SOS-Water – Earth Observation prototype technical notes

[D3.2, Improved EO application prototypes]

Description

This document outlines the Earth Observation Prototypes (EOP) developed as part of Task 3.2 of the SOS-Water project. It includes technical notes detailing the developed algorithms, tools, and methods used. The document consists of five distinct technical reports, each corresponding to a different prototype, which collectively form Deliverable 3.2. The associated source code, maintained by Eawag (Partner No. 11) and FutureWater-NL (Partner No. 6.1), is available in the SOS-Water – Work package 3 GitHub code repository at gitlab.eawag.ch/surf/remote-sensing/sos-water/sosw_wp3.

Documents

Document	Partner organisation	Partner no
EOP1: Suspended Particulate Matter (SPM) retrieval prototype technical notes	Eawag	11
EOP2: Lake and River ice mapper prototype technical notes	Eawag	11
EOP3: SWOT-based altimetry prototype technical note	FUTUREWATER-NL	6.1
EOP4: Surface water temperature (SWT) mapper prototype technical notes	Eawag	11
EOP5: Crop water productivity mapping prototype technical notes	FUTUREWATER-NL	6.1

January 23, 2026

SOS-Water – EOP1: Suspended Particulate Matter (SPM) retrieval prototype technical notes

[D3.2, Improved EO application prototypes]

Authors

Authors	Partner no	Partner organisation	Name of author
Main author	11	Eawag	Michael Brechbühler
Contributing author	11	Eawag	Daniel Odermatt

Document history

Date	Version	Chapters affected	Description of change	Author	Document status
12/08/2024	0.1	All	Initial version outlining document	Michael Brechbühler	FIRST DRAFT
23/09/2024	0.2	All	Internal review	Daniel Odermatt	FIRST DRAFT
23/09/2024	1	All	Adaptions and additions	Michael Brechbühler	FINAL
23/01/2026	1.1	3.5	Clarifications added to Section 3.5 to explicitly define the Generalized Index. Figure 3 was updated with a sensor color legend and revised caption/text to clarify its role in the calibration procedure rather than validation.	Michael Brechbühler	FINAL (REV)

Table of Contents

Abbreviations

ADCP	Acoustic Doppler Current Profiler
CFMASK	C Function of Mask
DSM	Digital Surface Model
EO	Earth Observation
EOP	Earth Observation Prototype
ETM+	Enhanced Thematic Mapper Plus
GEE	Google Earth Engine
MAIN	Modified Atmospheric correction for INland waters
MRC	Mekong River Commission
MSI	Multispectral Instrument
NDSSI	Normalized Difference Suspended Sediment Index
NDTI	Normalized Difference Turbidity Index
NDWI	Normalized Difference Water Index
OLI	Operational Land Imager
OMI	Ozone Monitoring Instrument
PCA	Principal Component Analysis
Rrs	Remote Sensing Reflectance
RS	Remote Sensing
SMOGN	Synthetic Minority Over-Sampling Technique for Regression
SPM	Suspended Particulate Matter
SWM	Small Waterbodies Mapping
TOMS	Total Ozone Mapping Spectrometer
VMD	Vietnamese Mekong Delta

1 Background

This document provides technical details for an Earth Observation (EO) application prototype developed for the SOS-Water project, for Suspended Particulate Matter (SPM) mapping. It includes a detailed collection of technical notes that describe the methodology and validation of the prototype created as part of Task 3.2. The associated source code, maintained by Eawag (Partner No. 11), is available in the SOS-Water – Work package 3 GitHub code repository at gitlab.eawag.ch/surf/remote-sensing/sos-water/sosw_wp3. SPM refers to solid particles suspended in water, including fine dust, soot, organic particulate matter, and microscopic organisms. This mixture comprises both inorganic materials (such as silt, clay, and mineral particles) and organic materials (like detritus). When the inorganic fraction dominates, e.g. in the case of rivers, SPM is closely correlated with turbidity.

This SPM retrieval application prototype is designed to indicate the concentration of suspended sediment in lakes and rivers using Sentinel-2 MSI and Landsat-7/8/9 ETM+/OLI/OLI-2 data. This prototype addresses the data gaps related to sediment flows, particularly in the Mekong basin, which are critical for maintaining ecosystem health and delta stability. The model offers high-resolution (30 to 10 meters) mapping capabilities and supports both monitoring applications and historical analysis dating back to April 1999 with Landsat 7.

The primary objective of the EO Prototype 1 (EOP1) is to develop and validate a methodology for accurately estimating sediment concentration and sediment flux at a high spatiotemporal resolution using EO data from both Landsat and Sentinel-2 satellites. This methodology is applied to the Mekong River Basin as the primary case study, demonstrating that SPM can be reliably extracted using EO-based methods across a large and complex river basin. The Mekong River Basin is an especially relevant case study due to its significant ecological and economic importance, coupled with the recent surge in dam development across the basin. These dams have drastically reduced sediment transport through the process of sediment trapping, which has profound implications for downstream sediment dynamics, delta formation, and overall river health. Understanding and quantifying these alterations is critical for managing the basin's natural resources effectively.

The ultimate purpose of EOP1 is to quantify sediment alterations caused by dam operations, specifically focusing on trapping efficiency. To achieve this, we plan to couple our EO-derived sediment concentration estimates with modelled discharge data from the VICRes model (Dang et al., 2020). This integration will allow for the estimation of sediment flux alterations attributable to dam operations, offering insights into how these structures impact sediment dynamics within the basin. The sediment flux estimates generated through this approach are intended to be used for calibrating the trap efficiency of dams in the CASCADE model (Tangi et al., 2019), which will further enhance our understanding of sediment transport and deposition processes. The universal design of our methodology ensures that it can be applied to other river basins, making it a valuable tool for broader applications within the project.

2 Data

2.1 In-situ data

For the calibration and validation of our SPM model, we compiled available in-situ datasets that include measurements of suspended sediment concentration obtained from the Mekong River Commission (MRC) data portal, as well as other relevant literature and sources. Table 1 summarizes the relevant datasets that were considered for calibration and validation purposes.

The most extensive dataset available is from the MRC's Water Quality Monitoring Program (WQMP), which has been collecting sediment concentration data since 1985 across 87 stations throughout the entire Mekong basin. These measurements are based on surface grab samples taken at a depth of 0.3 to 0.5 meters in the middle of the active channel (Des E. Walling, 2008). Concerns have been raised regarding the representativeness of these measurements across the river cross-section and the non-isokinetic nature of the sampling (Beveridge, 2020). Additionally, the relatively low sampling frequency, generally monthly, limits the number of potential matchups with satellite imagery. We also found it difficult to find accurate information on sampling locations. Due to these concerns and initial poor matchup results, we opted not to use this dataset for our analysis.

Instead, we focused on datasets that provide more representative and frequent measurements. The MRC's Discharge Sediment Monitoring Project (DSMP), conducted from 2009 to 2014, offers high-quality sediment concentration data from 15 stations along the entire Mekong mainstream. This dataset utilizes D-96 depth-integrating samplers and follows the Equal Distance Increment (EDI) method, ensuring isokinetic and cross-sectionally representative measurements. The DSMP dataset's higher temporal resolution (~1/3-monthly) made it a valuable reference for our model.

Additionally, two high-frequency datasets were provided by our project partner, the Southern Institute for Water Resource Planning (SIWRP), covering the Vietnamese Mekong Delta (VMD). The SIWRP Mekong Delta 2016-2020 dataset includes daily sediment concentration measurements from 5 stations, with increased sampling frequency during the wet season (twice per day). This dataset also employs depth-integrating samplers similar to the D-96, ensuring high-quality, representative data. The SIWRP Mekong Delta 2014 dataset further complements this with sediment concentration data for 36 stations, measured twice per year during the dry and wet seasons, across 10 transects.

Moreover, we utilized the dataset from the Environmental Information Data Centre (EIDC) - Water and Suspended Sediment Discharges for the Mekong Delta, Vietnam (Darby et al., 2020), which provides daily sediment concentration, sediment flux, and discharge data for 4 stations in the VMD from 2005 to 2015. This dataset uses an Acoustic Doppler Current Profiler (ADCP) with backscatter calibration based on sediment samples.

Ultimately, we selected the SIWRP Mekong Delta 2016-2020 dataset, the EIDC dataset, and the MRC DSMP dataset for our calibration and validation efforts. These datasets were chosen due to their high temporal resolution, methodological rigor, and coverage, with the DSMP covering the entire Mekong mainstream and the SIWRP and EIDC datasets focusing on the Vietnamese Mekong Delta. These datasets provided the necessary temporal frequency and spatial coverage to ensure robust calibration and validation of our SPM model.

Table 1 – Overview of relevant in-situ datasets

Dataset	Parameter	Stations	Temporal Extent	Temporal Resolution	Method	Comments
Mekong River Commission - Water Quality Monitoring Program (WQMP)	sediment concentration	87 stations (entire Mekong basin)	1985 to present	monthly	Surface grab samples taken at 0.3-to-0.5-meter depth in the middle of active channel (Des E. Walling, 2008)	measurements are not representative of cross-section and not isokinetic (Beveridge, 2020)
Mekong River Commission - Discharge Sediment Monitoring Project (DMSP)	sediment concentration	15 stations (Mekong mainstream)	2009 to 2014	~1/3-monthly	D-96 depth-integrating sampler	Depth-integrated Equal-distance-increment (EDI) and isokinetic sampling
SIWRP - Mekong Delta 2016-2020 dataset	sediment concentration	5 stations (VMD)	2016 to 2020	daily (twice per day in wet season)	Depth-integrating sampler similar to D-96	Depth-integrated Equal-distance-increment (EDI) and isokinetic sampling, very high temporal frequency
SIWRP - Mekong Delta 2014 dataset	sediment concentration for 10 transects	36 stations (VMD)	2014	twice per year (dry/wet season)	Depth-integrating sampler similar to D-96	Depth-integrated Equal-distance-increment (EDI) and isokinetic sampling, low temporal frequency
EIDC - Water and suspended sediment discharges for the Mekong Delta, Vietnam (Darby et al., 2020)	sediment concentration, sediment flux, discharge	4 stations (VMD)	2005-2015	daily	Acoustic Doppler Current Profiler (ADCP) with backscatter calibration based on sediment samples	Depth-integrated, very high temporal frequency

2.2 EO data

The EO input data for our model are multispectral measurements from Sentinel-2 Multispectral Instrument (MSI) from both the S2A and S2B platform, Landsat-7 Enhanced Thematic Mapper Plus (ETM+), Landsat-8 Operational Land Imager (OLI) 1 and Landsat-9 OLI-2. We use Level 1 data provided in terms of Top-of-Atmosphere Reflectance and Radiance for Sentinel-2 and Landsat sensors, respectively.

Table 2 – Overview of relevant EO sensors

Sensor	Processing level	Temporal Extent	Spatial Resolution
Sentinel-2A/B (MSI)	L1C (Top-of-atmosphere	2015 to present	10
Landsat-8/9 (OLI-1/2)	L1C2 (Radiance)	2013 to present	30
Landsat-7 (ETM+)	L1C2 (Radiance)	1999 to 2022	30

Both Sentinel-2 and Landsat satellites are in a near-polar, sun-synchronous orbit, which means they pass over the same area of the Earth at the same local solar time each day. The satellites cross the equator between 10:00 AM and 10:15 AM local solar time during their descending pass when image acquisitions are acquired. Figure 1 shows the achievable data coverage based on our multi-sensor approach. The coverage was calculated for the period between 2022-2023 over the Vietnamese Mekong Delta (VMD) with data loss due to clouds and cloud shadows taken into account. The. assessment shows that a median monthly coverage of 15.0% and median frequency of 7.8 days can be achieved with this sensor configuration. Seasonally, the achieved frequencies are 6.6 days in the dry and 10.4 days in the wet season.

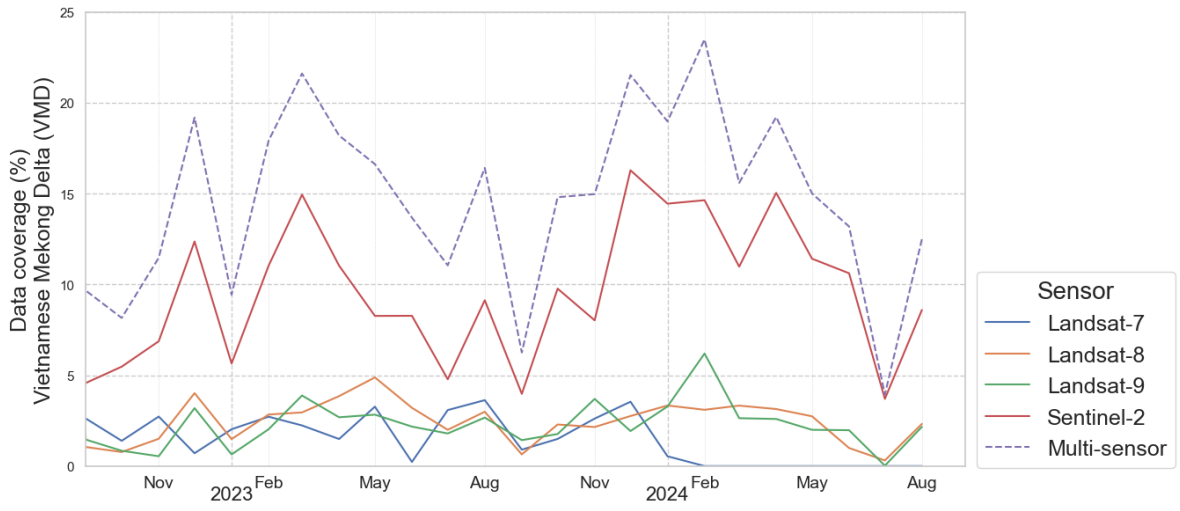

Figure 1 – Data coverage assessment of multi-sensor approach in the Vietnamese Mekong Delta. Coverage is calculated as monthly mean aggregates of spatially aggregated daily data availability. For example, for daily pixels in February 2024 there is a 24% chance that the data is available.

3 Methodology

Our SPM algorithm leverages datasets from the Google Earth Engine (GEE) platform, focusing on Sentinel-2 MSI L1C and Landsat-7 ETM+, Landsat-8 OLI, and Landsat-9 OLI2 L1C2 datasets. The process involves resampling, harmonization, cloud masking, water masking, and atmospheric correction steps. All these computationally expensive processing steps are executed within the Google Earth Engine environment. GEE's cloud-based infrastructure allows for efficient processing of large datasets, enabling rapid execution of these algorithms across extensive spatial and temporal scales. This efficiency is key to producing timely and reliable SPM data for monitoring and analysis. Figure 2 shows a simplified workflow of EOP1.

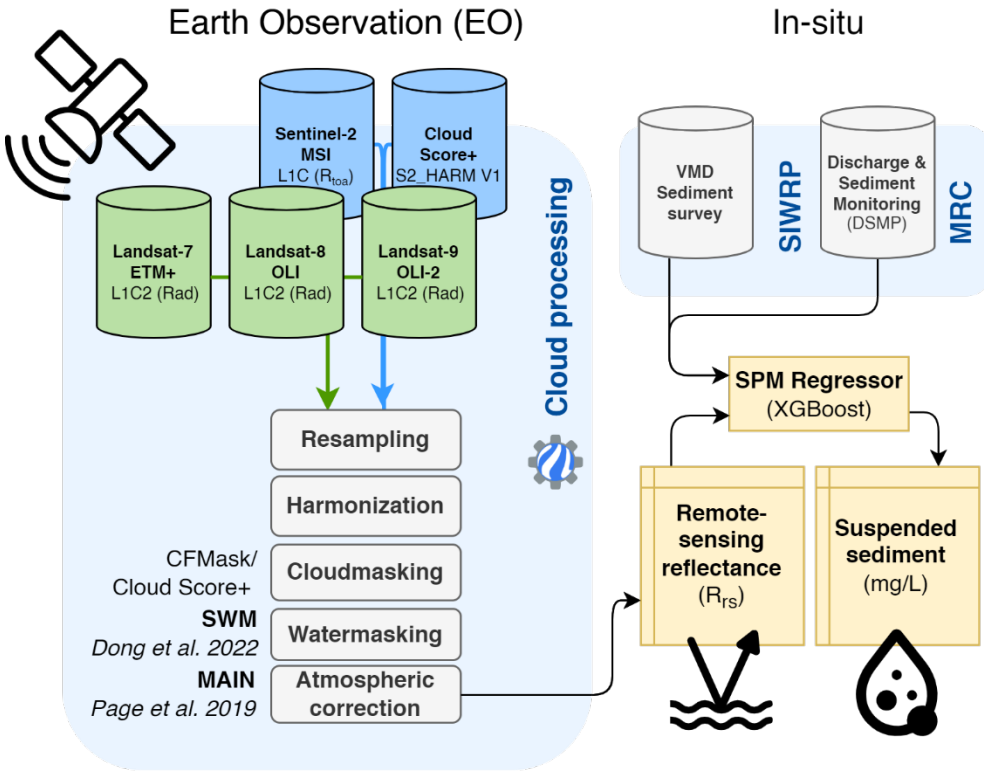

Figure 2 - Main processing steps of EOP1 algorithm

The first step involves resampling the EO datasets to a common spatial resolution of 30 meters, which is the native resolution of the Landsat bands. This ensures consistency across the Sentinel-2 and Landsat datasets. After resampling, we harmonize the datasets by applying consistent band labels (blue to swir2) across the different sensors. The applied harmonization extends beyond labelling; it includes adjusting sensor wavelengths during the later atmospheric correction phase to ensure that the data from different sensors can be directly compared.

To remove cloud and cloud shadow contaminated pixels, we apply the CFMASK algorithm (Foga et al., 2017) for Landsat datasets and the Cloud Score+ algorithm (Pasquarella et al., 2023) for Sentinel-2 data. In the water masking step, we identify water pixels. This step can be either done using the

provided water flags within the datasets or by employing our GEE implementation of the small water bodies mapping (SWM) algorithm. The SWM algorithm integrates spatial information to accurately identify small water bodies in optical remote sensing images. It combines spectral indices, like the Normalized Difference Water Index (NDWI), with contextual spatial information to enhance the detection of water bodies, especially in regions where water features are small or seasonally variable.

The critical atmospheric correction step utilizes the MAIN algorithm (Page et al., 2019) to convert the Level-1 data (top-of-atmosphere reflectance for Sentinel-2 and radiance for Landsat) into a harmonized remote sensing reflectance (Rrs). This approach addresses the challenges of atmospheric effects on satellite imagery by correcting for aerosols, water vapor, and ozone. The MAIN algorithm corrects the wavelengths of different sensors to a common reference, ensuring that the reflectance data from Sentinel-2, Landsat-7, -8, and -9 are consistent and comparable. This step is crucial for accurate SPM estimation, as atmospheric conditions can significantly affect the reflectance values used to calculate SPM concentrations.

Ultimately, we used a masking approach with a combination of CS+ and SWM and compared three different TSM retrieval algorithms described below, among which the XGBoost algorithm performed best.

3.1 Cloud masking

The CFMASK (C Function of Mask) algorithm (Foga et al., 2017) is part of the Landsat Collection 2 processing and identifies clouds and cloud shadows in the Landsat imagery. For the masking we remove pixels by chaining logical OR operators on active quality-assessment flags “Dilated Cloud” (bit 1), “Cirrus (high confidence)” (bit 2), “Cloud” (bit 3), “Cloud Shadow” (bit 4). An additional cloud-buffer can be supplied in the masking function to oppose under detection. For validation purposes we set this value to 30-meters.

The Cloud Score+ (CS+, Pasquarella et al. 2023) algorithm leverages a combination of spectral bands, particularly the blue, red, and shortwave infrared bands, to calculate a 10-meter cloud probability score based on Sentinel-2 MSI imagery. It uses a weakly supervised deep learning approach to grade the quality of individual pixels through time and provides a robust way to detect and mask clouds and cloud shadows. For the purpose of the SPM extraction, we use the CS+ quality assessment band cs_cdf with a threshold value of 0.6 where 0 is occluded and 1 represents clear.

3.2 Water masking

The Small Waterbodies Mapping (SWM) algorithm (Fan et al., 2022) is designed to accurately identify and map small water bodies, such as ponds, small lakes, and narrow rivers, which are often challenging to detect using conventional water index methods. We rebuilt the algorithm with the Earth Engine Python API. It operates in three main stages:

Stage 1: Identification of “Potentially Water” (P_1) and “Certainly Water” (C_1) Classes:

- Certainly Water (C_1): Pixels that are likely to represent pure water surfaces are classified as "certainly water". These pixels have high water index values that significantly differ from non-water pixels. The classification is based on a threshold T_{pure}, which is calculated using the maximum of an initial threshold T_s and a value derived from the median (M_1) and standard deviation (std_1) of the higher range of water index values:

$$C_1 = \{ i \mid WI_i > T_{pure} \}$$

$$T_{pure} = max(T_s, M_1 + 0.5 \times std_1)$$

- Potentially Water (P_1): Pixels that could represent water but are less definitive are classified as 'potentially water.' These pixels have water index values between two thresholds: $T_{mixed}$and T_{pure}. The T_{mixed} threshold is determined similarly to T_{pure} but considers the lower range of water index values:

$$P_1 = \{ i \mid T_{mixed}. < WI_i \leq T_{pure} \}$$

$$T_{mixed} = min(Ts, M_2 - 0.5 \times std2)$$

Stage 2: Refinement of Water Classes:

- Refining 'Potentially Water' (P_2): To reduce false positives, the algorithm refines the 'potentially water' class by removing non-water materials using the Shortwave Infrared (SWIR) band. Pixels with SWIR reflectance below a specified threshold ($T_{SWIR} = 0.1$) are retained in the refined 'potentially water' class (P_2):

$$P_2 = \{ i \in P_1 \mid \rho_{SWIR,i} < T_{SWIR} \}$$

- Refining 'Certainly Water' (C_2): The algorithm further refines 'certainly water' by expanding its spatial distribution within the 'potentially water' class. This is achieved by calculating the Water Index Roughness (WIR), which measures the local variability of the water index within a neighbourhood window (size 5). Pixels in P_2 with WIR values exceeding a threshold ($T_{WIR} = 0.4$) are reclassified as 'certainly water' (C_2):

$$WIR_i = \max_{j \in N_i} \left(\left| WI_i - WI_j \right| \right)$$

$$C_2 = \{ i \in P_2 \mid WIR_i > T_{WIR} \}$$

Stage 3: Multi-scale Water Mapping:

The final step involves combining the refined 'certainly water' (C_2) and 'potentially water' (P_2) classes to produce the final water mask. A connected-component labelling algorithm is applied to segment spatially connected pixels into regions. If a region contains both 'certainly water' and 'potentially water' pixels that are spatially connected, it is labelled as surface water.

For our implementation we used the Modified Normalized Difference Water Index (mNDWI, Xu 2006) for our water index (WI) and defined our initial threshold (T_s) using scene-based Otsu thresholding.

3.3 Atmospheric correction

Atmospheric correction is crucial in assessing water parameters like SPM because the radiance measured by satellite sensors is predominantly influenced by the land and the atmosphere, with only a small fraction actually coming from the water surface. Without correction, the observed reflectance values would be inaccurate, leading to errors in estimating water parameters. Remote sensing reflectance (Rrs) values, obtained after atmospheric correction over water surfaces, isolate the water-leaving radiance, providing a more accurate representation of the water's optical properties. These corrected Rrs values are essential for reliably deriving water quality metrics such as SPM, ensuring that the data reflects the actual conditions in the water rather than atmospheric influences.

To isolate the water-leaving radiance we used the MAIN (Modified Atmospheric correction for INland waters) (Page et al., 2019). This method is designed to extract harmonized remote sensing reflectance (Rrs) products from Sentinel-2/MSI and Landsat-8/OLI sensors, particularly for inland water bodies. We expanded the original algorithm, which is provided as part of the GEOAquaWatch project, to further include the Landsat-7/ETM+ and Landsat-9/OLI-2 sensors.

Key components of the MAIN method include:

- Rayleigh Correction: This step corrects for scattering caused by air molecules, which is significant in the blue wavelengths.
- Aerosol Correction: Aerosol effects are corrected using the shortwave infrared (SWIR) bands, which have minimal water-leaving signals. This correction adjusts for the influence of aerosols on the reflectance data, which is particularly important for inland waters with variable optical properties.
- Reflectance Calculation: The water-leaving reflectance (ρw) is computed after atmospheric corrections and then converted to remote sensing reflectance (Rrs), which is used for subsequent water quality analysis.

Furthermore, the method incorporates auxiliary datasets listed in Table 2, which are ozone levels from the TOMS and OMI Merged Ozone Data dataset and digital elevation models (DEMs), to enhance the accuracy of the atmospheric correction.

Table 2 - Overview of auxiliary datasets used for the Rrs estimation

Dataset	Earth Engine Asset	Description
TOMS and OMI Merged Ozone Data	`"TOMS/MERGED"`	Estimate of total column ozone in Dobson units
ALOS DSM: Global 30m v3.2	`"JAXA/ALOS/AW3D30/V3_2"`	Global Digital Surface Model (DSM)

3.4 Empirical XGBoost model

This empirical SPM retrieval algorithm is based on XGBoost (Chen & Guestrin, 2016), a scalable end-to-end tree boosting system, which is used widely by data scientists to achieve state-of-the-art results on many machine learning (ML) challenges.

We setup an XGBoost random forest (RF) regressor with 400 trees. The input data was first augmented by adding additional spectral indices listed in Table 3. These act as additional input features for the regression model. Sampling aggregation is applied to extract matchup samples at the ground stations and to create a local-side dataset for model generation.

Table 3 - Overview of spectral indices used in SPM Model (XGBoost)

Spectral index	Formula
Normalized Suspended Material Index (NDSMI)	$\frac{Rrs_{red} + Rrs_{green} - Rrs_{blue}}{Rrs_{red} + Rrs_{green} + Rrs_{blue}}$
Normalized Difference Suspended Sediment Index (NDSSI)	$\frac{Rrs_{blue} - Rrs_{NIR}}{Rrs_{blue} + Rrs_{NIR}}$
Normalized Difference Turbidity Index (NDTI)	$\frac{Rrs_{red} - Rrs_{green}}{Rrs_{red} + Rrs_{green}}$

An optional Principal Component Analysis (PCA) can be applied to reduce the feature vector by lowering the number of input bands. Furthermore, due to the unbalanced distribution of EO samples between wet and dry season, the Synthetic Minority Over-Sampling Technique for Regression with Gaussian Noise (SMOGN) algorithm (Branco et al., 2017). The application of SMOGN can help in construction of a robust classifier by counteract training data imbalances with artificial samples.

The augmented feature vector is finally split with a 70%/30% train-test split. The resulting datasets are used for training and validation of the ML SPM model.

3.5 Empirical line fitting model

This empirical SPM retrieval algorithm based on spectral indices and empirical weights was originally proposed by Yu et al. (2019). An adapted version was later successfully applied in the Mekong basin by Guan et al. (2022).

