a 2D raster-based hydraulic model for simulating fluvial or coastal flood spreading, with output consisting of raster maps of values for several flood water parameters such as depth, water surface elevation, velocity etc. in each grid cell at each time step. In the case of fluvial flooding, it also outputs predicted stage and discharge hydrographs at the outlet of the reach and other specified locations. It can simulate the dynamic propagation of flood waves over fluvial, coastal and estuarine floodplains. It is designed for research purposes as part of an effort

to improve the fundamental understanding of flood hydraulics, flood inundation prediction and flood risk assessment. The model has been employed in flood hazard mapping and flood risk assessment at a range of scales, including national and continental scales (e.g. Bates et al. 2021; Alfieri et al. 2014; Quinn et al. 2019).

3.9.2. Mathematical background

LISFLOOD-FP employs the 2D shallow-water equations using an explicit finite difference scheme. It solves the mass continuity in each cell and continuity of momentum between cells. For river flows, it employs either the kinematic or diffusive wave approximation of the one-dimensional St. Venant equations to simulate the passage of a flood wave along a channel reach. When bankfull depth is exceeded, water moves from the channel to adjacent floodplain sections and two-dimensional flood spreading is simulated using a storage cell concept applied over a raster grid.

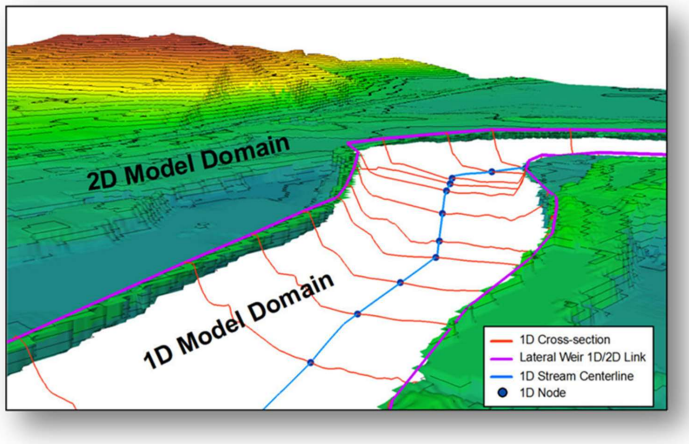

Figure 3.9.1: LISFLOOD-FP model schematics

The model offers an option of choosing different solvers with different degrees of complexity.

Floodplain solvers

The least complex solver based on the shallow water equations uses an approximation of the diffusive wave equations based on the Manning's equation by including only the friction and water slope terms. It uses a user-defined fixed time step. To ensure model stability, a flow limiter is introduced setting a limit on the volume of water allowed to flow between cells during a single time step as a function of flow depth, grid size, and time step. The 'adaptive' solver employs a similar formulation but allows a time step that varies throughout the simulation instead of using a flow limiter.

The acceleration solver uses all terms of the shallow water equations except for the convective acceleration term. It is first order in space and explicit in time but uses a semi-implicit treatment for the friction term to aid stability. The time step varies throughout the simulation according to the Courant-Friedrichs-Lewy condition, which is related to the cell size and water depth.

A solver option that uses the full shallow water equations is also available and is based on the Godunov approach and uses an approximate Riemann solver based on the TRENT model from Villanueva

and Wright (2006). It is also first order in space and explicit in time. It solves the full shallow water equations with a shock capturing scheme and uses a point-wise friction based on Manning's equation, while domain boundary/ internal boundary (wall) uses the ghost cell approach. It employs the Courant-Friedrichs-Lewy condition to ensure stability.

LISFLOOD-FP also offers a simple 'routing' option that is applied only to cells containing either very shallow water or where water slopes are very high. Water flows with a fixed flow velocity from the specified cell into whichever neighboring cell has the lowest elevation. It has the effect of reducing runtime and allowing water to flow over terrain discontinuities without destabilizing the solution.

