



Mapping water in a Dynamic World: A hybrid landcover-NDWI approach enhances surface water dynamics characterization from Sentinel-2

Leo Helling¹, Satbir Singh², Louis Rey¹, Hervé Parmentier¹, Mathis Loïc Messenger^{1,3}, Hervé Piégay¹,
5 Barbara Belletti⁴

¹ENS de Lyon, UMR 5600 - CNRS Environnement Ville Société, Lyon, France

²Indian Institute of Technology (BHU), Varanasi, India

³RiverLy Research Unit, National Research Institute for Agriculture, Food and Environment (INRAE), Villeurbanne, France

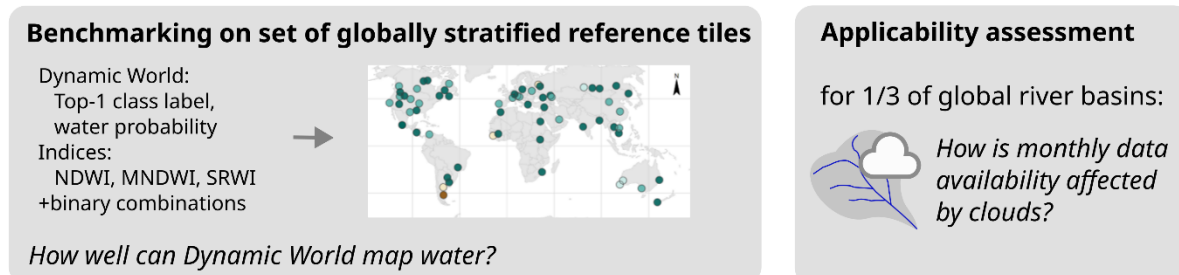
⁴Université Jean-Monnet, UMR 5600 - CNRS Environnement Ville Société, Saint-Étienne, France

10 Correspondence to: Leo Helling (leo.helling@ens-lyon.fr)

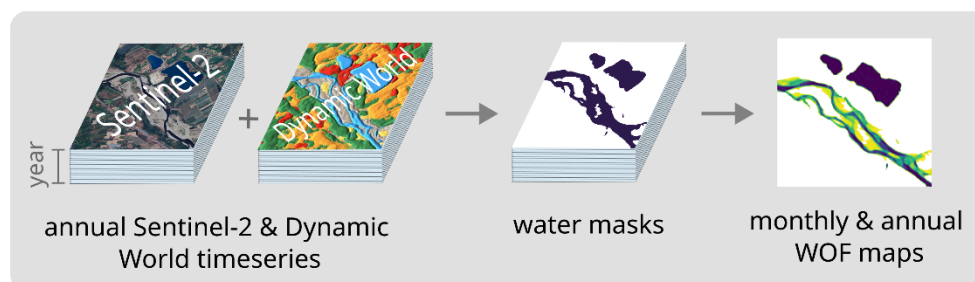
Abstract. Accurate information on the spatial and temporal dynamics of surface water is essential for monitoring freshwater ecosystems. However, existing satellite-based studies are often limited by coarse spatio-temporal resolution or constraints in their underlying classification algorithms. Here, we evaluate the Dynamic World land cover model for mapping inland water across globally diverse conditions using Sentinel-2 imagery. Cross-validation demonstrates that Dynamic World enables
15 accurate ($F1=0.90\pm0.04$) and computationally efficient water mapping, with performance further improved by combining it with the Normalized Difference Water Index (NDWI; $F1=0.91\pm0.04$), particularly for smaller water bodies and challenging conditions such as terrain shadows. Building on these results, we propose a computationally efficient and scalable workflow for mapping surface water dynamics from full annual Sentinel-2 time series. Our approach derives monthly and annual Water Occurrence Frequency (WOF) maps through automated processing on Google Earth Engine with advanced cloud and terrain-
20 shadow filtering. Applied to the Rhône-Mediterranean region in France — a hydro-morphologically complex mountainous region — the WOF product captures 21% more stable water surface and 71% more river network length than Landsat-based datasets. It better represents ephemeral water bodies and smaller streams, while significantly increasing precision over DSWE, the only other global 10-meter WOF product, which relies solely on the DW Top-1 label. We also assessed global applicability, revealing that cloud cover substantially reduces valid observations for several months in tropical and mountainous regions,
25 impacting the information density in WOF maps. The enhanced spatio-temporal detail of the method unlocks opportunities for characterizing inland water dynamics, including river geomorphology, floodplain connectivity, and wetland ecosystem functioning. The workflow is publicly available for use in other studies.



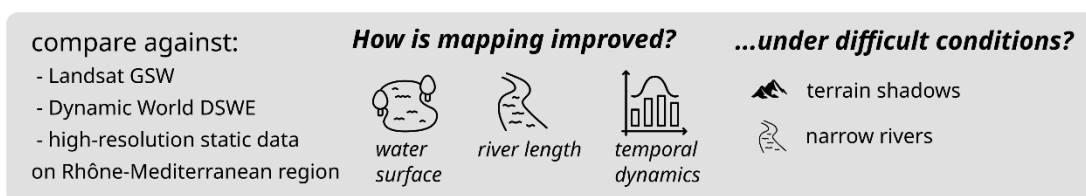
(1) Global evaluation of Dynamic World's water mapping capabilities



(2) Calculation of *water occurrence frequencies (WOF)*



(3) Regional evaluation of WOF output



1 Introduction

Most water bodies on Earth are dynamic systems whose extent varies in space and time (Pekel et al., 2016; Stegen et al., 2025). River networks, lakes and wetlands expand and contract in response to hydro-climatic, geological or human-driven processes (Messenger et al., 2021; Pi et al., 2022; Prancevic et al., 2025; Wu et al., 2023). Along river courses, floods temporarily inundate floodplains, whereas in deltas and floodplains river channels migrate laterally and undergo avulsions or cut-offs, creating complex space-time mosaics of water bodies (Jarriel et al., 2021). New permanent lakes emerge, while others dry up or drain (Pi et al., 2022). Human societies have profoundly altered the water cycle as well as the distribution and persistence of freshwater at the global scale (Cooley et al., 2021; Fluët-Chouinard et al., 2023). Knowing where, when, and how frequently



surface water occurs is crucial to distinguish human impacts from natural processes (Radeloff et al., 2024; Sogno et al., 2022). However, traditional methods like field surveys, *in situ* sensors, aerial photography and historical maps lack in spatial coverage or temporal continuity for large-scale monitoring of surface water dynamics.

40 Multispectral satellite programs such as MODIS, Landsat, and Sentinel-2 offer imagery at regular intervals with global coverage that can be used to map the extent of surface water and its changes (C. Huang et al., 2018). One method to characterize these dynamics is to calculate the water occurrence frequency (WOF) for each pixel as the ratio of the number of water observations to total observations. This metric indicates the probability that an individual pixel was occupied by water over a defined time period. When computed over multiple decades, WOF can reveal the long-term spatial imprint of water bodies and wetlands. For instance, WOF maps derived from MODIS (GLOBMAP SWF, (Y. Liu et al., 2022)) and Landsat time series (GSW Seasonality, Pekel et al., 2016; GLAD Annual Water Percent, Pickens et al., 2020) have been employed to map geomorphic patterns and change rates in rivers (Boothroyd et al., 2025; Gao et al., 2021; Luo, Park, et al., 2026) or to study inundation dynamics in floodplain wetlands (Allen, 2016; Dong & Hu, 2021; Luo, Feng, et al., 2026; Pal & Sarda, 2020). Nonetheless, these datasets are limited by the coarse spatial (250-500 m for MODIS) or temporal (16 days for Landsat) resolutions of the underlying satellite imagery. Sentinel-2's 10-m resolution and 5-day revisit time represent a significant improvement over previous sensors. Leveraging this improved spatio-temporal detail could extend the analysis to smaller water bodies and enable characterization of surface water dynamics at annual or seasonal scales. WOF maps at these scales could more effectively track hydro-geomorphological variability and anomalies (Li et al., 2023; Schwedhelm et al., 2026), or identify changes and trends (Song et al., 2013; Yao et al., 2023) throughout the functional hydrological cycle of water bodies, by comparing successive years or seasons.

However, existing attempts to extract water areas from Sentinel-2 imagery do not enable their large-scale application for high-precision WOF mapping. Current methods either rely on spectral index-based approaches (Ma et al., 2019; C. Zhao et al., 2025) or machine learning (Sogno et al., 2022). Spectral indices have traditionally been used due to their simplicity and physical interpretability, but are prone to errors under complex background environments such as artificial surfaces, building shadows, and snow (H. Liu et al., 2022; C. Zhao et al., 2025; Zhou et al., 2017). Machine learning methods, and especially deep learning (DL), have significantly improved classification through their ability to learn nonlinear relationships and incorporate spatial context (Sagan et al., 2020; Sogno et al., 2022). Yet, training and applying such models requires immense amounts of manually labeled data and computing capacities, thus limiting their geographic scalability. Until today, no dedicated DL model has been developed to map surface water dynamics from Sentinel-2 imagery on a global scale.

65 Probabilistic land use and land cover (LULC) classifications from the Dynamic World model (Brown et al., 2022) may provide a promising solution to efficiently map surface water at high spatiotemporal resolution. Dynamic World automatically applies a DL image segmentation model trained across global conditions to all Sentinel-2 images to provide LULC classifications in near-real time via Google Earth Engine (GEE). It supplies both a categorical classifier predicting the most probable LULC class (Top-1 label), and the specific probability estimate for the water class for each pixel. Using this model might improve water detection accuracy compared to traditional methods while minimizing computational effort. Initial studies have already



used Dynamic World for both short- and long-term WOF mapping (Jahangeer et al., 2026; Khare et al., 2025), for static water mapping (Mukherjee et al., 2024; Naboureh et al., 2025), or as inputs to more advanced water detection algorithms (Li et al., 2023; Lin et al., 2025; Y. Wang et al., 2025). However, it remains unclear how the water probability band compares to the more commonly used Top-1 label and which threshold should be applied to the probability band to most accurately identify water pixels under global conditions.

While Dynamic World excels as a generic LULC classifier, its performance for water mapping could be further improved through targeted enhancements. Its direct linkage to Sentinel-2 enables the integration of spectral indices, which have been shown to enhance classification accuracy when combined with DL segmentation results (Jahangeer et al., 2026; Kaplan & Avdan, 2017; Li et al., 2022). In addition, several post-processing steps could address known sources of artefacts in optical satellite imagery related to cloud cover and topographic shadows. Although these enhancements require relatively little computational power, they have not been tested in previous water mapping studies with Dynamic World.

The objectives of this study are therefore to **(i)** evaluate the capabilities of the Dynamic World model in comparison with commonly used spectral indices, and to identify how its use can be optimized, for mapping surface water areas worldwide from Sentinel-2 imagery; **(ii)** to investigate how the identified method improves WOF mapping in a hydro-morphologically complex region in France; and **(iii)** to assess its potential for global application.

2 Data and Methods

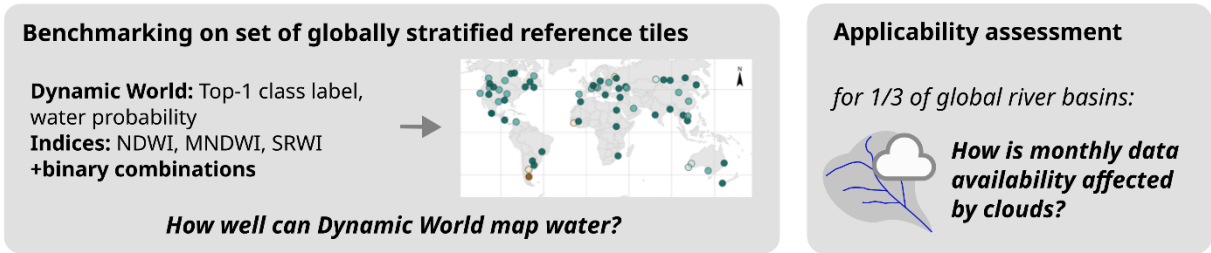
2.1 Methodological overview

First, we evaluated the surface water detection performance of Dynamic World (Top-1 label & water probability bands) on a global set of pre-labeled Sentinel-2 tiles (Figure 1; Wieland et al., 2024). We compared it to three common water indices calculated from Sentinel-2 images and explored combinations of each of the indices with the water probability band to determine the most effective global water detection method. To quantify mapping-related uncertainty and account for distribution bias in the reference (pre-labeled) dataset, we applied a globally stratified and area-weighted 5-fold cross-validation scheme.

In a second step, the resulting water detection method was integrated into a workflow to automatically map water occurrence frequencies (WOFs) from annual Sentinel-2 time series. We applied the workflow to the hydro-geomorphologically diverse French Rhône-Mediterranean region for 2021 and compared the resulting raster with three datasets: the Dynamic Surface Water Extent (DSWE) dataset (Khare et al., 2025) based on the Dynamic World Top-1 label, the Landsat-based Global Surface Water (GSW) Seasonality dataset (Pekel et al., 2016), and a high-resolution national land cover map of France. To understand

the behavior of these datasets under challenging conditions encountered across diverse freshwater ecosystems, we verified their performance on a selection of independent test-locations within the basin. Finally, we evaluated the impact of cloud inference on the availability of Dynamic World data at both the global and regional scale. To do this, we quantified the monthly and yearly coverage of Dynamic World predictions across more than 30% of river basins worldwide, and then assessed its availability in more detail in the Rhône-Mediterranean region.

(1) Global evaluation of Dynamic World's water mapping capabilities



(2) Calculation of water occurrence frequencies (WOF)



(3) Regional evaluation of WOF output

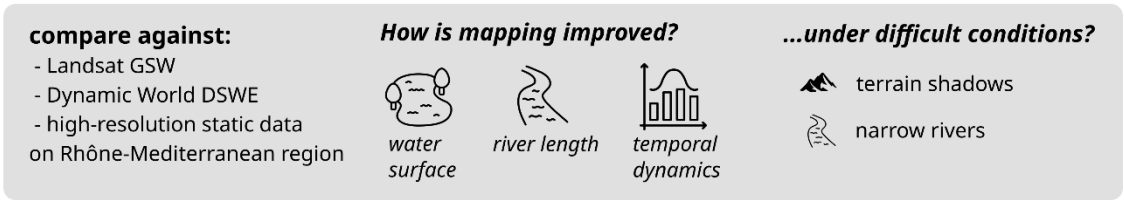


Figure 1: Overview of the methodology of this study illustrating the three key stages: (1), the global assessment of Dynamic World’s water detection capabilities, (2), the creation of a workflow to map annual water occurrence frequencies (WOF) based on the previously estimated optimal method for mapping water surface from Sentinel-2 imagery, and (3), the evaluation of the output against other datasets.



2.2 Data

2.2.1 Sentinel-2 and Dynamic World

Sentinel-2 is a global multi-spectral imaging satellite program operated by the European Space Agency. Its record going back to 2015, high spatial resolutions of 10-20 meters and recurrence time of a few days (3-5 at the equator) make Sentinel-2 a valuable resource for mapping surface water dynamics globally (Drusch et al., 2012).

Dynamic World is a land cover classification model for Sentinel-2 imagery based on a convolutional neural network (CNN), developed by Brown et al. (2022) and hosted on GEE (Gorelick et al., 2017). Dynamic World is applied automatically in near-real time to all available Sentinel-2 imagery. Unlike spectral index-based approaches that operate with a few spectral bands on a pixel-by-pixel basis, the Dynamic World's image segmentation technique relies on contextual spatial patterns (e.g., boundaries, texture, shapes) and the full Sentinel-2 spectral feature space (Brown et al., 2022). The dataset provides both a Top-1 class label, the most probable land cover class, and class-specific probability bands for each Sentinel-2 observation (pixel-date combination). The reported global accuracy of Dynamic World exceeds 90% for surface water when evaluated using the Top-1 label (Brown et al., 2022). However, the performance of the water class probability band for water extraction has not been explicitly assessed. Furthermore, the training and validation of the model was based on tiles with features of homogeneous land cover patches of at least 5×5 pixels, which may limit its sensitivity to small rivers or isolated floodplain water bodies. In the Dynamic World model pipeline, all Sentinel-2 tiles with more than 35% cloud cover are excluded from processing, and areas in the remaining images that may be affected by clouds and their shadows are removed through exhaustive modeling. However, the authors state that residual cloud and shadow contamination may persist in the model outputs (Brown et al., 2022). These characteristics highlight the need for a systematic evaluation of Dynamic World-derived water cover information, the consequences of exhaustive cloud mask removal, as well as its global application potential.