The core of the algorithm is the Generalized Index, which represents a composite, weighted spectral index designed to generalize suspended sediment retrieval across a wide range of turbidity conditions by integrating multiple reflectance band ratios into a single index. These wavelengths correspond to different parts of the visible and near-infrared spectrum: 469 nm, 555 nm, 645 nm, and 859 nm. The algorithm combines these band ratios using a formula that includes weight factors

(ω1, ω2) and adjustable coefficients (c0, c1, c2), which are tuned during the calibration process. Figure 3 shows an example of the fitted GI for DSMP samples.

$$GI_{SPM} = c_0 \times \frac{R_{555}}{R_{469}} \times c_1 \times \omega_1 \times \frac{R_{645}}{R_{555}} \times c_2 \times \omega_2 \times \frac{R_{859}}{R_{555}}$$

$$\omega_1 = \frac{R_{645}}{R_{645} + R_{859}} \qquad \omega_2 = \frac{R_{859}}{R_{645} + R_{859}}$$

The benefit of this approach is that using band ratios enhances the algorithm's robustness against atmospheric contamination. The algorithm's flexibility is further improved by the incorporation of weights, which enable it to adjust its sensitivity to different spectral bands. Rather than requiring separate algorithms for varying conditions, the weights dynamically influence the contribution of each spectral band based on the observed reflectance. For instance, if more shortwave infrared (SWIR) reflectance is detected, the algorithm can give more weight to the coefficient associated with the SWIR band. This allows the algorithm to adapt to different SPM levels and water conditions.

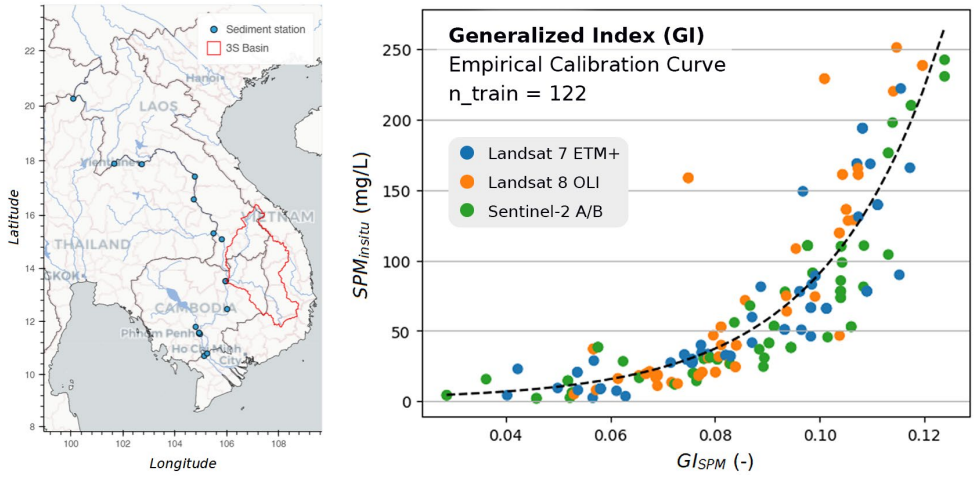

Figure 3 – Map of the Discharge and Sediment Monitoring Project (DSMP) ground stations used for calibration and validation of the EO prototype in the Mekong basin (left). And empirical calibration curve for Generalized Index (GI_{SPM}) to SPM conversion (right) derived from ± 1 day matchups of the satellite observations with depth-integrated in-situ measurements of suspended sediment concentration (SPM_{insitu}).

In this approach, we apply a 4-fold train–test split to the input data after subsampling and binning the dataset to obtain a balanced distribution and avoid bias in the curve fitting. For each fold, a training subset (approximately 25% of the balanced data) is used to calibrate the model coefficients, optimising the algorithm for the specific optical characteristics of the case study, while the remaining data are used to evaluate model performance. Figure 3 shows the training data for fold 1, comprising 122 matchups, together with the resulting empirically derived calibration curve. Each calibration curve is then evaluated on its corresponding held-out test subset, and the resulting performance metrics are aggregated to produce the final validation results.

3.6 Single-band semi-empirical model

We also implemented the Nechad single-band SPM algorithm (Nechad et al., 2010). This algorithm is a semi-empirical method that relates the reflectance observed by satellites to SPM concentrations using a predefined relationship. It is one of the most widely used algorithms for SPM estimation due to its simplicity and effectiveness across different water bodies.

$$C_{SPM} = \frac{A^{\rho}\,\rho_w}{1 - \rho_w / C^{\rho}} + B^{\rho}$$

$$\rho_w = \pi Rrs$$

The coefficients A^{ρ} , $B^{\rho}$and C^{ρ}rely on the selected band and are provided in look-up tables, with the offset coefficient B^{ρ} being 0 for satellite imagery.

Variations in sediment composition can cause shifts in the spectral reflectance. To take this into account we implemented the Dogliotti switching algorithm (Dogliotti et al., 2015) that combines two single-band algorithms and a linear weighting function to ensure more accurate estimation of SPM across different water types and sediment conditions.

$$C_{SPM} = \begin{cases} C_{SPM,645} & if\ \rho_{w,645} < 0.05 \\ (1-\omega)C_{SPM,645} + \omega C_{SPM,859} & if\ 0.05 < \rho_{w,645} < 0.07 \\ C_{SPM,859} & if\ \rho_{w,645} > 0.07 \end{cases}$$

$$\omega = \begin{cases} 0 & when\ \rho_{w,645} = 0.05 \\ 1 & when\ \rho_{w,645} = 0.07 \end{cases}$$

This algorithm generally provides a reliable tool for estimating SPM concentrations from remote sensing data. The Nechad algorithm was calibrated using a large dataset of in situ measurements that account for standard inherent optical properties (IOPs). Therefore, it is very important that the measurements of Rrs used in the algorithm are accurate and free from atmospheric contamination to ensure accurate estimates of SPM.

4 Results

1,146 algorithm validation matchups between satellite-derived SPM estimates and in situ measurements were generated using a ±1-day time window. When multiple satellite acquisitions occurred within this window, only the one closest in time to the in-situ measurement was selected to maximize temporal correlation. The match-ups were sampled within a polygonal region of interest (ROI), defined as a 100-meter-wide strip spanning the width of the river, centred on the sampling site. This approach accounts for potential signal variations across the river's width.

To evaluate the accuracy and performance of the SPM retrieval, we calculated several statistical measures based on these match-ups to provide a comprehensive understanding of the model's performance. The resulting performance metrics are shown in Table 4.

Table 4 - Validation results for SPM matchups

Model	R	R^2	Accuracy (±1σ)	RMSE
XGBoost	0.911	0.830	10.4 ± 12.9 mg/L	16.6 mg/L
Empirical line fitting	0.862	0.743	12.9 ± 16.9 mg/L	21.3 mg/L
Single-band semi-empirical model (Nechad)	0.775	0.60	16.1 ± 21.0 mg/L	26.5 mg/L

Our prototype validation reveals that the ML approach utilizing XGBoost outperformed the other tested methods for SPM extraction. ML regression models based on RF regression are generally recognized for their accuracy and robustness, requiring minimal model tuning. Conversely, we encountered significant challenges in properly tuning the Empirical line fitting approach. When applied solely to DSMP samples, it achieved an R^2 of 0.88 (refer to Figure 3). However, incorporating samples from additional stations diminished the validation performance, unlike the other approaches, suggesting that this method may lack robustness and offer lower generalization performance. The Nechad algorithm yielded reasonable results, and it is well-documented in the literature for generally achieving higher accuracies than was obtained in our approach. This reduced performance results on one hand from the depth integration of our reference measurements, as opposed to targeted sampling near the water surface. Furthermore, most matchup comparisons require reference samples to be taken on the same day, while we allow +/- 1 day.

In conclusion, our validation confirms that the SPM prototype, EOP1, is fully operational and performs within the required accuracy standards for suspended sediment extraction. When used with the best performing algorithm, the prototype demonstrates strong reliability in capturing SPM, delivering accurate results and robust performance, making it suitable for continued use within the project.

References

Beveridge, C. (2020). *Application-oriented approaches to modeling and satellite-based monitoring of watershed sediment dynamics*.

Branco, P., Ribeiro, R. P., Torgo, L., Krawczyk, B., & Moniz, N. (2017). SMOGN: a Pre-processing Approach for Imbalanced Regression. In *Proceedings of Machine Learning Research* (Vol. 74).

Chen, T., & Guestrin, C. (2016). XGBoost. *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 785–794. https://doi.org/10.1145/2939672.2939785

Dang, T. D., Vu, D. T., Chowdhury, A. F. M. K., & Galelli, S. (2020). A software package for the representation and optimization of water reservoir operations in the VIC hydrologic model. *Environmental Modelling & Software*, *126*, 104673. https://doi.org/10.1016/j.envsoft.2020.104673

Darby, S. E., Hackney, C. R., Parsons, D. R., & Tri, P. D. V. (2020). *Water and suspended sediment discharges for the Mekong Delta, Vietnam (2005-2015)*. NERC Environmental Information Data Centre. https://doi.org/10.5285/ac5b28ca-e087-4aec-974a-5a9f84b06595

Des E. Walling. (2008). The Changing Sediment Load of the Mekong River. *AMBIO: A Journal of the Human Environment*, *37*(3), 150–157.

Dogliotti, A. I., Ruddick, K. G., Nechad, B., Doxaran, D., & Knaeps, E. (2015). A single algorithm to retrieve turbidity from remotely-sensed data in all coastal and estuarine waters. *Remote Sensing of Environment*, *156*, 157–168. https://doi.org/10.1016/j.rse.2014.09.020

Fan, L., Dong, Y., Zhao, J., Geib, C., Wang, L., & Taubenbock, H. (2022). Mapping of Small Water Bodies with Integrated Spatial Information for Time Series Images of Optical Remote Sensing. *International Geoscience and Remote Sensing Symposium (IGARSS)*, *2022-July*, 6224–6227. https://doi.org/10.1109/IGARSS46834.2022.9883460

Foga, S., Scaramuzza, P. L., Guo, S., Zhu, Z., Dilley, R. D., Beckmann, T., Schmidt, G. L., Dwyer, J. L., Joseph Hughes, M., & Laue, B. (2017). Cloud detection algorithm comparison and validation for operational Landsat data products. *Remote Sensing of Environment*, *194*, 379–390. https://doi.org/10.1016/j.rse.2017.03.026

Guan, Q., Feng, L., Tang, J., Park, E., Ali, T. A., & Zheng, Y. (2022). Trends in River Total Suspended Sediments Driven by Dams and Soil Erosion: A Comparison Between the Yangtze and Mekong Rivers. *Water Resources Research*, *58*(10). https://doi.org/10.1029/2022WR031979

Nechad, B., Ruddick, K. G., & Park, Y. (2010). Calibration and validation of a generic multisensor algorithm for mapping of total suspended matter in turbid waters. *Remote Sensing of Environment*, *114*(4), 854–866. https://doi.org/10.1016/j.rse.2009.11.022

Page, B. P., Olmanson, L. G., & Mishra, D. R. (2019). A harmonized image processing workflow using Sentinel-2/MSI and Landsat-8/OLI for mapping water clarity in optically variable lake systems. *Remote Sensing of Environment*, *231*. https://doi.org/10.1016/j.rse.2019.111284

Pasquarella, V. J., Brown, C. F., Czerwinski, W., & Rucklidge, W. J. (2023). *Comprehensive quality assessment of optical satellite imagery using weakly supervised video learning*.

Tangi, M., Schmitt, R., Bizzi, S., & Castelletti, A. (2019). The CASCADE toolbox for analyzing river sediment connectivity and management. *Environmental Modelling & Software*, *119*, 400–406. https://doi.org/10.1016/j.envsoft.2019.07.008

Xu, H. (2006). Modification of normalised difference water index (NDWI) to enhance open water features in remotely sensed imagery. *International Journal of Remote Sensing*, *27*(14). https://doi.org/10.1080/01431160600589179

Yu, X., Lee, Z., Shen, F., Wang, M., Wei, J., Jiang, L., & Shang, Z. (2019). An empirical algorithm to seamlessly retrieve the concentration of suspended particulate matter from water color across ocean to turbid river mouths. *Remote Sensing of Environment*, *235*, 111491. https://doi.org/10.1016/j.rse.2019.111491

Disclaimer

Acknowledgement of funding

This project has received funding from the European Union's Horizon Europe research and innovation programme under grant agreement No 101059264.

This work has received funding from the Swiss State Secretariat for Education, Research and Innovation (SERI)

Project funded by

Schweizerische Eidgenossenschaft
Confédération suisse
Confederazione Svizzera
Confederaziun svizra

Swiss Confederation

Federal Department of Economic Affairs, Education and Research EAER
State Secretariat for Education, Research and Innovation SERI

SOS-Water – EOP2: Lake and River ice mapper prototype technical notes

[D3.2, Improved EO application prototypes]

Authors

Authors	Partner no	Partner organisation	Name of author
Main author	11	Eawag	Michael Brechbühler
Contributing author(s)	-	-	-

Document history

Date	Version	Chapters affected	Description of change	Author	Document status
12/08/2024	0.1	All	Initial version outlining document	Michael Brechbühler	FIRST DRAFT
23/09/2024	1.0	All	Edits and additions	Michael Brechbühler	FINAL

Table of Contents

Abbreviations

EO	Earth Observation
EOP	Earth Observation Prototype
ETM+	Enhanced Thematic Mapper Plus
GEE	Google Earth Engine
GDW	Google Dynamic World
GLCM	Gray-Level Co-occurrence Matrix
K-FOLD CV	K-fold Cross-validation
LEDAPS	Landsat Ecosystem Disturbance Adaptive Processing System
LIC	Lake Ice Coverage
LIP	Lake Ice Phenology
LOLO-CV	Leave-One-Lake-Out Cross-validation
LULC	Land Use Land Cover
mIoU	Mean Intersection over Union
MSI	Multispectral Instrument
OLI	Operational Land Imager
RIP	River Ice Phenology
RIC	River Ice Coverage
SAR	Synthetic Aperture Radar
SMILE	Statistical Machine Intelligence and Learning Engine
TIRS	Thermal Infrared Sensor

1 Background

This document provides technical details for an Earth Observation application prototype (EOP) developed for the SOS-Water project, for Lake and River ice mapping. It includes a detailed collection of technical notes that describe the methodology and validation of the prototype created as part of Task 3.2. The associated source code, maintained by Eawag (Partner No. 11), is available in the SOS-Water – Work package 3 GitHub code repository at gitlab.eawag.ch/surf/remote-sensing/sos-water/sosw_wp3.

Lake ice coverage (LIC) and river ice coverage (RIC) refer to the extent of the surface ice layer in waterbodies. These variables can be mapped as spatio-temporal timeseries data. To simplify the data and extract meaningful insights, aggregating the measurements to reflect relative waterbody coverage is commonly used. Additionally, processing the data into phenological attributes, referred to as lake ice phenology (LIP) and river ice phenology (RIP), is a standard approach. These estimates are crucial for understanding long-term trends and the impacts of climatic changes on ice coverage. However, ice records are often limited in both spatial and temporal coverage. Therefore, satellite remote sensing plays a critical role in providing frequent, comprehensive, and cost-effective measurements of ice cover.

Ice phenology measurements play a crucial role in assessing hydrological models, as they directly influence temperature distribution and the thermal stability of waterbodies. Additionally, river ice formation can lead to ice jams, which cause variations in water flow that are often overlooked in traditional hydrological modelling. Due to climatic conditions, ice-related variables in the SOS-Water project are particularly relevant for the Rhine and Danube basins—the former containing a significant number of high-altitude lakes, while the latter experiences river flows that tend to freeze during cold witnters. Thus, we have enhanced our existing algorithms for monitoring of ice cover and phenology. These algorithms utilize data from Sentinel-1 C-SAR, Sentinel-2 MSI, and Landsat-7/8/9 ETM+/OLI/OLI-2/TIRS/TIRS-2. Our updated multi-sensor machine learning model can estimate ice cover and phenology across the Alpine region from 2015 to the present. The model achieves a 20-meter spatial resolution and a daily to weekly temporal resolution, closing in on optimal target specifications for the Essential Climate Variable (ECV) (Global Climate Observation System, 2016).

2 Data

2.1 Webcam data

As calibration and validation dataset, high-resolution webcam imagery was collected for a total of 21 study lakes listed in Table 1, with daily images captured at 12:00 PM local time. The data was sourced from publicly available webcams that cover lakes within the study area. Only high-resolution imagery that provided a clear and interpretable representation of the lake's freezing state was used. The selection of training lakes was based on the availability of such webcams, with efforts made to include lakes of varying sizes, elevations, and geographic distribution to represent a broad range of conditions within the study area.

The observation period for the training lakes spans from the earliest start date of September 26, 2015, to the latest end date of January 5, 2023, providing an overall data collection period of over eight wintes. Across all lakes, a total of 117 winter cycles have been observed, where each cycle corresponds to a potential winter freezing cycle. In total, over 42,000 images have been collected for the analysis.

Table 1 - Table of the training lakes used for model generation and validation

ID	Lakename	Elevation (m a.s.l.)	Lake Area (km²)	Lat	Lon	Observation start	Observation end	Nr. of Seasons	EU-Hydro ID
1	Eibsee	973	1.77	47.456	10.973	2018-09-06	2023-01-04	5	IW35015086
2	Silsersee	1797	4.1	46.422	9.737	2017-11-01	2023-01-02	6	IW35016644
3	Silvaplanersee	1791	2.7	46.449	9.793	2020-11-04	2023-01-02	3	IW35016646
6	Schlegeisspeicher	1782	2.2	47.032	11.705	2017-01-02	2023-01-04	6	IW35016658
7	Turrachsee	1763	0.19	46.919	13.876	2016-01-04	2023-01-05	7	IW35000836
8	Stausee Mooserboden	2036	1.6	47.159	12.713	2015-09-26	2022-12-16	7	IW35016679
9	Stausee Wasserfallboden	1672	1.5	47.184	12.720	2020-11-11	2022-12-22	2	IW35016682
10	Zeller See	750	4.55	47.323	12.806	2016-01-01	2023-01-05	7	IW35015047
11	Lac d'Emosson	1930	3.27	46.084	6.914	2018-11-22	2023-01-04	5	IW11004981
12	Lac de Salanfe	1925	1.62	46.141	6.957	2020-07-01	2023-01-04	3	IW11003945
13	Lac de Tseuzier	1777	0.85	46.351	7.430	2021-07-12	2023-01-04	2	IW11003946
14	Hopfensee	783	1.94	47.601	10.672	2014-12-14	2023-01-03	9	IW35015142
15	Lago Serrù	2275	0.54	45.458	7.125	2017-10-15	2023-01-05	6	IW24000007
16	Lünersee	1970	1.12	47.053	9.753	2016-11-17	2023-01-05	7	IW26009826
17	Lago di Alleghe	998	0.5	46.407	12.016	2017-10-29	2023-01-05	6	IW24002264
18	Silvretta-Stausee	2030	1.31	46.907	10.094	2017-08-16	2023-01-05	6	IW26010769
19	Stubenbergersee	360	0.4	47.230	15.799	2018-11-22	2023-01-05	5	IW35015053
20	Ritzensee	747	0.06	47.419	12.847	2016-08-30	2023-01-05	7	IW35000850
22	Lac Blanc	2352	0.29	45.113	6.105	2019-10-12	2023-01-05	4	IW11002913
23	Lac de Montriond	1060	0.32	46.209	6.729	2017-11-17	2023-01-05	6	IW11003936
25	Lac de Goillet	2517	0.48	45.929	7.665	2015-07-09	2023-01-05	8	IW24000795

2.2 EO data

The EO input data for our model are multispectral, thermal and synthetic aperture radar (SAR) data from the Landsat-7/-8/-9, Sentinel-1A and 1B and Sentinel-2A and 2B satellite platforms. The used datasets and their specifications are listed in Table 2.

Table 2 - Overview of the used GEE EO datasets in the ice extraction workflow.
**TIR bands are up-sampled to 30 meters from a coarser native resolution*

Mission(s)	Landsat-7	Landsat-8	Landsat-9	Sentinel-1A and 1B	Sentinel-2A and 2B
Sensor(s)	Enhanced Thematic Mapper Plus (ETM+)	Operational Land Imager (OLI) 1 and Thermal Infrared Sensor (TIRS) 1	OLI-2 and TIRS-2	C-band Synthetic Aperture Radar (C-SAR)	Multispectral Instrument (MSI)
GEE identifier	LANDSAT/ LE07/C02/T1_L2	LANDSAT/ LC08/C02/T1_L2	LANDSAT/ LC09/C02/T1_L2	COPERNICUS/ S1_GRD_FLOAT	COPERNICUS/ S2_SR
Dataset start	28 May 1999	18 March 2013	31 October 2021	3 April 2014	28 March 2017
Dataset end	L1C2 (Radiance)	ongoing	ongoing	ongoing	ongoing
Spatial resolution (meters)	30 60(TIR)*	30 100(TIR)*	30 100(TIR)*	10	10 (Red, Green, Blue) 20 (others) 60 (Aerosol, Cirrus)
Repeat cycle	16 days	16 days	16 days	6 days (each)	5 days (each)
Estimated actual observation frequency	44.6 days	33.8 days	35.5 days	3.80 days	7.17 days
Spectral bands	7 visible to SWIR, 1 thermal	7 visible to SWIR, 1 thermal	7 visible to SWIR, 1 thermal	2 C-band	13 visible to SWIR

All used satellite platforms are in a near-polar, sun-synchronous orbit. This means they pass over the same area of the Earth at the same local solar time each day. For the Landsat and Sentinel-2 satellites this crossing at the equator occurs between 10:00 AM and 10:30 AM local solar time, since acquisitions are only made during the descending node. On the other hand, Sentinel-1 C-SAR has a passing at around 06:00 PM and 06:00 AM for ascending and descending node, respectively. Figure 1 shows the achievable data coverage based on the merging of the data with our multi-sensor approach. The coverage was calculated for the period of 2022-2024, measured over the Alpine Convention Perimeter, the extent of the European Alps. The achieved per-sensor actual observation frequencies are listed in Table 1. Data loss due to clouds and cloud shadows was taken into account in multispectral and thermal imagery. The assessment shows that an estimated median monthly coverage of 39.7% and estimated median frequency of 2.38 days can be achieved with this sensor constellation over the study area.

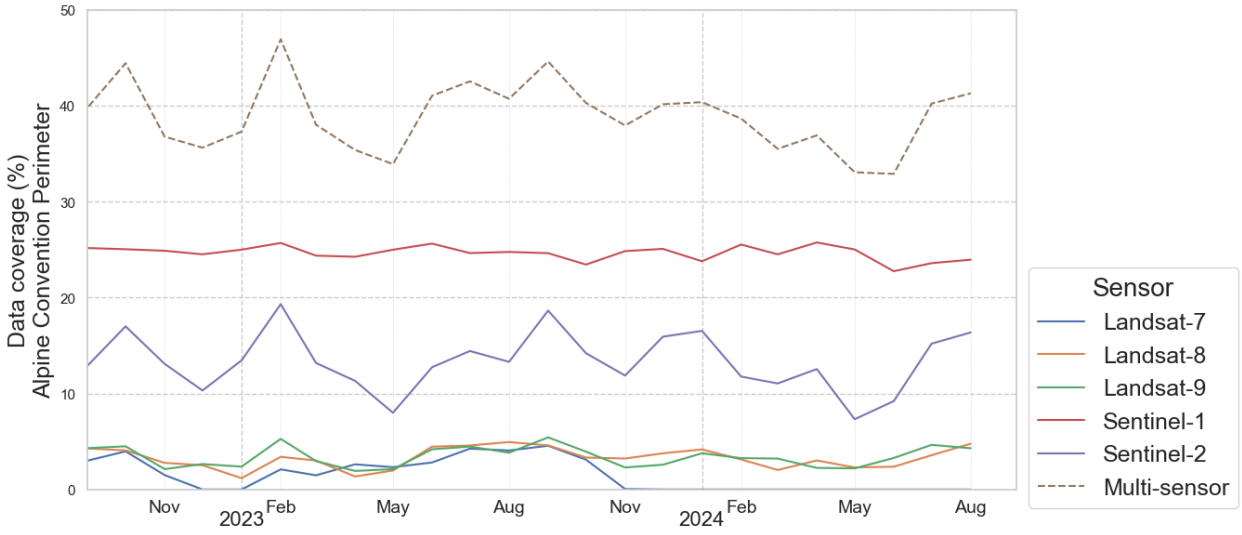

Figure 1 – Data coverage assessment of multi-sensor approach in the Alpine Convention Perimeter. Coverage is calculated as monthly mean aggregates of spatially aggregated daily data availability. For example, for daily pixels in February 2024 there is a 38% chance that the data is available.

2.3 Auxiliary data

As auxiliary data, we utilize several datasets to enhance the analysis listed in Table 3. ERA5-hourly, hosted on Google Earth Engine, provides global, hourly estimates of atmospheric variables. In this study, we specifically use temperature and wind data, which are key factors influencing ice cover dynamics and its remotely sensed signal. To accurately delineate waterbodies, we rely on the vector data from the EU-Hydro dataset, which offers high-precision geographic boundaries of rivers and lakes across Europe, ensuring precise identification of waterbody areas for ice monitoring. Additionally, to account for seasonal variability, we incorporate Google Dynamic World v1, a near real-time dataset based on Sentinel-2 imagery that offers global land cover classifications at a 10-meter resolution.

Table 3 - Overview of the used auxiliary datasets

Dataset	Description	Type	GEE identifier
ERA5-Land Hourly - ECMWF Climate Reanalysis	Reanalysis dataset providing land variables over several decades.	raster	ECMWF/ERA5_LAND/HOURLY
Google Dynamic World v1	10m near-real-time (NRT) Land Use/Land Cover (LULC) dataset based on Sentinel-2	raster	GOOGLE/DYNAMICWORLD/V1
EU-Hydro River Network Database 2006-2012 v1.3 (European Environment Agency, 2019)	Vector dataset of photo-interpreted river network, consistent of surface interpretation of water bodies (lake and rivers)	vector	-

3 Methodology

Our ice mapping algorithm leverages EO datasets from the Google Earth Engine (GEE) platform. The process involves a large variety of steps such as resampling, cloud masking, water masking, texture generation, image classification and data aggregation. All these computationally expensive processing steps are executed within the Google Earth Engine environment. GEE's cloud-based infrastructure allows for efficient processing of large datasets, enabling rapid execution of these algorithms across extensive spatial and temporal scales. This efficiency is key to producing timely and reliable ice coverage data for monitoring and analysis purposes. Figure 2 shows a detailed and almost comprehensive workflow of EOP2.

3.1 Algorithm description

The process begins by preparing the imagery from various EO data inputs. For multispectral and thermal data, cloud masking is applied to remove pixels affected by clouds. SAR data undergoes pre-processing steps such as Refined Lee speckle noise filtering and border noise removal to improve data quality. In addition, texture features are generated from SAR data using the Gray-Level Co-occurrence Matrix (GLCM) and neighbourhood-based metrics to capture spatial patterns indicative of ice and water. Temporal features are also integrated, with lag and lead variables added to account for temporal changes in ice cover and temporal aggregates to help further refine the detection of ice events over time.