Channel flow solvers

The simplest of the channel flow models is a 1D kinematic wave approximation of the shallow water equations, which assumes all terms except the friction and bed gradient are negligible ("kinematic" solver). The bed gradient is a simplification of the water slope term which considers the effect of changes in bed height with distance but not changes in the water free surface height. In contrast, the "diffusive" solver uses the 1D diffusive wave equation which includes the water slope term and thus can predict backwater effects. Using the 1D channel solvers, once channel water depth reaches bankfull height, water is routed onto adjacent floodplain cells to be distributed as per the chosen floodplain solver. Note: there is no transfer of momentum between the channel and floodplain, only mass.

The most recently developed method for representing rivers is as sub-grid channels, embedded with the 2D domain. The flow between channel segments is calculated based on the friction and water slopes, and local water acceleration (i.e. using the 'acceleration' solver). For any cell containing a sub-grid channel segment, the solver calculates the combined flow of water within the cell, contained both within the channel located in that cell and across the adjacent floodplain. The model is designed to operate over large data sparse areas where limited channel section data are available.

Most of the solvers assume flow to be gradually varied. The simple routing solver, however, can be used for cases of very shallow flow over steep gradients or discontinuities. The solver that uses the full shallow water equation may also handle flows that vary rapidly in time.

3.9.3. Inputs

Setting up LISFLOOD-FP requires a set of static data such as DEM, channel bathymetric information, and friction, as well as structures. For DEM, grid resolutions of approximately 25-100m would seem appropriate for most rural floodplain applications, but smaller resolutions are preferable in urban areas. Vertical accuracy of the DEM should generally be less than ±0.25m. Channel bathymetric information (channel slope, width and bankfull depth) can be derived from the DEM or taken from surveyed cross sections. They can be set individually for each point on the channel vector if necessary. Channel and floodplain friction are also required, and these can be user defined and can be set individually for each grid cell. Recently, SMHI has adopted a methodology to infer the necessary channel bathymetry from water surface elevation data along the river using a 1d hydraulic model assuming gradually varied flow. The model was successfully applied on the Torne river use case using the FABDEM elevation data, water surface elevation data based on the SWOT raster product, and river discharge from the Swedish national hydrological model analysis (S-HYPE). The same methodology will be used with the EO4FLOOD data as input (FathomDEM, SWOT water surface elevation profiles, EO based river discharge).

In addition to static data, boundary conditions are required, which can be specified as inflow discharge hydrographs that enter the model domain through the upstream channel cell forming the first location on each river channel, or flow across the domain edge, or point sources within the domain.

3.9.4. Output product

LISFLOOD-FP produces a number of output products at user specified time intervals. These include water depth, maximum water surface elevation, discharge and velocity values for each pixel; channel water surface profile, discharge and velocity values in the channels; inundated area, volume of water in the domain. All gridded outputs are in ARC ascii raster format. For the EO4FLOOD use case, we will focus on a) daily river surface water elevation profiles along the simulated rivers at 200m resolution, to compare with and utilize the corresponding data from the SWOT satellite, and b) flood extent at the spatial resolution of the model to compare with the corresponding EO data, as well as the water depth maps for the flood risk assessment.

4. References

Alfieri, L., Salamon, P., Bianchi, A., Neal, J., Bates, P., and Feyen, L. (2014). Advances in pan-European flood hazard mapping. Hydrol. Process., 28(13), 4067–4077, https://doi.org/10.1002/hyp.9947

Altenau, E. H., Pavelsky, T. M., Durand, M. T., Yang, X., Frasson, R. P. D. M., & Bendezu, L. (2021). The Surface Water and Ocean Topography (SWOT) Mission River Database (SWORD): A global river network for satellite data products. Water Resources Research, 57(7), e2021WR030054. https://doi.org/10.1029/2021WR030054

Azimi, S., Dariane, A. B., Modanesi, S., Bauer-Marschallinger, B., Bindlish, R., Wagner, W., & Massari, C. (2020). Assimilation of Sentinel 1 and SMAP—based satellite soil moisture retrievals into SWAT hydrological model: The impact of satellite revisit time and product spatial resolution on flood simulations in small basins. Journal of Hydrology, 581, 124367. https://doi.org/10.1016/j.jhydrol.2019.124367