2.2.2 Pre-processing datasets

Cloud cover impacts ground visibility in many satellite images, especially in tropical and mountainous regions (Sudmanns et al., 2020) and around rivers (Langhorst et al., 2024), while topographic shadows in rugged terrain make the reflectance of diverse materials resemble that of water, causing systematic misclassification (C. Zhao et al., 2025). We therefore enhanced the Sentinel-2 Surface Reflectance collection with the Cloud Score+ collection (Pasquarella et al., 2023) to remove cloud-affected pixels. Cloud Score+ is a continuously updated quality assessment collection that provides a pixel-wise quality classifier for each Sentinel-2 image to remove cloud and cloud-shadow artefacts (Pasquarella et al., 2023). For WOF mapping, we masked topographic shadows with the *Forest And Buildings removed Digital Elevation Model* (FABDEM; Hawker et al., 2022) on GEE. FABDEM is a global digital elevation model at one arc-second resolution (approximately 30m at the equator). It was produced by removing building and tree height biases from the Copernicus GLO 30 Digital Elevation Model, acquired by the TanDEM-X mission between 2011 and 2015 (Hawker et al., 2022).



2.2.3 Validation and comparison datasets

To evaluate the water detection performance of Dynamic World on independent data, we used the peer-reviewed global benchmark dataset S1S2-Water (Wieland et al., 2024). This dataset contains 63 Sentinel-2 tiles (each 110x110 km) distributed across 29 countries and covering 18 predominant land cover types worldwide across a wide range of topographic, climatic, and seasonal conditions (Figure 2). The images were captured between May 2018 and November 2020 and are accompanied by reference water masks created by three independent image interpretation experts. To classify these tiles into globally representative strata, we utilized the Ecoregions 2017 dataset (Dinerstein et al., 2017), which defines 14 terrestrial biomes based on biological distinctness and climate. These biomes were aggregated into five global eco-strata—Tropical, Temperate & Mediterranean, Grasslands & Flooded, Arid/Desert, and Cold/Boreal—to which we assigned each of the tiles, ensuring sufficient sample sizes within hydrological and climatic zones (Table 1). To only retain inland waters for the scope of this study, we removed marine areas using the Natural Earth Ocean layer (Kelso & Patterson, 2010) and additional manual refinement, resulting in an inland water class that represents 3.77% of the total analyzed area (651,585 km²; inland waters plus terrestrial surface).

Water classification algorithms can be prone to errors over land cover types with similar spectral reflectance properties such as urban areas, snow and ice or terrain-induced shadow (C. Huang et al., 2018; Naboureh et al., 2025). To study the behavior of the evaluated methods under over land cover types, the ESRI Sentinel-2 Annual Land Cover maps (Karra et al., 2021) were used as a secondary reference for the reference tiles. To specifically analyze behavior under terrain-shadow conditions, we used a separate set of terrain-shadow masks from the S2-GOLD dataset (De Felice et al., 2026), including 686 polygons across 53 globally distributed Sentinel-2 tiles over the years 2018 to 2021.

Table 1: Overview of global eco-strata used for stratified cross-validation, including the constituent biomes from Dinerstein et al. (2017), the number reference tiles in each stratum, and their respective global area-weight

Stratum	Constituent Biomes	Tile Count	Area Weight (%)
Tropical Zone	<i>Tropical & Subtropical Moist Broadleaf Forests; Tropical & Subtropical Dry Broadleaf Forests; Tropical & Subtropical Coniferous Forests; Mangroves</i>	10	13.79
Grasslands & Flooded	<i>Tropical & Subtropical Grasslands, Savannas & Shrublands ; Temperate Grasslands, Savannas & Shrublands ; Flooded Grasslands & Savannas ; Montane Grasslands & Shrublands</i>	12	24.06
Cold/Boreal Zone	<i>Boreal Forests/Taiga; Tundra</i>	8	30.43
Arid/Desert	<i>Deserts & Xeric Shrublands</i>	5	16.76
Temperate & Mediterranean	<i>Temperate Broadleaf & Mixed Forests; Temperate Conifer Forests; Mediterranean Forests, Woodlands & Scrub</i>	31	14.96

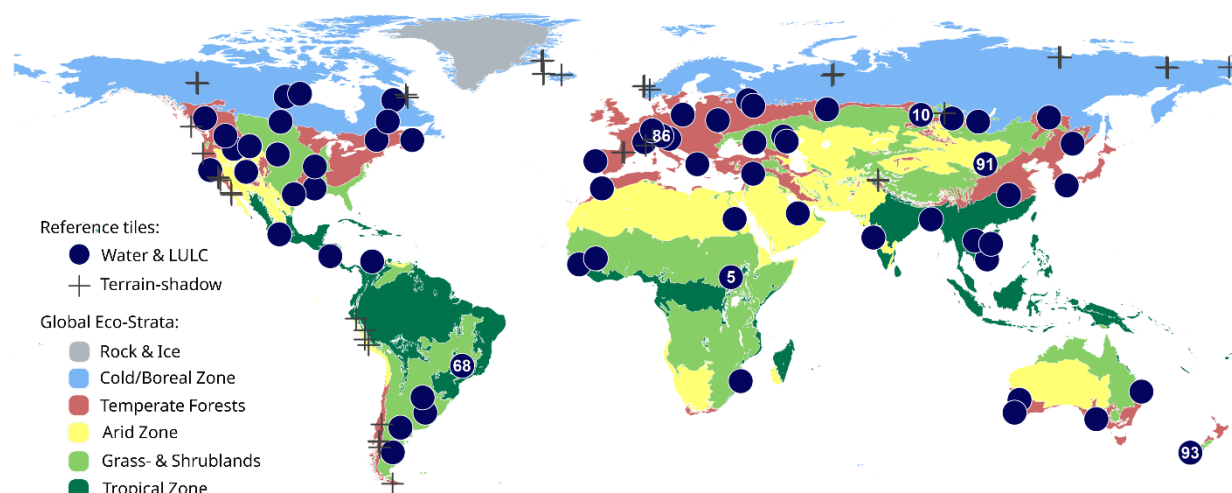


Figure 2: Locations of reference tiles for global water classification analysis. Including the additional locations of terrain-shadow masks and the global Eco-Strata. Numbers indicate locations of classification examples of Figure 7.

We used the GSW Seasonality (Pekel et al., 2016) and Dynamic World DSWE Occurrence (Khare et al., 2025) datasets for regional cross-comparison of the WOF output for the year 2021. These two datasets are, to our knowledge, the only available global WOF products at a spatial resolution of 30 meters or lower. GSW Seasonality quantifies pixel-wise yearly water occurrence as the number of months (0-12) with water from Landsat imagery at 30-meter resolution. DSWE Occurrence is based on the Dynamic World Top-1 label band and expresses the average monthly WOF (monthly number of water observations divided by the monthly number of valid observations). Including this dataset also enabled us to evaluate the impact of using the Dynamic World water probability instead of the Top-1 label on WOF calculation. Yang et al. (2020) also provide a Sentinel-2 WOF dataset for France. However, as this was unavailable in 2021, it was not included.

A vector-based hydrographic layer from the Mapd'O datasuite (Dunesme et al., 2025) was used as a high-resolution benchmark for comparison. It was derived from the French national topographic database (BD TOPO®), which presents a static snapshot of land cover conditions from aerial photographs created via automated deep learning interpretation and expert verification. The database includes water bodies with a minimum area of 200 m², or a length of at least 20 m for standing water bodies, or a 5 m width for flowing water bodies. Despite the data being designated as representing the year 2021, the majority of areas are only updated every 3–5 years (IGN, 2021). This static dataset does not include temporarily inundated areas in river corridors, such as secondary flow channels, unvegetated bars, and flooded floodplains areas. To better understand whether satellite-derived water areas that do not match Mapd'O potentially represent such conditions, we also extracted the active channel sediment bar surface and connected floodplain areas from the “lateral continuity” layer of the Mapd'O datasuite (Dunesme et al., 2025). The dataset is accompanied by a vector river network, a modified version of the French BD Carthage® hydrographic network (Dunesme et al., 2025), which is separated into 200-meter segments and ordered by Strahler stream order. In this hierarchical system, the smallest headwater streams are assigned an order of 1, and when two streams of the same



order intersect, they form a higher-order stream (order 2 in this case), with the order increasing by one with each confluence of same-order streams (Strahler, 1957).

In order to study the impact of cloud cover on the availability of Dynamic World data worldwide, we selected a homogeneously distributed sample of over 30% (1,504 in total) of all global river basins from the HydroBASINS dataset (level 5, with an average surface area of 27,484 km²; Lehner & Grill, 2013).

2.3 Methods

2.3.1 Global water classification analysis

We used the global reference dataset S1S2-Water to evaluate the water detection capability of eight methods (see Table 2). Besides the Dynamic World Top-1 label (DW_{Label}) and water class probability (DW_{WP}) bands, we selected three common water indices to compare them individually and in combination with DW_{WP} : the Normalized Difference Water Index (NDWI, McFeeters, 1996), the Modified Normalized Difference Water Index (MNDWI, Xu, 2006), and the Symbolic Regression Water Index (SRWI, Chrysostomou et al., 2026). The first two are the most popular and extensively evaluated water indices, shown to be effective under global conditions for Sentinel-2 imagery (Miura et al., 2025; C. Zhao et al., 2025). We have also included the recently proposed SRWI because its combination of short-wave infrared (SWIR) and near-infrared (NIR) bands was shown to yield more robust water classification under varied conditions worldwide, including built-up and shaded areas. All indices were scaled for the purpose of this study to values 0-1. We tested a simple a product combination with DW_{WP} per index. Prior to evaluation, we removed all cloudy or low-quality pixels with a Cloud Score+ of at least 0.6.

Table 2: Overview of the water detection methods evaluated in this study

Method	Description	Calculation
DW_{Label}	Dynamic World Top-1 labels representing the land cover class with the highest probability	-
DW_{WP}	Dynamic World-assigned probabilities for the water class	-
$NDWI_{0-1}$	Normalized Difference Water Index, calculated from the Green (G) and Near-Infrared (NIR) bands and scaled to values 0 – 1	$\left(\frac{G - NIR}{G + NIR} + 1\right) \times \frac{1}{2}$
$MNDWI_{0-1}$	Modified Normalized Difference Water Index, calculated from the Green (G) and Short-Wave-Infrared (SWIR1) bands and scaled to 0 – 1	$\left(\frac{G - SWIR1}{G + SWIR1} + 1\right) \times \frac{1}{2}$
$SRWI_{0-1}$	Symbolic Regression Water Index, calculated from the Green (G), Blue (B), Near-Infrared (NIR), Short-Wave Infrared (SWIR1) bands and scaled to 0 – 1	$\left(\frac{(G + B) - (NIR + SWIR1)}{(G + B) + (NIR + SWIR1)} + 1\right) \times \frac{1}{2}$



DW_{WP}×NDWI;	<i>Product of each index and the DW_{WP} band</i>	$(Ind \times DW_{WP})$
DW_{WP}×MNDWI;		<i>Ind</i>
DW_{WP}×SRWI		<i>= either NDWI, MNDWI, or SRWI</i>

For each method aside from DW_{Label}, the optimal separation boundary (the index value beyond which a pixel is labeled as water) was defined by the maximum F β score (Eq. (3), van Rijsbergen, 1974). This criterion provides the best equilibrium between maximizing the detection of true water pixels (high Recall, Eq. (1)) and reducing commission errors (high Precision Eq (2)). We selected a value of 1.2 for β to give slightly higher importance to the Recall. The scores are calculated based on the true positive (TP), false negative (FN), and false positive (FP) counts:

$$Recall = \frac{TP}{TP+FN} \quad (1)$$

$$Precision = \frac{TP}{TP+FP} \quad (2)$$

$$F\beta = (1 + \beta^2) \times \frac{Recall * Precision}{Recall + (\beta^2 * Precision)} \quad (3)$$

To obtain a globally unbiased measure of the performance of the different methods, we implemented a globally stratified and area-weighted 5-fold cross-validation scheme. Based on the defined eco-strata, we distributed all tiles into five folds, ensuring each fold maintained a similar share of each eco-stratum to function as a representative sample. For each fold, the optimal separation boundary of each method was derived exclusively from four training folds, and performance was subsequently assessed on the held-out fold. Following the recommendations of (Olofsson et al., 2014), we applied global area weights derived from the total area of each eco-stratum. During training, stratum-specific histograms were normalized by their valid pixel counts and multiplied by area weights to find an optimal global threshold that represents the true probability distribution of the world's biomes rather than the frequency of tiles in the dataset. In the evaluation phase, raw error matrices for the held-out folds were converted into unbiased area proportions using these weights to calculate the global Recall, Precision, and F1-score (Eq. (4)). This yielded five independent performance estimates per method. To quantify estimation uncertainty, we averaged the accuracy metrics and reported them together with the standard error.

$$F1 = 2 \times \frac{Recall * Precision}{Recall + Precision} \quad (4)$$

For estimating performance across global biomes, we averaged the value distributions among the tiles within each biome and determined accuracy (F1, Recall, Precision) of the classification methods using the average thresholds determined from the 5-fold cross-evaluation.

Water classification algorithms can be prone to errors over specific land cover types with similar spectral properties as surface water (C. Huang et al., 2018; Naboureh et al., 2025). To evaluate the methods under such conditions, we compared their value distributions in the water masks with that for non-water land cover classes in the reference tiles, including urban areas, forest,



rangeland, agriculture, bare soil as well as snow/ice. These classes were defined based on the consensus of the Dynamic World
Top-1 label and the ESRI Annual Land Cover map (Karra et al., 2021) corresponding to the year of the image. To ensure better
representability of global conditions, we implemented area-weighted stratified normalization to the value distributions within
each landcover class as well. To assess the impact of shadows, we additionally retrieved value distributions within the masks
of the “Topographic shadow” class of the S2-GOLD dataset (De Felice et al., 2026).

2.3.2 Water Occurrence Frequency (WOF) mapping workflow

Workflow creation

Based on the water detection method with the highest F1 score, we developed an automated workflow to derive annual WOF
from Sentinel-2 imagery entirely executed on GEE (from hereon, the S2-DW WOF workflow; Figure 3, Helling, 2026). For
each region of interest and year, the workflow generates a 10-m resolution raster file containing the monthly and annual WOF
values and ancillary bands.

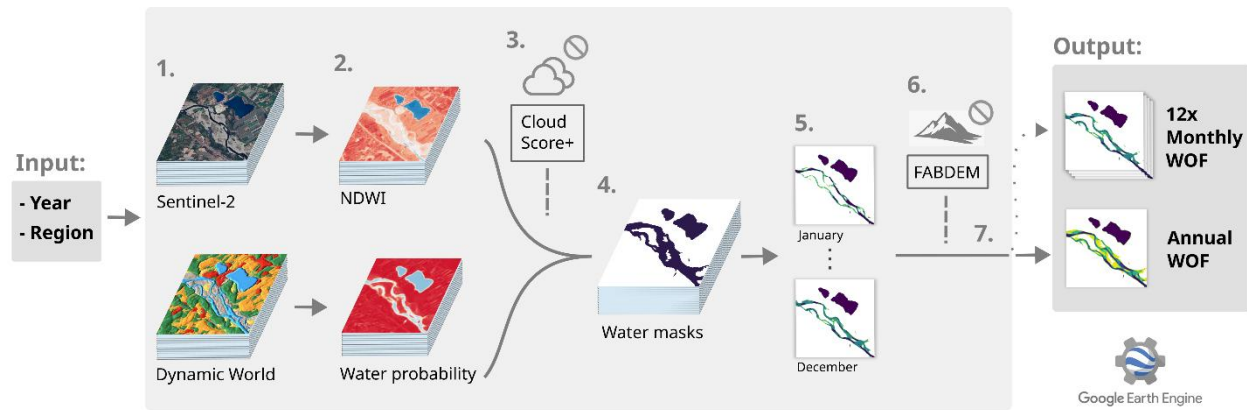


Figure 3: Workflow for creating monthly and annual water occurrence frequency (WOF) maps with Google Earth Engine (Helling, 2026). Steps: 1. Filtering of Sentinel-2 and Dynamic World collections by year and region (Input); 2. Calculation of Normalized Difference Water Index (NDWI) and linkage with Dynamic World water probability; 3. Removal of cloud pixels with Cloud Score+ collection; 4. Extraction of water pixels (via combined NDWI-Dynamic World water probability method); 5. Aggregation to monthly WOF; 6. Removal of shadow pixels with digital elevation model (FABDEM); 7. Aggregation to yearly WOF.

