High resolution satellite products improve hydrological modeling in northern Italy

Lorenzo Alfieri[1], Francesco Avanzi[1], Fabio Delogu[1], Simone Gabellani[1], Giulia Bruno[1], Lorenzo Campo[1], Andrea Libertino[1], Christian Massari[2], Angelica Tarpanelli[2], Dominik Rains[3], Diego G. Miralles[3], Raphael Quast[4], Mariette Vreugdenhil[4], Huan Wu[5,6], Luca Brocca[2]

[1]CIMA Research Foundation, University Campus of Savona, Savona, 17100, Italy
[2]National Research Council, Research Institute for Geo-Hydrological Protection, Perugia, 06128, Italy
[3]Hydro-Climate Extremes Lab (H-CEL), Ghent University, Gent, 9000, Belgium
[4]Department of Geodesy and Geoinformation, TU Wien, Vienna, 1040, Austria
[5]Southern Marine Science and Engineering Laboratory (Zhuhai), Zhuhai, 519082, China
[6]School of Atmospheric Sciences, Sun Yat-sen University, Guangzhou, 510275, China

Correspondence to: Lorenzo Alfieri (Lorenzo.Alfieri@cimafoundation.org)

Abstract. Satellite Earth observations (EO) are an accurate and reliable data source for atmospheric and environmental science. Their increasing spatial and temporal resolution, as well as the seamless availability over ungauged regions, make them appealing for hydrological modeling. This work shows recent advances in the use of high-resolution satellite-based Earth observation data in hydrological modelling. In a set of experiments, the distributed hydrological model Continuum is set up for the Po River Basin (Italy) and forced, in turn, by satellite precipitation and evaporation, while satellite-derived soil moisture and snow depths are ingested into the model structure through a data-assimilation scheme. Further, satellite-based estimates of precipitation, evaporation and river discharge are used for hydrological model calibration, and results are compared with those based on ground observations. Despite the high density of conventional ground measurements and the strong human influence in the focus region, all satellite products show strong potential for operational hydrological applications, with skillful estimates of river discharge throughout the model domain. Satellite-based evaporation and snow depths marginally improve (by 2% and 4%) the mean Kling-Gupta efficiency (KGE) at 27 river gauges, compared to a baseline simulation ($KGE_{mean}=0.51$) forced by high-quality conventional data. Precipitation has the largest impact on the model output, though the satellite dataset on average shows poorer skills compared to conventional data. Interestingly, a model calibration heavily relying on satellite data, as opposed to conventional data, provides a skillful reconstruction of river discharges, paving the way to fully satellite-driven hydrological applications.

1 Introduction

Remote sensing of the Earth from space is a ripe yet ever growing sector, with countless applications and users worldwide. Hydrological sciences have already benefited enormously from Earth observation data (see e.g., McCabe et al., 2017; Chen and Wang, 2018; Alfieri et al., 2018), thanks to global and independent datasets of the different components of the water and

energy cycles as well as anthropogenic processes such as irrigation (Massari et al., 2021). Hydrological models play a crucial role for monitoring and forecasting, thanks to their ability to reproduce the physical processes governing the water cycle. Their successful implementation is strongly conditioned by the availability of consistent, accurate, and seamless hydro-

35   meteorological datasets for the considered focus region, space/time resolution and period of interest. Conventional data including ground observations and weather radars are traditionally favourite sources of dynamic data to force these models. Yet, they are not viable options for the still vast ungauged regions of the world. Satellite products offer a range of alternatives to fill such gaps, thanks to their massive contribution to the atmospheric reanalyses (see e.g., Hersbach et al., 2020) as well as with independent products. Hydrological models can benefit from dynamic data (either ground or satellite-

40   based) in various forms: 1) as forcing datasets, 2) as assimilation datasets, 3) as benchmark data for model training and calibration, and 4) to investigate process understanding.

Forcing data are mandatory input for hydrological models. Key variables are precipitation, air temperature and evaporation or, alternatively, the meteorological variables needed to estimate them. Their influence in hydrological modeling was assessed, for instance, by Wu et al. (2017) and Beck et al. (2017) for precipitation datasets, Dembélé et al. (2020a) for

45   evaporation datasets, and Dembélé et al. (2020b) for combinations of temperature and precipitation datasets. The latter found a reduced influence of the choice of temperature datasets on the output discharge, though these can significantly impact evaporation and soil moisture estimates. Data assimilation methods are designed to merge measurements of any type with estimates from geophysical models (Reichle, 2008), to compensate for errors in the forcing data, model structural deficiencies, and update their state variables at the initial or intermediate simulation steps (Spaaks and Bouten, 2013).