Barbetta S., Coccia G., Moramarco T., Brocca L., and Todini E. (2017). Improving the effectiveness of real-time flood forecasting through Predictive Uncertainty estimation: the multi-temporal approach, J. Hydrol., 51, pp. 555-576. https://doi.org/10.1016/j.jhydrol.2017.06.030

Bates, P.D. and De Roo, A.P.J. (2000). A simple raster-based model for floodplain inundation. J. Hydrol., 236, 54-77. https://doi.org/10.1016/S0022-1694(00)00278-X

Bates, P. D., Quinn, N., Sampson, C., Smith, A., Wing, O., Sosa, J., et al. (2021). Combined modeling of US fluvial, pluvial, and coastal flood hazard under current and future climates. Water Resources Research, 57, e2020WR028673. https://doi.org/10.1029/2020WR028673

Bechtold, M., Modanesi, S., Lievens, H., Baguis, P., Brangers, I., Carrassi, A., ... & De Lannoy, G. (2024). Assimilation of Sentinel-1 backscatter into a land surface model with river routing and its impact on streamflow simulations in two Belgian catchments. Journal of Hydrometeorology, 24(12), 2389-2408 https://doi.org/10.1175/JHM-D-22-0198.1

Bosch, W., Dettmering, D., & Schwatke, C. (2014). Multi-Mission Cross-Calibration of Satellite Altimeters: Constructing a Long-Term Data Record for Global and Regional Sea Level Change Studies. Remote Sensing, 6(3), 2255–2281. https://doi.org/10.3390/rs6032255

Brocca, L., Ciabatta, L., Massari, C., Moramarco, T., Hahn, S., Hasenauer, S., Kidd, R., Dorigo, W., Wagner, W., and Levizzani, V.: Soil as a natural rain gauge: Estimating global rainfall from satellite soil moisture data, J. Geophys. Res.-Atmos., 119, 5128–5141, https://doi.org/10.1002/2014JD021489, 2014.

Brocca, L., Filippucci, P., Hahn, S., Ciabatta, L., Massari, C., Camici, S., Schüller, L., Bojkov, B., and Wagner, W.: SM2RAIN–ASCAT (2007–2018): global daily satellite rainfall data from ASCAT soil moisture observations, Earth Syst. Sci. Data, 11, 1583–1601, https://doi.org/10.5194/essd-11-1583-2019, 2019.

Carroll, M. L., DiMiceli, C. M., Townshend, J. R. G., Sohlberg, R. A., Elders, A. I., Devadiga, S., ... & Levy, R. C. (2017). Development of an operational land water mask for MODIS Collection 6, and influence on downstream data products. International Journal of Digital Earth, 10(2), 207-218. https://doi.org/10.1080/17538947.2016.1232756

Chew, C.C., Small, E.E., Huelsing, H. (2023). Flooding and inundation maps using interpolated CYGNSS reflectivity observations. Remote Sensing of Environment, 293. https://doi.org/10.1016/j.rse.2023.113598.

Chew, C. C., & Small, E. E. (2018). Soil moisture sensing using spaceborne GNSS reflections: Comparison of CYGNSS reflectivity to SMAP soil moisture. Geophysical Research Letters, 45(9), 4049-4057. https://doi.org/10.1029/2018GL077905

Coccia, G., Todini, E. (2011). Recent development in predictive uncertainty assessment based on the model conditional processor approach, Hydrol. Earth Syst. Sci., 15, 3253-3274, https://doi.org/10.5194/hess-15-3253-2011

Collischonn, W., Allasia, D., Da Silva, B. C., & Tucci, C. E. (2007). The MGB-IPH model for large-scale rainfall—runoff modelling. Hydrological Sciences Journal, 52(5), 878-895. https://doi.org/10.1623/hysj.52.5.878

Dallaire, C. O., Lehner, B., Sayre, R., & Thieme, M. (2019). A multidisciplinary framework to derive global river reach classifications at high spatial resolution. Environmental Research Letters, 14(2), 024003. https://doi.org/10.1088/1748-9326/aad8e9

Elmi, O., Tourian, M. J., Bárdossy, A., & Sneeuw, N. (2021). Spaceborne river discharge from a nonparametric stochastic quantile mapping function. Water Resources Research, 57(12), e2021WR030277. https://doi.org/10.1029/2021WR030277