First, the Sentinel-2 Surface Reflectance and Dynamic World collections are filtered for the specified year and region. Second,
the $NDWI_{0-1}$ is calculated and retained together with the Dynamic World water probability DW_{WP} and Top-1 label (DW_{Label})
bands. Pixel-quality screening is applied using the Cloud Score+ collection with a predefined quality threshold of 0.6. Then,
the optimal water detection method $DW_{WP} \times NDWI$ is applied to identify all water pixels in the collection. The water pixels are
then aggregated to monthly water occurrence frequencies ($WOF_{m,i}$) via Eq. (5). These are calculated as the pixel-wise
proportion of observations classified as water ($WO_{m,i}$) relative to the total number of valid observations ($VO_{m,i}$), expressed as
a percentage.

$$WOF_{m,i} = \frac{WO_{m,i}}{VO_{m,i}} * 100 \quad (5)$$



Finally, pixels affected by topographic shadows are excluded from the processed data using the GEE-native *hillShadow* modelling function on the FABDEM dataset (Hawker et al., 2022) with the month's sun azimuth and zenith. Each monthly water occurrence frequency is stored in a separate raster band. Months with zero observations are labeled as 'NA' and are therefore excluded from the calculation of the annual water occurrence frequency (WOF_y), which is the average of all monthly frequencies and is stored in a different raster band (Eq. (6)).

$$WOF_y = \frac{1}{n} \sum_i^n WOF_{m,i} \quad (6)$$

To remove remaining artifact water pixels due to topographic shadow, we masked all pixels with $WOF_y < 40$ over terrain slopes of more than 15° .

Regional benchmarking against other WOF datasets

Regional comparison on the Rhône-Mediterranean region was conducted for 2021, the most recent year with complete data availability across all four datasets used in this study.

First, we compared the output of our S2-DW WOF workflow to the GSW Seasonality and DSWE Occurrence by quantifying their intersection with the water areas from Mapd'O. To understand their distribution outside these areas, we also intersected them with river floodplain features (active channel including sediment, naturally connected floodplain and disconnected floodplain areas). We compared three key aspects across the datasets: surface water area, network length, and WOF value distribution. Total surface water area was computed as the area of all pixels with an annual water occurrence frequency (WOF_y) greater than zero (or at least one month for GSW data at monthly scale). To study the usefulness of the datasets for mapping river networks, we analyzed their intersection with the Mapd'O vector river network, extracting the accumulated river length per Strahler stream order. For that, we summed the length of all 200-meter segments whose 10-meter-wide buffer touched at least nine water pixels of the benchmarked dataset (three in the case of GSW's 30-meter resolution). To study structural differences in frequency values, we compared the value distribution of the annual WOF_y and water months from the S2-DW WOF dataset with the DSWE Occurrence and GSW Seasonality estimates.

In a second step, we compared the behavior of all three WOF datasets over a set of test-locations, with water-mapping-related dynamic landcover conditions. For every location, we manually validated each observation and derived water occurrence frequencies over the year 2021. For that, 120 circular areas with a 20-meter diameter were selected in the Rhône Mediterranean Basin, distributed among five surface cover types (see Figure 4):

1. **always water:** perennial water surfaces from large river channels, lakes or reservoirs
2. **never water:** areas that are never under water, e.g., agricultural fields, urban areas, or forests outside of pluvial and fluvial flood zones
3. **intermittent:** areas alternating between dry and wet conditions, e.g., sediment bars or sparsely vegetated riparian zones with recurring inundation
4. **narrow river:** water surfaces in river channels with a visible water surface narrower than 40 meters



5. **shadow-affected:** locations covered by topographic shadow at least once a year

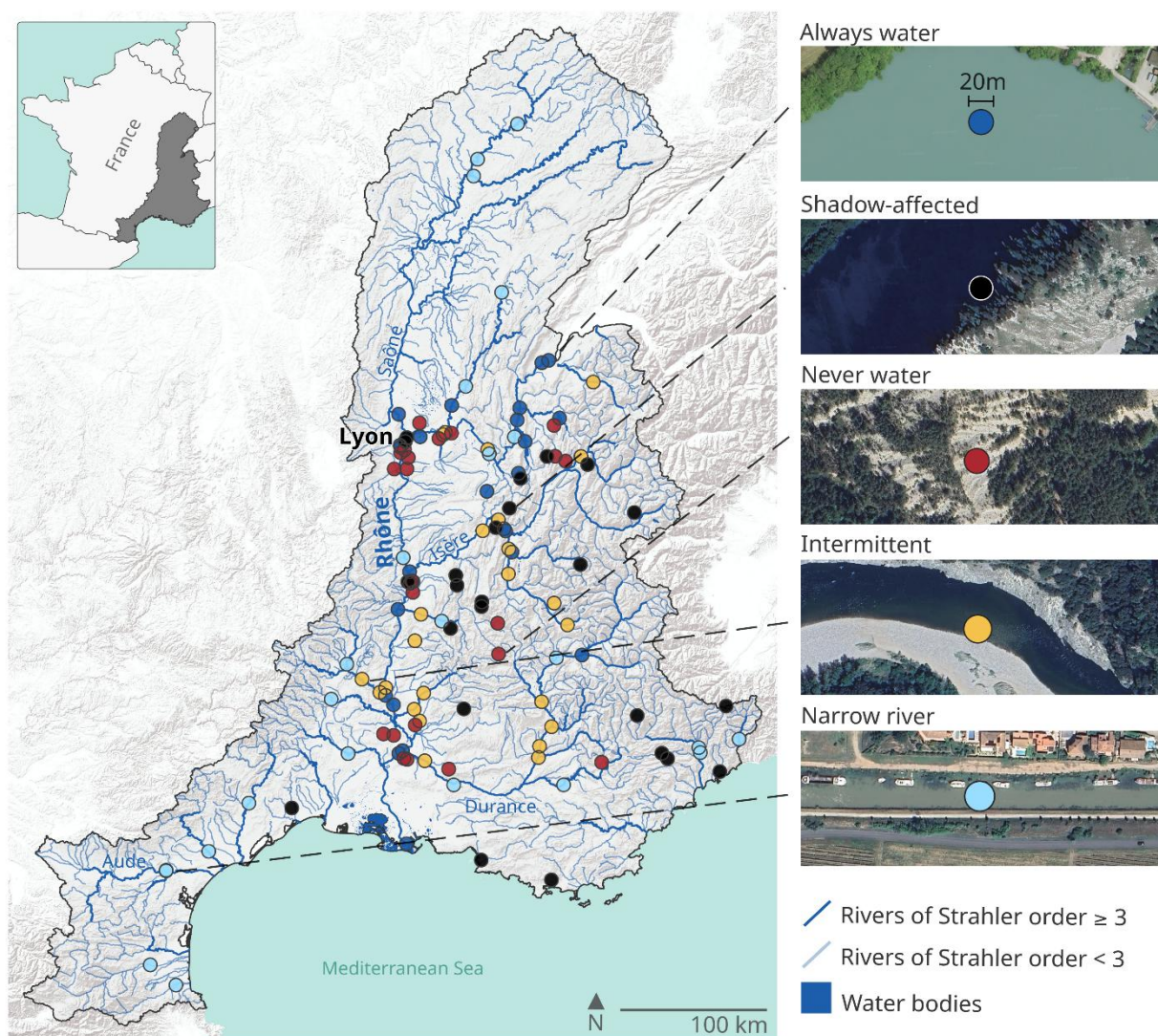
These classes were selected to assess the methods over clearly distinguishable surfaces (always water and never water), and their limits under more uncertain conditions (classes 3, 4 and 5). Mixed pixels that contain different types of materials (e.g., soils, rock, leaves) in the ground area covered by the pixel are particularly common at the interface of water channels with their riparian zones. This motivated the inclusion of narrow river and intermittent classes. The median values of the described bands (DW_{Label} , DW_{WP} , $NDWI_{0-1}$, $NMDWI_{0-1}$, $SRWI_{0-1}$) for all pixels with their center within the 20-meter diameter (typically 2-3 pixels) were extracted for each valid Sentinel-2 overpass between January and December 2021, resulting in an average of 35.5 observations per year and location. For observations in the intermittent and narrow river classes, we manually verified the water/not-water condition of the entire series of 1708 observations, distinguishing between clear water presence or absence and uncertain conditions. Based on this, we calculated the manually verified monthly and yearly WOF values and associated uncertainty for each location. This enabled the comparison of the manually validated WOF value to each of the three WOF datasets, supporting a global standard-error comparison between methods.

2.3.3 Assessment of observation availability

To examine how the different cloud- and shadow-masking steps would influence data availability for WOF calculation, we undertook both a global analysis on 1504 selected river basins and a regional analysis on the Rhône-Mediterranean region. At the global scale, we calculated for each basin the proportion of pixels with at least one Dynamic World prediction for each month and for the entire year of 2021 using GEE. In the Rhône-Mediterranean region (step 2.3.2), we quantified the number of available observations and months per pixel in 2021.

2.4 Evaluation Region

The regional comparison of WOF maps is conducted in the Rhône-Mediterranean region, which spans 121,219 km² and contains all French mainland rivers draining into the Mediterranean Sea (Figure 4). The region includes pronounced topographic gradients and diverse hydroclimatic regimes, resulting in strong seasonal and spatial variability in surface water dynamics (Olivier et al., 2022). This variability is heavily altered by a long history of human intervention, including channelization and flood protection works, hydroelectric dams, and urban and agricultural development (Olivier et al., 2022; Parrot, 2015). The region also contains important wetland systems such as the Dombes or the Camargue delta. Together, these characteristics make the region a particularly relevant test case for satellite-derived WOF, as surface water presence varies strongly across space and time and reflects both natural hydrological processes and anthropogenic controls.



320 **Figure 4: Map of the Rhône-Mediterranean region, this study's validation region, with locations of the 120 test locations, examples of each of the evaluated surface classes, and the comparison river network (Mapd'O). Background map: ESRI | Powered by Esri. Imagery © 2026 IGN.**

3 Results

3.1 Global water classification comparison

325 3.1.1 Global classification performance

The three combined approaches— $DW_{WP} \times NDWI$, $DW_{WP} \times SRWI$, and $DW_{WP} \times MNDWI$ —emerged as the top performers based on the 5-fold optimization and cross-validation applied to the global reference dataset. They achieved the highest F1 scores:



DW_{WP}×NDWI 0.913 ± 0.038 , DW_{WP}×SRWI 0.911 ± 0.040 , and DW_{WP}×MNDWI 0.905 ± 0.074 (Table 3). These methods demonstrated particularly high Recall scores while maintaining consistently high Precision. The Dynamic World water probability (DW_{WP}) followed closely, with comparable F1 and Precision but a slightly lower Recall. For all four of these methods, the average values fall within each other's respective standard deviations.

The NDWI₀₋₁ index exhibited a balanced trade-off between Recall (0.872 ± 0.076) and Precision (0.861 ± 0.089), resulting in a higher F1 score (0.861 ± 0.038) than DW_{Label}, SRWI₀₋₁, and MNDWI₀₋₁. The latter two spectral indices yielded significantly lower F1 scores (0.843 ± 0.061 and 0.818 ± 0.081 , respectively), despite similar Recall values (0.875 ± 0.046 and 0.871 ± 0.052). However, their Precision differed markedly (0.827 ± 0.158 and 0.677 ± 0.122). Finally, the Dynamic World Top-1 label (DW_{Label}) achieved the highest Recall (0.966 ± 0.018) but the lowest Precision (0.677 ± 0.122), indicating a high false-positive rate and the lowest F1 score (0.791 ± 0.090).

Table 3: Average performance of the evaluated water detection methods on the S1S2-Water reference dataset, based on globally stratified 5-fold cross-evaluation. Boundaries were estimated on the training-folds (80% of tiles) and performance evaluated on test-folds (20% of tiles). ± columns show the standard deviation of the metrics across folds.

Method	F1		Recall		Precision	
DW _{WP} ×NDWI	0.913	±0.038	0.933	±0.021	0.897	±0.074
DW _{WP} ×SRWI	0.911	±0.040	0.935	±0.023	0.891	±0.080
DW _{WP} ×MNDWI	0.905	±0.039	0.934	±0.026	0.882	±0.081
DW _{WP}	0.904	±0.038	0.918	±0.033	0.892	±0.064
NDWI ₀₋₁	0.861	±0.038	0.872	±0.076	0.861	±0.089
SRWI ₀₋₁	0.843	±0.061	0.875	±0.046	0.827	±0.131
MNDWI ₀₋₁	0.818	±0.081	0.871	±0.052	0.790	±0.158
DW _{Label}	0.791	±0.090	0.966	±0.018	0.677	±0.122

The optimal thresholds, estimated via 5-fold optimization on the training folds, are relatively consistent among the combined methods (0.157–0.184) and among the spectral indices (0.438–0.468; Table 4). For the latter, when converted back from normalization to the (-1;1) scale, these thresholds correspond to values closely below 0, e.g., NDWI₀₋₁ = 0.468 translates to an NDWI threshold of -0.064. The Dynamic World water probability (DW_{WP}) band achieves optimal results with a threshold of 0.504.

The relative threshold widths—the difference between maximum and minimum threshold above which a specific F1-score is attained—reflect the sensitivity of the estimated thresholds to changes. Based on the results, the DW_{WP} band is particularly robust, achieving the widest ranges across all three F1-score benchmarks (0.524 for F1 > 0.8, 0.410 for F1 > 0.85, and 0.346 for F1 > 0.9; Table 4). The three water indices exhibit the lowest relative threshold widths, ranging from 0.167 (NDWI₀₋₁) to 0.212 (SRWI₀₋₁) for F1 > 0.8 and from 0.110 (NDWI₀₋₁) to 0.137 (MNDWI₀₋₁) for F1 > 0.85. The behavior of each individual method is reflected in their combinations. The combined methods exhibit wider ranges, with similar threshold widths for F1 >



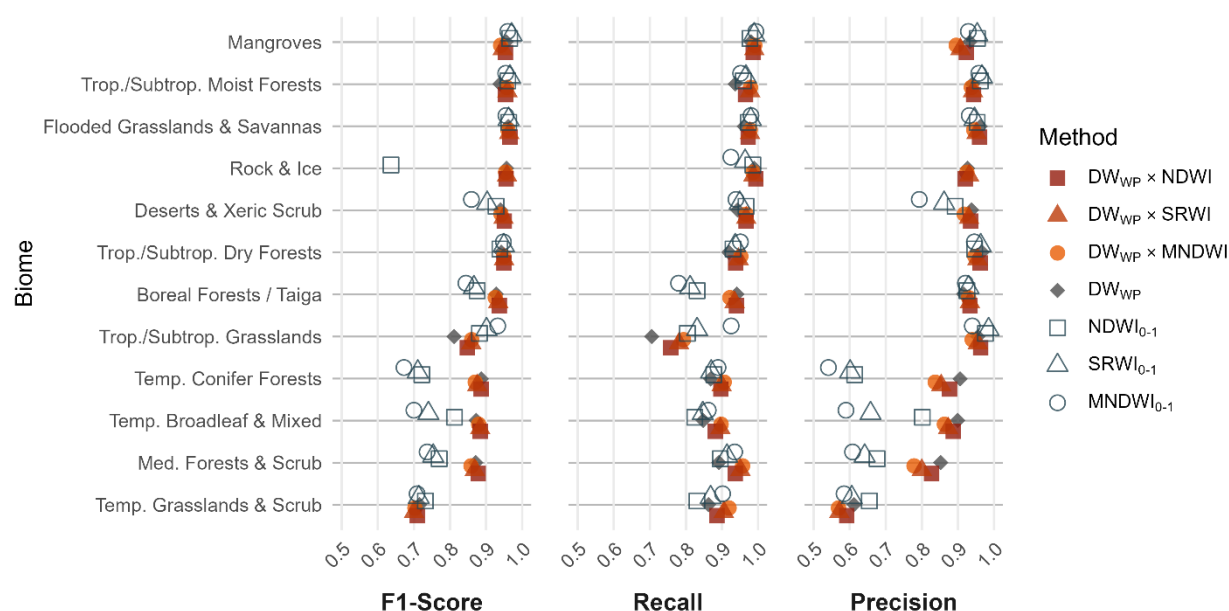
0.8 (0.360–0.370) and $F1 > 0.85$ (0.284–0.300). Among them, $DW_{WP} \times NDWI$ demonstrates a significantly larger range for $F1 > 0.9$ (0.246) compared to the other combined methods (0.166 for $DW_{WP} \times SRWI$ and 0.198 for $DW_{WP} \times MNDWI$).