50   Relevant applications of assimilating satellite products in hydrological modeling include soil moisture (Massari et al., 2015; Wanders et al., 2014), water storage (Li et al., 2012), snow cover (Thirel et al., 2013), evaporation (Hartanto et al., 2017), land surface temperature (Silvestro et al., 2013), water levels (Paiva et al., 2013), discharge (Ishitsuka et al., 2021), water extent (Revilla-Romero et al., 2016; Hostache et al., 2018), and multi variable combinations (Wongchuig-Correa et al., 2020). Hydro-meteorological data has also been used as a benchmark to train the model parameters through machine

55   learning techniques (Mosaffa et al., 2022) or calibration techniques based on minimization of cost functions computed between simulated and observed variables (Pechlivanidis et al., 2011). Satellite estimation of river levels also shows promising applications in the field. It has been tested in the calibration of hydrological (Getirana et al., 2013; Dhote et al., 2021) and hydraulic (Domeneghetti et al., 2021) models.

As part of the Green Deal and the Digital Strategy, the European Commission recently launched the Destination Earth

60   program[1], a joint effort involving key European institutions to develop a very high precision digital model, or "Digital Twin", of the Earth to monitor and predict environmental change and human impacts, to ultimately support sustainable development. The present work strives in that direction, by contributing to the development of a Digital Twin Earth focused on the water cycle and hydrological processes. It highlights the potential of high-resolution satellite products in describing

---

1   https://digital-strategy.ec.europa.eu/en/policies/destination-earth

the water cycle and monitoring hydrological extremes and water resources. Through various dedicated experiments, we test
65   the influence of five new high-resolution satellite-derived datasets on the performance of the distributed hydrological model
Continuum (Silvestro et al., 2013) set up for the entire Po River Basin, in northern Italy. These include (1) GPM-SM2RAIN
(Massari et al., 2020) precipitation and (2) GLEAM (Miralles et al., 2011) evaporation as dynamic forcing; data assimilation
of (3) C-SNOW (Lievens et al., 2019) snow depth and (4) RT1 (Quast et al., 2019) soil moisture; and model calibration
using (5) satellite-based river discharge (Tarpanelli et al., 2020) as a benchmark. By comparing results with observed river
70   discharge and with a simulation forced by conventional data, we investigate the relative impact of these high-resolution
satellite products. Further, we take the first steps towards hydrological modelling fully relying on satellite data, by
calibrating and subsequently running the model using satellite precipitation and evaporation as forcing, and satellite-based
estimates of river discharge as benchmark data for the calibration.

2 Data

75   ### 2.1 Static data

We used the Digital Elevation Model (DEM) from the global USGS Hydrologic Derivatives for Modeling and Analysis
(HDMA, Verdin, 2017) at 3 arc-second spatial resolution (about 90 m at the equator), which comes with pre-computed and
corrected hydrological derivatives including channel network and macro basins. The DEM was upscaled at the chosen model
resolution of 1 km to define the computational grid and compute the necessary hydrological derivatives (flow accumulation,
80   drainage directions and channel network). The river network is defined by cells with an upstream area larger than 240 km^2.
To improve its spatial representation, the DEM was carved with a high-resolution stream network of the main rivers, while
dikes were manually placed at specific locations, especially in flat areas.