Entekhabi, D. et al., "The Soil Moisture Active Passive (SMAP) Mission," in Proceedings of the IEEE, vol. 98, no. 5, pp. 704-716, May 2010, https://doi.org/10.1109/JPROC.2010.2043918

Filippucci, P., Brocca, L., Bonafoni, S., Saltalippi, C., Wagner, W., Tarpanelli, A.: Sentinel-2 high-resolution data for river discharge monitoring, Remote Sensing of Environment, 281, 113255, https://doi.org/10.1016/j.rse.2022.113255, 2022.

Filippucci, P., Andreadis, K., David, C. H., Tarpanelli, A.: A Tool for Surface Reflectance Retrieval from Google Earth Engine for the estimation of river flow variation, submitted to Water Resource Research, 2025a

Filippucci, P., Sahoo, D. P., Tarpanelli, A.: Two decades of river discharge from multi-mission multi-spectral data, Remote Sensing of Environment, 329, 114919, doi: https://doi.org/10.1016/j.rse.2025.114919, 2025b

Friedl, M., Gray, J., & Sulla-Menashe, D. (2019). MCD12Q2 MODIS/Terra+ Aqua land cover dynamics yearly L3 global 500m SIN grid V006.

Funk, C., Peterson, P., Landsfeld, M. et al. The climate hazards infrared precipitation with stations—a new environmental record for monitoring extremes. Sci Data 2, 150066 (2015). https://doi.org/10.1038/sdata.2015.66

Gascoin, S., Grizonnet, M., Bouchet, M., Salgues, G., and Hagolle, O.: Theia Snow collection: high-resolution operational snow cover maps from Sentinel-2 and Landsat-8 data, Earth Syst. Sci. Data, 11, 493–514, https://doi.org/10.5194/essd-11-493-2019, 2019.

Hall, D. K., Riggs, G. A., Salomonson, V. V., DiGirolamo, N. E., and Bayr, K. J.: MODIS snow-cover products, Remote Sens. Environ., 83, 181–194, https://doi.org/10.1016/S0034-4257(02)00095-0, 2002.

Hersbach, H., Bell, B., Berrisford, P., Hirahara, S., Horányi, A., Muñoz-Sabater, J., Nicolas J., Peubey, C., Radu, R., … & Thépaut, J. N. (2020). The ERA5 global reanalysis. Quarterly Journal of the Royal Meteorological Society 146, no. 730 (2020): 1999-2049. https://doi.org/10.1002/qj.3803

Huang, Z., Hejazi, M., Li, X., Tang, Q., Vernon, C., Leng, G., ... & Wada, Y. (2018). Reconstruction

of global gridded monthly sectoral water withdrawals for 1971–2010 and analysis of their spatiotemporal patterns. Hydrology and Earth System Sciences, 22(4), 2117-2133. https://doi.org/10.5194/hess-22-2117-2018

Huffman, G.J., D.T. Bolvin, D. Braithwaite, K. Hsu, R. Joyce, C. Kidd, E.J. Nelkin, S. Sorooshian, E.F. Stocker, J. Tan, D.B. Wolff, P. Xie, 2020: Integrated Multi-satellitE Retrievals for the Global Precipitation Measurement (GPM) mission (IMERG). Chapter 19 in Adv. Global Change Res., Vol. 67, Satellite Precipitation Measurement, V. Levizzani, C. Kidd, D. Kirschbaum, C. Kummerow, K. Nakamura, F.J. Turk (Ed.), Springer Nature, Dordrecht, ISBN 978-3-030-24567-2 / 978-3-030-24568-9 (eBook), 343-353. doi:/10.1007/978-3-030-24568-9_19.

Jarvis, A., Reuter, H., Nelson, A., & Edith, G. (2008). Hole-filled seamless SRTM data V4. Tech. Rep., International Centre for Tropical Agriculture (CIAT). Cali, Columbia.