Table 4: Optimal thresholds and relative threshold width of the evaluated water detection methods. Thresholds were estimated from the training-folds (80% of tiles) by maximizing the $F_{1,2}$ -score. Relative threshold width represents the range (maximum – minimum value) above which the F1-score is above 0.8, 0.85, 0.9, respectively. $\pm$ columns show the standard deviation of the metrics across folds.

Method	Optimal Threshold		Relative Threshold Width		
			$F1 > 0.8$	$F1 > 0.85$	$F1 > 0.9$
$DW_{WP} \times NDWI$	0.184	± 0.010	0.360 ± 0.050	0.300 ± 0.058	0.246 ± 0.051
$DW_{WP} \times SRWI$	0.157	± 0.008	0.361 ± 0.033	0.289 ± 0.051	0.166 ± 0.105
$DW_{WP} \times MNDWI$	0.165	± 0.007	0.370 ± 0.031	0.284 ± 0.047	0.198 ± 0.041
DW_{WP}	0.504	± 0.016	0.524 ± 0.100	0.410 ± 0.162	0.346 ± 0.076
$NDWI_{0-1}$	0.468	± 0.007	0.167 ± 0.047	0.110 ± 0.041	0.024 ± 0.023
$SRWI_{0-1}$	0.438	± 0.006	0.212 ± 0.044	0.130 ± 0.046	-
$MNDWI_{0-1}$	0.461	± 0.007	0.201 ± 0.068	0.137 ± 0.034	-

3.1.2 Performance across biomes

The combined methods ($DW_{WP} \times NDWI$, $DW_{WP} \times SRWI$, and $DW_{WP} \times MNDWI$) together with DW_{WP} consistently achieve the highest performance across most biomes, particularly in F1-score, where they exceed 0.9 in seven out of twelve biomes. The DW_{WP} method alone performs similarly strong and works particularly well in difficult biomes such as rock & ice or different forest-types (temperate coniferous, temperate broadleaf & mixed, Mediterranean forest & scrub) where it outperforms the spectral indices in Precision. In contrast, the spectral indices ($NDWI_{0-1}$, $SRWI_{0-1}$, and $MNDWI_{0-1}$) exhibit higher accuracy only in tropical and subtropical grasslands, where DW_{WP} has a lower Recall. In both cases, the combination of DW_{WP} with the indices balances out the weaknesses of one of the methods.



370 **Figure 5: Performance of methods across global biomes. Performance scores were derived by applying the estimated best thresholds for each method (see Table 4) to all S1S2-Water reference tiles and averaging results for each biome.**

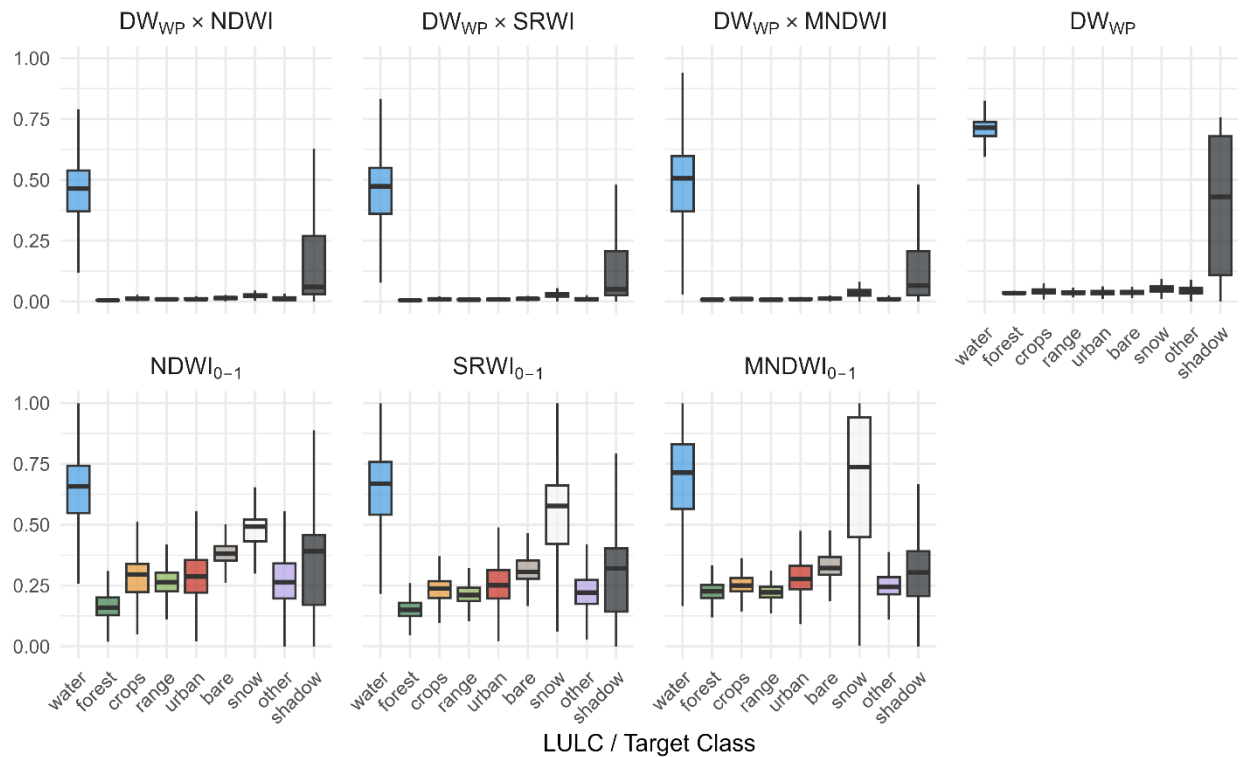
3.1.2 Landcover-specific distribution

Compared to the spectral indices, the Dynamic World water probability (DW_{WP}) band exhibits different behavior across global land-cover classes (Figure 6). DW_{WP} provides clear differentiation between water surfaces (median ~ 0.7) and other land-cover types (medians ~ 0.05), with only terrain shadows nearly overlapping with water surfaces (median ~ 0.42). In contrast, the spectral indices exhibit less pronounced distinctions between water (median around ~ 0.65) and other land-cover types (medians ranging from 0.15 to 0.38), except snow/ice. While $NDWI_{0-1}$ mostly distinguishes snow/ice from water, $SRWI_{0-1}$ and $MNDWI_{0-1}$ show clear overlap between these two classes. Generally, the spectral indices perform better in distinguishing water from shadows, but worse at separating it from snow/ice. As a result, the index- DW_{WP} combinations leverage the individual

375 types (medians ~ 0.05), with only terrain shadows nearly overlapping with water surfaces (median ~ 0.42). In contrast, the spectral indices exhibit less pronounced distinctions between water (median around ~ 0.65) and other land-cover types (medians ranging from 0.15 to 0.38), except snow/ice. While $NDWI_{0-1}$ mostly distinguishes snow/ice from water, $SRWI_{0-1}$ and $MNDWI_{0-1}$ show clear overlap between these two classes. Generally, the spectral indices perform better in distinguishing water from shadows, but worse at separating it from snow/ice. As a result, the index- DW_{WP} combinations leverage the individual

380 discrimination strengths of their constituent parts. They suppress the influence of other land-cover classes compared to the spectral indices alone, preserve high water values, and enhance separation from shadows by increasing the distance between water and shadow distributions compared to DW_{WP} .

This behavior is reflected when applying the methods with the estimated optimal thresholds (Figure 7). In most cases, the $DW_{WP} \times NDWI$ method outperforms DW_{WP} and $NDWI_{0-1}$, effectively eliminating most false positives such as topographic shadows (subplot f), snow and ice (subplots b, d), and urban areas (subplot d). Additionally, it often increases detail by better resolving small features compared to DW_{WP} (subplots a, c, d). However, there are also instances where DW_{WP} identifies rivers surface not detected by $NDWI_{0-1}$, which is also reflected in the combined method (subplot a). Nevertheless, not all water areas are captured when they were only partially classified by one of the two methods (subplot a, d, e, f).



390 **Figure 6: Behavior of evaluated water classification methods over specific landcover classes. Boxplots represent the globally stratified value distributions over classified pixels in the S1S2-Water reference dataset. The (terrain-)shadow class was included from the S2-GOLD dataset.**

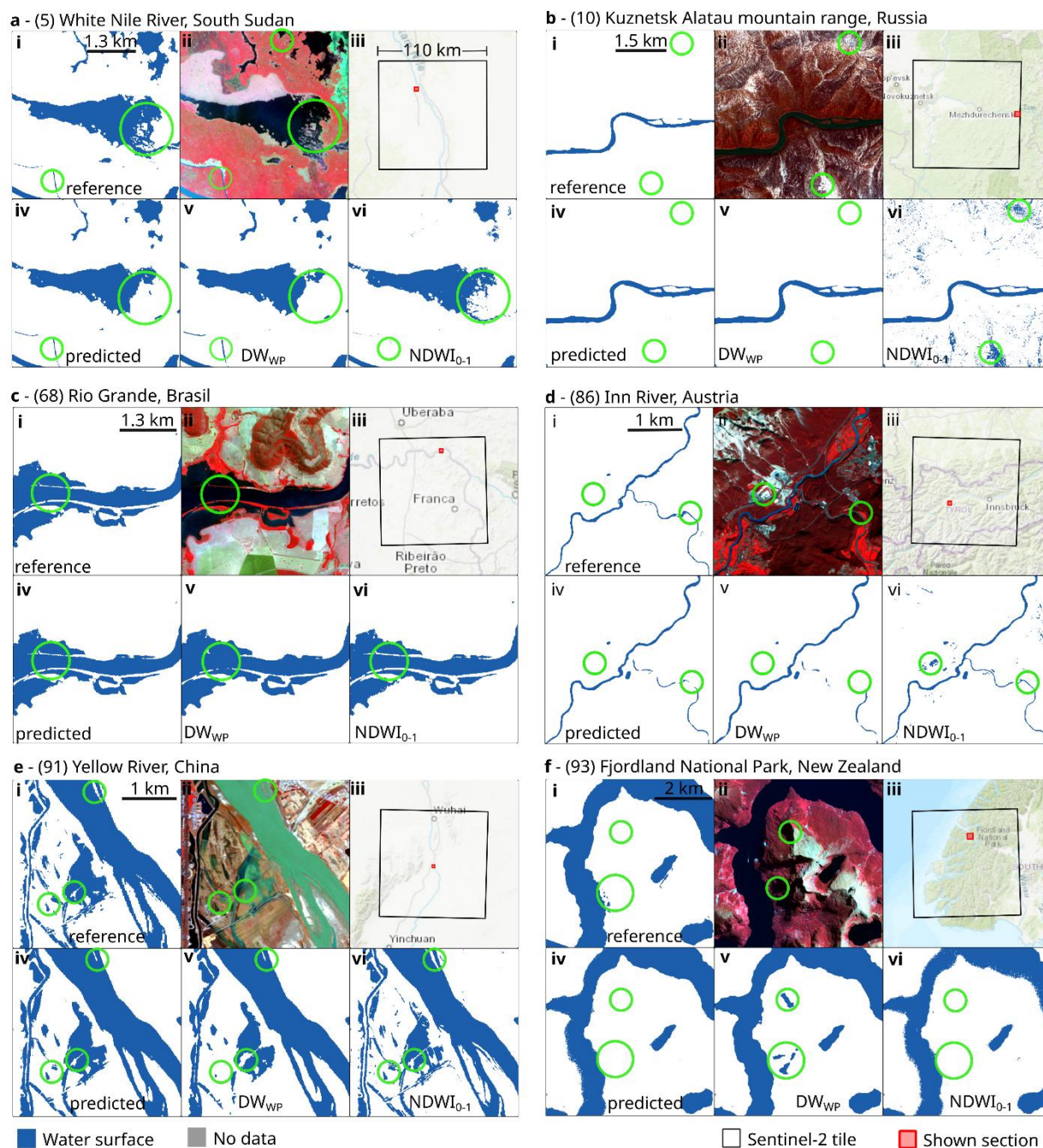


Figure 7: Examples of the predicted water areas with the methods $DW_{WP} \times NDWI$ (iv), DW_{WP} (v), and $NDWI_{0-1}$ (vi) from selected tiles, together with the reference water mask (i), the original Sentinel-2 false-RGB image (ii), the location of the tile and example section (iii). Green circles highlight discontinuities between the methods.



3.2 Annual surface water dynamics mapping: regional analysis

3.2.1 Surface water area, river length, and occurrence frequencies

400 None of the satellite-derived products fully captures the surface water extent depicted by the high-resolution dataset Mapd'O across the Rhône-Mediterranean region (Figure 8a). The two Sentinel-based products (DSWE Occurrence from Khare et al. (2025) and this study's S2-DW WOF) more closely match the surface water and active channel extent estimates of Mapd'O than the Landsat based GSW Seasonality (Pekel et al., 2016). However, the share of matched area (i.e., predicted as water and mapped as water or active channel in Mapd'O) from the total area of the S2-DW WOF is 66.5 %. The 30-meter GSW
405 Seasonality product shows a higher share at 78.6 %, while predicting the overall lowest surface water area, while DSWE contains large parts of water predictions outside Mapd'O water surfaces, presenting the lowest ratio at 34.9 %. A large part of these mismatched areas from the S2-DW WOF and GSW Seasonality datasets can be found within river floodplains (305 and 176 km², respectively). However, the DSWE predicts more than half of its water areas (2052 km²) outside river floodplain systems and lakes included in Mapd'O.