The Curve Number map used in Continuum to model direct runoff and infiltration from rainfall excess, was derived from the
ESA-CCI 2018 Land Cover map (ESA, 2017) at 300 m resolution. A hydrologic soil type map was extracted from the
85   HYSOGs250m (Ross et al., 2018), while for soil capacity, we applied the USDA method for soil texture identification using
the ISRIC SoilGrids (Hengl et al., 2017) global maps of the fraction of sand and clay, combined with the ESA CCI
SoilMoisture (Dorigo et al., 2017) global map of soil porosity. Glacier areas used in the cryospheric model S3M (see Section
3.1) were taken from the Randolph Glacier Inventory (RGI) v6 (Raup et al., 2007). Vegetation coverage is taken from the
global land cover map ECOCLIMAP (Faroux et al., 2013).
90   Point information for a set of 99 reservoirs and the three major lakes (Maggiore, Como and Garda) was included in the
model setup (Figure 1). Information on the dams and the corresponding reservoirs was provided by the Italian Civil
Protection Department (DPC) and from the GranD database (Lehner et al., 2011). Data ingested by Continuum for each dam
include the maximum stored volume, initial volume, maximum non-damaging discharge at the outflow gates, weir length,
maximum storage level, outflow coefficient, and coordinates of the release point. For lakes, required metadata are the outlet
95   coordinates, minimum volume inducing outflow discharge, initial volume, and emptying coefficient.

2.2 Dynamic data

Continuum requires input maps of precipitation, air temperature, relative humidity, wind speed and incoming solar radiation. The baseline hydrological simulation uses conventional meteorological data as input. Precipitation fields were estimated with the Modified Conditional Merging (MCM) technique (Bruno et al., 2021), which incorporates precipitation gauges and

100   radar estimates. MCM is an improvement of the Conditional Merging proposed by Sinclair and Pegram (2005), which estimates the structure of covariance and the length of spatial correlation at every gauge, taking it from the cumulated radar precipitation fields. For the Po River basin, MCM is based on 1377 precipitation gauges and on the mosaic of the Italian weather radars.

Hourly maps of the weather variables needed by Continuum for the Po river basin ultimately include 1258 temperature

105   stations, 608 for relative humidity, 460 for wind speed and 278 for solar radiation. Temperature maps include an altitude correction algorithm with temperature gradients estimated at every time step by linearly interpolating available data at different elevations. They also include an outlier removal algorithm which discards station data with a deviation of more than 20°C from the corresponding temperature-elevation interpolating line.

Discharge data at 27 river gauging stations with hourly sampling frequency for the years 2016-2019 were provided by the

110   DPC and the regional hydrometeorological offices. 22 stations were selected for model calibration, while 5 were retained for validation only (Figure 1).

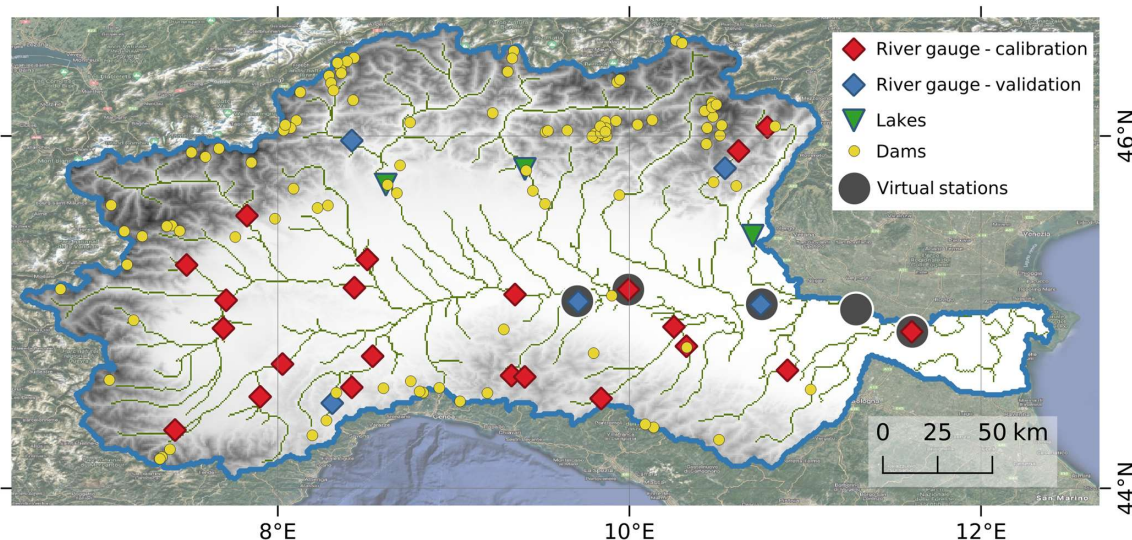



Figure 1: Simulated domain (blue line) and river network (dark green) of the Po river basin. Symbols show the point features considered in the hydrological model.