Jiang, L., Westphal Christensen, S., & Bauer-Gottwein, P. (2021). Calibrating 1D hydrodynamic river models in the absence of cross-section geometry using satellite observations of water surface elevation and river width. Hydrology and Earth System Sciences, 25(12), 6359-6379. https://doi.org/10.5194/hess-25-6359-2021

Kalnay E., Kanamitsu M., Kistler R., Collins W., Deaven D., Gandin L., Iredell M., Saha S., White G., Woollen J., Zhu Y., Leetmaa A., Reynolds R., Chelliah M., Ebisuzaki W., Higgins W., Janowiak J., Mo K. C., Ropelewski C., Wang J., Jenne R., and Joseph D., The NCEP/NCAR 40-year reanalysis project, Bulletin of the American Meteorological Society. (1996) 77, no. 3, 437–471, https://doi.org/10.1175/1520-0477(1996)077<0437:tnyrp>2.0.co;2.

Kerr, Y. H., Waldteufel, P., Richaume, P., Wigneron, J. P., Ferrazzoli, P., Mahmoodi, A., Al Bitar,A., Cabot, F., Gruhier, C., Juglea, S. E., Leroux, D., Mialon, A., and Delwart, S. (2012). The SMOSsoil moisture retrieval algorithm. 50(5):1384–1403. Conference Name: IEEE Transactions on Geoscienceand Remote Sensing. https://doi.org/10.1109/TGRS.2012.2184548

Kraft, B., Jung, M., Körner, M., Koirala, S., & Reichstein, M. (2021). Towards hybrid modeling of the global hydrological cycle. Hydrology and Earth system sciences discussions, 2021, 1-40. https://doi.org/10.5194/hess-26-1579-2022

Kristensen, K., Nielsen, A., Berg, C. W., Skaug, H., & Bell, B. M. (2016). TMB: automatic differentiation and Laplace approximation. Journal of statistical software, 70, 1-21. https://doi.org/10.18637/jss.v070.i05

Kubota, T., K. Aonashi, T. Ushio, S. Shige, Y. N. Takayabu, M. Kachi, Y. Arai, T. Tashima, T. Masaki, N. Kawamoto, T. Mega, M. K. Yamamoto, A. Hamada, M. Yamaji, G. Liu and R. Oki 2020: Global Satellite Mapping of Precipitation (GSMaP) products in the GPM era, Satellite precipitation measurement, Springer, https://doi.org/10.1007/978-3-030-24568-9_20.

Larnier, K., J. Monnier, P.-A. Garambois et J. Verley, River discharge and bathymetry estimation from SWOT altimetry measurements, Inverse problems in science and engineering, vol. 29, p. 759–789, 2021. https://doi.org/10.1080/17415977.2020.1803858

Larnier, K., Garambois, P. A., Emery, C., Pujol, L., Monnier, J., Gal, L., ... & Calmant, S. (2025). Estimating channel parameters and discharge at river network scale using hydrological-hydraulic models, SWOT and multi-satellite data. Water Resources Research, 61(7), e2024WR038455. https://doi.org/10.1029/2024WR038455

Lehner, B., & Grill, G. (2013). Global river hydrography and network routing: baseline data and new approaches to study the world's large river systems. Hydrological Processes, 27(15), 2171-2186. https://doi.org/10.1002/hyp.9740

Lindström, G., Pers, C.P., Rosberg, R., Strömqvist, J. and Arheimer, B. (2010). Development and test of the HYPE model – A water quality model for different spatial scales. Hydrol. Res., 41, 295-319 https://doi.org/10.2166/nh.2010.007

Luojus, K.; Venäläinen, P.; Moisander, M.; Pulliainen, J.; Takala, M.; Lemmetyinen, J.; Derksen, C.; Mortimer, C.; Mudryk, L.; Schwaizer, G.; Nagler, T. (2024): ESA Snow Climate Change Initiative (Snow_cci): Snow Water Equivalent (SWE) level 3C daily global climate research data package (CRDP) (1979 - 2022), version 3.1. NERC EDS Centre for Environmental Data Analysis, 07 November 2024. https://dx.doi.org/10.5285/9d9bfc488ec54b1297eca2c9662f9c81

Maidment, D. R. (Ed.). (2002). Arc Hydro: GIS for water resources. ESRI, Inc..