For model generation and validation, we use high-resolution webcam imagery as ground truth data, manually labelled to reflect the freezing state of the lakes. These labels are then temporally matched with the pre-processed EO data to create training and test datasets. To ensure robust model validation, we employ a K-fold cross-validation approach using 10 fols. Given the class imbalance, where unfrozen features are more common than frozen ones, we apply class balancing by under sampling the majority class (unfrozen) to match the size of the minority class (frozen).

The ice classification itself is performed using three supervised Random Forest (RF) binary classifiers, built with the Statistical Machine Intelligence and Learning Engine (SMILE) library RF implementation in GEE. These classifiers are trained separately on Landsat, Sentinel-2, and Sentinel-1 imagery. We conduct hyperparameter tuning to find the optimal model parameters for each of the RF models. Once trained, the classifiers are applied to the prepared spatiotemporal EO data, where pixel-wise binary classification is performed to distinguish between water (class 0) and ice (class 1). Each pixel is assigned one of the classes, and a probability score is generated based on the voting process within the Random Forest network.

After the initial pixel-wise classification, the next stage involves aggregating the generated maps to obtain daily waterbody-level (lake or river) ice cover information. This step retrieves the GEE server-side processed data and pulls it to the local environment. For quality control, we filter out any datapoints where less than 30% of the waterbody area is covered by valid data. From the remaining data, an aggregated ice cover probability is calculated based on the pixel classification probabilities.

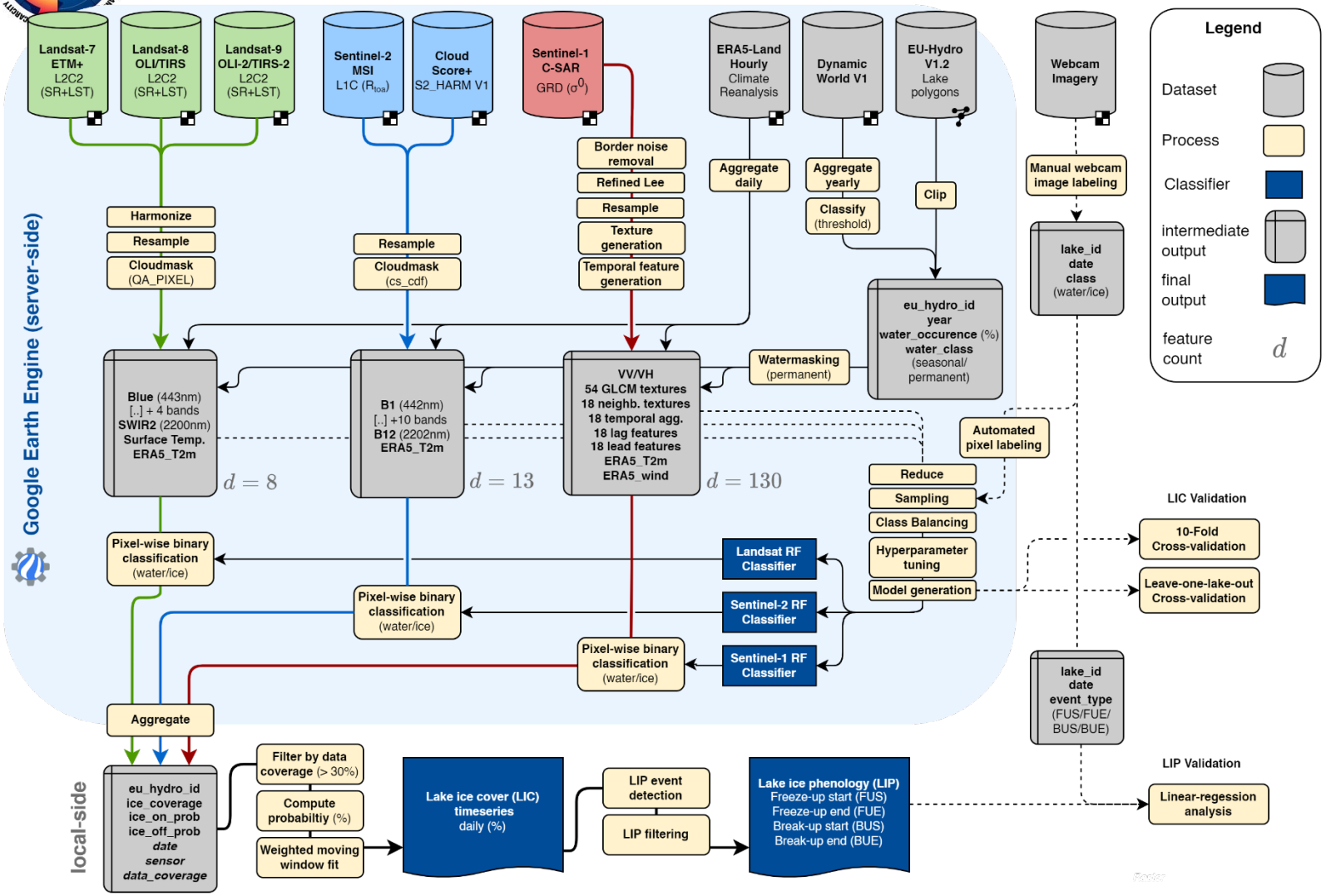

Figure 2 - Ice classification and extraction workflow

To refine this, we apply a weighted moving window fit. This process uses the probability of each data point, along with predefined sensor-specific weights, to extract a unified ice cover signal across all available data from the three EO datasets. The result is one of the two main outputs of our model: a time series representing the daily relative ice coverage for each lake.

In addition to ice coverage, we extract key phenological events using a Lake Ice Phenology (LIP) detection algorithm. This algorithm identifies crucial moments in each freezing cycle, including Freeze-up Start (FUS), Freeze-up End (FUE), Break-up Start (BUS), and Break-up End (BUE). After applying a post-filter step, this process yields our second main output: ice cycles, represented by the timing of these four phenological events for each freezing cycle.

3.2 Cloud masking

For the masking of Landsat data, we utilize the CFMASK (C Function of Mask) algorithm (Foga et al., 2017). CFMASK is part of the Landsat Collection 2 processing and identifies clouds and cloud shadows in the Landsat imagery. For the masking we remove pixels by chaining logical OR operators on active quality-assessment flags “Dilated Cloud” (bit 1), “Cirrus (high confidence)” (bit 2), “Cloud” (bit 3), “Cloud Shadow” (bit 4). An additional cloud-buffer can be supplied in the masking function to oppose under detection. For validation purposes we set this value to 30-meters.

For Sentinel-2 data we use the Cloud Score+ (CS+) dataset. The CS+ algorithm leverages a combination of spectral bands, particularly the blue, red, and shortwave infrared bands, to calculate a 10-meter cloud probability score based on Sentinel-2 MSI imagery. It uses a weakly supervised deep learning approach to grade the quality of individual pixels through time and provides a robust way to detect and mask clouds and cloud shadows. For our prototype we use the CS+ quality assessment band cs_cdf with a threshold value of 0.6 where 0 is occluded and 1 represents clear.

3.3 Water masking

The waterbody boundaries provided by the used EU-Hydro dataset are static, which means that seasonal changes in water levels within these polygons can introduce land-contaminated pixels. This poses a challenge for our binary classifier, which is trained only on water and ice-covered water surfaces. To address seasonal water variations, we use the dynamic water classification from the Google Dynamic World (GDW) v1 dataset. The dataset is aggregated into yearly image stacks, and for each stack, the mean occurrence of water pixels is calculated. This produces an output representing the mean water occurrence over each year. To account for potential misclassification, a 90% threshold is applied to distinguish between seasonal and permanent waters. An example for Lake Grimsel is shown in Figure 3. During this process, the snow/ice class from the GDW classification is masked out. This approach mirrors the methodology used in the JRC Global Surface Water Mapping Layers (Pekel et al., 2016), which was initially considered but ultimately not used due to infrequent

updates and possible discontinuation of the dataset.

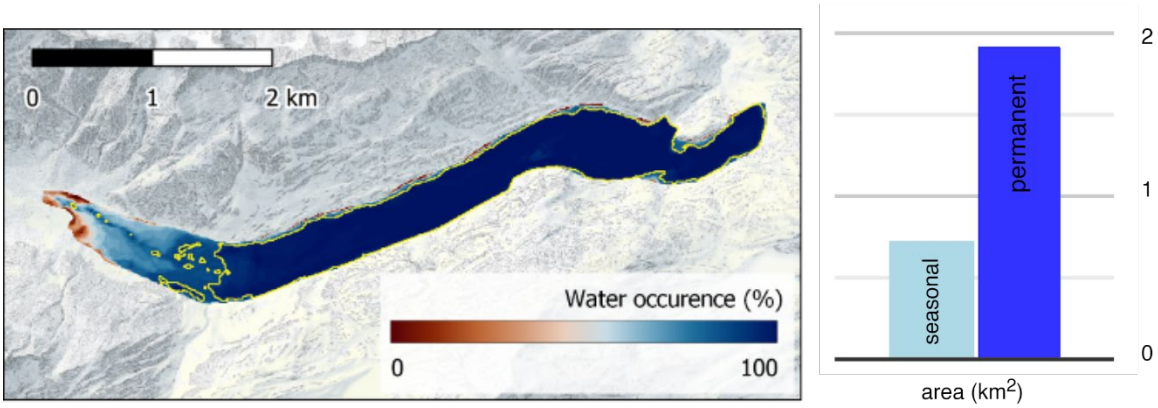

Figure 3 – Example of the water occurrence at Lake Grimsel (Guttannen, CH) calculated based on Google Dynamic World v1. Histogram (right) shows the associated area of seasonal and permanent water (indicated by yellow boundary) using a 90% occurrence threshold.

3.4 Validation

The final step of the workflow focuses on validating the model's performance for both ice coverage and phenological event detection. For ice coverage, we evaluate the performance of the Random Forest classifier using standard metrics such as accuracy and mean Intersection over Union (mIoU). We perform a 10-fold cross-validation to assess the model's overall performance, along with a Leave-One-Lake-Out Cross-Validation (LOLOCV) to test how well the model generalizes across different lakes.

To validate the phenology outputs, we utilize the previously generated ground truth labels from webcam imagery. The labels are used to identify phenlogy events and are compared with the model's estimated event dates. This allows us to assess the accuracy of the phenological event detection by matching observed freezing and thawing dates with those predicted by the model.

3.5 River ice cover

To evaluate whether our lake ice model can be adapted for the assessment of river ice coverage, we applied it to a river segment near the Tulcea station (Tulcea, Romania). Literature on historical river ice coverage and ice jams in the Danube basin shows that an obstruction at this segment occurred during the winter of 2016-2017, lowering the discharge of the Saint George branch of the Danube River (Ionita et al., 2018). This assessment provides a qualitative evaluation of our lake-trained prototype's potential for extracting river ice information.

4 Results

4.1 Ice coverage validation

Table 4 shows the k-fold cross-validation results of the binary classifier for the lake ice extraction model across the three sensor datasets. This validation assesses the overall model performance by applying a 10-fold procedure with repeated random train-test splits on the test lakes. A selection of widely used and well-established validation metrics for classifier models was applied. The overall accuracy reflects the proportion of correctly classified ice and no-ice observations. Kappa provides a measure of agreement between the predicted and actual classifications, accounting for chance. Sensitivity, or the true positive rate, indicates the classifier's ability to correctly identify ice-covered areas, while specificity, or the true negative rate, measures its ability to correctly classify non-ice areas. Lastly, the Intersection over Union (IoU) assesses the overlap between predicted and actual ice cover, offering a robust metric for spatial classification performance.

Table 4 - Validation results for k-fold cross-validation

Sensor dataset	Overall accuracy (%)	Kappa	Sensitivity (%)	Specificity (%)	IoU (%)
Landsat	99.7	0.99	99.7	99.7	99.5
Sentinel-1	85.4	0.71	83.0	87.8	74.7
Sentinel-2	98.3	0.97	97.8	98.8	96.6

The Leave-One-Lake-Out cross-validation (LOLOCV) evaluates the model's ability to generalize to unseen lakes by testing its performance on a validation set, where each fold corresponds to data from individual lakes. The model achieved mean Intersection over Union (IoU) scores of 96.3%, 92.2%, and 71.7% for Landsat, Sentinel-2, and Sentinel-1, respectively. These results indicate strong generalization performance, ranging from excellent to good across the datasets. Figure 3 displays the LOLO-CV assessment using the unaggregated IoU metric for each individual lake.

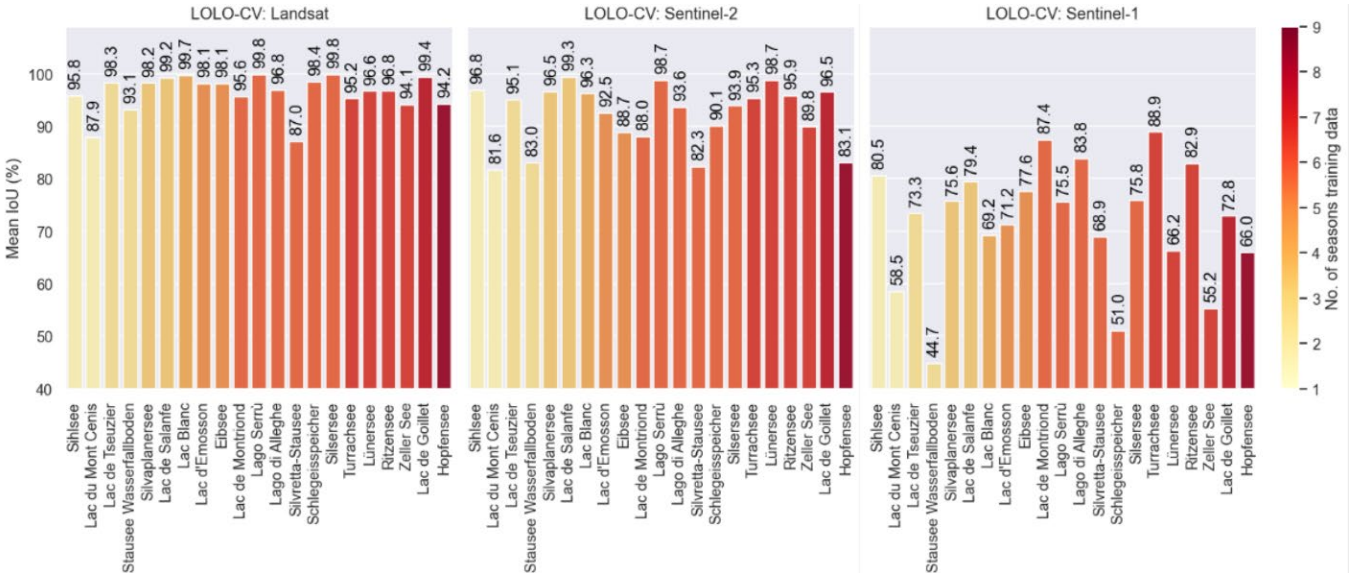

Figure 4 - Leave one lake out cross-validation of LIC

4.2 Ice phenology validation

We were able to create match-ups with 78 observed LIP events extracted from webcam imagery with our remotely sensed estimates. Our error assessment showed an event detection accuracy of -0.17 ± 4.3 days (1σ). Figure 5 shows a histogram of the deviations between estimated and observed LIP events.

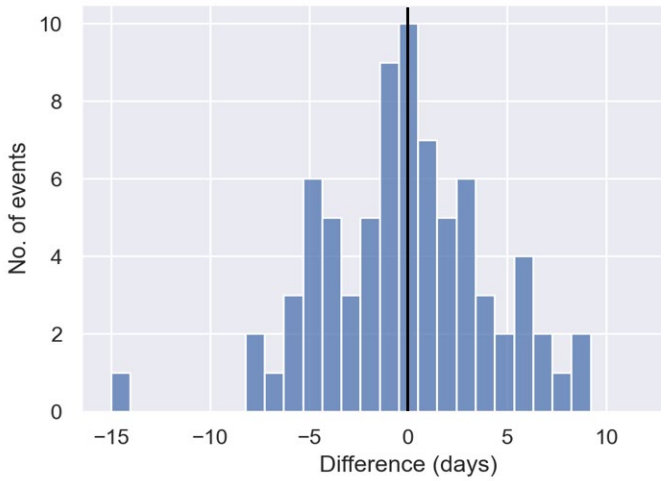

Figure 5 - Histogram of match-up deviations of ice phenology events

The linear regression analysis result, shown in Figure 4, yielded per-event R^2 values ranging from 0.93 to 0.97, indicating excellent model performance on all phenological event types. When assessing all event types combined, the fitted R^2 reached 0.98.

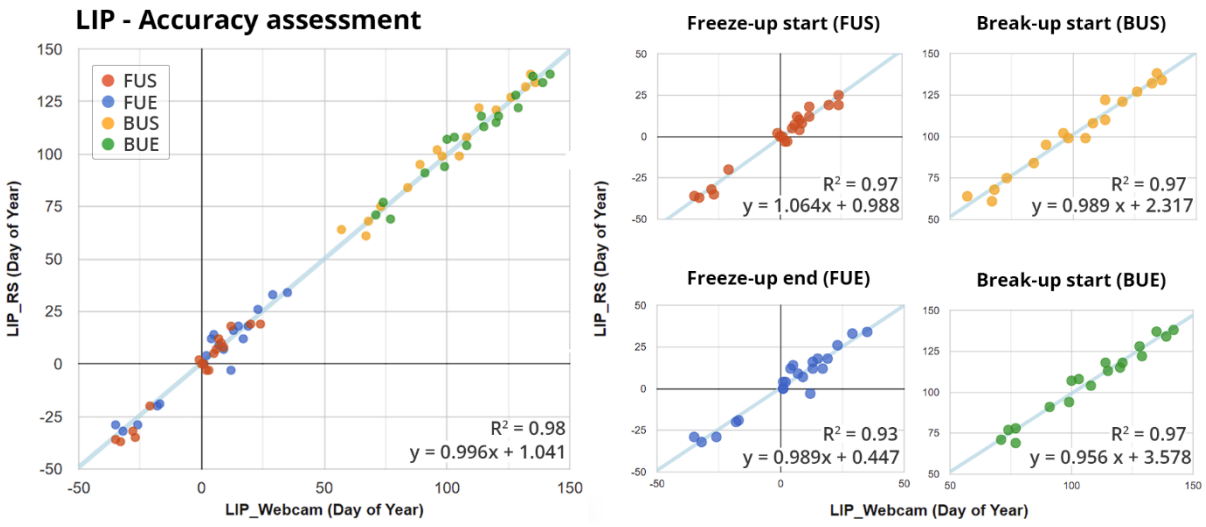

Figure 6 - Linear regression analysis on match-ups between observed and estimated phenology event dates

4.3 River ice cover

Figure 6 displays the aggregated river ice coverage over the Saint George Branch of the Danube River at Tulcea station (Tulcea, Romania) for the 2016-2017 winter. The results show a positive detection of the ice obstruction, indicating that our prototype can be effectively transferred to detect river ice coverage.

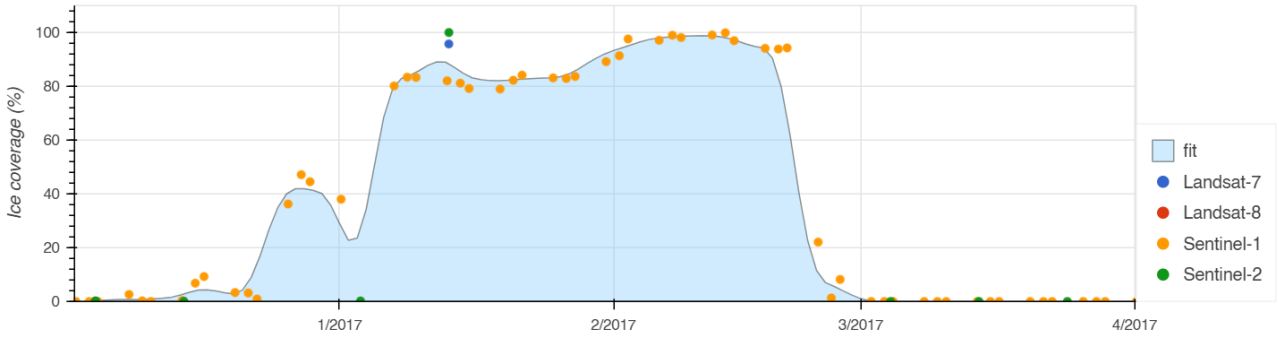

Figure 7 - River ice coverage at Tulcea station (Tulcea, Romania) river segment

References

European Environment Agency. (2019). *EU-Hydro River Network Database 2006-2012 (vector), Europe - version 1.3, Nov. 2020*. https://doi.org/10.2909/393359a7-7ebd-4a52-80ac-1a18d5f3db9c

Foga, S., Scaramuzza, P. L., Guo, S., Zhu, Z., Dilley, R. D., Beckmann, T., Schmidt, G. L., Dwyer, J. L., Joseph Hughes, M., & Laue, B. (2017). Cloud detection algorithm comparison and validation for operational Landsat data products. *Remote Sensing of Environment*, *194*, 379–390. https://doi.org/10.1016/j.rse.2017.03.026

Global Climate Observation System. (2016). *The Global Observing System for Climate: Implementation Needs GCOS-200 (GOOS-214)*. https://gcos.wmo.int/en/publications/gcos-implementation-plan2022

Ionita, M., Badaluta, C. A., Scholz, P., & Chelcea, S. (2018). Vanishing river ice cover in the lower part of the Danube basin-signs of a changing climate. *Scientific Reports*, *8*(1). https://doi.org/10.1038/s41598-018-26357-w

Pekel, J. F., Cottam, A., Gorelick, N., & Belward, A. S. (2016). High-resolution mapping of global surface water and its long-term changes. *Nature*, *540*(7633), 418–422. https://doi.org/10.1038/nature20584

Disclaimer

Acknowledgement of funding

This project has received funding from the European Union's Horizon Europe research and innovation programme under grant agreement No 101059264.

This work has received funding from the Swiss State Secretariat for Education, Research and Innovation (SERI)

Project funded by

Schweizerische Eidgenossenschaft
Confédération suisse
Confederazione Svizzera
Confederaziun svizra

Swiss Confederation

Federal Department of Economic Affairs, Education and Research EAER
State Secretariat for Education, Research and Innovation SERI

SOS-Water – EOP3: SWOT-based altimetry prototype technical note

[D3.2, Improved EO application prototypes]

Authors

Authors	Partner no	Partner organisation	Name of author
Main author	6.1	FutureWater-NL	Jorge Portugués
Contributing author(s)	6.1	FutureWater-NL	Corjan Nolet

Document history

Date	Version	Chapters affected	Description of change	Author	Document status
28/08/2024	0.1	All	Initial version outlining document	Jorge Portugués, Corjan Nolet	DRAFT
23/09/2024	1.0	All	Final version	Jorge Portugués, Corjan Nolet	FINAL

Table of Contents

Abbreviations

API	Application Programming Interface
CNES	Centre National d'Études Spatiales (French Space Agency)
EO	Earth Observation
EOP3	Earth Observation Prototype 3
EPSG	European Petroleum Survey Group (coordinate reference system)
ETRS89	European Terrestrial Reference System 1989
EVRF2007	European Vertical Reference Frame 2007
GE	Germany
GIS	Geographic Information System
GRAND	Global Reservoir and Dam Database
GROD	Global River Obstruction Database
GRWL	Global River Widths from Landsat
IGN	Institut Géographique National (French mapping agency)
IMS	Impact Models
IWMS	Integrated Water Modelling System
JPL	Jet Propulsion Laboratory
KARIN	Ka-band Radar Interferometer
L2_HR_LAKESP	Level 2 High-Rate Lake Single-Pass Vector Product
L2_HR_RIVERSP	Level 2 High-Rate River Single-Pass Vector Product
L2_HR_PIXC	Level 2 High-Rate Pixel Cloud Product
NAP	Normal Amsterdam Level
NIHWM	National Institute of Hydrology and Water Management (Romania)
PLD	Prior Lake Database
PO.DAAC	Physical Oceanography Distributed Active Archive Center
PRD	Prior River Database
RS	Remote Sensing
SWORD	SWOT River Database
SWOT	Surface Water and Ocean Topography
WFS	Web Feature Service
WSE	Water Surface Elevation
U-RMSE	Unbiased Root Mean Square Error
UCLA	University of California, Los Angeles
WGS84	World Geodetic System 1984

1 Maintainer and code access

This document provides the technical notes for Earth Observation (EO) Prototype 3 (EOP3) – SWOT-based altimetry. It includes a detailed collection of technical notes that describe the methodology and validation of the prototype created as part of Task 3.2 of the SOS-Water project. The associated source code, maintained by FutureWater-NL (Partner No. 6.1), is available in the SOS-Water – Work package 3 GitHub code repository at gitlab.eawag.ch/surf/remote-sensing/sos-water/sosw_wp3.

2 General description

The prototype developed for handling Surface Water and Ocean Topography (SWOT) data is designed to streamline the visualization and extraction of satellite-derived information related to rivers and lakes hydrological features over time. This approach facilitates a programmatic workflow that enables the user's ability to quickly access and analyze satellite data without the need for complex processing. The process begins when a user defines a specific area of interest by inputting a bounding box, which outlines the coordinates of the region. This bounding box is then used to query a pre-existing database that contains unique identifiers (IDs) and geometry details for rivers and lakes. The application matches the bounding box with the relevant IDs and retrieves the corresponding water body information, ensuring that only such features within the defined area are considered.

The same ID allows making an additional request to an Application Programming Interface (API) that gathers relevant hydrological SWOT data in the area of interest, providing measurements for a period of choice. The prototype delivers detailed key hydrological metrics that can be visualized interactively, showing changes in river and lake hydrology over time, or they can be downloaded for further analysis. The core of the process involves accessing pre-processed SWOT data products, which have been refined by specialized algorithms designed to interpret the raw satellite radar measurements. These products are often in the form of tiles or granules, which require ample filtering and processing for a specific feature. This automated, programmatic approach significantly reduces the complexity and time involved in accessing this data, providing users with a straightforward approach to visualize, download, and further analyze hydrology parameters such as water surface elevation (WSE), river width and slope, and lake/reservoir area.

The SWOT mission is continually updating its data products with new version releases, which incorporate improvements and refinements based on ongoing research and advancements in processing techniques. In future releases, river data products will be expanded to include discharge measurements, while lake data products will offer insights into storage change, enhancing the application's capabilities in monitoring water resources.

3 Objective

The objective of the prototype is to facilitate the integration of SWOT hydrological measurements to improve EO-based indicators in a remote sensing (RS) context for accurate parametrization of the Integrated Water Modelling System (IWMS). Enhancing the knowledge and applying cutting-edge EO data streams is within the expected outcomes of SOS-Water, for which the prototype aims to identify which parameters can be used to improve validation and benchmarking of the impact models (IMS) per case study. These include one location for the Rhine and the Danube Delta, while three locations have been identified for the Upper Danube to validate the results.