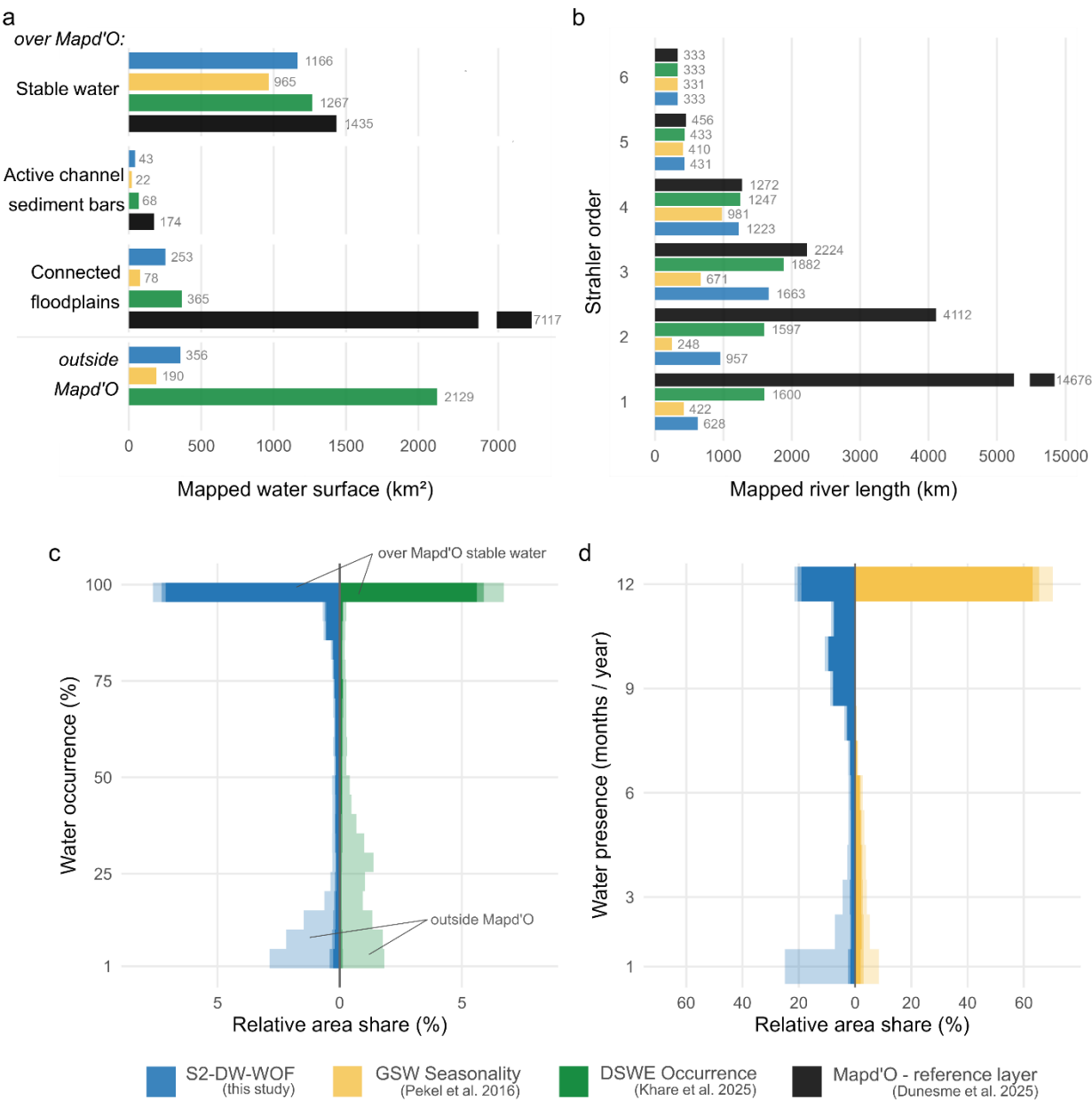
410 The length of the Mapd'O hydrographic network exponentially increases with lower Strahler order, with more than half of the total length stemming from Strahler 1 rivers (Figure 8b). This is not reproduced by the satellite-derived datasets which only match Strahler orders 4 to 6. That said, compared to the Landsat-based GSW dataset, the Sentinel-derived products furthermore enable the mapping of large parts of Strahler 3 rivers depicted in Mapd'O (2224 km). They capture 1664 km (S2-DW WOF) and 1822 km (DSWE) — more than twice as much as the GSW dataset (671 km). While DSWE is able to represent about 40%
415 of the Strahler 2 network, S2-DW WOF and GSW show substantially lower length. Finally, none of the satellite products map Strahler 1 rivers well, even though DSWE captures 11 % of network length in Mapd'O, compared to 4 % (S2-DW WOF) and 3 % (GSW).

Structural differences can be observed among the datasets in their distribution of occurrence frequency/water month values. Most water pixels identified by DSWE and S2-DW WOF are predicted to be nearly perennial (WOF >95%) and align with
420 Mapd'O water areas (Figure 8c). For both products, areas with WOF values below 50% most often represent pixels located outside water and active channel areas in Mapd'O. DSWE exhibits two peaks within this range of values: one at 30% and one around 1–10%. In contrast, most non-perennial S2-DW WOF pixels are below WOF values of 25%, with a peak at 1–5%. This structure is similar when looking at water occurrence months: More than 50 % of pixels in S2-DW WOF and GSW have water over the whole year (Figure 8d). While the S2-DW WOF contains pixels over all values with a second peak around 2-3 months,
425 GSW has no pixels with 9-11 water months and a second peak at 1 month.

Sentinel-2's resolution enables more detailed water mapping compared to the Landsat-derived GSW product, particularly of narrow features like rivers (Figure 8). However, the DSWE product does not include large parts of temporarily flooded zones (Figure 9-1), erroneously maps large terrain shadows and streets as water (Figure 9-3), and predicts significantly more water surface in complex systems such as the Dombes ponds and the Camargue lagoons (Figure 9-2, 9-5) — not matching the other



430 datasets. Beyond spatial details, satellite products better depict variably inundated zones compared to the static Mapd'O dataset (Figure 9-1, 9-2, 9-4, 9-5).



435 **Figure 8: Cross-comparison of different Water Occurrence Frequency (WOF) datasets for the French Rhône-Mediterranean region and the year 2021: S2-DW WOF (this study, blue); DSWE Occurrence from Khare et al. (2025, green), GSW Seasonality from Pekel et al (2016, yellow); Mapd'O, the static high-resolution land cover dataset from Dunesme et al. (2025) (black). a shows the total water areas each satellite product predicts over the specific landcover classes derived from Mapd'O: stable water surface, active channel sediment bars, connected floodplains, as well as all water predictions outside Mapd'O. b presents the mapped river network length**



440

by Strahler order based on the Mapd'O hydrographic network. Total mapped river lengths are: 23,073 km Mapd'O, 7,090 km DSWE, 5,233 km S2-DW WOF, and 3,062 km GSW Seasonality. c compares the total WOF value distribution of S2-DW WOF with that from DSWE Occurrence. d compares the total counted water month value distribution of S2-DW WOF with GSW Seasonality.

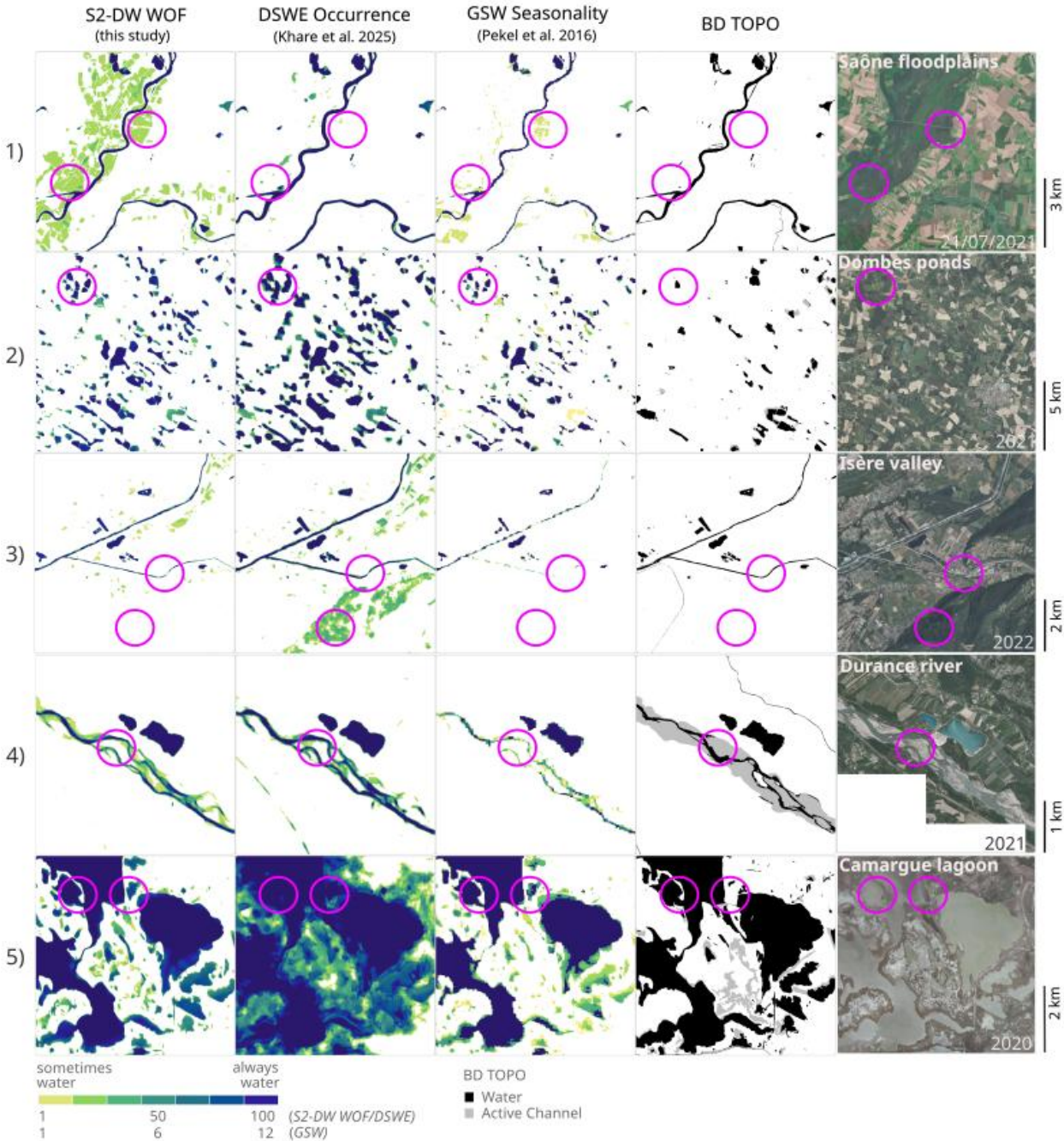


Figure 9: Output examples from the water occurrence frequency (WOF) maps for the year 2021 in the Rhône-Mediterranean region. Maps from the 4 compared products for specific locations: the developed dataset of this study S2-DW WOF, the Dynamic World



based DSWE product by Khare et al. (2025), the Landsat-based GSW product from Pekel et al. (2016), and the static high-resolution map Mapd'O. Imagery © 2026 IGN.

3.2.2 Performance over surface-water-specific landcover conditions

All satellite-derived datasets contain prediction errors, leading to deviations in WOF and water-month frequency relative to manually validated values (Figure 10). While all three datasets correctly identify "Always water" and "Never water" locations, only S2-DW WOF and GSW Seasonality avoid errors in terrain shadows—DSWE misclassifies these as water, overestimating WOF by >20% on average (with some errors reaching 50%; subplot a). For "Narrow rivers" (subplot b), GSW almost entirely misses water occurrences, whereas S2-DW WOF underestimates them by 20% (or ~2 months). DSWE exhibits near-zero average error (i.e., bias) but occasionally over- and underestimates occurrence. Intermittent pixels remain challenging: S2-DW WOF shows spread but balanced errors, DSWE slightly overestimates water occurrence, and GSW tends to underestimate it.

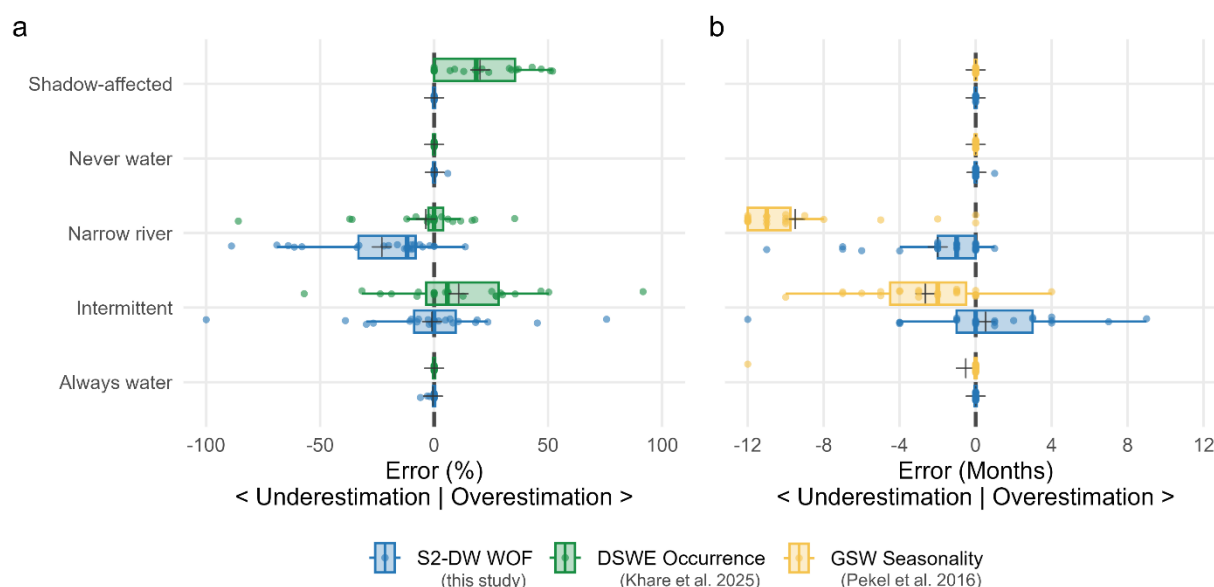


Figure 10: Errors in predicted WOF values of the test datasets over different landcover classes on 120 test-locations (20 m diameter) with manually validated WOF for the year 2021. Subplot a shows the absolute error in frequencies for the S2-DW WOF and the DSWE datasets. Subplot b presents the comparison of water occurrence months between the S2-DW WOF and the GSW Seasonality dataset.

3.3 Assessment of observation availability

Strong spatial contrasts in the availability of Dynamic World predictions are evident across the world's basins (Figure 11), reflecting the combined effects of upstream tile filtering and additional cloud-affected pixel processing in the Dynamic World modeling pipeline. High data availability (>10 months) is observed predominantly in arid and semi-arid regions, including large parts of northern and southern Africa, the Middle East, Australia, North America, and portions of South America and southern Russia. In contrast, persistently cloud-affected regions such as the tropical rainforests of the Amazon and Congo



basins, and Southeast Asia exhibit substantially reduced availability, frequently below six and locally below two months. Mountainous regions, including the Himalayas and Andes also show pronounced reductions. Moderate coverage of seven to eight months can be found in regions with arctic influence such as northern Russia and Canada, as well as tropical zones such as northern Brazil, East Africa, and monsoon-exposed regions of Southern Asia. These spatial patterns demonstrate that Dynamic World is subject to strong regional variability in observation density, resulting in truncated time series for predicting WOF, particularly in humid tropical and complex topographic environments.

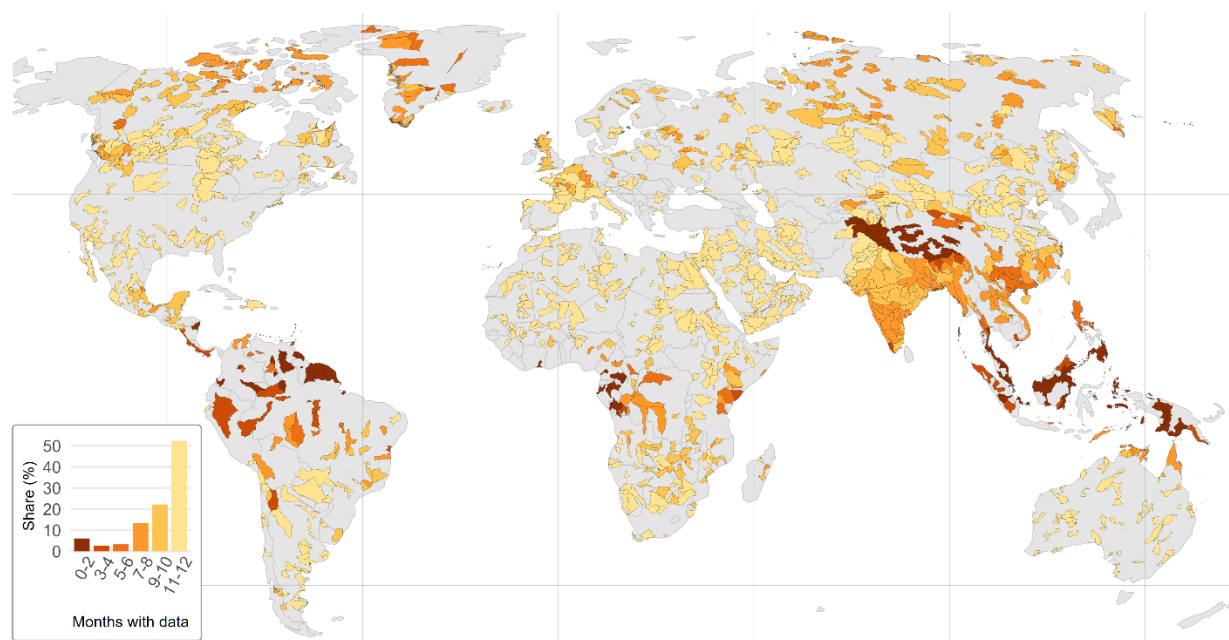


Figure 11: Yearly availability of Dynamic World predictions in 2021 for selected river basins worldwide (HydroBASIN level 5, n=1504), expressed as the number of months with more than 50% spatial coverage

Significant differences in the availability of underlying data can also be observed across the Rhône-Mediterranean region. Coverage is lower in mountainous regions in the east, where valid data are available only five to eight months in 2021 for most pixels. In contrast, the Rhône and Saône valleys and the Mediterranean coast mostly have pixels with 11–12 months of coverage (Figure 12a, b). On average, 42 observations per pixel contribute to the S2-DW WOF for 2021, with a large proportion of pixels having over 60 yearly observations (Figure 12c).