2.3 Satellite products and validation

115 **1.2.1 Precipitation**

The precipitation dataset used in this work and referred to as SM2RAIN (Figure 2) merges SM2RAIN-ASCAT (Brocca et al., 2019) and the Global Precipitation Measurement Mission IMERG-LR (Huffman et al., 2015) datasets, both available at 10km spatial resolution. Unlike in Massari et al. (2020), where the fusion of the two datasets was based on an Optimal Interpolation Technique, here we relied upon a Triple Collocation (TC)-based merging using the Signal to Noise Ratio (SNR), as in Gruber et al (2017). In particular, to derive the merged dataset we seek the optimality in a least squares sense, so that the variance of residual random errors is minimized. This leads to a weighted average between SM2RAIN-ASCAT and IMERG-LR, i.e.,

120

$$P_{DTE} = w_1 P_{SM2RAIN-ASCA} + w_2 P_{IMERG-LR} \tag{1}$$

where the weights w_1 and w_2 are calculated as:

125

$$w_1 = \frac{SNR_1}{SNR_1 + SNR_2}, w_2 = \frac{SNR_2}{SNR_1 + SNR_2} \tag{2}$$

and the subscripts 1 and 2 refer to the SM2RAIN-ASCAT and IMERG-LR datasets, respectively. Under the assumption that the two datasets are independent (as also required by TC), the random error of the merged time series is lower than those of the individual input datasets.

Prior to the merging, the two datasets are rescaled to a common reference to remove the relative systematic differences

130 between the products. TC was applied here between SM2RAIN-ASCAT, IMERG-LR and the MCM radar-gauge precipitation dataset. The fusion of the two datasets was only done for the time steps where IMERG-LR was greater than zero; due to the high sensitivity of the GPM mission, values with zero precipitation in IMERG-LR were set to zero. Furthermore, hourly data were obtained by imposing the sub-daily temporal pattern of IMERG-LR to the merged dataset.

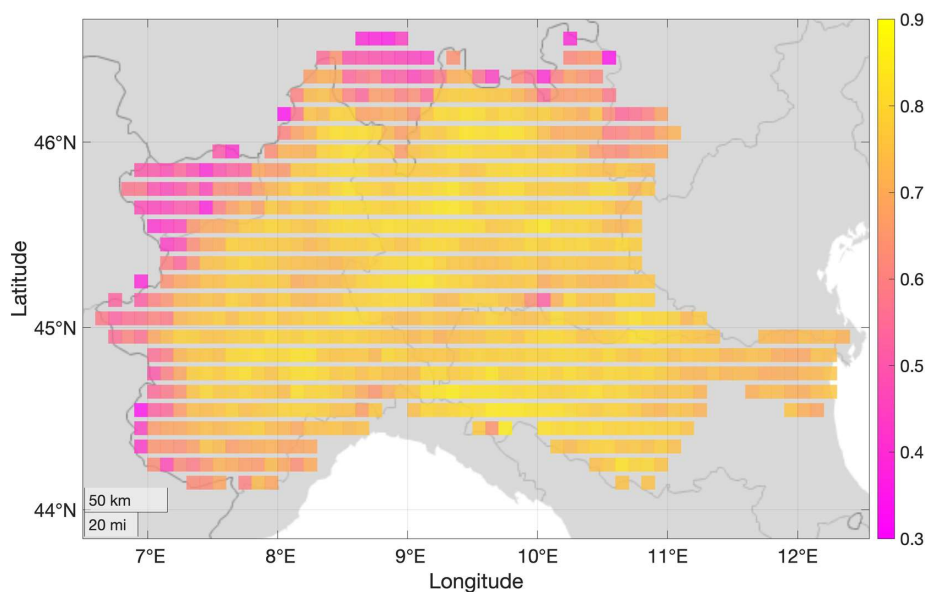



Figure 2: Daily Pearson correlation coefficient between SM2RAIN precipitation and the MCM (radar-gauge) precipitation dataset. Median correlation $r_{SM2RAIN}$=0.76 largely improves that of the two individual products, $r_{SM2RAIN-ASCAT}$=0.66 and $r_{IMERG-LR}$=0.67.