The application aims to serve the explicit data needs of the modelling community.

4 Data

4.1 In situ data

The in-situ data used to validate the application are hydrological stations in the Danube and Rhine Rivers that take measurements of water level in centimeters, relative to a gauge zero or zero level point reference placed under the riverbed.

4.1.1 Upper Danube

Donauwörth, Germany

The first station in the Upper Danube is located in the town of Donauwörth in Germany and managed by the Water Management Authorities in Bavaria. With coordinates EPSG:4937 ETRS89 Lon 10°48'5.623" Lat 48°42'39.125", this station aggregates water level daily mean relative to its gauge zero of 394.77 meters, referenced to EVRF2007.

Wilhering, Austria

The second station is located in the town of Wilhering in Upper Austria and managed by via donau – Austrian Waterways GmbH. With coordinates EPSG:4937 ETRS89 Lon 14°10'52.942" Lat 48°19'47.786", this station measures the water level every 15 minutes relative to a zero level point of 248.76 meters, referenced to EVRF2007.

Linz, Austria

The third station is located in the city of Linz in Upper Austria and is also managed by via donau – Austrian Waterways GmbH. With coordinates EPSG:4937 ETRS89 Lon 14°17'4.866" Lat 48°18'24.919", this station measures the water level every 15 minutes relative to a zero level point of 247.38 meters, referenced to EVRF2007.

4.1.2 Danube Delta

Isaccea, Romania

The station selected for the Danube Delta case study is located in the town of Isaccea in Romania and managed by the Romanian National Institute of Hydrology and Water Management (NIHWM). With coordinates EPSG:4937 ETRS89 Lon 28°27'13.32" Lat 45°17'20.4", it measures the water level once per day at 5:00 PM relative to its gauge zero of 0.703, referenced to EVRF2007.

4.1.3 Rhine

Lobith, The Netherlands

The station used for the Rhine River case study is located in the town of Lobith, where the river enters The Netherlands. It is managed by the Ministry of Infrastructure and Water Management of the country. Located at coordinates EPSG:4937 ETRS89 Lon 6°5'29" Lat 51°51'16", the in situ water level data is provided every ten minutes and referenced to Normal Amsterdam Level (NAP).

4.2 EO data

There are two sources of EO data accessed by the prototype. The first is the prior databases, where ID and feature geometry are found and available in the Theia catalog, maintained by CNES. The second is granules of river and lake products that are generated from a prior pixel cloud of SWOT measurements. These products are maintained by the Physical Oceanography Distributed Active Archive Center (PO.DAAC).

4.2.1 Prior Databases

The first database utilized in the SWOT mission is the static Prior River Database (PRD), known as the SWOT River Database (SWORD). Developed by the SWOT Science Team, (Altenau et al., 2021), SWORD serves as the foundation for identifying and analyzing river features. It contains river reach and node geometry and metadata. It is built upon several datasets, including the Global River Widths from Landsat (GRWL), which provides measurements of river widths worldwide, MERIT-Hydro, a high-resolution hydrological model that offers detailed river network information, and the Global River Obstruction Database (GROD), which catalogs dams and other man-made river blockages.

The second database is the Prior Lake Database (PLD), providing a global inventory of lakes and reservoirs, including polygons and metadata that help track and analyze changes over time. The PLD is generated by integrating multiple global and regional hydrographic datasets into a unified product. It draws lake features and names from sources such as OpenStreetMap, IGN Carthage, the Global Lakes and Wetlands Database (GLWD), and the vMap0 database. Additionally, reservoirs are incorporated from the Global Reservoir and Dam (GRanD) database. The PLD is compatible with the SWOT Prior River Database (SWORD), which ensures consistency across SWOT's hydrological measurements. The current version is primarily based on the CIRCA-2015 lake water extent map, created by UCLA.

Table 1 – Databases accessed

Data source	Product
University of North Carolina	SWOT Prior River Database (PRD) - SWORD
CNES	Prior Lake Database - PLD

4.2.2 SWOT Products

Using interferometry, SWOT measures surface water elevation with great accuracy, surpassing previous altimetry missions (Biancamaria et al., 2016). With a duration of 3 years, the science orbit has a 21-day repeat cycle to cover the entirety of the globe. On-board sensors of SWOT include KaRIn, a Ka-band synthetic aperture radar (SAR) interferometric system with two swaths of 50 km each (10-60 km from nadir) as seen in Figure 1 (Fjortoft et al., 2014). This system measures inland water features.

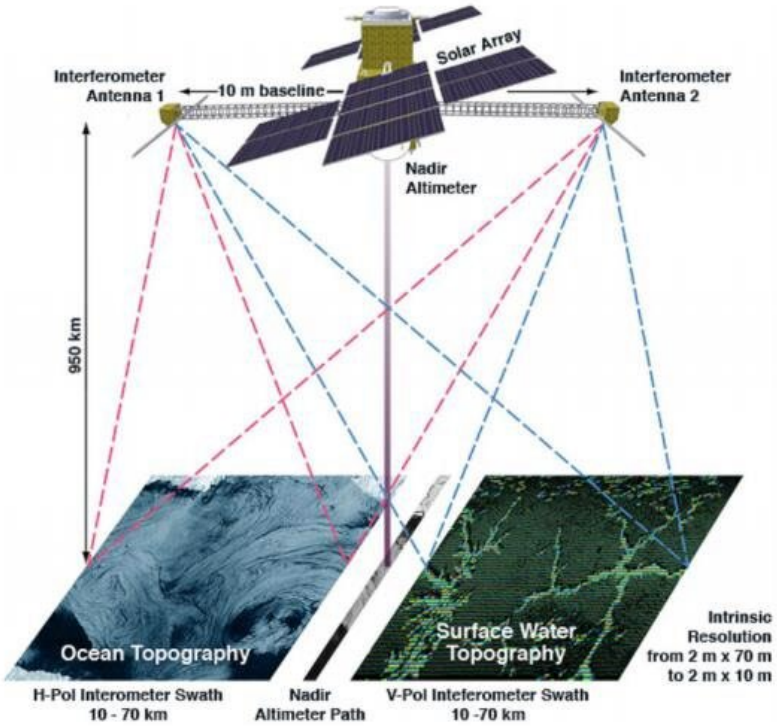

Figure 1 – Overview of the SWOT satellite and its instruments

Because of near-nadir geometry of SAR measurements, ground range resolution varies from 70 m near swath to 10 m far swath. Surface Water Topography measurements for SWOT aim to provide a global inventory of all terrestrial water bodies whose surface area exceeds 0.0625 km^2 for lakes, reservoirs, and wetlands, and >10 km long river reaches whose width exceeds 100 meters (JPL D-109532, 2024). SWOT was launched on the 16th of December of 2022 into a 1-day repeat orbit for calibration and fast-sampling of the mission. After this phase, SWOT entered its science 21-day repeat orbit in August 2023. There are several level 2 standard science products made available since August 2023 by the Physical Oceanography Distributed Active Archive Center (PO.DAAC) that are available for analysis, with the most relevant to hydrology and water management applications described in the next sections.

L2_HR_RiverSP

The Level 2 Ka-band Radar Interferometer (KaRIn) high-rate river single pass vector product provides key hydrological measurements such as water surface elevation (WSE), slope, and width for river reaches and nodes. These reaches are identified by the SWOT River Database (SWORD), a static prior river database (PRD). The location of each reach is defined by centerlines in the PRD, and updates to this database result in new versions of the L2_HR_RiverSP product. The product's granules cover a 128

km wide swath in the cross-track direction, and data is provided in Esri GIS shapefile format, either as reach or node shapefiles. Each reach is divided into static nodes spaced 200 meters apart, with a unique identifier (ID) for each reach and node based on the Pfafstetter coding system (Verdin & Verdin, 1999). These identifiers help locate reaches and match them to hydrological features.

Measurements in the L2_HR_RiverSP product are derived from pixels from a prior pixel cloud product (L2_HR_PIXC) (JPL D-56411JPL, 2023). It contains arrays of measured height, geolocation, and classification data from KaRIn, and aggregated in the aforementioned pixel cloud. This information is used to aggregate data to the vector products, where pixels are associated with known river features from the PRD and assign each measurement accordingly. Pixel cloud data is aggregated to node location first, and then corresponding node attributes are further aggregated to generate reach attributes. More information about how pixel cloud data is converted to river vector products can be found in the Level 2 KaRIn High-Rate River Single Pass Algorithm Theoretical Basis Document (ATBD) (JPL D-105505, 2023).

SWOT measurements include WSE, river width, water surface slope, and discharge, which are aggregated from pixel-level data. Measurements are made over river reaches generally spanning 10 km, although this may vary between 5 and 20 km. Each reach has a unique identifier and key parameters such as WSE and slope are computed at the reach centerline using polynomial fits. In addition to the core hydrological measurements, the product includes uncertainty estimates (random and systematic) for each parameter and quality flags indicating data reliability. Other parameters, such as the fraction of dark water not detected by SWOT, are also included. Calculations of WSE and slope are based on the EGM2008 geoid model, referenced to the ellipsoid WGS84, and include corrections for atmospheric delays and tides.

L2_HR_LakeSP

The L2_HR_LakeSP product provides high-resolution measurements of water surface elevation (WSE) and surface area for lakes and lake-like features. Lake objects are reported as polygons associated with a measured WSE and area, computed from an algorithm that determines which pixels from the L2_HR_PIXC product are considered lake objects and aggregated accordingly. It classifies lakes into three categories**:** Observed**,** Prior**,** and Unassigned shapefiles. The Observed and Prior shapefiles are derived from the Prior Lake Database (PLD). The Unassigned category includes lake-like features that do not correspond to known lakes in the PLD. Due to file size management, the API used in the prototype repackages the Prior lakes only, using their center points as reference measurements.

The L2_HR_LakeSP product is distributed in Esri GIS shapefile format, with spatial and temporal information based on granules covering a 128 km swath in the cross-track direction. Observations may be missing, degraded, or flagged over the central 20 km of the swath. Lakes may be only partially observed if located near the edges of the swath. Measurements in the L2_HR_LakeSP product are derived from pixels from the prior pixel cloud product (L2_HR_PIXC) (JPL D-56411JPL, 2023). It contains arrays of measured height, geolocation, and classification data from the KaRIn instrument, and aggregated in the aforementioned pixel cloud. Vectorizing the pixel cloud edges after height-constrained geolocation derives lake polygons if they are not previously associated in the PLD, except for lakes that are connected to a river topology, which are also identified with a reach feature in the

L2_HR_RiverSP product. More information on how the pixel cloud data is converted to lake shapefiles can be found in the ATBD (JPL D-105505, 2023).

The product offers key hydrological parameters like WSE, water surface area, and dark water fraction. Each lake is assigned a unique obs_id and lake_id, which link the observed features to the PLD. Additionally, this product provides uncertainty estimates (random and systematic) and quality flags for each parameter. As the mission progresses, the L2_HR_LakeSP product will also include volume and storage change estimates, comparing current observations to reference data for PLD-oriented lakes.

The L2_HR_LakeSP product is key for monitoring hydrological features and changes in lake conditions, with a latency of up to 45 days for data consolidation and calibration.

Table 2 – SWOT products

Data source	Product
PO.DAAC	SWOT Level 2 River Single-Pass Vector Data Product
PO.DAAC	SWOT Level 2 Lake Single-Pass Vector Data Product

5 Methodology

5.1 EO application to extract time series of SWOT data

This application is implemented as a Jupyter Notebook, a versatile tool widely used for interactive computing and data analysis. Jupyter Notebooks offer a flexible environment where Python code, visualizations, and explanatory text are integrated into a single document. This setup supports the development and execution of code while enhancing the presentation of data-driven workflows, making it easy to visualize insights and export analysis-ready data. In this application, Python is employed to interact with various hydrological databases and products, generate visualizations, and export time series data. It leverages standard Python libraries for geospatial analysis and API interaction.

5.1.1 *HydroFeature* Class and Subclasses

The core functionality of the application revolves around the *HydroFeature* class and its subclasses, Reach and Lake. These classes contain methods for querying hydrological features and retrieving associated data. Its design allows the integration of a node subclass if made available by CNES.

***HydroFeature* Class:** This abstract class defines the foundational methods and properties common to all hydrological features. It includes methods for querying the PRD and the PLD using a Web Feature Service (WFS), and querying the Hydrocron API to extract time-series data (Altenau et al., 2023). Hydrocron repackages hydrology datasets from the Surface Water and Ocean Topography (SWOT) satellite into formats that make time-series analysis easier. SWOT data is archived as individually timestamped shapefiles, which would otherwise require users to perform multiple file IO operations per river or lake feature to view the data as a time series. Hydrocron makes this possible with a single API call, facilitating the creation of analysis-ready dataframes of the hydrological parameters from the

SWOT Level 2 River Single-Pass Vector Data Product and the SWOT Level 2 Lake Single-Pass Vector Data Product. Finally, HydroFeature contains methods to visualize and export spatiotemporal data. These can be adjusted to visualize different parameters depending on user needs.

Subclasses *Reach* and *Lake*: The subclasses of *HydroFeature* provide methods to extract river reach and features from the SWOT Prior River Database SWORD and the PLD. These subclasses allow users to target and retrieve spatiotemporal data on the body of water of choice. They contain core functionalities that can be adopted for other feature types.

- *get_features*: This method queries the WFS to extract hydrological features (e.g., river reaches or lake polygons) within a defined bounding box. It retrieves feature IDs and spatial data for subsequent analysis.
- *query_hydrocron*: After identifying the features, this method queries the Hydrocron API to request time series data for the specified ID of hydrological features. It retrieves data such as water surface elevation, flow rates, or lake surface area for both rivers and lakes. Can be adjusted to retrieve the parameters of choice.
- *visualize_features*: This method generates interactive maps displaying the spatial distribution of hydrological features within the region of interest.
- *plot_results*: After fetching time series data, this method creates plots to visualize water surface elevation. Can be adjusted in the *HydroFeature* class to display different parameters.

Hydrocron API already supports operational methods for nodes, still, the ability to extract nodes directly from the SWOT Prior River Database, maintained by CNES, is not yet available. It should be made available in the near future. This expansion will allow users to extract individual river point parameters in the same application.

5.1.2 Workflow

The workflow of this application is designed to extract, visualize, and export hydrological time-series data from specified inland bodies of water. It begins by defining a geographic area of interest through an interactive map where the user draws a bounding box. This bounding box defines the spatial extent for querying hydrological features such as rivers or lakes. Once the area is selected, the application queries either the SWOT Prior River Database or the SWOT Prior Lake Database, retrieving relevant features and their geospatial data within the boundaries. The data extracted includes the necessary feature IDs and spatial coordinates, which are essential for subsequent steps in the process.

Once the hydrological features are identified, the next step involves querying the Hydrocron API to retrieve time series data for each identified feature. Parameters such as the time range, and hydrological variables (eg., water surface elevation, slope, or width), are defined and used to make API requests. The time series data is fetched for each feature, organized into data structures, and stored for analysis. The application then visualizes this data, providing spatial context and the time-series parameter. The spatial distribution of features is displayed through interactive maps, while the time series plots display the recorded measurement of SWOT products at different timesteps.

The final part of the workflow involves exporting the processed data to CSV files. For each hydrological feature, the corresponding time series data can be saved for further use or external analysis. This

export ensures that the data is not only analyzed within the application but is also accessible for broader usage.

5.2 Validation

The validation of river reaches and node water surface elevation (WSE) in this study involves a comparative analysis between in situ measurements and those recorded by the Surface Water and Ocean Topography (SWOT) satellite mission. The goal is to evaluate the accuracy and consistency of the SWOT-derived data by checking how correlated the time series are between SWOT and locations in the case studies: the Upper Danube, Danube Delta, and the Rhine.

The process begins by preparing the in situ data for comparison. In situ water surface elevation measurements are typically recorded at gauging stations in centimeter units relative to a level zero point near the riverbed. If the in situ data is not already in the form of daily averages, it is aggregated accordingly to ensure temporal alignment with the SWOT observations. To standardize the in situ measurements, they are converted from centimeters over the riverbed zero-level point to orthometric height in meters. This adjustment requires transforming the raw measurements from centimeters to meters and adding the level zero point after vertical datum transformation since zero level points are generally referenced to national vertical datums. The target transformation was the European Vertical Reference System (EVRS2007), using the CRS-EU tool. SWOT data is referenced to the WGS84 reference ellipsoid as a first step, and in a second step referenced to the EGM2008 geoid that references WGS84 for the vertical datum. Transformation of local data to the EGM2008 geoid would further refine the accuracy of the validation, PO.DAAC has examples of this vertical datum transformation for North American references.

Once the in situ data is standardized, the next step is to identify and extract the corresponding SWOT measurements. SWOT records water surface elevation for river reaches and nodes at specific timesteps. Only the dates that coincided between the in situ and SWOT observations were retained for analysis. To improve the reliability of the data, outliers were removed from the SWOT dataset. Specifically, values with Z-scores greater than 3 or lesser than -3 were excluded, as these Z-scores indicate data points that are significantly deviated from the mean, suggesting they may be erroneous or unrepresentative of typical conditions. Additionally, measurements flagged as bad quality were also removed. For the node-level data, Hydrocron API was used to specify the node attribute ID from the SWORD database, identified through GIS software.

For statistical comparison, Pearson's correlation coefficient was calculated to measure the linear relationship between the in situ and SWOT-derived WSE measurements. The significance of this correlation is assessed using the p-value. Low values (typically $p < 0.05$) suggest that the correlation is statistically significant, implying a strong relationship between the datasets. Alongside the correlation coefficient, the bias between the two datasets was calculated to assess systematic discrepancies. Additionally, the unbiased Root Mean Square Error (u-RMSE) was calculated, which isolates random differences by excluding the mean bias. These metrics provide an assessment of the relationship between in situ and SWOT measurements.

6 Results

The results present reach-level and node-level metrics for five case study locations.

Reach-level metrics (Table 3) show varying correlations between in situ and observed data across locations, with values ranging from -0.14 to 0.96. The p-values indicate statistical significance for Wilhering, Isaccea, and Lobith. Bias and u-RMSE values suggest differing levels of error and variability in the different locations.

Node-level metrics (Table 4) generally show higher correlations than reach-level metrics, with strong statistical significance for most locations. Isaccea, RO, shows the highest correlation (0.97), while Linz, AU, has the lowest (0.54). Bias and u-RMSE values are relatively low, indicating better SWOT measurements at the node level.

Table 3 – Reach-level metrics for the case studies locations

Location	Correlation	p-value	Bias	u-RMSE
Donauwörth, GE	0.32	0.067	-1.15	1.21
Wilhering, AU	0.59	9.6×10^{-7}	-0.25	0.96
Linz, AU	-0.14	0.335	-0.09	0.57
Isaccea, RO	0.96	6.6×10^{-25}	-0.53	0.28
Lobith, NE	0.70	1.1×10^{-6}	0.25	0.93

Table 4 – Node-level metrics for the case studies locations

Location	Correlation	p-value	Bias	u-RMSE
Donauwörth, GE	0.89	3.8×10^{-7}	-0.06	0.36
Wilhering, AU	0.91	6.4×10^{-19}	-0.16	0.96
Linz, AU	0.54	4.8×10^{-4}	-0.16	0.36
Isaccea, RO	0.97	1.1×10^{-20}	-0.50	0.25
Lobith, NE	0.69	2.6×10^{-6}	0.26	1.02

The validation shows promising performance across all case studies when comparing in situ and observed data. Using node-level time series data significantly improves accuracy in the Upper Danube locations. As shown in Table 4, node-level metrics display high correlation values (e.g., 0.89 in Donauwörth and 0.91 in Wilhering), with statistically significant p-values and reduced bias compared to the reach-level metrics in Table 3. This improvement is due to the precise representation of water surface elevation at specific nodes, avoiding the averaging effect seen over longer reaches. In the

Upper Danube, where the elevation difference between the start and end points is substantial over the 10-kilometer stretches, reach-level averaging can obscure finer-scale variations in WSE, leading to discrepancies between in situ and observed data.

The node-level analysis, by providing a more localized perspective, captures these WSE dynamics more accurately, as shown by the lower u-RMSE values in the node-level metrics. In the Danube Delta and Rhine locations, where topographic variations are less pronounced, node-level improvements are less evident. Here, the flat terrain results in minimal elevation changes over the reaches, making the reach-level analysis sufficient for accurately representing WSE.

Linz is an exception, showing poor results between in situ and observed data in both reach and node-level comparisons. At the reach level, the correlation is negative (-0.14), and even at the node level, the correlation is moderate (0.54), indicating challenges in capturing the WSE accurately. This is likely due to the topography and urban features surrounding the water surface in Linz, which may lead to inconsistencies between in situ measurements and observed data.

The following figures display the time series and regression plot of SWOT measurements against in situ observations for each location:

Upper Danube:

Donauwörth, GE

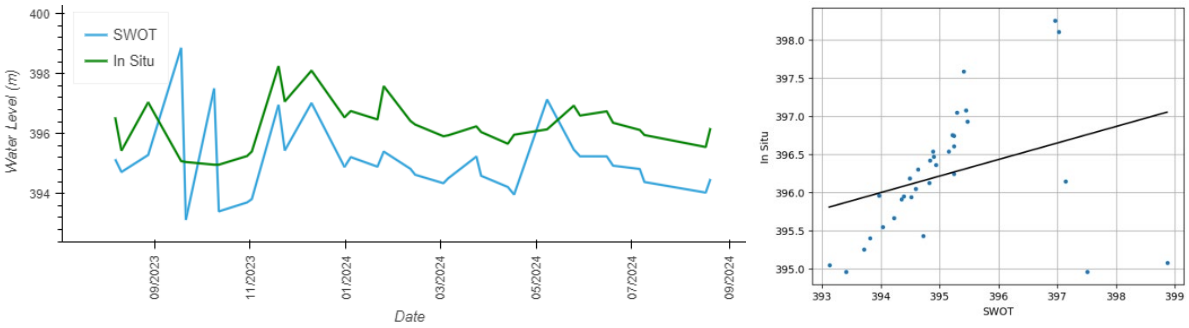

Figure 2 – Time Series and Scatter Plot with Regression Line for Donauwörth, GE, Reach: Comparison of in situ and observed data

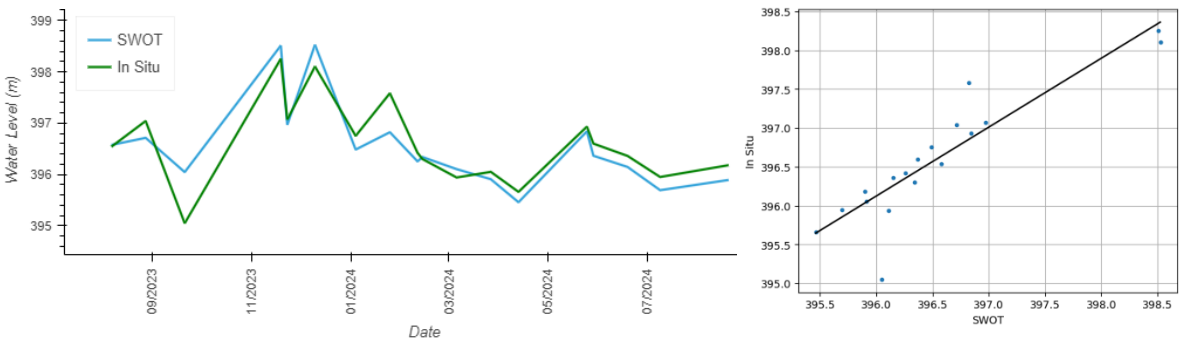

Figure 3 – Time Series and Scatter Plot with Regression Line for Donauwörth, GE, Node: Comparison of in situ and observed data

Wilhering, AU

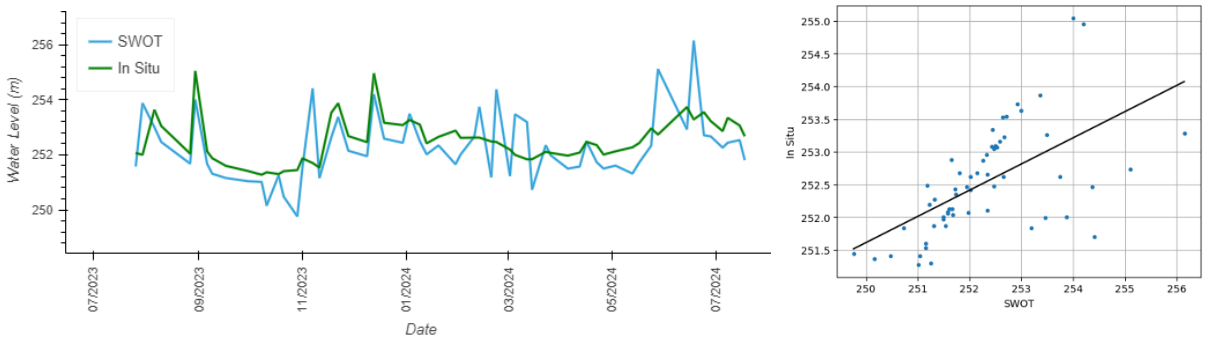

Figure 4 – Time Series and Scatter Plot with Regression Line for Wilhering, AU, Reach: Comparison of in situ and observed data

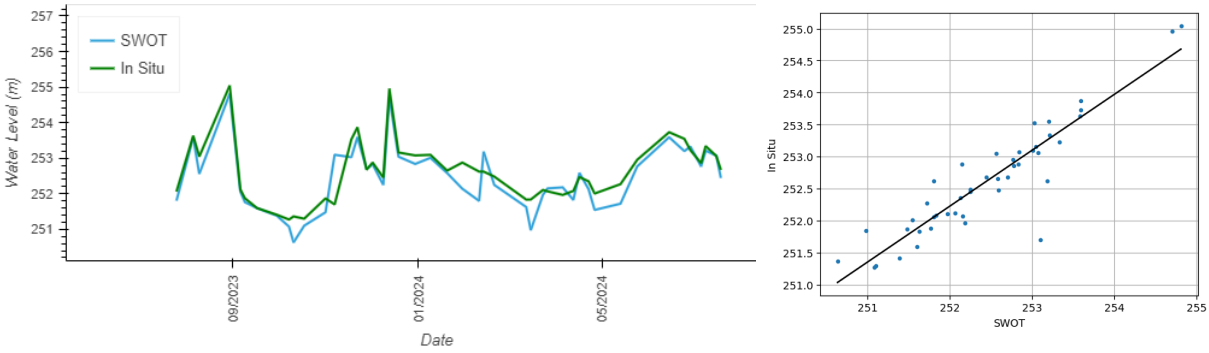

Figure 5 – Time Series and Scatter Plot with Regression Line for Wilhering, AU, Node: Comparison of in situ and observed data