Malou, T., P.-A. Garambois, A. Paris, J. Monnier et K. Larnier, 2021, Generation and analysis of stage fall-discharge laws from coupled hydrological-hydraulic river network model integrating sparse multi-satellite data. Journal of Hydrology, vol. 603, p. 126993. https://doi.org/10.1016/j.jhydrol.2021.126993

Massari, C., Brocca, L., Pellarin, T., Abramowitz, G., Filippucci, P., Ciabatta, L., Maggioni, V., Kerr, Y., and Fernandez Prieto, D., 2020. A daily 25 km short-latency rainfall product for data-scarce regions based on the integration of the Global Precipitation Measurement mission rainfall and multiple-satellite soil moisture products, Hydrol. Earth Syst. Sci., 24, 2687–2710, https://doi.org/10.5194/hess-24-2687-2020

Massari, C., Brocca, L., Tarpanelli, A., & Moramarco, T. (2015). Data assimilation of satellite soil moisture into rainfall-runoff modelling: A complex recipe? Remote Sensing, 7(9), 11403–11433. https://doi.org/10.3390/rs70911403

Messager, M.L., Lehner, B., Grill, G., Nedeva, I., Schmitt, O. (2016). Estimating the volume and age of water stored in global lakes using a geo-statistical approach. Nature Communications, 7: 13603. https://doi.org/10.1038/ncomms13603

Metsämäki, S., Pulliainen, J., Salminen, M., Luojus, K., Wiesmann, A., Solberg R. and Ripper, E. 2015. Introduction to GlobSnow Snow Extent products with considerations for accuracy assessment. Remote Sensing of Environment, 156, 96–108. https://doi.org/10.1016/j.rse.2014.09.018

Metsämäki, S. ; Mattila, O. - P. ; Pulliainen, J. ; Niemi, K. ; Luojus, K. ; Böttcher, K. An optical reflectance model-based method for fractional snow cover mapping applicable to continental scale. Remote Sensing of Environment 2012, 123, 508 - 521. https://doi.org/10.1016/j.rse.2012.04.010

Muñoz Sabater, J. (2019): ERA5-Land hourly data from 1950 to present. Copernicus Climate Change Service (C3S) Climate Data Store (CDS) [Dataset]. https://doi.org/10.24381/cds.e2161bac

Murray, A. M., Jørgensen, G. H., Godiksen, P. N., Anthonj, J., & Madsen, H. (2023). DHI-GHM: Real-time and forecasted hydrology for the entire planet. Journal of Hydrology, 620, 129431. https://doi.org/10.1016/j.jhydrol.2023.129431

Nagler, T., Rott, H., Ripper, E., Bippus, G., and Hetzenecker, M., 2016. Advancements for Snowmelt Monitoring by Means of Sentinel-1 SAR, Remote Sens., 8, 348, https://doi.org/10.3390/rs8040348

Ortenzi, S., C. Massari, L. Ciabatta, C. Cencetti, M. Dionigi, I. Marchesini, M. Stelluti, and L. Di Matteo, 2025: Leveraging SMAP-Based Soil Moisture to Identify Runoff Thresholds on Flash Flood–Prone Basins in the Mediterranean Region. J. Hydrometeor., 26, 975–990, https://doi.org/10.1175/JHM-D-24-0143.1.

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. https://doi.org/10.1002/9781119657002.ch18

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 modelled discharge in the Amazon basin. Water Resources Research, 52(5), 3787-3814. https://doi.org/10.1002/2014WR016618

*Pekel, JF., Cottam, A., Gorelick, N. et al. High-resolution mapping of global surface water and its long-term changes. Nature **540**, 418–422 (2016). https://doi.org/10.1038/nature20584*

Pujol, L., Garambois, P. A., Finaud-Guyot, P., Monnier, J., Larnier, K., Mose, R., ... & Calmant, S. (2020). Estimation of multiple inflows and effective channel by assimilation of multi-satellite hydraulic signatures: The ungauged anabranching Negro River. Journal of Hydrology, 591, 125331. https://doi.org/10.1016/j.jhydrol.2020.125331