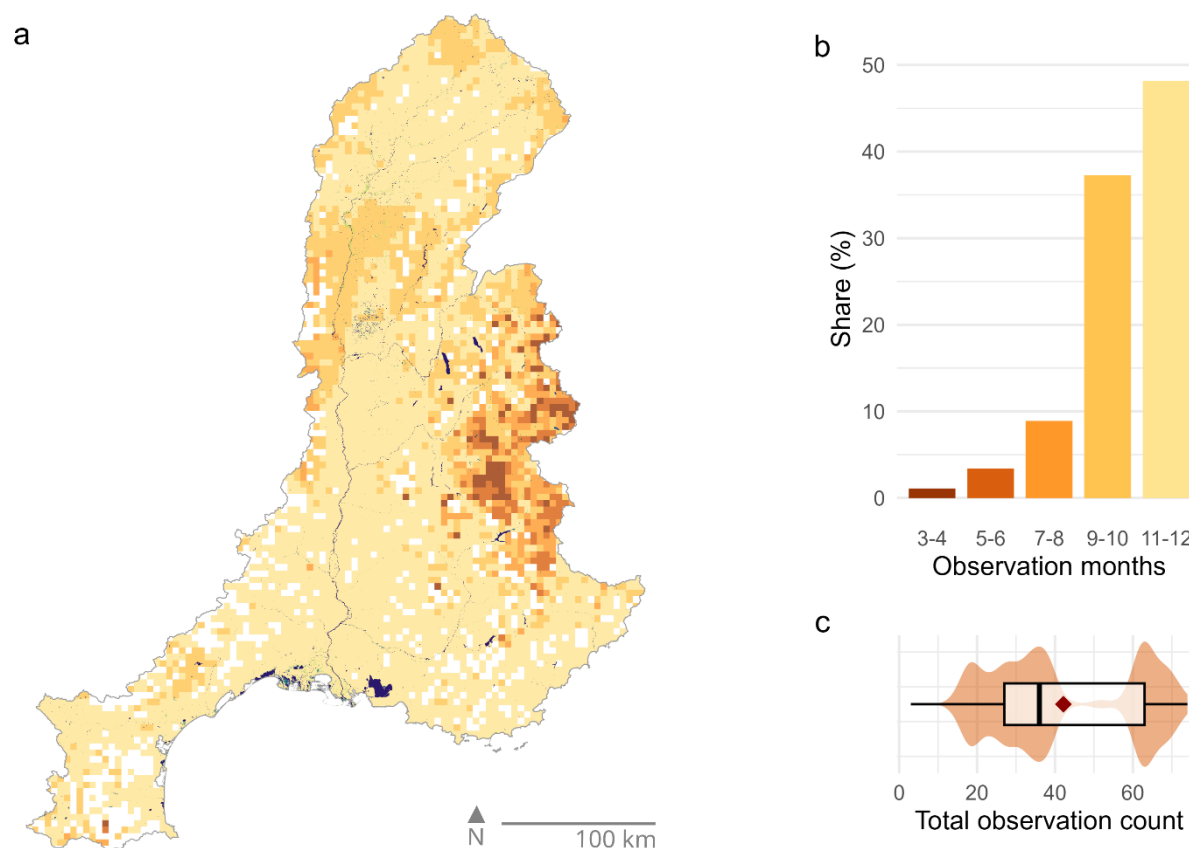


Figure 12: Availability of Dynamic World data underlying the S2-WOF computation for 2021 for the Rhône-Mediterranean Region: a – Map of 5 km-average observation months, b – Histogram of 5 km-average of observation months count, c – Total value distribution of annual observation counts of mapped water areas.

4 Results

4.1 Dynamic World's global water mapping capabilities

This study's global evaluation demonstrates that the Dynamic World model provides a robust method to map inland surface waters from Sentinel-2 imagery, particularly when combined with spectral indices such as the NDWI. Dynamic World's Top-1 classifier tends to prioritize commission over omission (i.e., erroneously overestimating water cover;), making it only suitable for screening or maximum-extent mapping. Thus, this method should not be used for quantitative water area estimation, particularly for narrow or morphologically complex features such as rivers and streams. The water probability band DW_{WP} outperforms both the simple Top-1 label classifier and common spectral indices (e.g., NDWI, MNDWI, SRWI) across most global biomes (Table 3), when used alone. It achieves a balanced trade-off between Recall (0.918 ± 0.033) and Precision (0.892 ± 0.064) with a threshold of 0.504. Its low threshold sensitivity (Table 4) further underscores its robustness, likely due to training on an extensive global dataset that enables better separation of water from other land-cover classes. Notably, it



correctly identifies challenging surfaces such as urban areas, snow, and ice (Figure 6: Behavior of evaluated water classification
495 methods over specific landcover classes. Boxplots represent the globally stratified value distributions over classified pixels in
the S1S2-Water reference dataset. The (terrain-)shadow class was included from the S2-GOLD dataset.), which are often
misclassified by spectral indices (C. Zhao et al., 2025). However, it commonly misclassifies topographic shadows as water
and exhibits reduced accuracy for small features (e.g., narrow rivers and streams), probably due to its 5×5 minimum
homogeneous surface area training constraint (Brown et al., 2022). All evaluated indices (NDWI, MNDWI, SRWI) perform
500 less effectively under global conditions, as they struggle to distinguish water from other land-cover classes, confirming other
global studies (C. Zhao et al., 2025). However, they provide better separation for topographic shadows and capture smaller
features.

The different modeling approach underlying DW_{WP} likely explains the superior performance of the proposed combinations of
Dynamic World water probabilities with spectral indices. Such hybrid approaches that integrate spectral indices with deep
505 learning architectures have recently gained attention in water mapping applications (Rajeswari & Rathika, 2024). For example,
using segmentation models, Kaplan & Avdan (2017) leveraged spectral indices via direct combination; Li et al. (2022) used
them for preparational image filtering; while J. Wang et al. (2022) augmented the model's feature inputs. Here, the hybrid
 $DW_{WP} \times NDWI$ approach achieves the best trade-off between omission and commission errors, leveraging the complementary
strengths of both methods: NDWI captures interpretable spectral water signals at the pixel level, while Dynamic World's
510 segmentation model integrates multi-band information and contextual patterns. This combination results in globally improved
performance ($F1=0.913 \pm 0.038$, $Recall=0.933 \pm 0.021$, $Precision=0.897 \pm 0.074$), reducing errors over shadow or snow/ice
affected areas (Figure 6), and enhancing detail for smaller features such as rivers (Figure 7). These results suggest that the
hybrid $DW_{WP} \times NDWI$ method can address the need for globally transferable water mapping applications, as expressed by
(Sogno et al., 2022), because it is reliable in complex land cover conditions and can be applied efficiently via GEE.

515 4.2 New possibilities for large-scale characterization of surface water dynamics

Our study shows that the proposed WOF workflow (S2-DW WOF, Figure 3) significantly improves the spatial and temporal
detail of WOF mapping by leveraging the $DW_{WP} \times NDWI$ method and advanced erroneous pixel masking. While a global
assessment was not possible due to the lack of spatio-temporally dense Sentinel-2 validation data, we conducted a multi-stage
evaluation in the hydro-dynamically complex Rhône-Mediterranean region. As suggested by Rajib et al. (2024), we undertook
520 a cross-comparison with other WOF datasets, including the popular Landsat-based GSW Seasonality dataset and the newly
proposed DSWE dataset. We evaluated these against the static high-resolution landcover map Mapd'O and manually validated
WOF benchmarks for the year 2021.

Within the Rhône-Mediterranean region, the S2-DW WOF workflow captures 21 % more stable water area (+201 km²) and
71 % more river network length (+ 2,171 km) than the GSW in 2021 (as mapped by the static high-resolution Mapd'O dataset;
525 Figure 8). Large parts of the mapped water areas outside the stable water surface defined by Mapd'O can be found within river
floodplains, suggesting that they represent inundations of riparian areas (e.g., Figure 9-1,3). These short-term events seem to



be better captured by Sentinel-2's higher recurrence frequency. The GSW product provides data for up to 8 months only (Yang et al., 2020), explaining the significantly smaller mapped water surface within floodplains and the absence of pixels with 9-11 water months (Figure 8d). Furthermore, S2-DW WOF especially outperforms GSW for mapping medium- to small-size rivers of Strahler order 3 and 4 (Figure 8b). The DSWE Occurrence dataset, based on Dynamic World Top-1-labels, also maps a significantly greater water surface and river network length than GSW, mostly out of river floodplains. However, the underlying Top-1 predictor overestimates water extent (Figure 9-5) and produces errors from terrain shadows (Figure 9-3). Visual evaluation confirms that large parts of the water surface predicted outside river floodplains (>50 %, 2052 km²) stem from this cause. Simultaneously, DSWE fails to capture large portions of the 2021 Saône flood event (Figure 9-1) which may be explained by its outlier-filtering step (Khare et al., 2025). Finally, our evaluation of WOF estimates over manually validated test locations confirms the advantages of S2-DW WOF over DSWE and GSW. It balances accurate mapping—e.g., avoiding erroneous classification of shadows as water—with consistency in smaller or dynamic features (e.g., narrow rivers, intermittent classes), while delivering superior spatio-temporal detail compared to GSW (Figure 8).

GSW and other Landsat-based WOF products have shown great potential for mapping the dynamics of lentic waterbodies such as lakes and reservoirs (Pi et al., 2022; G. Zhao et al., 2022, 2025), wetlands (Olefeldt et al., 2021; Tootchi et al., 2019) or pond aquaculture (Ottinger et al., 2026), and of river systems (Boothroyd et al., 2025; Schwedhelm et al., 2026; Wu et al., 2023). They also enable the study of interactions between these two types of ecosystems (Dong & Hu, 2021; Pal & Sarda, 2020) and support additional applications such as coastline monitoring (R. Huang et al., 2025), human flooding exposure assessment (Tellman et al., 2021), and river network delineation (Chen et al., 2020). The S2-DW WOF approach could advance these efforts by enhancing the analysis of smaller, ephemeral or rapidly changing lentic water bodies and river systems. It represents a significant methodological advance towards the reliable and consistent mapping of surface water occurrence (Radeloff et al., 2024), even though the inconsistencies in data availability that affect all optical WOF products should be carefully taken into account at regional and global scales (see Figure 11).

By publishing the workflow in ready-to-use form (Helling, 2026) and providing direct access to its results via our publicly accessible GEE-app, we hope that users can leverage the WOF maps with up to 10 years of record (2016-2025). The ancillary data provided give additional information about the underlying uncertainty as well as further possibilities to filter the data. For example, one might distinguish flooded areas inside the active channel from those outside by masking with the annual land cover predictions. Processing the ~120,000 km² Rhône-Mediterranean region, including cloud and terrain-shadow masking, required approximately 10 hours, demonstrating its scalability and computational efficiency.

4.3 Limitations of this study

Our study comes with several limitations. Our global evaluation of Dynamic World relies on a single dataset, though we used stratified and area-weighted cross-validation to strengthen generalizability. This enabled analysis across diverse biomes and land-cover classes, but did not include distinctions between water types (e.g., sediment- and vegetation-rich vs. clear water,



lotic vs. lentic). Due to the limited number of reference tiles (65 in total), stratification of folds was based on manually defined
560 global eco-strata rather than all available biomes.

Although Sentinel-2 has a nominal 3–5-day revisit time, image availability is strongly impacted by cloud cover. Our analysis
reveals highly heterogeneous data availability at regional and global scales caused by Dynamic World’s conservative upstream
removal of cloud-affected areas. Pronounced deficits occur in tropical regions, mountainous terrain, and monsoon-affected
basins, with monthly data availability frequently falling below 50%. This aligns with previous studies (Sudmanns et al., 2020)
565 and limits the reliability of WOF values and seasonal analyses. Furthermore, with fewer observations, sampling errors increase,
making frequency estimates less stable.

The length of the Sentinel-2 record (since 2016) limits applications for long-term dynamics analysis. For multi-decadal studies,
Landsat remains the most suitable source. While Sentinel-2’s 10-meter spatial resolution allows more detailed mapping than
Landsat, it remains limited for small waterbodies, particularly narrow and intermittent rivers, due to mixed pixels. We did not
570 examine how mixed pixels influence the WOF distribution along river axes.

Due to the lack of global ground-truth WOF values, we evaluated only water detection globally and limited WOF-analysis to
the Rhône-Mediterranean Region. Moreover, while we evaluated spatial agreement across different datasets, dataset
differences (DSWE: %, GSW: months, Mapd’O: static 5-year) complicate a true one-to-one comparison. Given the region’s
diverse surface water conditions, we expect similar performance gains elsewhere owing to Sentinel-2’s superior spatio-
575 temporal resolution compared to Landsat as well as our method’s advantages over the Top-1 label approach used by DSWE.
However, this hypothesis requires empirical testing in other hydro-climatic settings. For example, although we tested the
robustness of the underlying water detection method over ice and snow cover, the performance of the WOF algorithm in Arctic
regions has yet to be evaluated.

Finally, our method depends on Dynamic World, so future model updates must be carefully considered as they could shift
580 probability distributions. Reliance on Google Earth Engine (GEE) enables global deployment without local computational
resources but raises data sovereignty concerns and depends on service continuity (Stürmer et al., 2021), particularly in light of
recent policy changes (Google, 2026). While the shadow-masking step is platform-specific, the Dynamic World model is
transferable to local processing environments.

585 4.4 Future directions

Future studies could explore biome-wise threshold selection or local adaptive thresholding methods (e.g., Otsu) to further
refine water detection, though computational trade-offs must be considered. To validate the global transferability of the
spatiotemporal accuracy of the S2-DW WOF, empirical testing in diverse hydro-climatic settings beyond the Rhône-
Mediterranean region and under different cloudiness conditions would be important. To address data availability challenges,
590 integrating radar observations which are unaffected by clouds, like Sentinel-1, represents a promising direction for WOF
mapping (Li et al., 2023; Martinis et al., 2022; Tottrup et al., 2022; Vanderhoof et al., 2023). Recent advances in generating



synthetic Sentinel-2 and Dynamic World scenes from Sentinel-1 (Carbonneau, 2025; Yu et al., 2026) could enable robust, all-
weather monitoring of surface water dynamics at regional to global scales. Additionally, higher-resolution imagery, such as
Planet's 3-meter data (Cooley et al., 2017; Tottrup et al., 2022), could enhance detail, but its commercial licensing hinders
595 adoption for large-scale research.