Linz, AU

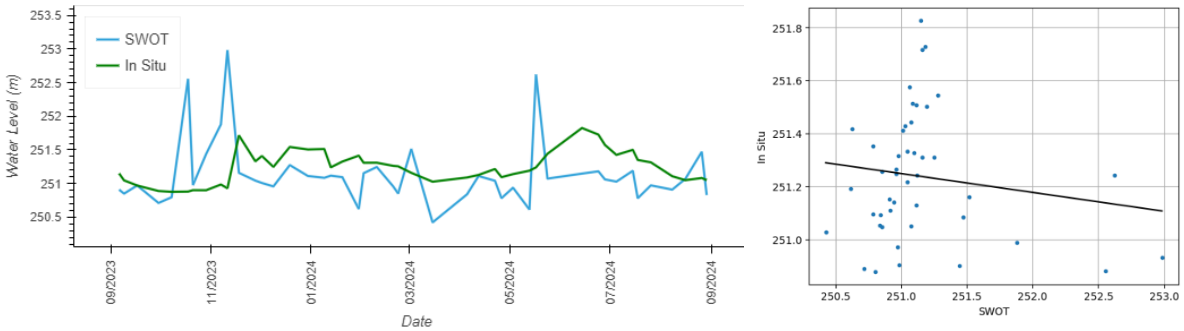

Figure 6 – Time Series and Scatter Plot with Regression Line for Linz, AU, Reach: Comparison of in situ and observed data

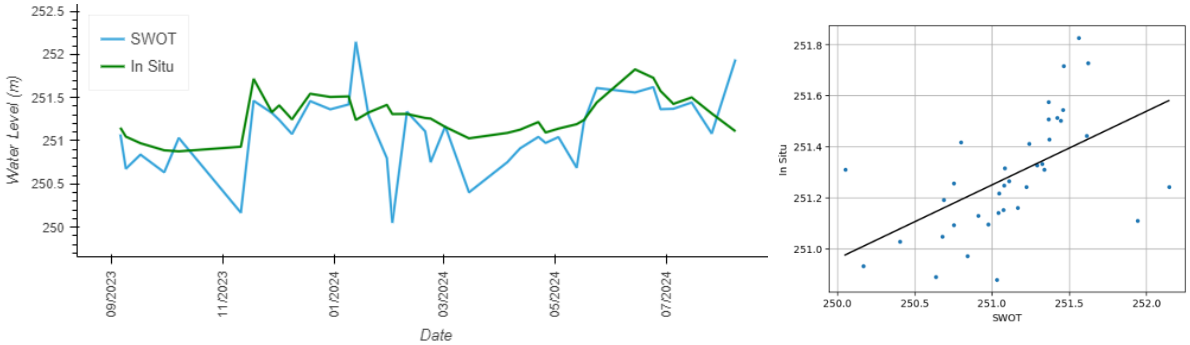

Figure 7 – Time Series and Scatter Plot with Regression Line for Linz, AU, Node: Comparison of in situ and observed data

Danube Delta:

Isaccea, RO

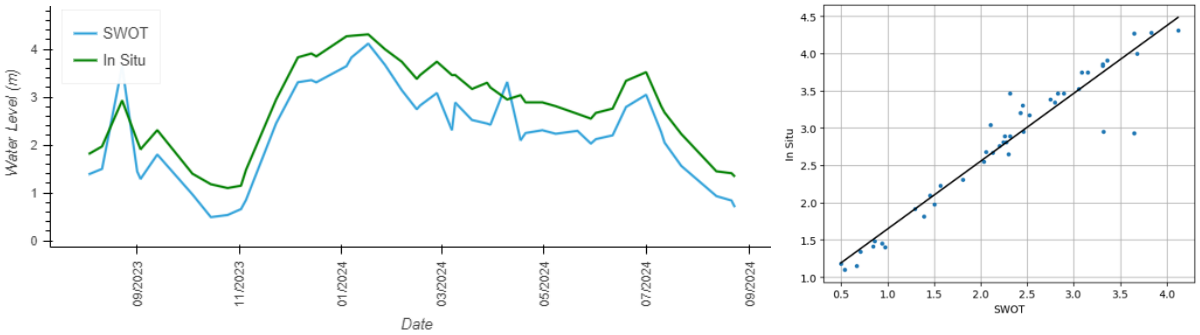

Figure 8 – Time Series and Scatter Plot with Regression Line for Isaccea, RO, Reach: Comparison of in situ and observed data

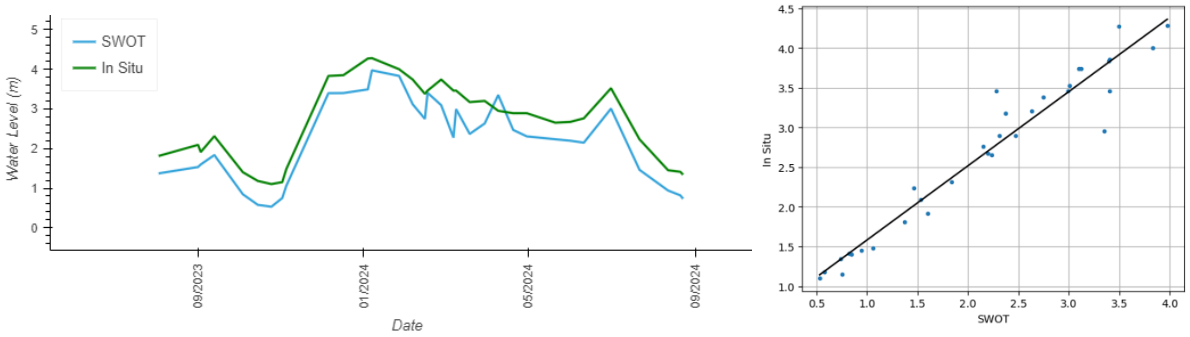

Figure 9 – Time Series and Scatter Plot with Regression Line for Isaccea, RO, Node: Comparison of in situ and observed data

Rhine River:

Lobith, NE

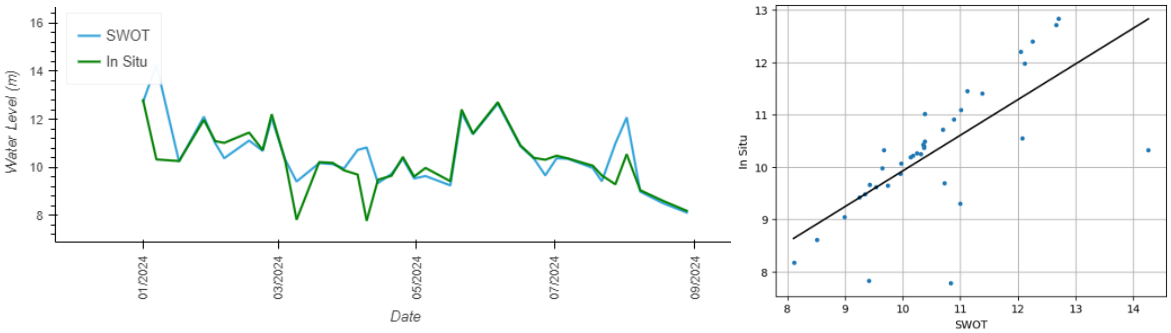

Figure 10 – Time Series and Scatter Plot with Regression Line for Lobith, NE, Reach: Comparison of in situ and observed data

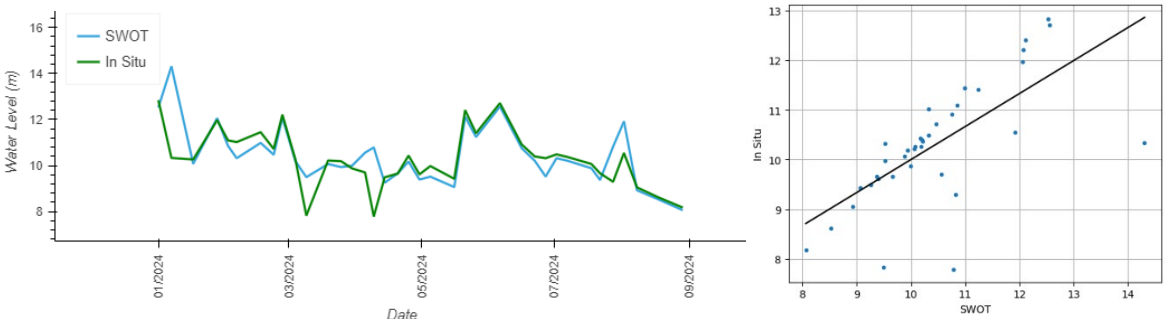

Figure 11 – Time Series and Scatter Plot with Regression Line for Lobith, NE, Node: Comparison of in situ and observed data

The findings highlight the role of location in ensuring the accuracy for further analysis and its inclusion in models. Choosing the appropriate method based on the unique characteristics of each location is essential, as factors such as topography and urban features can significantly influence the reliability of comparisons between in situ and observed data. Further information on the phenomenology that may influence SWOT measurements can be found in the SWOT Data User Handbook.

References

Altenau, E. H., Pavelsky, T. M., Durand, M. T., Yang, X., Frasson, R. P. de 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). https://doi.org/10.1029/2021WR030054

Biancamaria, S., Lettenmaier, D. P., & Pavelsky, T. M. (2016). The SWOT Mission and Its Capabilities for Land Hydrology. *Surveys in Geophysics*, *37*(2), 307–337. https://doi.org/10.1007/s10712-015-9346-y

Elizabeth H. Altenau, Tamlin M. Pavelsky, Michael T. Durand, Xiao Yang, Renato P. d. M. Frasson, & Liam Bendezu. (2023). SWOT River Database (SWORD) (Version v16) [Data set]. Zenodo. https://doi.org/10.5281/zenodo.10013982

Fjortoft, R., Gaudin, J.-M., Pourthie, N., Lalaurie, J.-C., Mallet, A., Nouvel, J.-F., Martinot-Lagarde, J., Oriot, H., Borderies, P., Ruiz, C., & Daniel, S. (2014). KaRIn on SWOT: Characteristics of Near-Nadir Ka-Band Interferometric SAR Imagery. *IEEE Transactions on Geoscience and Remote Sensing*, *52*(4), 2172–2185. https://doi.org/10.1109/TGRS.2013.2258402

JPL D-56411JPL. (2023). *Product Description Document, Level 2 KaRIn High Rate Water Mask Pixel Cloud Product (L2_HR_PIXC)*.

JPL D-105505. (2023). *Algorithm Theoretical Basis Document (ATBD) for Level 2 KaRIn High Rate River Single Pass Science Algorithm Software*.

JPL D-109532. (2024). *SWOT Science Data Products User Handbook*.

Verdin, K. L., & Verdin, J. P. (1999). A topological system for delineation and codification of the Earth's river basins. *Journal of Hydrology*, *218*(1–2), 1–12. https://doi.org/10.1016/S0022-1694(99)00011-6

Disclaimer

Acknowledgement of funding

This project has received funding from the European Union's Horizon Europe research and innovation programme under grant agreement No 101059264.

SOS-Water – EOP4: Surface water temperature (SWT) mapper prototype technical notes

[D3.2, Improved EO application prototypes]

Authors

Authors	Partner no	Partner organisation	Name of author
Main author	11	Eawag	Michael Brechbühler
Contributing author(s)	11	Eawag	Daniel Odermatt

Document history

Date	Version	Chapters affected	Description of change	Author	Document status
12/08/2024	0.1	All	Initial version outlining document	Michael Brechbühler	FIRST DRAFT
18/09/2024	0.2	All	Internal review	Daniel Odermatt	FIRST DRAFT
18/09/2024	1	All	Adaptions and additions	Michael Brechbühler	FINAL

Table of Contents

Abbreviations

CFMASK	C Function of Mask
EO	Earth Observation
EOP	Earth Observation Prototype
ETM+	Enhanced Thematic Mapper Plus
GEE	Google Earth Engine
IWMS	Integrated Water Modelling System
LST	Land Surface Temperature
mNDWI	modified Normalized Difference Water Index
PCR-GLOBWB	PCRaster GLOBal Water Balance model
PLD	Prior Lake Database
RS	Remote Sensing
SPM	Suspended Particulate Matter
SWM	Small Waterbodies Mapping
SWORD	SWOT River Database
SWOT	Surface Water and Ocean Topography
SWT	Surface Water Temperature
TIRS-1	Thermal Infrared Sensor-1
TIRS-2	Thermal Infrared Sensor-2
TM	Thematic Mapper

1 Background

This document provides technical details for an Earth Observation application prototype developed for the SOS-Water project, for surface water temperature mapping. It includes a detailed collection of technical notes that describe the methodology and validation of the prototype created as part of Task 3.2. The associated source code, maintained by Eawag (Partner No. 11), is available in the SOS-Water – Work package 3 GitHub code repository at gitlab.eawag.ch/surf/remote-sensing/sos-water/sosw_wp3. Surface water temperature (SWT) refers to the surface layer temperature of water bodies, which is a critical factor in hydrological processes, aquatic ecosystems, and climate studies. It is considered a key indicator of water quality and has a direct influence on habitat conditions for aquatic organisms, as well as overall ecological stability.

The SWT retrieval prototype is designed to estimate the surface water temperature of lakes and rivers using satellite imagery from Landsat Collection 2 Level 2 data, which was acquired using thermal instruments such as TIRS-2 (Thermal Infrared Sensor-2) aboard Landsat-9. The prototype provides an independent source for validating the temperature data modelled as part of the Integrated Water Modelling System (IWMS) in SOS-Water. Specifically, it was created with the target to validate modelled surface water temperature outputs from PCRaster GLOBal Water Balance model (PCR-GLOBWB) 2 which is being applied in the Rhine case study within SOS-Water. These outputs are particularly important for understanding thermal dynamics in water bodies.

The prototype delivers high-resolution data, offering 30-meter mapping capabilities that are down sampled from the original 60- to 120-meter thermal data. It supports both real-time monitoring and historical analysis dating back to the year 1982, leveraging Landsat 4, 5, 7, 8, and 9 data. This historical perspective allows long-term climate and environmental assessments, especially in regions where water temperature plays a key role in biodiversity and water management.

2 Data

2.1 In situ data

To validate our SWT estimates, we collected in situ datasets that include measurements on river temperature available from local authorities along the Rhine River and its outflows. Table 1 summarizes the relevant data providers that were considered for validation purposes.

Table 1 – Overview of relevant in-situ data providers of Rhine water temperature

Country	Provider	Stations	Comments
Switzerland	Federal Office for the Environment (FOEN)	7	Only one station at river segment with river width >200 m
Germany	directly disseminated by federal states	7	
Netherlands	Ministry of Infrastructure and Water Management	2 (Waal), 2 (Lower Rhine), 2 (Lek), 3 (Scheur), 2 (Oude Maas)	Stations besides Waal likely not modelled in SOS-Water IWMS

From the collected in situ datasets we found 5 stations (Table 2) with open access to historical records and an adequate sampling frequency of daily to sub-daily to generate sufficient match-ups for a comparison.

Table 2 – Stations used for SWT validation

Station id	Location	Lat.	Long.	Approx. river width (m)	Temporal extent
BS_WAR	Weil am Rhein, Switzerland	47.61582	7.57376	210	2002-01 to 2024-04
CXX147	Rheinfelden, Germany	47.57718	7.81360	240	1993-07 to 2024-04
CXX334	Iffezheim, Germany	48.83126	8.10969	200	1992-02 to 2024-04
CXX359	Karlsruhe, Germany	49.01138	8.29631	240	1988-01 to 2024-04
LOBHVN	Lobith, Netherlands	51.85846	6.08474	400	2012-04 to 2023-12

Figure 1 shows the location of the selected stations within the Rhine basin.

Figure 1 – Overview map of Rhine basin with ground stations used for the validation of EOP4

2.2 EO data

The Landsat Collection 2 Level-2 (L2C2) dataset provides Land surface temperature (LST) data derived from the thermal infrared bands of Landsat satellites, namely Landsat-4/-5/-7/-8 and -9. The L2C2 LST product is generated using the Top-of-Atmosphere (TOA) brightness temperatures and applies atmospheric correction and emissivity adjustments in a single-band algorithm to calculate the surface temperature. These TIR bands measure the thermal radiation emitted by the surface in a spatial resolution of 60 to 120-meters as indicated in Table 1. This allows for the calculation of surface temperatures after accounting for atmospheric effects and surface emissivity. Landsat satellites are in a near-polar, sun-synchronous orbit, which means they pass over the same area of the Earth at the same local solar time each day. The satellites cross the equator between 10:00 AM and 10:15 AM local solar time during their descending pass when image acquisitions are made.

Table 2 – Overview of relevant EO sensors and their GEE datasets
** Band 11 in parenthesis is not used for LST estimation*

Sensor	Temporal Extent	Band *	Wavelength (µm)	Native spatial resolution (meters)
Landsat-4 TM	1982-08 to 1993-06	6	10.40-12.50	120
Landsat-5 TM	1984-03 to 2012-05	6	10.40-12.50	120
Landsat-7 ETM+	1999-05 to 2024-01	6	10.40-12.50	60
Landsat-8 TIRS-1	2013-03 to present	10 (11)	10.60-11.19 (11.50-12.51)	100
Landsat-9 TIRS-2	2021-10 to present	10 (11)	10.60-11.19 (11.50-12.51)	100

For Landsat missions 4 through 9, each satellite revisits the same location every 16 days. The overlapping missions allow for combined multi-sensor data, focusing on sensor harmonization. For instance, combining data from the active Landsat-8 and Landsat-9 satellites effectively reduces the revisit time to 8 days (excluding cloud cover). This enables more frequent observations of Earth's surface.

Landsat has been delivering continuous multispectral Earth observation data since 1972, making it the longest-running satellite mission for surface monitoring (Figure 1). Beginning with Landsat-4 in 1982, the program introduced thermal data collection. This extensive archive is invaluable and unique for tracking changes in SWT over time. Looking ahead, the Landsat Next mission, which is expected to launch in 2030/2031, will ensure the continuation of Landsat's long-term observation capabilities of SWT.

Figure 2 – Timeline of Landsat missions with thermal sensors, adapted from USGS (2024)

3 Methodology

Our SWT algorithm leverages datasets from the Google Earth Engine (GEE) platform, focusing on Landsat Level 2 Collection 2 data. This data is comprised of atmospherically corrected Land Surface Temperature (LST) data from Landsat-4 Thematic Mapper (TM), Landsat-5 TM, Landsat-7 Enhanced Thematic Mapper Plus (ETM+), Landsat-8 Thermal Infrared Sensor-1 (TIRS-1) and Landsat-9 TIRS-2 sensors. Our processing algorithm shown in Figure 1, involves harmonization, cloud masking, water masking, pure water pixel selection steps. All these processing steps are executed within the GEE environment. GEE's cloud-based infrastructure allows for efficient processing of large datasets, enabling rapid execution of these algorithms across extensive spatial and temporal scales. This efficiency is key to producing timely and reliable SWT data for monitoring and analysis.

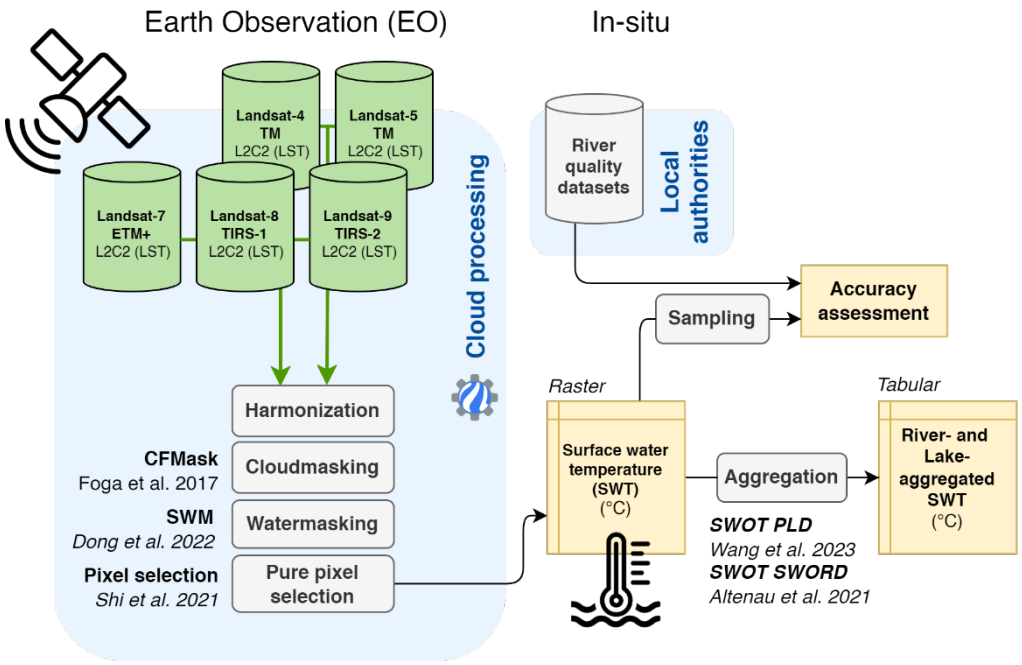

Figure 3 – Main processing steps of EOP4 algorithm

The first step involves harmonization of the datasets by applying consistent band labels across the different thermal bands of all sensors. To remove cloud and cloud shadow contaminated pixels, we apply pixel masking based on the CFMASK (C Function of Mask) flags (Foga et al., 2017). This algorithm is part of the Landsat Collection 2 processing and identifies clouds and cloud shadows in the Landsat imagery. In the following masking step, we identify water pixels. This step can be either done using the provided water flags within the datasets by CFMASK or by employing our GEE implementation of the small water bodies mapping (SWM) algorithm (Fan et al., 2022).

All the available Landsat 30-meter TIR data were smoothed and resampled from a coarser original sensor data, ranging from 60 m for Landsat 7 ETM+ up to 120 m for Landsat 5 TM. Thus, river–land boundary effects beyond the 30-meter proximity of the shore have to be considered. To identify pixels with only minimal contamination by land surfaces we apply an additional pure water pixel selection (Martí-Cardona et al., 2019) following the water-detection. Although this reduces the overall data coverage it enhances the SWT estimation quality.

Algorithms for direct data aggregation from the raster data to a tabular format are included in the codebase. In this aggregation step standard point sampling can be used. Alternatively, it is also possible to directly aggregate the retrieved SWT data to river reaches as well as lakes based on the global NASA Surface Water and Ocean Topography (SWOT) Prior Lake Database (PLD) (Wang et al., 2023) and SWOT River Database (SWORD) (Altenau et al., 2021), respectively.

3.1 Cloud masking

The CFMASK (C Function of Mask) (Foga et al., 2017) algorithm is part of the Landsat Collection 2 processing and identifies clouds and cloud shadows in the Landsat imagery.

For the masking we remove pixels by chaining logical OR operators on active quality-assessment flags “Dilated Cloud” (bit 1), “Cirrus (high confidence)” (bit 2), “Cloud” (bit 3), “Cloud Shadow” (bit 4). An additional cloud-buffer can be supplied in the masking function to oppose under detection. For validation purposes we set this value to 30-meters.

3.2 Water masking and pixel selection

Identification of water pixels and masking of land surfaces is done using a two-step approach. In the first step an initial water mask is created. This can be done by either using the quality assessment flags based on CFMask (Foga et al., 2017) or the Small Waterbodies Mapping (SWM) algorithm (Fan et al., 2022). In a secondary step pure water pixels are selected from the computed water mask. This step removes boundary effects in the SWT signal from land contamination. This is especially crucial when extracting data from narrow rivers systems as land contamination can generally bias the retrieved signals towards higher temperatures in summer and lower temperatures in winter.

The SWM algorithm is designed to accurately identify and map small water bodies, such as ponds, small lakes, and narrow rivers, which are often challenging to detect using conventional water index methods. We implemented this algorithm with the Earth Engine Python API. A detailed description of how the algorithm is applied can be found in the “SOS-Water – EOP1: Suspended Particulate Mapper (SPM) mapper prototype technical notes”. As more efficient, but in most cases less effective alternatives, two additional approaches based on the CFMASK flags as well as modified Normalized Difference Water Index (mNDWI) are implemented in the codebase as well.

All the available Landsat 30-meter TIR data were resampled from a coarser original sensor data, ranging from 60-meter for Landsat 7 ETM+ up to 120 m for Landsat 5 TM. Additionally, a 16x16 cubic convolution kernel is applied on the 30-meter data to smooth the retrieved radiance signal. Thus, water-land boundary effects beyond the 30-meter proximity of the shore have to be considered. This is especially crucial when extracting data from narrow rivers systems as land contamination can generally bias the retrieved signals towards higher temperatures in summer and lower temperatures in winter. To identify pure pixels with only minimal contamination by land surfaces we apply an additional pure water pixel selection following the water-detection. Although this reduces the overall data coverage it enhances the quality of retrieved SWT estimates.

We differentiate between two different types of contaminated pixels in the water-land boundary with an erroneous temperature value. Boundary Effect 1 (BE1) and Boundary Effect 2 (BE2) as shown in Figure 3. BE1 represents the 30-meter waterbody pixels, which are mixed with the land pixels in the up sampling from the native TIR sensor spatial resolution (60/100/120m, depending on sensor). While BE2 are the pixels with an error caused by the cubic convolution smoothing. The temperature of these pixels is more tending to the land nearby and thus share an excessive error. Therefore, both will severely influence the result’s accuracy if they were sampled. This pixel selection method was initially developed by Martí-Cardona et al. (2019) for Landsat 8 TIRS. We used the algorithm description to reconstruct this approach for the additional Landsat sensors in the GEE Python API.

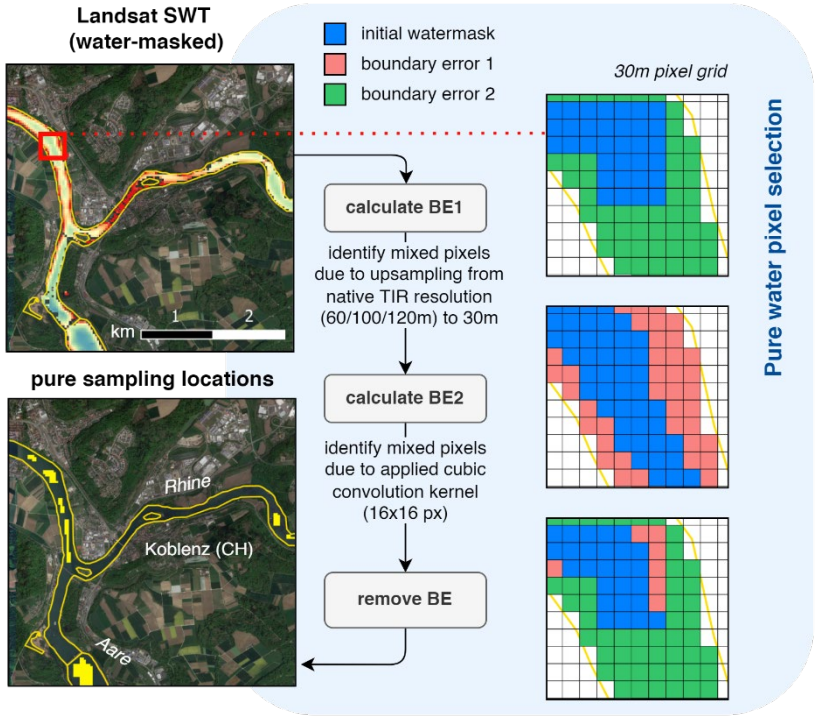

Figure 4 – Process of pure water pixel selection and the boundary-effect removal shown with the example of a Landsat-9 TIRS-2 scene at the Aare-Rhine confluence in Koblenz, CH

To calculate the mask for BE1, we down sample the data to a native resolution grid. From this grid, the pixels falling entirely within the water mask are determined. Only the downscaled 30m pixels falling entirely within these pure 100m pixels are selected to remove any pixels affected by BE1.