Pulliainen, J. 2006. Mapping of snow water equivalent and snow depth in boreal and sub-arctic zones by assimilating space-borne microwave radiometer data and ground-based observations. Remote Sensing of Environment, 101, 257-269, https://doi.org/10.1016/j.rse.2006.01.002

Quinn, N., Bates, P. D., Neal, J., Smith, A., Wing, O., Sampson, C., et al. (2019). The spatial dependence of flood hazard and risk in the United States. Water Resour. Res., 55, 1890–1911. https://doi.org/10.1029/2018WR024205

Rodríguez-Fernández, N. J., Muñoz Sabater, J., Richaume, P., de Rosnay, P., Kerr, Y. H., Albergel, C., ... & Mecklenburg, S. (2017). SMOS near-real-time soil moisture product: processor overview and first validation results. *Hydrology and Earth System Sciences*, *21*(10), 5201-5216. *https://doi.org/10.5194/hess-21-5201-2017*

Ross, C. W., Prihodko, L., Anchang, J., Kumar, S., Ji, W., & Hanan, N. P. (2018). HYSOGs250m, global gridded hydrologic soil groups for curve-number-based runoff modeling. Scientific data, 5(1), 1-9. https://doi.org/10.1038/sdata.2018.91

Salomonson, V. V. and Appel, I.: Estimating fractional snow cover from MODIS using the normalized difference snow index, Remote Sens. Environ., 89, 351–360, https://doi.org/10.1016/j.rse.2003.10.016, 2004.

Scherer, D., Schwatke, C., Dettmering, D., & Seitz, F. (2023). ICESat-2 river surface slope (IRIS): A global reach-scale water surface slope dataset. Scientific Data, 10(1), 359. https://doi.org/10.1038/s41597-023-02215-x

Scherer, D., Schwatke, C., Dettmering, D., & Seitz, F. (2024). Monitoring river discharge from space: An optimization approach with uncertainty quantification for small ungauged rivers. Remote Sensing of Environment, 315, 114434. https://doi.org/10.1016/j.rse.2024.114434

Schwaizer, G.; Nagler, T.; Hetzenecker, M.; Mölg, N.; Metsämäki, S.; Nemec, J. (2024): ESA Snow Climate Change Initiative (Snow_cci): Daily global Snow Cover Fraction - snow on ground (SCFG) from MODIS (2000-2022), version 3.0. NERC EDS Centre for Environmental Data Analysis, 15 October 2024. https://doi.org/10.5285/80567d38de3f4b038ee6e6e53ed1af8a

Schwatke, C., Dettmering, D., Bosch, W., & Seitz, F. (2015). DAHITI - An innovative approach for estimating water level time series over inland waters using multi-mission satellite altimetry. Hydrology and Earth System Sciences, 19(10), 4345–4364. https://doi.org/10.5194/hess-19-4345-2015

Schwatke, C.; Dettmering, D.; Seitz, F. Volume Variations of Small Inland Water Bodies from a Combination of Satellite Altimetry and Optical Imagery. Remote Sens. 2020, 12, 1606. https://doi.org/10.3390/rs12101606

Schwatke, C., Dettmering, D., Passaro, M., Hart-Davis, M., Scherer, D., Müller, F. L., Bosch, W., & Seitz, F. (2023). OpenADB: DGFI-TUM's Open Altimeter Database. Geoscience Data Journal. https://doi.org/10.1002/gdj3.233

Shangguan, W., Hengl, T., Mendes de Jesus, J., Yuan, H., & Dai, Y. (2017). Mapping the global depth to bedrock for land surface modeling. Journal of Advances in Modeling Earth Systems, 9(1), 65-88. https://doi.org/10.1002/2016MS000686

Simons, G., Koster, R., & Droogers, P. (2020). Hihydrosoil v2. 0-high resolution soil maps of global hydraulic properties. Future Works.[online] Available from https://www.futurewater.eu/projects/hihydrosoil

Solberg, R., G. Schwaizer, T. Nagler, M. Hetzenecker, S. Wunderle, X. Xiao, C. Neuhaus, A. Wiesmann, K. Luojus, M.Takala, J. Pulliainen, J. Lemmetyinen, M. Moisander, and N. Mölg (2025) ESA CCI+ Snow ECV: Climate Research Data Package, version 4.1, January 2025.