Finally, new studies may investigate how well Sentinel-2's increased temporal density compared to Landsat captures
hydrological dynamics in rivers. Here, the S2-DW WOF may enable the development of novel indicators of hydro-
geomorphological processes, critical for ecosystem monitoring and management. Further research could also examine how
mixed pixels affect WOF distributions and derived indicators along narrow and intermittent river axes.

600 5 Conclusions

Characterizing surface water dynamics over large scales is essential for understanding hydrological, geomorphological, and
ecological processes and for informing ecosystem management and restoration. While Sentinel-2 imagery provides a suitable
spatial (10 m) and temporal (3–5 days) resolution for monitoring medium and large-sized water bodies, existing mapping
methods have been constrained by limitations in water classification algorithms.

605 This study demonstrates that the Dynamic World model enables accurate ($F1 = 0.904 \pm 0.038$) and computationally efficient
water mapping from Sentinel-2 across globally diverse conditions. Performance is further improved by combining Dynamic
World with NDWI ($F1 = 0.913 \pm 0.038$), particularly for smaller features and challenging conditions such as terrain shadows.
When integrated into an automated GEE workflow, this hybrid approach enables large-scale mapping of surface water
dynamics from full annual Sentinel-2 archives, delivering higher spatio-temporal detail and accuracy than existing datasets.
610 Applied to the Rhône-Mediterranean region, the resulting water occurrence frequency (WOF) map captures substantially
greater detail than Landsat-based datasets, mapping 21% more stable water surface and 71% more river network length, while
better representing ephemeral water bodies and smaller streams. However, temporal consistency is strongly influenced by
observation availability, underscoring the need to account for cloud-related data gaps and motivating future multi-sensor data
integration.

615 This approach enables new opportunities for large-scale applications requiring temporally resolved surface water information,
such as assessing river dynamics, floodplain connectivity, and wetland ecosystem functioning.

Code and data availability

The underlying code of this study is accessible at: https://gitlab.in2p3.fr/lh/paper_s2dw-wof_analysis. The annual S2-DW
WOF workflow has also been implemented in a publicly available GEE app, with which users can interactively visualize the
620 monthly and yearly WOF maps for any location and year (2016–2025), restricted to areas of 60 x 60 km to enable efficient on-
the-run processing: <https://ee-dw-water.projects.earthengine.app/view/s2-dw-wof-app>. A Jupyter notebook, directly



executable via Google Colab, provides access to the full workflow, enabling users to process larger areas and download the data into Google Drive: https://github.com/jlhelling/s2-dw-wof_workflow (Helling, 2026).

The S1S2-Water reference dataset was obtained from <https://zenodo.org/records/11278238>; the GSW Seasonality data from <https://global-surface-water.appspot.com/download>; the S2-GOLD dataset from <http://data.europa.eu/89h/03625cf9-5d2a-49ac-a536-7979fbc131a8>; the Natural Earth Global Oceans layer from <https://www.naturalearthdata.com/downloads/10m-physical-vectors/>; the global Eco-regions from <https://ecoregions.appspot.com/>; the DSWE Occurrence data from <https://colab.research.google.com/drive/1nRI4LHYfyJsIOSZFY4CvN1Hs2POt31vh>; the Mapd'O river network from <https://doi.org/10.57745/V7D5RZ>; the Mapd'O LULC map from <https://doi.org/10.48579/PRO/NCDQM7>.

630 Author contributions

Leo Helling: Conceptualization, Methodology, Software, Validation, Formal analysis, Investigation, Data curation, Writing – original draft, Writing – review and editing, Visualization, Project administration. **Satbir Singh:** Software, Validation, Formal analysis, Investigation, Visualization. **Louis Rey:** Validation, Formal analysis, Investigation. **Hervé Parmentier:** Validation, Formal analysis, Investigation. **Mathis Loïc Messenger:** Conceptualization, Methodology, Writing – review and editing, Visualization, Supervision. **Hervé Piégay:** Conceptualization, Methodology, Writing – review and editing, Visualization, Supervision, Funding acquisition. **Barbara Belletti:** Conceptualization, Methodology, Validation, Investigation, Writing – review and editing, Visualization, Supervision, Funding acquisition

Competing interests

The authors declare that they have no conflicts of interest.

640 Disclaimer

Copernicus Publications remains neutral with regard to jurisdictional claims made in the text, published maps, institutional affiliations, or any other geographical representation in this paper. While Copernicus Publications makes every effort to include appropriate place names, the final responsibility lies with the authors. Views expressed in the text are those of the authors and do not necessarily reflect the views of the publisher.

645 Acknowledgements

The main author would like to thank Layla Roesler for her constructive feedback on the writing. AI assistance was utilized in this study: Mistral was partly used to correct, format and comment analysis code, and DeepL was used for language and grammar checking of the manuscript text. The authors take full responsibility for the accuracy of all final content.



Financial support

650 This work was supported by the CPJ Hydrosystèmes (ANR-23-CPJ1-0148-01), the ANR GloUrb project (ANR-22-CE03-
0005), and the EUR H2O'Lyon (ANR-17-EURE-0018) of Université de Lyon, within the programme “Investissements
d'Avenir” operated by the French National Research Agency (ANR). S.S. was supported by an India-France International
Research Project (IRP) funded by the French CNRS. M.L.M. was supported by the Lyon-Saint Etienne Research Federation
on Biodiversity, Water, Environment, City, Health (FR3728 BioEEnViS) and H2O'Lyon (Université de Lyon, ANR-17-
655 EURE-0018).

Review statement

The review statement will be added by Copernicus Publications listing the handling editor as well as all contributing referees
according to their status anonymous or identified.



660 References

- Allen, Y. (2016). Landscape Scale Assessment of Floodplain Inundation Frequency Using Landsat Imagery. *River Research and Applications*, 32(7), 1609–1620. <https://doi.org/10.1002/rra.2987>
- Boothroyd, R. J., Williams, R. D., Hoey, T. B., Brierley, G. J., Tolentino, P. L. M., Guardian, E. L., Reyes, J. C. M. O., Sabillo, C. J., Quick, L., Perez, J. E. G., & David, C. P. C. (2025). Big data show idiosyncratic patterns and rates of geomorphic
665 river mobility. *Nature Communications*, 16(1), 3263. <https://doi.org/10.1038/s41467-025-58427-9>
- Brown, C. F., Brumby, S. P., Guzder-Williams, B., Birch, T., Hyde, S. B., Mazzariello, J., Czerwinski, W., Pasquarella, V. J., Haertel, R., Ilyushchenko, S., Schwehr, K., Weisse, M., Stolle, F., Hanson, C., Guinan, O., Moore, R., & Tait, A. M. (2022). Dynamic World, Near real-time global 10 m land use land cover mapping. *Scientific Data*, 9(1), Article 1. <https://doi.org/10.1038/s41597-022-01307-4>
- 670 Carbonneau, P. E. (2025). Style Transfer from Sentinel-1 to Sentinel-2 for Fluvial Scenes with Multi-Modal and Multi-Temporal Image Fusion. *Remote Sensing*, 17(20). <https://doi.org/10.3390/rs17203445>
- Chen, H., Liang, Q., Liang, Z., Liu, Y., & Ren, T. (2020). Extraction of connected river networks from multi-temporal remote sensing imagery using a path tracking technique. *Remote Sensing of Environment*, 246, 111868. <https://doi.org/10.1016/j.rse.2020.111868>
- 675 Chrysostomou, C., Neophytides, S. P., Mavrovouniotis, M., & Hadjimitsis, D. G. (2026). Optimized spectral indices for global vegetation and water mapping using Sentinel-2. *Scientific Reports*, 16(1), 4491. <https://doi.org/10.1038/s41598-025-34720-x>
- Cooley, S. W., Ryan, J. C., & Smith, L. C. (2021). Human alteration of global surface water storage variability. *Nature*, 591(7848), 78–81. <https://doi.org/10.1038/s41586-021-03262-3>
- 680 Cooley, S. W., Smith, L. C., Stepan, L., & Mascaro, J. (2017). Tracking Dynamic Northern Surface Water Changes with High-Frequency Planet CubeSat Imagery. *Remote Sensing*, 9(12). <https://doi.org/10.3390/rs9121306>
- De Felice, L., Pekel, J.-F., Ugolotti, R., & Kempeneers, P. (2026). S2-GOLD [Dataset]. European Commission, Joint Research Centre (JRC). <https://doi.org/10.2905/JRC.NZQ1FAG>
- Dinerstein, E., Olson, D., Joshi, A., Vynne, C., Burgess, N. D., Wikramanayake, E., Hahn, N., Palminteri, S., Hedao, P., Noss, R., Hansen, M., Locke, H., Ellis, E. C., Jones, B., Barber, C. V., Hayes, R., Kormos, C., Martin, V., Crist, E., ...
685 Saleem, M. (2017). An Ecoregion-Based Approach to Protecting Half the Terrestrial Realm. *BioScience*, 67(6), 534–545. <https://doi.org/10.1093/biosci/bix014>
- Dong, X., & Hu, C. (2021). A new method for describing the inundation status of floodplain wetland. *Ecological Indicators*, 131, 108144. <https://doi.org/10.1016/j.ecolind.2021.108144>
- 690 Drusch, M., Del Bello, U., Carlier, S., Colin, O., Fernandez, V., Gascon, F., Hoersch, B., Isola, C., Laberinti, P., Martimort, P., Meygret, A., Spoto, F., Sy, O., Marchese, F., & Bargellini, P. (2012). Sentinel-2: ESA's Optical High-Resolution



- Mission for GMES Operational Services. *Remote Sensing of Environment*, 120, 25–36.
<https://doi.org/10.1016/j.rse.2011.11.026>
- Dunesme, S., Helling, L., Manière, L., Melun, G., Piégay, H., & Vaudor, L. (2025, June). *Mapd'O - A new tool for visualization and hydromorphological analysis of river corridor networks at the national scale*. <https://hal.science/hal-05178499>
- 695 Fluet-Chouinard, E., Stocker, B. D., Zhang, Z., Malhotra, A., Melton, J. R., Poulter, B., Kaplan, J. O., Goldewijk, K. K., Siebert, S., Minayeva, T., Hugelius, G., Joosten, H., Barthelmes, A., Prigent, C., Aires, F., Hoyt, A. M., Davidson, N., Finlayson, C. M., Lehner, B., ... McIntyre, P. B. (2023). Extensive global wetland loss over the past three centuries. *Nature*, 614(7947), 281–286. <https://doi.org/10.1038/s41586-022-05572-6>
- 700 Gao, S., Li, Z., Chen, M., Lin, P., Hong, Z., Allen, D., Neeson, T., & Hong, Y. (2021). Spatiotemporal variability of global river extent and the natural driving factors revealed by decades of Landsat observations, GRACE gravimetry observations, and land surface model simulations. *Remote Sensing of Environment*, 267, 112725. <https://doi.org/10.1016/j.rse.2021.112725>
- Google. (2026, February 20). *Earth Engine Noncommercial Tiers | Google Earth Engine*. Google for Developers. https://developers.google.com/earth-engine/guides/noncommercial_tiers
- 705 Hawker, L., Uhe, P., Paulo, L., Sosa, J., Savage, J., Sampson, C., & Neal, J. (2022). A 30 m global map of elevation with forests and buildings removed. *Environmental Research Letters*, 17(2), Article 2. <https://doi.org/10.1088/1748-9326/ac4d4f>
- Helling, L. (2026). *S2-DW-WOF workflow* [Computer software]. Zenodo. <https://doi.org/10.5281/zenodo.19392257>
- 710 Huang, C., Chen, Y., Zhang, S., & Wu, J. (2018). Detecting, Extracting, and Monitoring Surface Water From Space Using Optical Sensors: A Review. *Reviews of Geophysics*, 56(2), 333–360. <https://doi.org/10.1029/2018RG000598>
- Huang, R., Zhuo, L., Jie, N., Zheng, J., & Tao, H. (2025). Improving spatial and temporal accuracy in coastal reclamation mapping using satellite image time series. *International Journal of Applied Earth Observation and Geoinformation*, 139, 104499. <https://doi.org/10.1016/j.jag.2025.104499>
- 715 IGN. (2021). *BD TOPO®* [Dataset]. <https://geoservices.ign.fr/bdtopo>
- Jahangeer, J., Kapoor, A., Joshi, P., & Tang, Z. (2026). Assessing inundation dynamics of conservation reserve program lands to quantify ecosystem services. *Scientific Reports*. <https://doi.org/10.1038/s41598-026-45281-y>
- Jarriel, T., Swartz, J., & Passalacqua, P. (2021). Global rates and patterns of channel migration in river deltas. *Proceedings of the National Academy of Sciences of the United States of America*, 118(46), e2103178118. <https://doi.org/10.1073/pnas.2103178118>
- 720 Kaplan, G., & Avdan, U. (2017). Object-based water body extraction model using Sentinel-2 satellite imagery. *European Journal of Remote Sensing*, 50(1), 137–143. <https://doi.org/10.1080/22797254.2017.1297540>
- Karra, K., Kontgis, C., Statman-Weil, Z., Mazzariello, J. C., Mathis, M., & Brumby, S. P. (2021). Global land use / land cover with Sentinel 2 and deep learning. *2021 IEEE International Geoscience and Remote Sensing Symposium IGARSS*, 4704–4707. <https://doi.org/10.1109/IGARSS47720.2021.9553499>
- 725



- Kelso, N. V., & Patterson, T. (2010). *INTRODUCING NATURAL EARTH DATA - NATURALEARTHDAT.COM* [Dataset].
https://technicalgeography.org/pdf/sp_i_2010/12_introducing_natural_earth_data__natureleart.pdf
- Khare, A., Gupta, B. C., Rajib, A., Vanderhoof, M., & Wu, Q. (2025). Estimates of global surface water dynamics harnessing near real-time land cover observations and open science geospatial capabilities. *Environmental Research Letters*.
730 <https://doi.org/10.1088/1748-9326/ae137b>
- Langhorst, T., Andreadis, K. M., & Allen, G. H. (2024). Global Cloud Biases in Optical Satellite Remote Sensing of Rivers. *Geophysical Research Letters*, 51(16), e2024GL110085. <https://doi.org/10.1029/2024GL110085>
- Lehner, B., & Grill, G. (2013). Global river hydrography and network routing: Baseline data and new approaches to study the world's large river systems. *Hydrological Processes*, 27(15), 2171–2186. <https://doi.org/10.1002/hyp.9740>
- 735 Li, J., Li, L., Song, Y., Chen, J., Wang, Z., Bao, Y., Zhang, W., & Meng, L. (2023). A robust large-scale surface water mapping framework with high spatiotemporal resolution based on the fusion of multi-source remote sensing data. *International Journal of Applied Earth Observation and Geoinformation*, 118, 103288. <https://doi.org/10.1016/j.jag.2023.103288>
- Li, J., Meng, Y., Li, Y., Cui, Q., Yang, X., Tao, C., Wang, Z., Li, L., & Zhang, W. (2022). Accurate water extraction using remote sensing imagery based on normalized difference water index and unsupervised deep learning. *Journal of*
740 *Hydrology*, 612, 128202. <https://doi.org/10.1016/j.jhydrol.2022.128202>
- Lin, Y., Fan, C., Liu, K., Liu, L., Deng, X., Xu, P., Zhan, P., Fu, X., & Song, C. (2025). Characterizing the capability of public DEMs for mapping global floodplain bathymetry. *Journal of Hydrology*, 133205.
<https://doi.org/10.1016/j.jhydrol.2025.133205>
- Liu, H., Hu, H., Liu, X., Jiang, H., Liu, W., & Yin, X. (2022). A Comparison of Different Water Indices and Band Downscaling
745 Methods for Water Bodies Mapping from Sentinel-2 Imagery at 10-M Resolution. *Water*, 14(17), 2696.
<https://doi.org/10.3390/w14172696>
- Liu, Y., Liu, R., & Shang, R. (2022). GLOBMAP SWF: A global annual surface water cover frequency dataset during 2000–2020. *Earth System Science Data*, 14(10), 4505–4523. <https://doi.org/10.5194/essd-14-4505-2022>
- Luo, Q., Feng, L., Park, E., Walling, D. E., Huang, L., Wang, M., Woolway, R. I., Fang, H., & Gong, J. (2026). Global net
750 increase in surface water connectivity in river–floodplain systems. *Nature Geoscience*, 1–9.
<https://doi.org/10.1038/s41561-026-01953-y>
- Luo, Q., Park, E., Latrubesse, E. M., Pi, X., Fang, H., Liu, Y., Switzer, A. D., Li, D., Qiu, W., Zheng, C., Liu, P., & Feng, L. (2026). Global alluvial channel patterns. *Nature Communications*, 17(1), 1950. <https://doi.org/10.1038/s41467-026-68569-z>
- 755 Ma, S., Zhou, Y., Gowda, P. H., Dong, J., Zhang, G., Kakani, V. G., Wagle, P., Chen, L., Flynn, K. C., & Jiang, W. (2019). Application of the water-related spectral reflectance indices: A review. *Ecological Indicators*, 98, 68–79.
<https://doi.org/10.1016/j.ecolind.2018.10.049>