To calculate the mask for BE2, we consider the 16x16 pixel cubic convolution kernel that is applied in the production of Level 2 Collection 2 Landsat data. This corresponds to a pixel neighbourhood of an area of 480x480-meters. This smoothing introduces additional mixing and errors in SWT estimation. Due to the applied cubic function, the degree of land contamination will depend on the relative position of the land pixels in the pixel neighbourhood and on the ratio of land to water thermal radiance. Pixels closer to the shore are much more affected. The same is the case if higher surface temperature differences occur between land and water. We simulate this radiance mixing introduced by the convolution.

To do this, the 30-meter water mask is resampled to a 10m resolution grid. Water pixels are assigned value 1, while overland pixels are given value 1.25, representative of a conservative value for the land to water radiance mixing ratio based on Shi et al. (2021). An image is simulated by resampling the modified water mask to native resolution (60/100/120m). Finally, bilinear resampling is then applied as a practical approximation of cubic convolution to resample the image back to 30m resolution. From this intermediate output we can estimate the introduced BE2 errors. By filtering pixel values above 1.05, we remove pixels with a thermal radiance affected by a maximum relative convolution error greater than 5%.

3.3 Sampling

In the data aggregation process, raster stacks from GEE imageCollection objects are converted into a tabular format. SWT data can be sampled at virtual stations either as point samples or within more complex polygon regions of interest (ROIs). For each image, various statistical metrics, such as mean, median, minimum, and maximum, are calculated over the ROI. Additionally, a data coverage indicator is generated, enabling a post-sampling filter to identify incomplete data coverage, which may be necessary due to issues like cloud contamination.

3.3.1 Lake- and River-aggregation

For a basin-wide extraction of data in a useful and processable format it is important to have a robust workflow. We provide an algorithm to directly aggregate the retrieved Landsat SWT to lake polygons and to river reaches, river sections with similar hydrologic conditions.

For the river reach aggregation, ROI polygons are generated from the SWOT River Database (SWORD) (Altenau et al., 2021) v15. The SWORD vector dataset was built to provide high-resolution river nodes (200 m) and reaches with attached hydrologic variables (width, slope, etc.) as well as a consistent topological system for global rivers 30-meters wide and greater. Our algorithm uses the width attributes of each node in a reach together with the relative position to the proceeding nodes to calculate node polygons that in union build a complete reach polygon. Figure 4 and Figure 5 present a schematic and an applied depiction of the reach polygon generation, respectively.

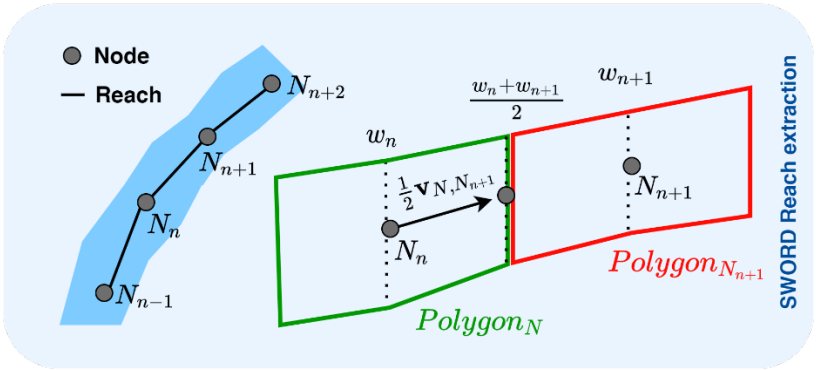

Figure 5 – Schematic representation of SWORD reach polygon generation

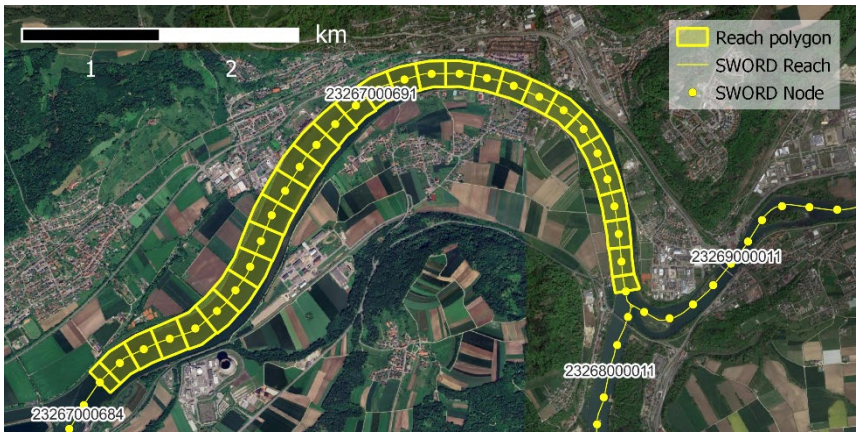

Figure 6 – Extracted polygon for SWORD reach ID2326700691 (Rhine, CH/DE)

For the aggregation of the extracted SWT data to lakes we added functionality to directly aggregate to the lake polygons of the SWOT Prior Lake Database (PLD) (Wang et al., 2023). This vector database was created for the SWOT mission to provide locations and extents for all global lakes ≥ 1 ha. It is one of the largest and most comprehensive lake databases and currently consists of over 6 million lakes.

Integrating the SWT mapping workflow with the SWORD and SWOT PLD databases enables the consolidation of large historical raster datasets, which are often substantial in size and challenging to manage, into aggregated tabular data organized by river segments and lakes. Converting raster stacks to tabular form simplifies data handling and makes it easier to present and analyse. Furthermore, this tabular data can be directly linked to other hydrological attributes and, in future implementations, potentially to satellite-based water level and discharge estimates retrieved from SWOT.

4 Results

4.1 Validation

To validate the Surface Water Temperature (SWT) retrieval model, we utilized in-situ measurements from ground-based water quality stations that provide high-frequency water temperature data. These stations are strategically located along rivers and lakes, ensuring comprehensive temporal and spatial coverage for the validation process.

Match-ups between satellite-derived SWT estimates and in-situ measurements were created using a ±0.5-day time window. This approach creates same-day matchups but accounts for temporal discrepancies between satellite overpasses and in-situ measurements. In cases where multiple satellite acquisitions occurred within the same day, the acquisition closest in time to the in-situ measurement was selected to minimize temporal differences. The matchups were sampled at a single clean pixel cantered over the unobstructed river centreline closest to the sampling station. Tests with larger sampling grids, such as 3x3 pixel (90x90-meter), resulted in worse agreement with ground data.

To evaluate the accuracy and performance of the SWT retrieval, we calculated several statistical measures based on these match-ups to provide a comprehensive understanding of the model's performance. The achieved validation metrics are shown in Table 3.

Table 3 – Validation results for SWT matchups

Station	n	Slope	R	R^2	Bias	Accuracy (±1σ)	MSE	RMSE
LOBHVN	288	1.077	0.99	0.981	< 0.001	0.71 ± 0.67 °C	0.951	0.975
CXX334	547	1.086	0.98	0.961	< 0.001	1.1 ± 0.92 °C	2.057	1.434
CXX359	562	1.051	0.974	0.949	< 0.001	1.1 ± 0.99 °C	2.2	1.483
BS_WAR	823	1.115	0.967	0.935	< 0.001	1.3 ± 1.3 °C	3.405	1.845
CXX147	456	1.109	0.959	0.921	< 0.001	1.4 ± 1.2 °C	3.576	1.891

The achieved R^2 values, ranging from 0.921 to 0.981, indicate strong to very strong performance of the SWT prototype. The estimated values align with the observed ground measurements, showing an accuracy between 0.71°C and 1.4°C. Bias values close to zero suggest no significant systematic bias toward under- or overestimation. However, the slightly positive slope in all linear regression fits indicates a tendency for underestimation at lower temperatures and overestimation at higher temperatures, as shown in the scatterplot in Figure 6. This suggests a subtle systematic bias in the SWT extraction. One potential source of this error could be residual contamination of sampled pixels by land signals. Another possibility is the skin effect, where the thin surface layer (micrometers to millimetres thick) is cooler due to heat loss from radiation, evaporation, and conduction. In calmer conditions or with strong solar heating, the reverse effect may also occur. The former explanation is supported by the fact that the best-performing station is located at a sufficient and substantially larger distance from the shore compared to other stations. A very small amount of strongly underestimated datapoints indicate some amount of cloud contamination. This can occur when the cloud-masking fails at the sampling station and the cold thermal signal of clouds is influencing the estimate.

In summary, our validation demonstrates that the SWT prototype, EOP4, is fully functional and performs well within the required accuracy ranges for surface water temperature extraction. The small discrepancies observed are expected given the fundamentally different measurement approaches between in situ observations and remote sensing. The prototype shows high reliability in capturing SWT, providing accurate results and robust performance, making it suitable for further use within the project.

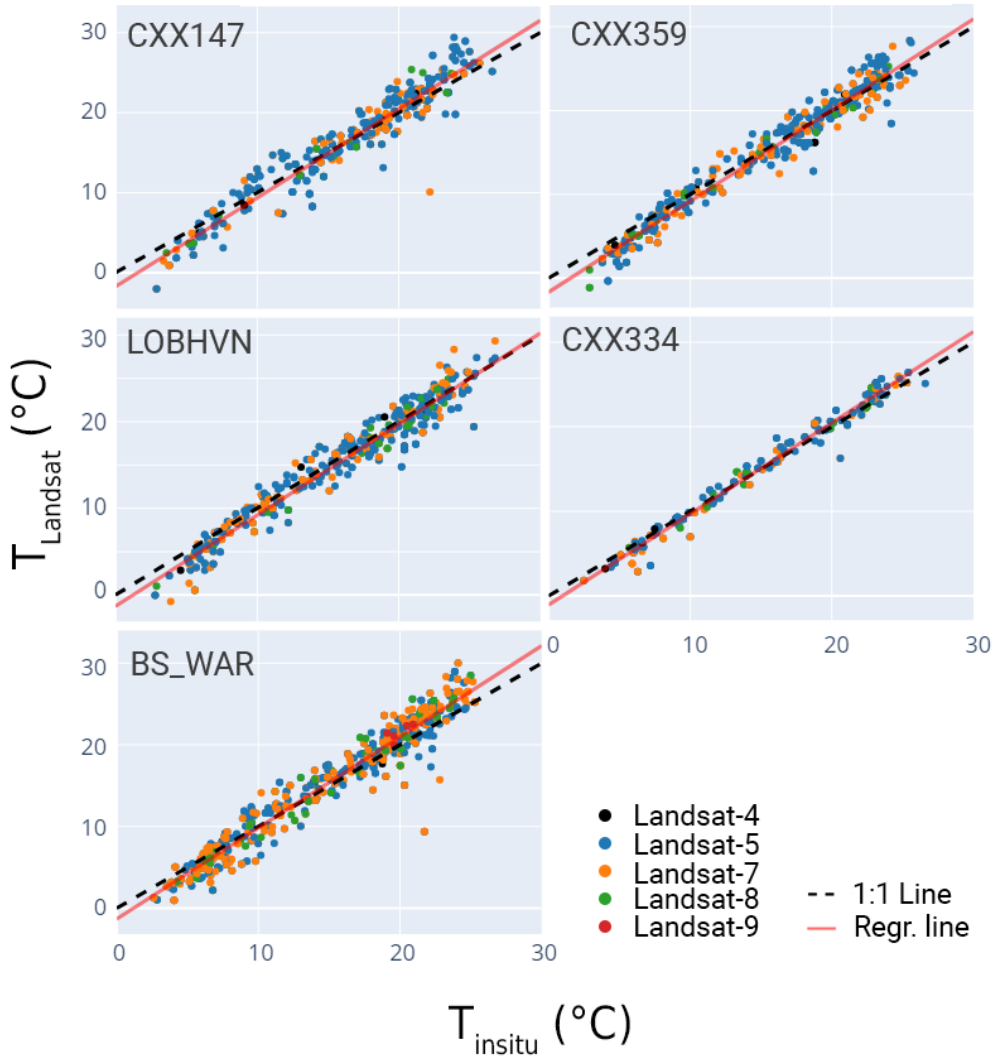

Figure 7 – Model evaluation of satellite-based SWT estimates against insitu surface water temperatures

References

Altenau, E. H., Pavelsky, T. M., Durand, M. T., Yang, X., Frasson, R. P. de 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). https://doi.org/10.1029/2021WR030054

Fan, L., Dong, Y., Zhao, J., Geib, C., Wang, L., & Taubenbock, H. (2022). Mapping of Small Water Bodies with Integrated Spatial Information for Time Series Images of Optical Remote Sensing. *International Geoscience and Remote Sensing Symposium (IGARSS)*, *2022-July*, 6224–6227. https://doi.org/10.1109/IGARSS46834.2022.9883460

Foga, S., Scaramuzza, P. L., Guo, S., Zhu, Z., Dilley, R. D., Beckmann, T., Schmidt, G. L., Dwyer, J. L., Joseph Hughes, M., & Laue, B. (2017). Cloud detection algorithm comparison and validation for operational Landsat data products. *Remote Sensing of Environment*, *194*, 379–390. https://doi.org/10.1016/j.rse.2017.03.026

Martí-Cardona, B., Prats, J., & Niclòs, R. (2019). Enhancing the retrieval of stream surface temperature from Landsat data. *Remote Sensing of Environment*, *224*, 182–191. https://doi.org/10.1016/j.rse.2019.02.007

Shi, X., Sun, J., & Xiao, Z. (2021). Investigation on river thermal regime under dam influence by integrating remote sensing and water temperature model. *Water (Switzerland)*, *13*(2). https://doi.org/10.3390/w13020133

USGS. (2024, February 6). *Landsat Missions: Imaging the Earth Since 1972*. https://www.usgs.gov/media/images/landsat-missions-timeline

Wang, J., Pottier, C., Cazals, C., Battude, M., Sheng, Y., Song, C., Sikder, S., Yang, X., Ke, L., Gosset, M., Reis, R., Oliveira, A., Grippa, M., Girard, F., Allen, G. H., Biancamaria, S., Smith, L., Crétaux, J.-F., & Pavelsky, T. M. (2023). *The Surface Water and Ocean Topography Mission (SWOT) Prior Lake Database (PLD): Lake mask and operational auxiliaries*. https://doi.org/10.22541/au.170258987.72387777/v1

Disclaimer

Acknowledgement of funding

This project has received funding from the European Union's Horizon Europe research and innovation programme under grant agreement No 101059264.

This work has received funding from the Swiss State Secretariat for Education, Research and Innovation (SERI)

Project funded by

Schweizerische Eidgenossenschaft
Confédération suisse
Confederazione Svizzera
Confederaziun svizra

Swiss Confederation

Federal Department of Economic Affairs, Education and Research EAER
State Secretariat for Education, Research and Innovation SERI

SOS-Water – Crop water productivity mapping prototype technical notes

[D3.2, Improved EO application prototypes]

Authors

Authors	Partner no	Partner organisation	Name of author
Main author	7	FutureWater	Sergio Contreras
Contributing author(s)	7	FutureWater	Amelia Fernández-Rodríguez Gijs Simons

Document history

Date	Version	Chapters affected	Description of change	Author	Document status
28/08/2024	0.1	All	Initial version outlining document	-	DRAFT
20/09/2024	1.0	All	Final version	-	FINAL
23/01/2026	1.1	5, 6, 7	Final version, with revisions	Sergio Contreras, Amelia Fernández-Rodríguez, Gijs Simons	FINAL (REV)

Table of Contents

Abbreviations

RS	Remote Sensing
SWOT	Surface Water and Ocean Topography
ATBD	Algorithm Theoretical Basis Document
SOS	Safe Operating Space
IWMS	Integrated Water Modelling System
JWRS	Júcar Water Resource System
UDA	Agrarian Demand Units

1 Maintainer and code access

This document describes the Earth Observation Prototype 5 (EOP5) – Crop water productivity mapping. It includes technical notes related with the tools and methods employed, the improvements generated, and how the prototype was implemented and tested in one SOS-Water pilot site. This prototype and implementation exercise was created as part of Task 3.2 of the SOS-Water project. The associated source code, maintained by FutureWater-NL (Partner No. 6.1), is available in the SOS-Water – Work package 3 GitHub code repository at https://gitlab.eawag.ch/surf/remote-sensing/sos-water/sosw_wp3.

2 General description

Pressure on water resources has been increasing worldwide in the last decades due to rising demand for food and energy, improved living standards, and the more complex regional water governance frameworks (Kahil et al., 2015).

The SOS-Water Project endeavours to set out the boundaries within which the Earth's capacity to provide life-support systems for humanity is not endangered, and humanity's capacity to adapt to environmental changes is not overburdened.

WP3 aims to: 1) identify the main data gaps and information needs that can be extracted from existing Earth Observation information sources, and 2) to customize and up-scale EO-based products (prototypes) which may be used for the definition of SOS and for the benchmarking of the Integrated Water Modelling System (IWMS) implemented within the case studies and relevant time scales.

The EO prototypes in the SOS project are linked to further enhancements of Copernicus and GEOSS services from an SOS-indicator point of view.

3 Objective

The EOP5 prototype aims to address the high-resolution quantification of water productivity in main croplands in the Jucar River Basin (Spain) by using the WaPOR modelling tool. Water productivity (WP) in crops is a measure of the amount of biomass or yield produced per unit of water consumed by a crop. It is a key performance indicator of agricultural productivity and sustainability, particularly in regions where water resources are limited or where efficient water management is essential. Improving WP is important for maximizing agricultural production while minimizing water consumption and environmental impacts. Strategies for enhancing WP in crops include adopting more efficient irrigation systems, optimizing planting densities, selecting drought-tolerant varieties, improving soil water retention, and implementing conservation practices (Pereira et al., 2012).

WaPOR[1] (Water Productivity Open-access of Remote sensing data) is the FAO's portal to assist countries in monitoring water productivity, identifying water productivity gaps, and proposing

[1] https://data.apps.fao.org/wapor/?lang=en

solutions to reduce these gaps (FAO, 2023). WaPOR offers a large dataset of continuous data from 2009 onwards at a 10-day basis across Africa and Middle East. The dataset includes energy/water-related variables like precipitation, evaporation, transpiration; and production/productivity-related variables like the net primary production, total biomass production, and gross and net biomass water productivity. The coverage of WaPOR datasets is restricted to Africa and Middle East, and data is provided at three levels or spatial resolutions: The continental level (250 m) covers all Africa and the Middle East (L1); the national level (100 m) covers 21 countries and 4 river basins (L2); the district level (30 m) covers eight irrigation areas (L3).

For the purposes of this study, WaPOR is enhanced from its original configuration and properly adapted for increasing the spatial resolution and the accuracy of the outputs for the pilot case of the Júcar Water Resource System (Figure 1, referred hereafter as JWRS). The improved model configuration, which includes a better characterization of land use through the ingestion of a local LUC dataset, is tested and evaluated. As final output, representative values of net biomass water productivity - as surrogate of crop productivity – are extracted for the most prominent cropland categories in the pilot area.

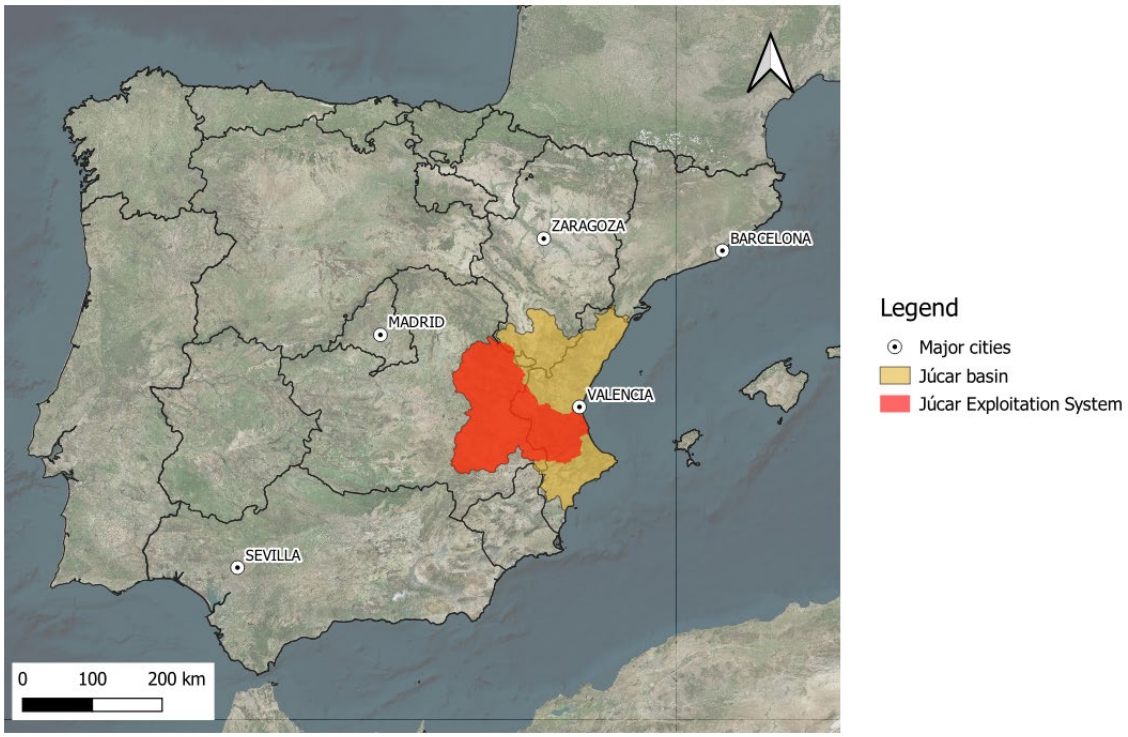

Figure 1. Location of the Júcar Water Resource System (thick black boundaries) in the Jucar River Basin (reddish area).

This document describes the methodology employed and the results derived, highlighting the adaptations addressed over the default configuration of the WaPOR algorithm to enhance its applicability within the context of the JWRS.

Intermediate objectives of this exercise include:

1) Evaluation of the technical feasibility of implementing WaPOR in a new pilot site not previously tested by the scientific community. This analysis contributes to test and demonstrate the adaptability and effectiveness of the WaPOR algorithm for being applied across different geographical settings.

2) To assess the flexibility of the algorithm for the assimilation of high-resolution local datasets that better represent land use-land cover and meteorological forcings.

4 Data

Data inputs or intermediate variables used by WaPOR consist of:

- EO datasets or products derived from visible-NIR[2] and thermal spectrum bands.
- Meteorological variables (wind speed, air temperature, relative humidity…) generated or derived from climate model/reanalysis datasets.
- Soil moisture estimates. These are internally estimated by WaPOR using the S_e model (Yang et al., 2015), which rest on the Trapezoid Method concept.
- Static input layers as elevation and slope maps generated from a Digital Elevation Model, or a landuse-landcover map.

After the collection of raw inputs from their native repositories, input data is pre-processed to ensure interoperability within the processing chain among different grids, spatial and temporal resolutions coming from the different sources. Moreover, input data is used in multiple pre-processing steps to produce intermediate data components needed for computing the final outcomes of the algorithm.

4.1 In situ and auxiliary data

No in-situ data was used for this prototype.

4.2 EO data

WaPOR products rest on the use of different EO data sources depending on level of spatial resolution desired. There are three levels of detail already implemented. Level 1 , in which each one requires the combination of a set of EO datasets and code versions. Each code version includes internal improvements that relate with the pre-processing chain and the ingestion of raw datasets that are used to force the WaPOR algorithms. The combination of code version and levels provides a list of WaPOR configurations. Table 1 shows the number of WaPOR configurations already available by default plus the enhanced configuration for this prototype (Level_3(z1*)).

[2] NIR = Near Infrared

Table 1. Levels and sources available in PyWaPOR

Type	Nature	Variables	Level 2 (z1)	Level 2 (z3)	Level_3 (z1)	Level_3 (z3)	Level_3 (z1*) (SOS-Water-Jucar)
Biomass	Dynamic	NDVI	PROBA-V (100m)	SENTINEL2 (60m)	LANDSAT 5,7,8,9 (30m)	SENTINEL2 (20m)	LANDSAT 5,7,8,9 (30m)
Radiation		Albedo					
Soil	Dynamic	LST		VIIRS (375m)	LANDSAT 5,7,8,9 (90m, sharp. to 30m)	VIIRS (375m)	LANDSAT 5,7,8,9 (90m, sharp. to 30m)
		Relative Root Zone Soil Moisture (RZSM)	SE_ROOT_MODEL	SE_ROOT MODEL	SE_ROOT MODEL	SE_ROOT MODEL	SE_ROOT MODEL
Climate	Dynamic	Precipitation	CHIRPS (5.5km)	CHIRPS (5.5km)	CHIRPS (5.5km)	CHIRPS (5.5km)	CHIRPS (5.5km)
Land	Static	Elevation above sea level	SRTM (30m)	GLO90	SRTM (30m)	SRTM (30m)	SRTM (30m)
Climate	Dynamic	Radiation	MERRA2 (55km)	ERA5 - agro (11km)	MERRA2 (55km)	MERRA2 (55km)	ERA5 - agro (11km)
		Air temperature	GEOS5 (34.7 x 27.8 km)		GEOS5 (34.7 x 27.8 km)	GEOS5 (34.7 x 27.8 km)	
		Air temperature max					
		Air temperature min					
		Specific humidity					
		Northward wind speed at 2 meter		ERA5 - reanalysis (27.8km)			ERA5 - reanalysis (27.8km)
		Wind speed at 2 meter					
		Air pressure					
		Air pressure at sea level					
Land	Static	**Land use classification**	GLOBCOVER (300m)	WaPOR2 (250m)	GLOBCOVER (300m)	GLOBCOVER (300m)	SIGPAC
		Minimal stomatal resistance	LookUp Table-based (LULC)				
		Maximum obstade height					
		Max Light Use Efficiency					
Radiation		Offset of the tau-term in the FAO-56 LW-radiation relationship	WaPOR2 (250m)				
		Slope of the tau-term in the FAO-56 LW-radiation relationship					
		Orographic roughness					
		Offset rn/g0-relation water					
		Slope rn/g0-relation water					
		Yearly air temperature amplitude					
		Optimum air temperature for plant growth					
		Vapour pressure stress curve slope					

Improvements with respect to Level3_z1

A primary goal during the SOS pilot implementation was to use and achieve the highest resolution for both the input forcings and outputs (Level_3). Consequently, all default configurations involving the Level_1 and Level_2 were excluded due to the low resolution of the outputs generated.