Takala, M., K. Luojus, J. Pulliainen, C. Derksen, J. Lemmetyinen, J.-P. Kärnä, J. Koskinen, B. Bojkov, 2011. Estimating northern hemisphere snow water equivalent for climate research through assimilation of space-borne radiometer data and ground-based measurements. Remote Sensing of Environment, 115, 12, 3517-3529, https://doi.org/10.1016/j.rse.2011.08.014

Tarpanelli A., Brocca L., Melone F., Moramarco T., Lacava T., Faruolo M., Pergola N., Tramutoli V. (2013) Toward the estimation of river discharge variations using MODIS data in ungauged basins. Remote Sensing of Environment, 136, 47–55. http://dx.doi.org/10.1016/j.rse.2013.04.010

Tarpanelli A., Amarnath G., Brocca L., Massari C., Moramarco T. (2017). Discharge estimation and forecasting by MODIS and altimetry data in Niger-Benue River, Remote Sensing of Environment, 195, 96-106. http://dx.doi.org/10.1016/j.rse.2017.04.015

Tarpanelli A., Iodice F., Brocca L., Restano M., Benveniste J. (2020). River flow monitoring by Sentinel-3 OLCI and MODIS: comparison and combination. Remote Sensing, 12(23), 3867. https://doi.org/10.3390/rs12233867

Townshend, J. (2016). Global forest cover change (GFCC) water cover 2000 global 30 m V001. NASA EOSDIS Land Processes Distributed Active Archive Center (DAAC) data set, GFCC30WC-001.

Venäläinen, P., K. Luojus, C. Mortimer, J. Lemmetyinen, J. Pulliainen, M. Takala, M. Moisander, and L. Zschenderlein, 2023. Implementing spatially and temporally varying snow densities into the GlobSnow snow water equivalent retrieval. The Cryosphere, 17, 2, 719-736, https://doi.org/10.5194/tc-17-719-2023

Villanueva I, and Wright NG. (2006). Linking Riemann and storage cell models for flood prediction. Proceedings of the Institution of Civil Engineers, Journal of Water Management 159: 27–33. https://doi.org/10.1680/wama.2006.159.1.27

Wagner, W., Hahn, S., Kidd, R., Melzer, T., Bartalis, Z., Hasenauer, S., Figa-Saldaña, J., de Rosnay, P., Jann, A., Schneider, S., Komma, J., Kubu, G., Brugger, K., Aubrecht, C., Züger, J., Gangkofner, U., Kienberger, S., Brocca, L., Wang, Y., Blöschl, G., Eitzinger, J., Steinnocher, K., 2013. The ASCAT Soil Moisture Product: A Review of its Specifications, Validation Results, and Emerging Applications. Meteorologische Zeitschrift 5–33. https://doi.org/10.1127/0941-2948/2013/0399

Wongchuig, S., Kitambo, B., Papa, F., Paris, A., Fleischmann, A. S., Gal, L., ... & Calmant, S. (2023). Improved modeling of Congo's hydrology for floods and droughts analysis and ENSO teleconnections. Journal of Hydrology: Regional Studies, 50, 101563. https://doi.org/10.1016/j.ejrh.2023.101563

Xiao, X.; Naegeli, K.; Neuhaus, C.; Salberg, A.-B.; Schwaizer, G.; Wiesmann, A.; Wunderle, S.; Nagler, T. (2024): ESA Snow Climate Change Initiative (Snow_cci): Daily global Snow Cover Fraction - snow on ground (SCFG) from AVHRR (1979 - 2022), version 3.0. NERC EDS Centre for Environmental Data Analysis, 15 October 2024. https://dx.doi.org/10.5285/56ff07acabab42888afe2d20b488ec49

Yamazaki, D., Ikeshima, D., Sosa, J., Bates, P. D., Allen, G. H., & Pavelsky, T. M. (2019). MERIT Hydro: A high-resolution global hydrography map based on latest topography dataset. Water Resources Research, 55(6), 5053-5073. https://doi.org/10.1029/2019WR024873