- Martinis, S., Groth, S., Wieland, M., Knopp, L., & Rättich, M. (2022). Towards a global seasonal and permanent reference water product from Sentinel-1/2 data for improved flood mapping. *Remote Sensing of Environment*, 278, 113077. <https://doi.org/10.1016/j.rse.2022.113077>
- McFeeters, S. K. (1996). The use of the Normalized Difference Water Index (NDWI) in the delineation of open water features. *International Journal of Remote Sensing*, 17(7), 1425–1432. <https://doi.org/10.1080/01431169608948714>
- Messenger, M. L., Lehner, B., Cockburn, C., Lamouroux, N., Pella, H., Snelder, T., Tockner, K., Trautmann, T., Watt, C., & Detry, T. (2021). Global prevalence of non-perennial rivers and streams. *Nature*, 594(7863), 391–397. <https://doi.org/10.1038/s41586-021-03565-5>
- Miura, Y., Shamsudduha, M., Suppasri, A., & Sano, D. (2025). A Global Multi-Sensor Dataset of Surface Water Indices from Landsat-8 and Sentinel-2 Satellite Measurements. *Scientific Data*, 12(1), 1253. <https://doi.org/10.1038/s41597-025-05562-z>
- Mukherjee, R., Policelli, F., Wang, R., Arellano-Thompson, E., Tellman, B., Sharma, P., Zhang, Z., & Giezdanner, J. (2024). A globally sampled high-resolution hand-labeled validation dataset for evaluating surface water extent maps. *Earth System Science Data*, 16(9), Article 9. <https://doi.org/10.5194/essd-16-4311-2024>
- Naboureh, A., Li, A., Bian, J., Lei, G., Nan, X., & Zhang, Z. (2025). A Comprehensive Analysis of the Different Remote Sensing-Based Approaches for Surface Water Mapping in Mountainous Areas. *Earth Systems and Environment*. <https://doi.org/10.1007/s41748-025-00782-0>
- Olefeldt, D., Hovemyr, M., Kuhn, M. A., Bastviken, D., Bohn, T. J., Connolly, J., Crill, P., Euskirchen, E. S., Finkelstein, S. A., Genet, H., Grosse, G., Harris, L. I., Heffernan, L., Helbig, M., Hugelius, G., Hutchins, R., Juutinen, S., Lara, M. J., Malhotra, A., ... Watts, J. D. (2021). The Boreal–Arctic Wetland and Lake Dataset (BAWLD). *Earth System Science Data*, 13(11), 5127–5149. <https://doi.org/10.5194/essd-13-5127-2021>
- Olivier, J.-M., Carrel, G., Lamouroux, N., Dole-Olivier, M.-J., Malard, F., Bravard, J.-P., Piégay, H., Castella, E., & Barthélemy, C. (2022). Chapter 11—The Rhône River Basin. In K. Tockner, C. Zarfl, & C. T. Robinson (Eds.), *Rivers of Europe (Second Edition)* (pp. 393–453). Elsevier. <https://doi.org/10.1016/B978-0-08-102612-0.00011-0>
- Olofsson, P., Foody, G. M., Herold, M., Stehman, S. V., Woodcock, C. E., & Wulder, M. A. (2014). Good practices for estimating area and assessing accuracy of land change. *Remote Sensing of Environment*, 148, 42–57. <https://doi.org/10.1016/j.rse.2014.02.015>
- Ottinger, M., Liu, K., Ullmann, T., Huth, J., Kuenzer, C., & Bachofer, F. (2026). Pond aquaculture dynamics in Asia: Satellite time series for analyzing the spatio-temporal development of coastal aquaculture. *Aquaculture*, 610, 742940. <https://doi.org/10.1016/j.aquaculture.2025.742940>
- Pal, S., & Sarda, R. (2020). Damming effects on the degree of hydrological alteration and stability of wetland in lower Atreyee River basin. *Ecological Indicators*, 116, 106542. <https://doi.org/10.1016/j.ecolind.2020.106542>
- Parrot, E. (2015). *Analyse spatio-temporelle de la morphologie du chenal du Rhône du Léman à la Méditerranée* [Université Jean Moulin (Lyon 3)]. <https://doi.org/10.70675/590864baz8d53z428aza9dezc5d31352ff17>



- Pasquarella, V. J., Brown, C. F., Czerwinski, W., & Rucklidge, W. J. (2023). *Comprehensive Quality Assessment of Optical Satellite Imagery Using Weakly Supervised Video Learning*. 2125–2135. <https://doi.org/10.1109/CVPRW59228.2023.00206>
- 795 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), Article 7633. <https://doi.org/10.1038/nature20584>
- Pi, X., Luo, Q., Feng, L., Xu, Y., Tang, J., Liang, X., Ma, E., Cheng, R., Fensholt, R., Brandt, M., Cai, X., Gibson, L., Liu, J., Zheng, C., Li, W., & Bryan, B. A. (2022). Mapping global lake dynamics reveals the emerging roles of small lakes. *Nature Communications*, 13(1), 5777. <https://doi.org/10.1038/s41467-022-33239-3>
- 800 Pickens, A. H., Hansen, M. C., Hancher, M., Stehman, S. V., Tyukavina, A., Potapov, P., Marroquin, B., & Sherani, Z. (2020). Mapping and sampling to characterize global inland water dynamics from 1999 to 2018 with full Landsat time-series. *Remote Sensing of Environment*, 243, 111792. <https://doi.org/10.1016/j.rse.2020.111792>
- Prancevic, J. P., Seybold, H., & Kirchner, J. W. (2025). Variability of flowing stream network length across the US. *Science*, 387(6735), 782–786. <https://doi.org/10.1126/science.ado2860>
- 805 Radeloff, V. C., Roy, D. P., Wulder, M. A., Anderson, M., Cook, B., Crawford, C. J., Friedl, M., Gao, F., Gorelick, N., Hansen, M., Healey, S., Hostert, P., Hulley, G., Huntington, J. L., Johnson, D. M., Neigh, C., Lyapustin, A., Lymburner, L., Pahlevan, N., ... Zhu, Z. (2024). Need and vision for global medium-resolution Landsat and Sentinel-2 data products. *Remote Sensing of Environment*, 300, 113918. <https://doi.org/10.1016/j.rse.2023.113918>
- Rajeswari, S., & Rathika, P. (2024). Emerging methodologies in waterbody delineation: An In-depth review. *International Journal of Remote Sensing*, 45(17), 5789–5819. <https://doi.org/10.1080/01431161.2024.2379518>
- 810 Rajib, A., Khare, A., Golden, H. E., Gupta, B. C., Wu, Q., Lane, C. R., Christensen, J. R., Zheng, Q., Dahl, T. A., Ryder, J. L., & McFall, B. C. (2024). A call for consistency and integration in global surface water estimates. *Environmental Research Letters: ERL [Web Site]*, 19, 1–6. <https://doi.org/10.1088/1748-9326/ad1722>
- Sagan, V., Peterson, K. T., Maimaitijiang, M., Sidike, P., Sloan, J., Greeling, B. A., Maalouf, S., & Adams, C. (2020). Monitoring inland water quality using remote sensing: Potential and limitations of spectral indices, bio-optical simulations, machine learning, and cloud computing. *Earth-Science Reviews*, 205, 103187. <https://doi.org/10.1016/j.earscirev.2020.103187>
- 815 Schwedhelm, H., Siegfried, T., Casas-Mulet, R., Dallmeier, A., & Rüther, N. (2026). A novel framework for analysing river planform characteristics in data-scarce regions using remote sensing data. *Earth Surface Processes and Landforms*, 51(3), e70276. <https://doi.org/10.1002/esp.70276>
- 820 Sogno, P., Klein, I., & Kuenzer, C. (2022). Remote Sensing of Surface Water Dynamics in the Context of Global Change—A Review. *Remote Sensing*, 14(10), Article 10. <https://doi.org/10.3390/rs14102475>
- Song, C., Huang, B., & Ke, L. (2013). Modeling and analysis of lake water storage changes on the Tibetan Plateau using multi-mission satellite data. *Remote Sensing of Environment*, 135, 25–35. <https://doi.org/10.1016/j.rse.2013.03.013>



- 825 Stegen, J., Burgin, A. J., Busch, M. H., Fisher, J. B., Ladau, J., Abrahamson, J., Kinsman-Costello, L., Li, L., Chen, X., Datry, T., McDowell, N., Tatariw, C., Braswell, A., Deines, J. M., Guimond, J. A., Regier, P., Rod, K., Bam, E. K. P., Fluet-Chouinard, E., ... Patel, K. F. (2025). Reviews and syntheses: Variable inundation across Earth's terrestrial ecosystems. *Biogeosciences*, 22(4), 995–1034. <https://doi.org/10.5194/bg-22-995-2025>
- Strahler, A. N. (1957). Quantitative analysis of watershed geomorphology. *Eos, Transactions American Geophysical Union*, 38(6), 913–920. <https://doi.org/10.1029/TR038i006p00913>
- 830 Stürmer, M., Nussbaumer, J., & Stöckli, P. (2021). *Security implications of digitalization: The dangers of data colonialism and the way towards sustainable and sovereign management of environmental data*. <https://doi.org/10.13140/RG.2.2.24791.80807>
- Sudmanns, M., Tiede, Dirk, Augustin, Hannah, & Lang, S. (2020). Assessing global Sentinel-2 coverage dynamics and data availability for operational Earth observation (EO) applications using the EO-Compass. *International Journal of Digital Earth*, 13(7), 768–784. <https://doi.org/10.1080/17538947.2019.1572799>
- Tellman, B., Sullivan, J. A., Kuhn, C., Kettner, A. J., Doyle, C. S., Brakenridge, G. R., Erickson, T. A., & Slayback, D. A. (2021). Satellite imaging reveals increased proportion of population exposed to floods. *Nature*, 596(7870), 80–86. <https://doi.org/10.1038/s41586-021-03695-w>
- 840 Tootchi, A., Jost, A., & Ducharne, A. (2019). Multi-source global wetland maps combining surface water imagery and groundwater constraints. *Earth System Science Data*, 11(1), 189–220. <https://doi.org/10.5194/essd-11-189-2019>
- Tottrup, C., Druce, D., Meyer, R. P., Christensen, M., Riffler, M., Dulleck, B., Rastner, P., Jupova, K., Sokoup, T., Haag, A., Cordeiro, M. C. R., Martinez, J.-M., Franke, J., Schwarz, M., Vanthof, V., Liu, S., Zhou, H., Marzi, D., Rudiyanto, R., ... Paganini, M. (2022). Surface Water Dynamics from Space: A Round Robin Intercomparison of Using Optical and SAR High-Resolution Satellite Observations for Regional Surface Water Detection. *Remote Sensing*, 14(10). <https://doi.org/10.3390/rs14102410>
- 845 van Rijsbergen, C. J. (1974). Foundation of Evaluation. *Journal of Documentation*, 30(4), 365–373. <https://doi.org/10.1108/eb026584>
- Vanderhoof, M. K., Alexander, L., Christensen, J., Solvik, K., Nieuwlandt, P., & Sagehorn, M. (2023). High-frequency time series comparison of Sentinel-1 and Sentinel-2 satellites for mapping open and vegetated water across the United States (2017–2021). *Remote Sensing of Environment*, 288, 113498. <https://doi.org/10.1016/j.rse.2023.113498>
- Wang, J., Chen, F., Zhang, M., & Yu, B. (2022). NAU-Net: A New Deep Learning Framework in Glacial Lake Detection. *IEEE Geoscience and Remote Sensing Letters*, 19, 1–5. <https://doi.org/10.1109/LGRS.2022.3165045>
- Wang, Y., Foody, G., Zhou, P., Li, Y., Li, X., Zhang, Y., Du, Y., & Li, X. (2025). Dynamic surface water fraction (DSWF): Global surface water fraction mapping at 10-meter spatial resolution with Sentinel-2 imagery in Google Earth Engine. *International Journal of Applied Earth Observation and Geoinformation*, 143, 104813. <https://doi.org/10.1016/j.jag.2025.104813>
- 855



- Wieland, M., Fichtner, F., Martinis, S., Groth, S., Krullikowski, C., Plank, S., & Motagh, M. (2024). S1S2-Water: A Global Dataset for Semantic Segmentation of Water Bodies From Sentinel- 1 and Sentinel-2 Satellite Images. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 17, 1084–1099. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*. <https://doi.org/10.1109/JSTARS.2023.3333969>
- Wu, Q., Ke, L., Wang, J., Pavelsky, T. M., Allen, G. H., Sheng, Y., Duan, X., Zhu, Y., Wu, J., Wang, L., Liu, K., Chen, T., Zhang, W., Fan, C., Yong, B., & Song, C. (2023). Satellites reveal hotspots of global river extent change. *Nature Communications*, 14(1), 1587. <https://doi.org/10.1038/s41467-023-37061-3>
- 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), 3025–3033. <https://doi.org/10.1080/01431160600589179>
- Yang, X., Qin, Q., Yésou, H., Ledauphin, T., Koehl, M., Grussenmeyer, P., & Zhu, Z. (2020). Monthly estimation of the surface water extent in France at a 10-m resolution using Sentinel-2 data. *Remote Sensing of Environment*, 244, 111803. <https://doi.org/10.1016/j.rse.2020.111803>
- Yao, F., Livneh, B., Rajagopalan, B., Wang, J., Crétaux, J.-F., Wada, Y., & Berge-Nguyen, M. (2023). Satellites reveal widespread decline in global lake water storage. *Science*, 380(6646), 743–749. <https://doi.org/10.1126/science.abo2812>
- Yu, H., Li, G., Liu, H., Zhu, S., Xu, J., Dong, W., Li, C., & Shi, J. (2026). Synergistic Fusion of Sentinel-1 and Sentinel-2 for Global LULC Mapping: The Multimodal Network LULC-Former and Dynamic World+ Dataset. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 19, 2511–2524. <https://doi.org/10.1109/JSTARS.2025.3641788>
- Zhao, C., Wei, H., Feyisa, G. L., De Castro Tayer, T., Ma, G., Wu, H., & Pan, Y. (2025). Evaluating spectral indices for water extraction: Limitations and contextual usage recommendations. *International Journal of Applied Earth Observation and Geoinformation*, 139, 104510. <https://doi.org/10.1016/j.jag.2025.104510>
- Zhao, G., Gao, H., Li, Y., Tang, Q., Woolway, R. I., Merder, J., Rosa, L., & Michalak, A. M. (2025). Decoupling of surface water storage from precipitation in global drylands due to anthropogenic activity. *Nature Water*, 3(1), 80–88. <https://doi.org/10.1038/s44221-024-00367-7>
- Zhao, G., Li, Y., Zhou, L., & Gao, H. (2022). Evaporative water loss of 1.42 million global lakes. *Nature Communications*, 13(1), 3686. <https://doi.org/10.1038/s41467-022-31125-6>
- Zhou, Y., Dong, J., Xiao, X., Xiao, T., Yang, Z., Zhao, G., Zou, Z., & Qin, Y. (2017). Open Surface Water Mapping Algorithms: A Comparison of Water-Related Spectral Indices and Sensors. *Water*, 9(4), 256. <https://doi.org/10.3390/w9040256>