Among the Level_3 candidates, the Level_3(z3) was tested in an initial phase but finally discarded due to difficulties found in accessing and downloading Sentinel-2 data - the collection of historical Sentinel-2 data has to be done through the Long-Term Archive (LTA) repository which significantly slows down the process -. For research that involves the retrieval of historical datasets, this fact may increase abruptly the computational time which would, at the end, reduce the attractiveness for using Sentinel data.

Level_3 (z1) was the most suitable solution found for testing the prototype. By default, it involves the combination of MERRA2-GEOSS (climate variables), LANDSAT (greenness and land surface temperature) and GLOBCOVER (land use – land cover) datasets. To increase the spatial resolution and accuracy of the outputs (water productivity in cropland), the default configuration was enhanced by:

1) replacing the climate forcing layers from the MERRA (0.5deg.) and GEOS5 (0.25deg.) datasets to the ERA5-Agro (0.1deg.) and ERA5-reanalysis (0.25deg.).
2) modifying the section of the code which is dedicated to the ingestion of the land cover map and the layers with the biophysical crop parameters.

The enhanced WaPOR configuration resulting from the adaptation of the Level_3(z1) was termed as Level_3(z1*) (Table 1).

5 Methodology

5.1 Algorithm description

WaPOR rests on a two-stage modelling process that involves: a) the ETLook-WaPOR model to compute the water-energy balance components, and b) the C-Fix model to compute the carbon balance and the total biomass production. These two models are available through an open-source python package called PyWaPOR, developed by FAO in order to provide users with the ability to generate tailored water productivity data and related outputs.

5.1.1 ETLook

The retrieval of evaporation, transpiration and interception fluxes in WaPOR rests on ETLook model (Pelgrum et al., 2010) which has been adapted to remote sensing input data. ETLook is a 2-source Surface-Energy-Balance model that rests on the Penman-Monteith equation which integrates the radiative and aerodynamic forcings as described in FAO-56 drainage paper (Allen, Pereira, Raes, & Smith, 1998). Transpiration and soil evaporation are computed separately using Equations (1) and (2).

$$\lambda T = \frac{\Delta(R_{n,canopy}) + \rho c_p \frac{\Delta_e}{r_{a,canopy}}}{\Delta + \gamma\left(1 + \frac{r_{canopy}}{r_{a,canopy}}\right)}$$

Equation 1

$$\lambda E = \frac{\Delta(R_{n,soil} - G) + \rho c_p \frac{\Delta_e}{r_{a,soil}}}{\Delta + \gamma\left(1 + \frac{r_{soil}}{r_{a,soil}}\right)}$$

Equation 2

where Δ is the slope of the saturation vapour pressure curve [mbar/K], Δ_e vapour pressure deficit [mbar], ρ is the air density [kg m-3], c_p is the specific heat of dry air [J/kg·K], γ is the psychrometric constant [mbar/K], G is the soil heat flux [W/m^2], $R_{n,canopy}$ and $R_{n,soil}$ [W/m^2] are the net radiation for canopy and soil respectively, r_{canopy} and r_{soil} [s/m] are the canopy and soil resistance respectively, and $r_{a,canopy}$ and $r_{a,soil}$ [s/m] are the aerodynamic canopy and soil resistance, respectively.

r_{soil} is a function of the relative soil moisture content in the topsoil, while r_{canopy} is a function of the Leaf Area Index (LAI), the minimal stomatal resistance ($r_{s,min}$) and a number of reduction factors. $r_{s,min}$ determines the minimum level of resistance offered by the stomata of crops to the exchange of water vapor with the atmosphere.

The aerodynamic resistance components ($r_{a,soil}$, $r_{a,canopy}$) are functions of the surface roughness, which depends on the maximum obstacle height ($Zobst_{max}$, [m]) and turbulence conditions. The maximum

obstacle height refers to the tallest feature in the surface (e.g. the maximum height of a crop or tree), while turbulence are simulated following the Monin-Obukhov similarity theory.

In the ETLook model, the minimal stomatal resistance ($r_{s,min}$) and the maximum obstacle height ($Zobst_{max}$) are landuse-landcover specific parameters that have to be calibrated and set up by user.

Finally, the third component of the water balance is interception, i.e. the fraction of the rainfall that is intercepted by the canopy before reaching the surface. It is computed according to the Braden's method (Equation 3)

$$I = 0.2 \cdot LAI \cdot \left(1 - \frac{1}{1 + \frac{FVC \cdot P}{0.2 \cdot LAI}}\right) \cdot \frac{\lambda}{86{,}400}$$

Equation 3

being, P the precipitation [mm/day], λ is the latent heat of evaporation [J/kg], and FVC and LAI are the Fractional Vegetation Cover [-] and LAI the Leaf Area Index [-], respectively, both computed from satellite values of Normalized Difference Vegetation Index (NDVI).

The WaPOR Wiki [link] can be visited for a more detailed description of the modelling algorithms used for computing energy and water fluxes at the surface.

5.1.2 C-Fix

The C-Fix model (Veroustraete et al., 2002) is a diagnostic model which rests on the Penman parametric equation which is forced driven by temperature, radiation, and the fraction of Absorbed Photosynthetically Active Radiation (fAPAR). C-fix computes the Net Primary Productivity at submonthly scale as:

$$NPP = fAPAR \cdot SM \cdot LUE_{max} \cdot NPP_{max}$$

Equation 4

Where, NPP is Net Primary Production [gC/m²·day], fAPAR is the Fraction of photosynthetically active radiation absorbed by green vegetation [JPA/JP], SM Soil moisture stress reduction factor, and LUE_{max} is the Light Use Efficiency at optimum conditions [kgDM/GJPA or gDM/MJ] which refers to the maximum amount of biomass (or carbon) that a plant can produce per unit of absorbed photosynthetically active radiation (PAR) under ideal conditions. This parameter is closely linked to the crop yield potential, and it is a landuse-landcover specific parameter that have to be calibrated and set up by user.

NPP_{max} = Maximum obtainable NPP for when fAPAR equals one [gC/m²/day]. It relies on the incoming shortwave radiation and two scalar factors that control how radiation is converted into dry matter and net primary productivity, and how this conversion is mediated by temperature and CO2 content stressors.

Once NPP is computed, Total Biomass Production in a year or season is generated by accumulating the NPP values fort the period of interest (Equation 5).

$$TBP = \sum_{dk=1}^{36} 22.222 \cdot NPP_{dk}$$

Equation 5

where, TBP is the Total Biomass Production [kgDM/ha], and NPP_{dk} = Dekad Net Primary Production [gC/m2].

5.1.3 Water productivity

Biomass water productivity [kgDM/ha], also known in literature as Water Use Efficiency, indicates how efficiently water is used in an agricultural system to produce biomass. It is the ratio between total biomass (aboveground + belowground) generated in a period of interest and the water consumed for its generation. It can be expressed as Gross Biomass Water Productivity when the total amount of water consumed is considered (GBWP = TBP/ETI), or as Net Biomass Water Productivity when only the transpiration loss is taken (NBWP = TBP/T).

The Net Biomass Water Productivity provides valuable insight into the effectiveness of water use in agricultural systems, and it is targeted within this exercise as the main output indicator to support the SOS definition in the pilot case.

Although it out of the scope of this prototype, the retrieval of values of Crop Water Productivity would require the conversion of the total biomass into crop yield, which involves the need for quantifying the fraction of the total biomass that is aboveground, and the fraction of the aboveground biomass which is harvestable. This conversion is influenced by the crop type but also by other agronomic and managerial factors that are out of the EO domain.

5.2 PyWaPOR

PyWaPOR[3] is the open-source Python package that provides users with the ability to generate water productivity and related outputs at any region and period of interest by using the same algorithms as the WaPOR database. PyWaPOR consists of two modules (Figure 2) the PRE ET-Look includes all those routines that aim to collect and prepare the input data into the right shape and format needed for running the ET-Look module in which the core algorithms for the retrieval of energy, water and productivity variables are located.

A configuration file is available for the establishment the user input parameters related with the simulation domain (region of interest, time period, and spatial aggregation), the usage of EO sources the model configuration, or the number and types of outputs for being generated.

[3] https://www.fao.org/aquastat/py-wapor/index.html

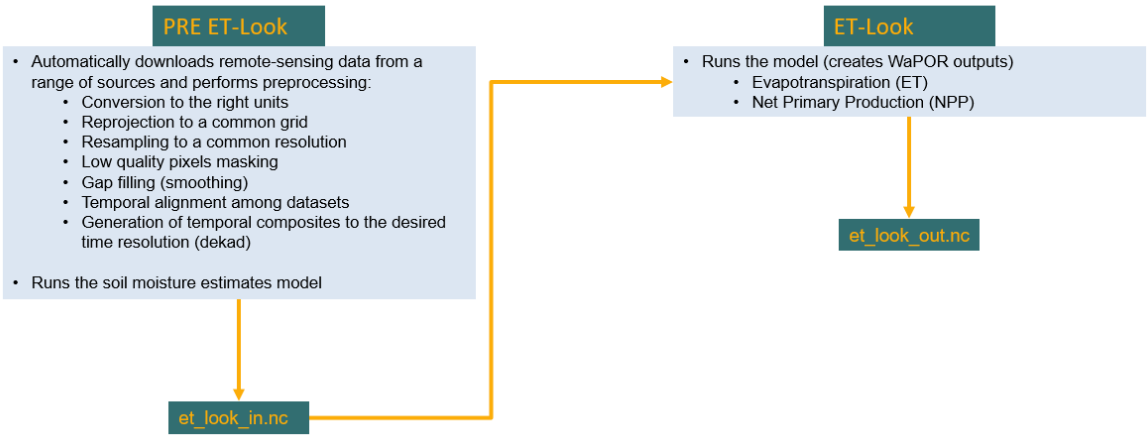

Figure 2. PyWaPOR workflow

5.3 Prototype improvements

As part of the SOS EOP5 prototype, the default PyWaPOR-Level 3 configuration has been enhanced for being able to ingest:

- Climate-forcing variables from ERA5 dataset.
  Default configuration of PyWaPOR–Level 3 combines climate forcings from MERRA (spatial resolution: 55 km) and GEOS5 (spatial resolution: ~30 km) datasets. The inclusion of ERA5 datasets (ERA5-agro and ERA5-reanalysis) increases the spatial resolution of the climate forcings. A new routine has been also included to compute specific air humidity from the ERA5 raw products.

- High-resolution LULC map
  The LULC characterization in pyWaPOR – Level 3 rests on the GLOBCOVER dataset (spatial resolution: 300 m). In this prototype, the ingestion of a new land use cover map has been tested by using the SIGPAC product generated by the Spanish Ministry of Agriculture[4]. A brief description of the SIGPAC and main advantages over the GLOBCOVER map is provided in Box 1.

[4] https://sigpac.mapa.gob.es/fega/visor/

Box 1. SIGPAC description.

SIGPAC is a high-res product regularly generated by the Spanish Ministry of Agriculture. SIGPAC is a high-res product at the parcel level which offers a more detailed and accurate characterization of cropland than other global products. Main improvements relate with: 1) a better discretization of cropland categories, e.g. 4 in GLOBCOVER vs 7 in SIGPAC, and level of mixture in mosaic parcels, 2) an improved granularity with a higher spatial resolution (300m in GLOBCOVER vs a parcel-based at 1:5000 scale which has been resampled to 30m for the purposes of this analysis). Large differences between GLOBCOVER (default WaPOR product) and SIGPAC are briefly illustrated in Figure B1.1. GLOBCOVER strongly overestimates the representativeness of rainfed crop categories, which is very far from reality. Additionally, GLOBCOVER is not able to differentiate well among different cropland categories which is also mandatory in this study.

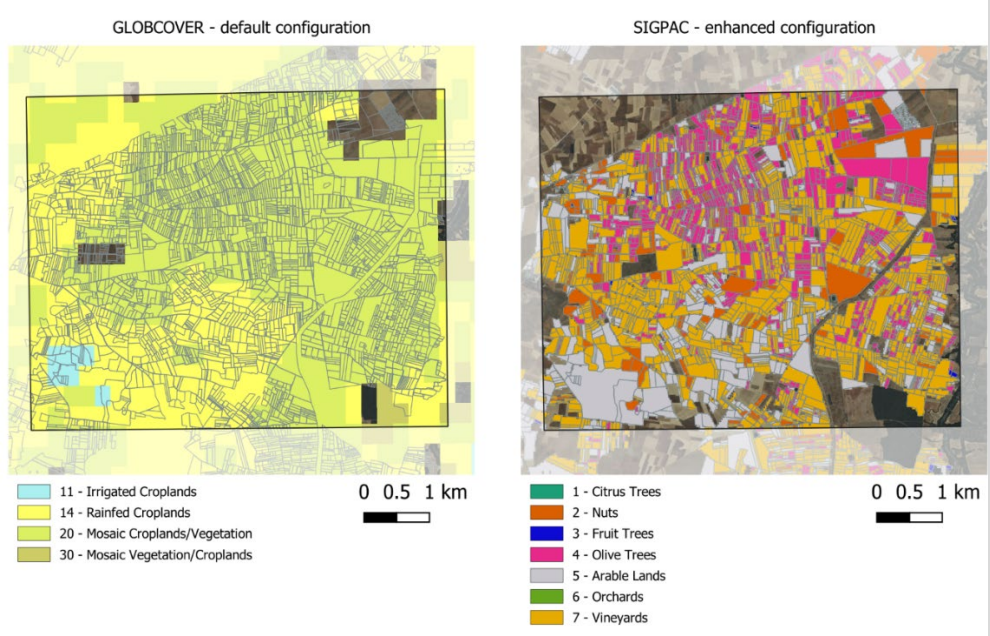

Figure B1.1. SIGPAC vs GLOBCOVER landcover categories in an agricultural district located in the Jucar River Basin

5.4 Pilot implementation and validation

The enhanced version of WaPOR within the JWRS was implemented in four moderate-sized (~50 km^2) test sites that all together provide a good representation of the most dominant cropland categories - including rainfed and irrigated systems - and spatial variability developed in the region. The selected test sites are located in three agricultural districts (A5030, A5150 and A5125) of the region (Figure 3). Table 2 and Figure 4 show the area of each test site, and the most dominant cropland categories present in each one.

Table 2. Area and dominant crop categories in selected pilot areas

Pilot_ID	Prov.*	Area (km²)	SIGPAC cropland categories											
			TA	TAi	FS	FSi	FY	FYi	OV	OVi	VI	VIi	CI	CIi
A5030-01	Cu	32.66	■	■					■		■	■		
A5030-03	Ab	73.83	■	■	■	■								
A5150-01	Va	32.78		■				■						■
A5125-01	Va	55.06			■						■	■		

Cropland categories: VI = Vineyards, OV = Olive trees, TA = Arable lands and others, FS = Nuts, CI = Citrus trees, FY = Fruit trees. Suffix "i" means irrigated.

* Province (NUTS3): Cu = Cuenca, Ab = Albacete, Va = Valencia

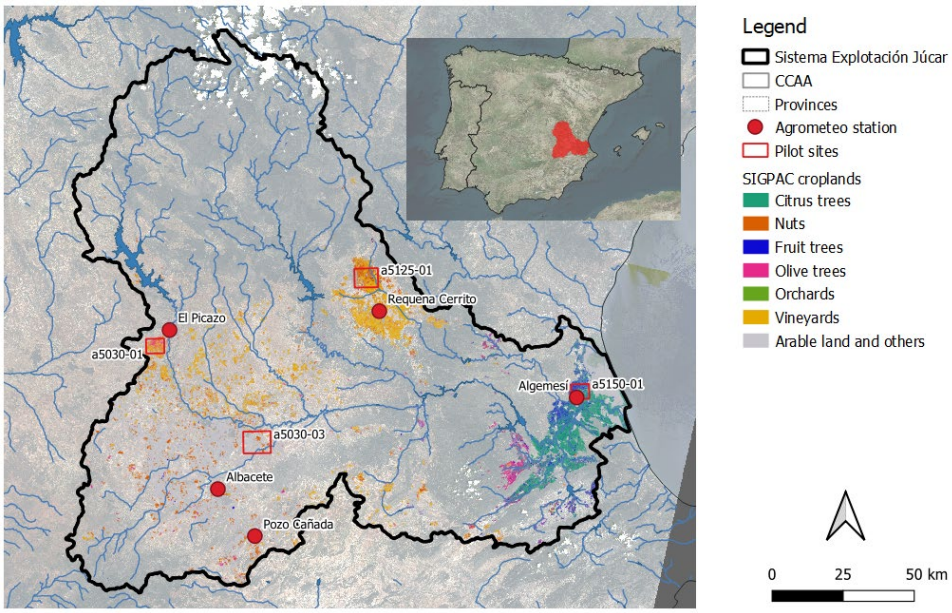

Figure 3. Location of test sites selected for WaPOR implementation

In each site, only pixels classified as "pure" in terms of cropland dominance were adopted for the retrieval of WaPOR variables. Pixel classified as mosaic or mix of cropland were excluded form the analysis.

To run the ETLook and C-fix models, a look-up table with values for three key LULC-specific parameters were set up for each cropland category in the study region (Table 3). $Zobst_{max}$ and LUE_{max} represents typical values in the region which have been agreed with UPV staff based on expert knowledge. For calibrating Rs_{min}, an internal sensitivity analysis was performed for each cropland category by testing a range of plausible values and selecting the ones that minimize the error in ET estimations when are compared against typical values found in literature (Appendix 1).

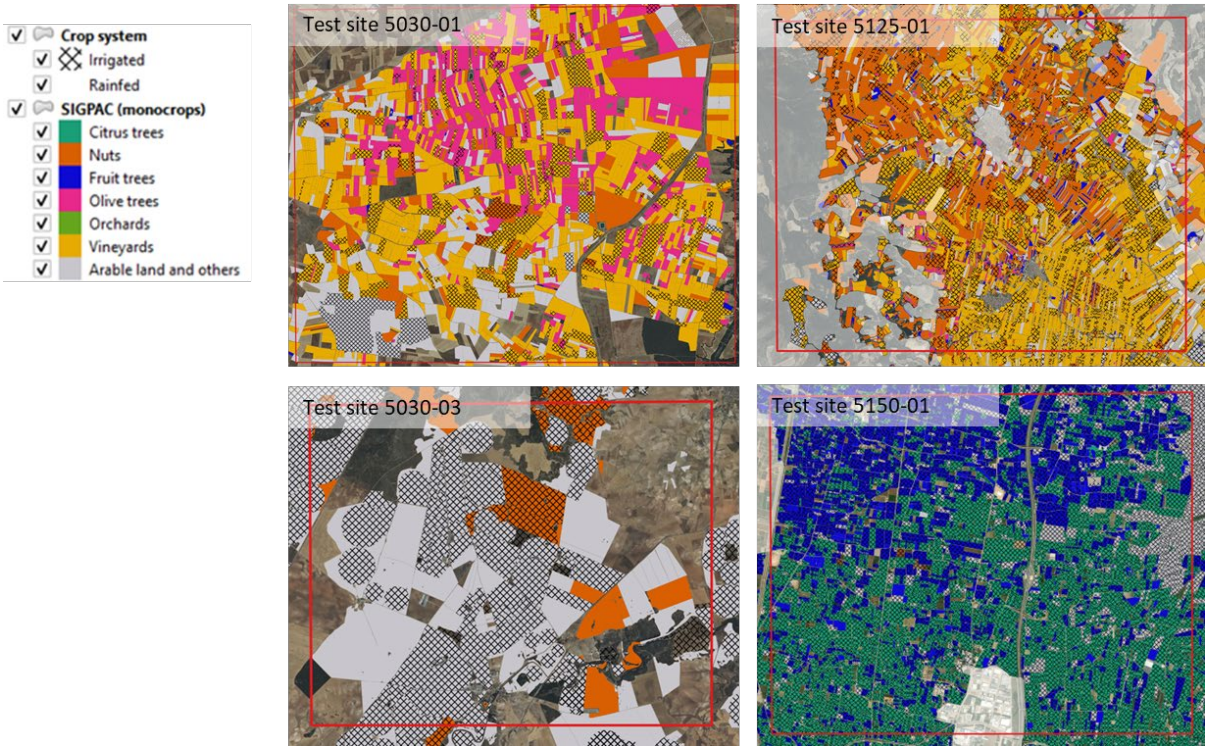

Figure 4. Croplands in pilot areas of implementation

Table 3. Selected SIGPAC crop categories and biophysical parameters used in the ETLook and C-fix models.

ID	Cropland category	Type of crops	Rs_{min} [s/m]	$Zobst_{max}$ [m]	LUE_{max} [gDM/MJ]
CI	Citrus trees	Orange, lemons, limes, grapefruits	250	4.0	2.70
FS	Nuts	Almond trees, walnut trees, hazelnut trees, pistachio	85	5.0	2.70
FY	Fruit trees	Pome fruits (apples, pears), stone fruits (peaches), tropical fruits (avocado)	100	4.0	2.70
OV	Olive trees		150	3.5	2.70
TA	Arable lands	Cereals, oilseeds, legumes, root crops, vegetables, industrial crops (cotton, sugarcane)	100	1.0	2.70
TH	Orchards	Vegetables (e.g. tomatoes, onions), aromatic herbs (e.g. cilantro, rosemary, lavender), small fruits (berries), medicinal plants	150	0.5	2.70
VI	Vineyards		100	2.0	2.70

Separate runs of the WaPOR algorithm were successfully implemented for the four test sites selected by covering the 2018-2022 simulation period, and average values of NBWP were computed for each crop category.

Due to the general lack of actual data needed for performing a comprehensive validation analysis for all the LUC+irrigation systems evaluated, WaPOR estimates for ET (primary water flux) and Total Biomass Production (TBP, the primary carbon flux) were verified in several ways:

ET WaPOR vs typical figures from literature review

The first level of verification consists of the comparison of annual values of ET from WaPOR with typical figures of ET, actual or simulated, reported for Mediterranean crops in technical and scientific literature.

ET WaPOR vs actual measurements

The second level of verification consists of a comparison between daily measurements of actual evapotranspiration (ET) obtained from an eddy-covariance flux tower operated by the UPV team and ET estimates derived from WaPOR for the same area. The flux tower footprint encompasses three rainfed almond farms located in the vicinity of the testing site *a5125-01* (Figure 5). The observation period spans from 30 March to 13 August 2025. WaPOR was run independently for the study area, and pixel-based ET estimates corresponding to the three farms were spatially averaged for comparison. Scatterplots of daily ET and dekadal averages were analysed for this land-use class.

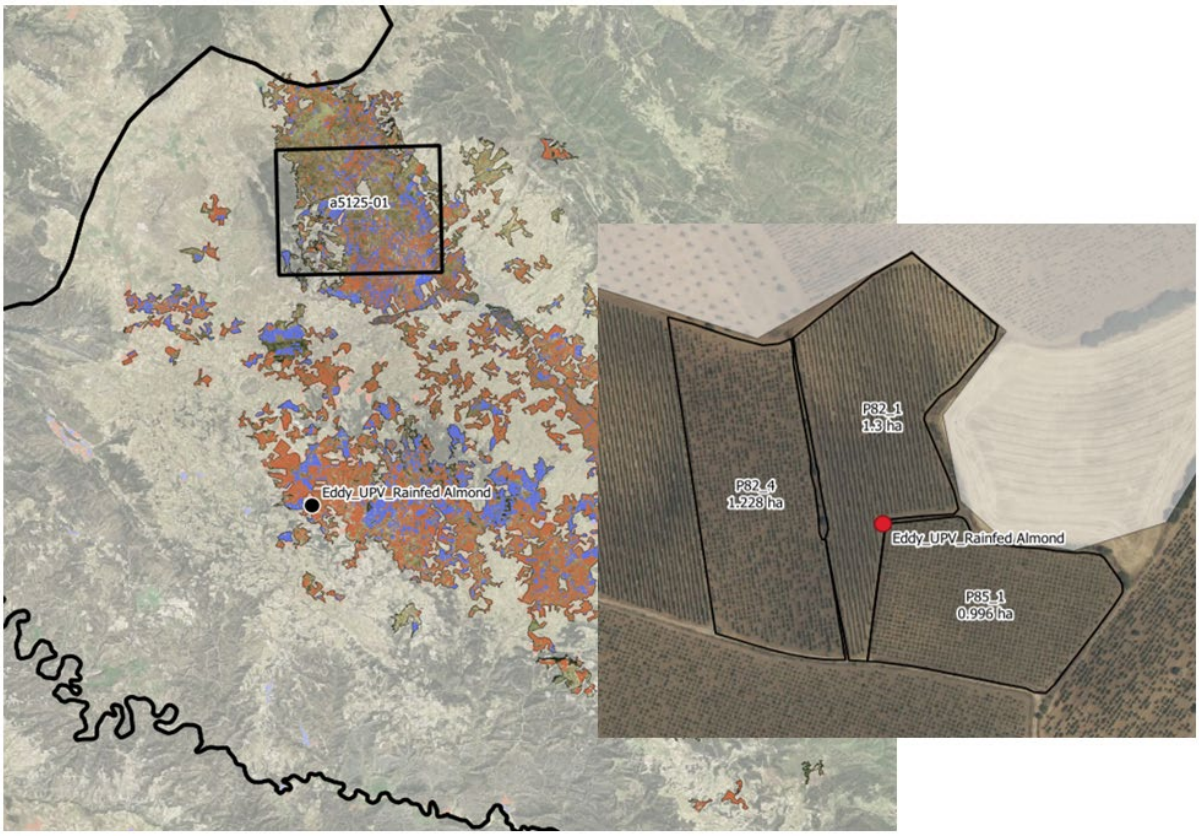

Figure 5. Location of the UPV's Eddy-covariance flux tower with measurements of actual evapotranspiration in an area dominated by rainfed almonds.

TBP WaPOR vs Crop Yield measurements

The conversion of the Total Biomass Production (TBP) into Crop Yield (CY) is modulated by the HBF scalar, as

$$CY = TBP \cdot HBF = TBP \cdot fAG \cdot HI$$

Equation 6

The harvestable biomass fraction (HBF) represents the fraction of total biomass converted into harvestable product. Conceptually, HBF is defined as the product of two dimensionless scalars: fAG, the aboveground biomass fraction, and HI, the harvest index, which represents the fraction of aboveground biomass allocated to the harvested product.

A strong linear correlation between WaPOR estimates of total biomass production (TBP) and observed crop yield (CY) can be considered a prerequisite for yield estimation and an indication that the WaPOR model captures production dynamics and relative ranking across crops and management systems reasonably well. However, such a relationship alone does not demonstrate the model's ability to quantitatively predict crop yield without further evaluation.

To evaluate the linear TBP-CY relationship as a prerequisite for yield estimation, annual spatially averaged TBP values derived from WaPOR for cropland categories and irrigation management were compared against annual observed crop yield values obtained from the ESYRCE dataset[5]. Crop yield data for the dominant crop within each cropland category were collected for the three NUTS-3 units in the study region and subsequently averaged to obtain a single annual value per cropland category and irrigation management system (Table 4). The strength of the relationship (R^2) and the slope of the linear regression -interpreted as the harvestable biomass fraction (HBF in Equation 6) - were then extracted.

[5] ESYRCE (Encuesta de Superficies y Rendimientos de Cultivos en España) is the official agricultural survey conducted annually by the Spanish Ministry of Agriculture (MAPA) to provide statistically representative, harmonised, and reproducible estimates of agricultural production (crop areas and crop yields) in Spain, and it serves as a core input for national statistics, EU reporting (Eurostat), and policy analysis. Crop yield values in ESYRCE represent observed or actual harvested yields, aggregated statistically to the national (NUTS-1) and regional levels (NUTS-2 and NUTS-3), rather than potential yields or biomass-based estimates. More info can be found here.

Table 4. Average number of plots surveyed per year and mean annual crop yield value observed in the 2018-2022 period observed for most dominant crops per cropland category in the study region.

Cropland category	Dominant crop	Rainfed		Irrigated	
		Plots*	Avg. Crop Yield ($kg \cdot ha^{-1}/y$)	Plots*	Avg. Crop Yield ($kg \cdot ha^{-1}/y$)
Arable	2-row barley	714	477	50	995
Nuts	Almond	344	119	122	234
Fruit	Orange	0	0	2663	4440
Olive	Olive (oil)**	423	320	193	668
Vineyard	Vineyard (wine)**	353	1375	185	1988

* mean annual of number of plots surveyed in the period 2018-2022.
** for oil or wine production

6 Results

Table 5 reports the mean annual values in the 2018-2022 period for the main primary outputs retrieved from WaPOR algorithm in the Jucar River Basin, including estimates of Transpiration (T), Total Evapotranspiration (ET), Net Primary Production (NPP), Total Biomass Production (TBP), and Net Biomass Water Productivity (NBWP). These figures represent the spatially- and timely averaged values per each cropland category and irrigation scheme evaluated.

The following conclusions can be derived:

- Total evapotranspiration values estimated in the region by WaPOR algorithm range from 182 mm/year (rainfed olive trees) to 939 mm/year (irrigated nuts).
- Lowest values of NBWP in the region during the simulation period was observed for rainfed arable lands and vineyards (1.60 – 1.75 kgDM/m^3, Figure 6) while the highest productivities were reached in irrigated Citrus trees with an mean annual value of 4.23 kgDM/m^3 (Table 5).
- Interannual variability of NBWP measured as the max-min range is highest for nuts (2.25 – 3), followed by vineyards (1.55 – 2.2) (Figure 6).
- Only in nuts, the impact of irrigation in net biomass water productivity was observed by a scale factor of 1.12. No relevant differences were detected for the other croplands due to lack of samples for comparison.

Table 5. Average values of main WaPOR variables for crop categories and irrigation schemes

Crop	Irrigation scheme	T (mm/y)	ET (mm/y)	NPP[1)] (gCm^{-2}/y)	TBP ($kgDM \cdot ha^{-1}/y$)	NBWP ($kgDM/m^3$)
TA - Arable land	Rainfed	164	269	150	3344	2.03
	Irrigated	353	483	328	7295	2.05
FS – Nuts	Rainfed	121	600	131	2930	2.54
	Irrigated	314	939	412	9173	2.84
FY – Fruit trees	Rainfed					
	Irrigated	460	570	411	9142	1.96
OV – Olive trees	Rainfed	89	182	111	2478	2.76
	Irrigated					
VI - Vineyards	Rainfed	157	230	137	3053	2.05
	Irrigated	183	256	160	3577	2.07
CI – Citrus trees	Rainfed					
	Irrigated	278	381	529	11761	4.23

1) annual mean of the dekad values

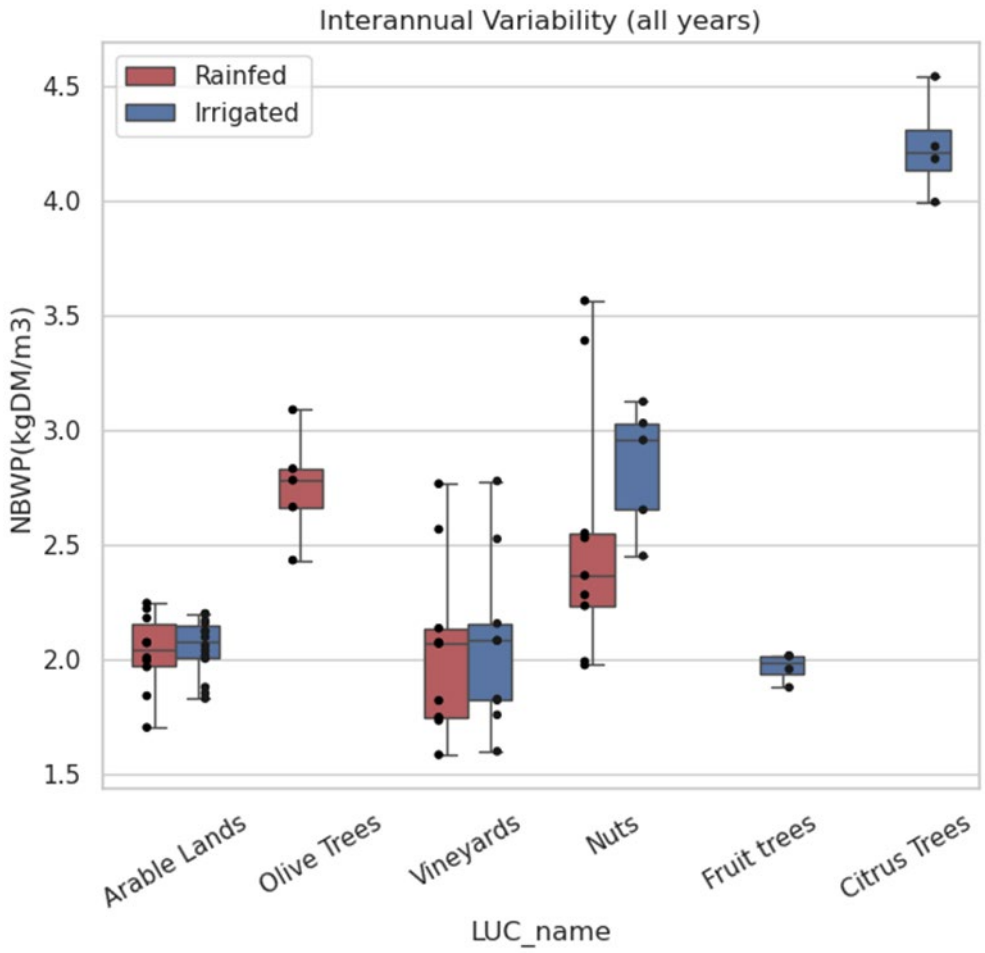

Figure 6. Interannual variability of annual values of NBWP per cropland category observed in the 2018-2022 period. Whiskers show maximum-minimum annual values.

Figure 7 shows the spatial variability of the mean annual NBWP values estimated in each cropland category. The tails in the violin plots were cropped to the 10th (lower tail) and 90th (upper tail) percentiles to exclude outliers. Highest spatial variability in the NBWP estimates was retrieved for nuts (80% of the values are in the range of 2.15 to 3.60 kgDM/m3 for the rainfed cases, and from 2.45 to 3.20 kgDM/m^3 for irrigated).

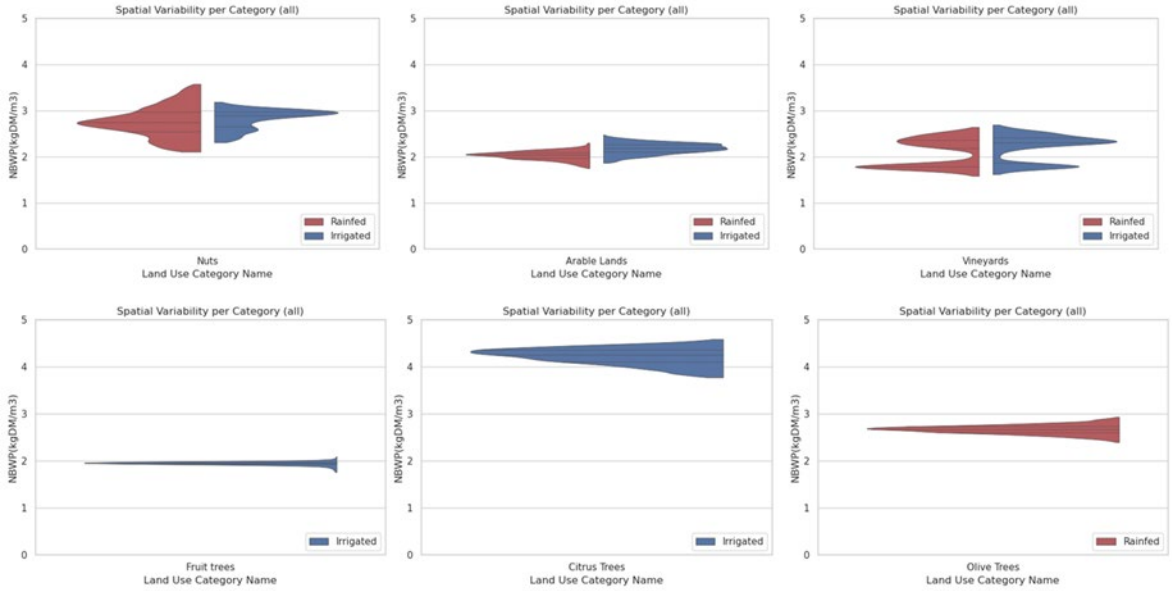

Figure 7. Spatial variability of interannual average (2018-2022) of NBWP values per cropland category

When comparing with other data sources, evapotranspiration (ET) values estimated using the locally calibrated WaPOR algorithm are consistently lower than those reported for the same or comparable crops and varieties in Mediterranean climates (see Appendix 1, which summarises typical crop ET values for California, USA, and the Mediterranean Basin).

Two main findings emerge from the comparison between observed actual ET at UPV's validation site and WaPOR simulations. First, WaPOR systematically underestimates actual ET for rainfed almond orchards under the adopted calibration settings (Figure 8, Figure 9), confirming conclusions drawn from the literature review. Second, WaPOR does not adequately capture the short-term dynamics of soil moisture and ET during rainfall-driven wet pulses, resulting in increasing bias with higher rainfall amounts, although this relationship remains relatively weak (Figure 10).

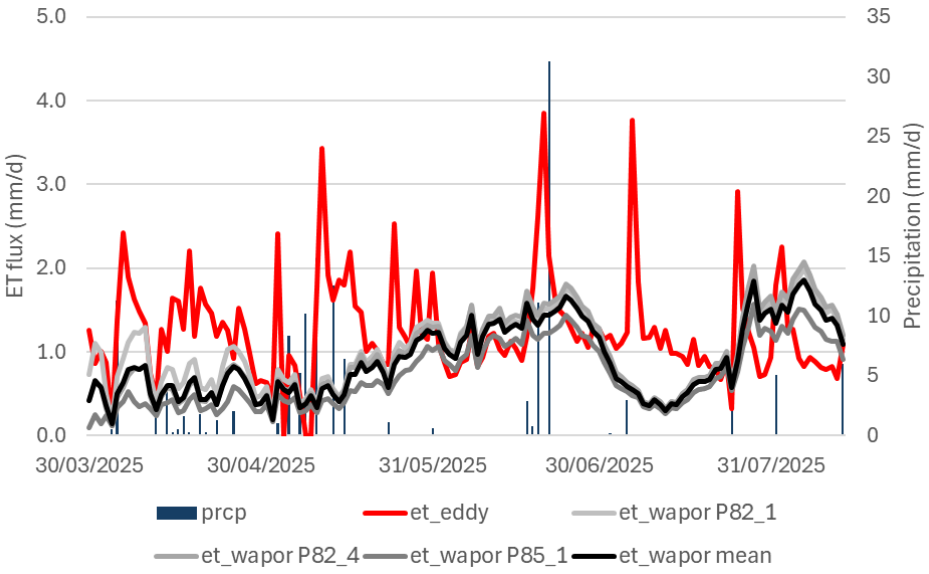

Figure 8. Temporal dynamic of daily evapotranspiration fluxes (mm/day, left axis) from eddy-covariance measurements (red line) and WaPOR algorithm (grey lines) at UPV's validation site with rainfed almonds.

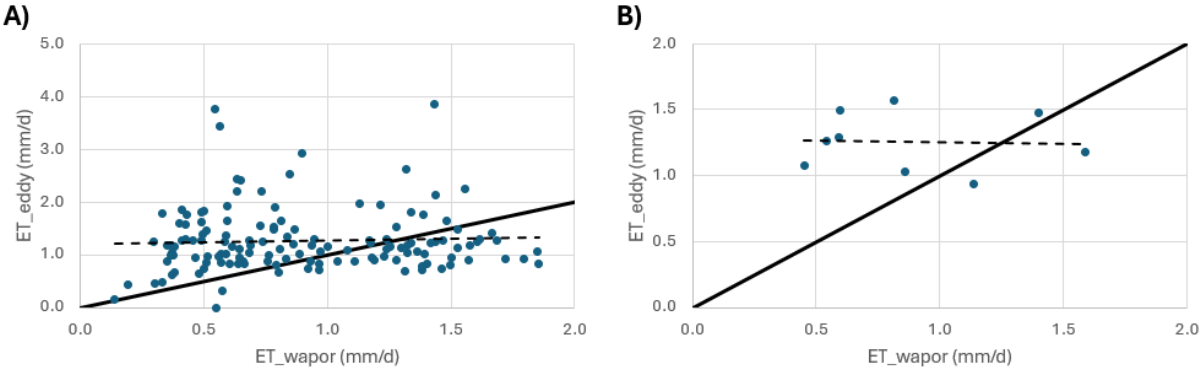

Figure 9. Comparison of evapotranspiration fluxes from Eddy-covariance measurements and WaPOR estimations in UPV's validation site (rainfed almond). A) daily values in the monitoring period, b) mean daily per dekad

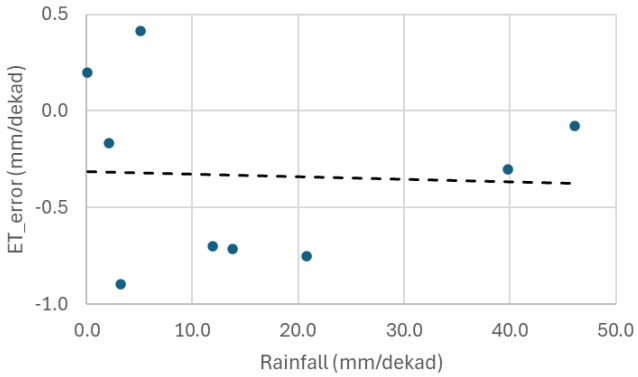

Figure 10. Dekadly error (accumulated) accounted in WaPOR ET estimation according to the total rainfall monitored at UPV's validation site.

The validation of the carbon component of the WaPOR algorithm is presented in Figure 11. Overall, strong linear correlations were obtained ($R^2 > 0.75$). The lowest performance was observed for the nuts category (almonds), which is consistent with the results derived from the validation against in-situ evapotranspiration measurements at UPV's testing site. WaPOR estimates of total biomass production show the strongest correlations for arable land (two-row barley) and vineyards, indicating that relative variability in production across time (interannual dynamics) and management conditions (rainfed versus irrigated systems) is well captured. However, the presence of a high correlation between total biomass production and observed crop yield alone is insufficient to confirm the correctness of absolute magnitudes, to identify the dominant stressors or processes involved (e.g. water, heat, or nutrient limitations), or to infer the harvestable biomass fraction represented by the slope of the relationship. This analysis should therefore be regarded as a first exploratory validation of WaPOR estimates in the region. A more comprehensive assessment, incorporating greater spatial variability and longer time series, is required to robustly evaluate quantitative yield accuracy and crop stress responses.

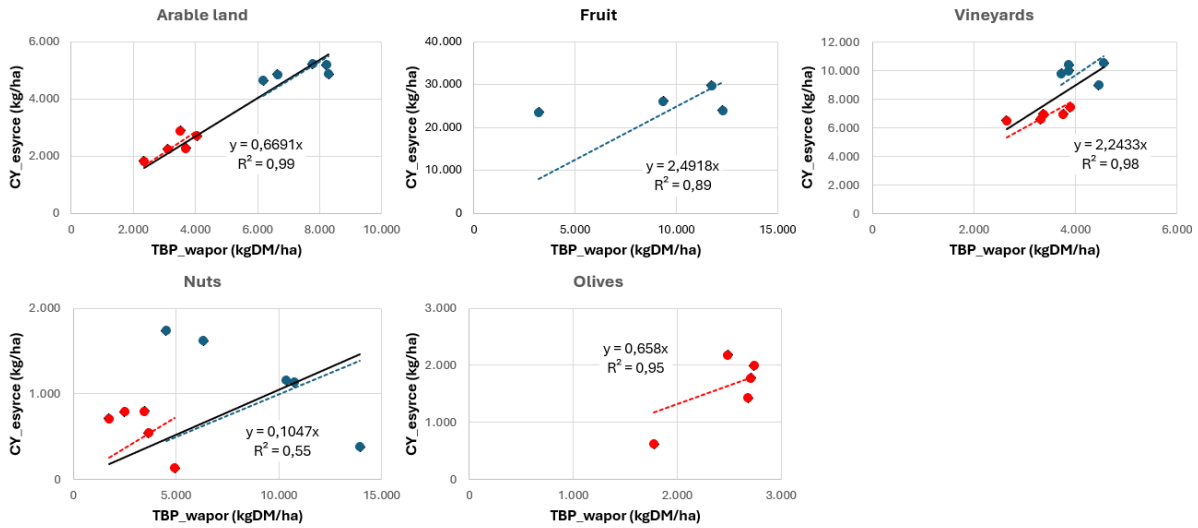

Figure 11. Relationships between estimated Total Biomass Production (TBP_wapor) and observed Crop Yield (CY_esyrce) in cropland categories. Blue dots for irrigated crops, red dots for rainfed crops.

7 Final remarks

The algorithm behind the WaPOR2 database was enhanced from its default configuration and properly adapted for increasing the spatial resolution and accuracy of the water productivity values in the Jucar River Basin. As final remarks, it is stated that:

1. The default configuration of PyWaPOR–Level 3 has been satisfactorily adapted and tested for a new pilot site, in order to solve the lack of WaPOR-Level 3 data in the official portal. This exercise demonstrates the adaptability of WaPOR for being used in different geographical settings. The implementation of the algorithm allows the user to choose the native sources from which remote sensing data are downloaded. Sideloading is available, so the user can even generate its own remote sensing datasets to be ingested prior to run the model. This level of customization is a very positive highlight found through the implementation of this exercise.
2. The flexibility of the algorithm for the assimilation of high-resolution local datasets has been tested by the ingestion of a local land use/land cover dataset, tailoring the analysis to the study area. Due to the complexity of the model and the multiple interdependencies between parameters, the ingestion of local datasets can be challenging in the sense that extra parametrization might be needed. Some advanced knowledge on the algorithm is needed in order to identify how the workflow may be affected by the replacement of a default dataset - in this particular exercise, the ingestion of a local land use dataset required the set up of three biophysical parameters.
3. Regarding the accuracy of the WaPOR results, there are some issues that might reduce it, e.g.: the high sensitivity of WaPOR to the presence of data gaps in land surface temperature (strongly affected by cloud cover, as it is derived from thermal bands), the weak characterization/lack of dynamic data for characterization of landcover, and coarse native sources of data (addition of uncertainty due to resampling and interpolation techniques to match the highest resolution source).

Nowadays, spatial and temporal resolutions and datasets already available through the WaPOR portal or PyWaPOR configurations may not fit all applications and user needs. The FAO's WaPOR developer team is working towards releasing the highest resolution possible at the global scale. This study lead by FutureWater has served to identify and fix bugs, optimize the code, and support technical improvements in the official versions of WaPOR. However, the open-source algorithm available is powerful and provides unprecedent opportunities for improvement and scaling up.

The interpretation of the data to be used in decision making needs advanced knowledge of the area of interest and additional validation with different datasets is needed in order to apply WaPOR in new pilot areas.

All the algorithms used in this exercise are included in the Github project where user can find all the algorithms and configuration files needed for running WaPOR-Level 3 (z1*) within the JWRS. A detailed README file guides the user on how to proceed.

- The main script is *pywapor_v343.p*y, which prepares the inputs and runs the model.
- For the extraction of spatially-aggregated values (.nc files and .xlsx tables) at the polygon level, the user needs to run pywapor_postprocessing.py script by filling *shp_name* and version for the area of interest. From this script, graphs like the ones presented in the results section can be obtained.
- Optionally, for the extraction of timeseries for specific sites, prepare a shapefile with the sites of interest (e.g. *area_valsites.shp*) and run *extract_tss4sites.py* by filling demo_site parameter.

References

FAO. (2023). Remote sensing determination of evapotranspiration – Algorithms, strengths, weaknesses, uncertainty and best fit-for-purpose. Em *Remote sensing determination of evapotranspiration*. FAO. https://doi.org/https://doi.org/10.4060/cc8150en

Kahil, M. T., Dinar, A., & Albiac, J. (2015). Modeling water scarcity and droughts for policy adaptation to climate change in arid and semiarid regions. *Journal of Hydrology*, *522*, 95–109. https://doi.org/10.1016/J.JHYDROL.2014.12.042

McCallum, I., Wagner, W., Schmullius, C., Shvidenko, A., Obersteiner, M., Fritz, S., & Nilsson, S. (2009). Satellite-based terrestrial production efficiency modeling. *Carbon balance and management*, *4*, 8. https://doi.org/10.1186/1750-0680-4-8

Momblanch, A., Andreu, J., Paredes-Arquiola, J., Solera, A., & Pedro-Monzonís, M. (2014). Adapting water accounting for integrated water resource management: The Jucar Water Resource System (Spain). *Journal of Hydrology*, *519*, 3369–3385. https://doi.org/10.1016/j.jhydrol.2014.10.002

Monin-Obukhov. (1954). *Monin-Obukhov Theory*. https://www.sciencedirect.com/topics/earth-and-planetary-sciences/monin-obukhov-theory

Pelgrum, H., Miltenburg, I., Cheema, M., Klaasse, A., & Bastiaanssen, W. (2010). ETLook a novel continental evapotranspiration algorithm. *Remote Sensing and Hydrology Symposium, Jackson Hole, Wyoming, USA*, *1085*, 1087.

Pereira, L. S., Cordery, I., & Iacovides, I. (2012). Improved indicators of water use performance and productivity for sustainable water conservation and saving. *Agricultural Water Management*, *108*, 39–51. https://doi.org/10.1016/j.agwat.2011.08.022

Ramos, T. B., Darouich, H., Oliveira, A. R., Farzamian, M., Monteiro, T., Castanheira, N., Paz, A., Gonçalves, M. C., & Pereira, L. S. (2023). Water use and soil water balance of Mediterranean tree crops assessed with the SIMDualKc model in orchards of southern Portugal. *Agricultural Water Management*, *279*. https://doi.org/10.1016/j.agwat.2023.108209

Rana, G., Katerji, N., & De Lorenzi, F. (2005). Measurement and modelling of evapotranspiration of irrigated citrus orchard under Mediterranean conditions. *Agricultural and Forest Meteorology*, *128*(3–4), 199–209. https://doi.org/10.1016/J.AGRFORMET.2004.11.001

Testi, L., Villalobos, F. J., Orgaz, F., & Fereres, E. (2006). Water requirements of olive orchards: I simulation of daily evapotranspiration for scenario analysis. *Irrigation Science*, *24*(2), 69–76. https://doi.org/10.1007/S00271-005-0011-Y/METRICS

Veroustraete, F., Sabbe, H., & Eerens, H. (2002). Estimation of carbon mass fluxes over Europe using the C-Fix model and Euroflux data. *Remote Sensing of Environment*, *83*(3), 376–399. https://doi.org/10.1016/S0034-4257(02)00043-3

Yang, Y., Guan, H., Long, D., Liu, B., Qin, G., Qin, J., & Batelaan, O. (2015). Estimation of Surface Soil Moisture from Thermal Infrared Remote Sensing Using an Improved Trapezoid Method. *Remote Sensing 2015, Vol. 7, Pages 8250-8270*, *7*(7), 8250–8270. https://doi.org/10.3390/RS70708250

Disclaimer

Views and opinions expressed are those of the author(s) only and do not necessarily reflect those of the European Union or CINEA. Neither the European Union nor the granting authority can be held responsible for them.

This prototype has been developed and tested using version 3.4.3 of the PyWaPOR algorithm. PyWaPOR is a recent tool that is being updated regularly, with significant changes between versions, fixed bugs and modifications in terms of syntax for the different algorithms. Moreover, its high dependence from different native sources of data makes the algorithm sensitive to changes in these external sources' APIs.

During the implementation of this prototype, the Copernicus Climate Data Store was in process of migration to a new architecture, which resulted in big disruptions of the service. The new architecture comes with a more intuitive and user-friendly API, but this could lead to incompatibilities with the pywapor version that has been used for this exercise.

For further developments of this prototype, the migration of the code to the updated version of pywapor will be required.

Acknowledgement of funding

This project has received funding from the European Union's Horizon Europe research and innovation programme under grant agreement No 101059264.

Appendix 1. Typical ET values in Mediterranean croplands

Table A1. Annual ET values retrieved in this study and collected from other databases and studies.

Category	Crops	This study	Californian crops [1]	Other studies
Arable lands		**270 – 480**		
	Alfalfa Hay and Clover		1150	
	Corn and Grain Sorghum		730	
	Cotton		800	
	Onions and Garlic		500	
	Potatoes, Sugar beets, Turnip etc..		850 - 880	
	Safflower and Sunflower		670	
Citrus tress		**380**		800 – 1200* 700 – 870 [2] 760 – 780 [4]
	Citrus (no ground cover)		900 - 1000	
Fruit trees		**570**		
	Apple, Pear, Cherry, Plum and Prune		980	800 – 1200*
	Avocado		950	900 – 1300*
	Peach, Nectarine and Apricots		935-970	
Nuts		**600 – 939**		
	Almonds		995	800 – 1200* 610 – 970 [2]
	Pistachio		880	900 – 1300*
	Walnuts		1050	
Olive trees		**182**		600 – 1000* 530 – 620 [2] 885 – 1070 [3]
Vineyards		**230 - 250**		500 – 800*
	Grape Vines with 80% canopy		700	

- [1] Typical values for crops with sprinkler or drip irrigation for an average-rainfall year (1997) - precipitation = 419 mm; grass reference ETo = 1345 mm). Source: Irrigation Training & Research Center (https://www.itrc.org/)
- [2] Source: (Ramos et al., 2023)
- [3] Source: (Testi et al., 2006)
- [4] Source: (Rana et al., 2005)
- * Range of values provided by AI.