2.3.2 Evaporation

Global Land Evaporation Amsterdam Model (GLEAM, Miralles et al., 2011) is a state-of-the-art methodology to derive evaporation and its various components (i.e. transpiration, bare soil evaporation, interception loss, snow sublimation and open-water evaporation). It combines global satellite observations of meteorological (precipitation, near-surface net radiation, and air temperature) and surface (soil and vegetation water content, and snow water equivalent) variables that are informative for the evaporation process. The model is based on the Priestley and Taylor (1972) equation to estimate potential evaporation. Those estimates are then constrained based on root-zone soil moisture, which results from a precipitation-driven running water balance in which satellite-based soil moisture can be assimilated. Interception loss is independently estimated through an adapted Gash analytical model (Miralles et al., 2010). Since its first version, GLEAM has been widely deployed at coarse resolution for climatic studies. In the past few years, it has been further developed to solve higher spatial and temporal resolutions. For instance, Martens et al. (2018) obtained accurate results in an implementation over the Netherlands at 100-meter resolution. For this work, GLEAM was applied over the entire Po River Basin to produce evaporation estimates at 1 km resolution.

Since measurements of evaporation in the focus region are limited, the performance of the 1-km evaporation dataset was inferred on the basis of the FluxNet IT-Tor site, located in the mountainous Val d'Aosta region in the NW part of the domain (Figure 3). While based on one station only, the performance (Pearson's correlation r=0.83) is in line with results obtained in

the high-resolution implementation across the Netherlands, where Martens et al. (2018) found a median temporal correlation

155   coefficient of 0.76 across 29 sites.

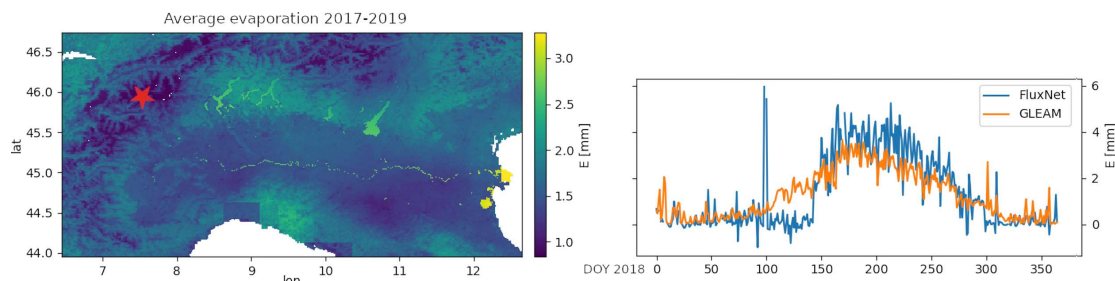



Figure 3: Average GLEAM daily evaporation for 2017-2019 in the Po river basin at 1 km resolution (left) and comparison with daily evaporation from the FluxNet site IT-Tor for 2018 (right). Pearson's correlation r=0.83. The location of the FluxNet site is marked with a red star on the left panel.

160   **2.3.3 Soil moisture**

High-resolution soil moisture was retrieved from incidence angle dependent Sentinel-1 backscatter measurements at 500m spatial sampling (~1km spatial resolution) (Bauer-Marschallinger et al., 2019) by using a time series based first-order radiative transfer modelling approach (RT1, see Quast and Wagner, 2016; Quast et al., 2019). The RT1 model uses auxiliary Leaf Area Index (LAI) time series provided by ECMWF ERA5-Land reanalysis dataset (Muñoz-Sabater et al., 2021) to

165   correct for effects induced by seasonal dynamics of vegetation. The retrieval is then performed via a non-linear least-squares regression that optimizes static and dynamic model-parameters to minimize the difference between measured and modelled backscatter for a set of ~300,000 pixels over a 4-year time-period (2016-2019). The resulting soil moisture product represents a percentage measure of the relative moisture saturation of the soil surface. The performance of the obtained soil moisture time series was validated with in-situ observations as well as compared to top-layer (0-7cm) soil moisture estimates

170   from ERA5-Land. In addition, the spatial distribution of the resulting auxiliary model parameters (single-scattering albedo, soil scattering directionality) was analysed with respect to CCI Landcover (ESA, 2017) classifications to assess the physical plausibility of the resulting parametrization. The observed spatial pattern of the parameters indicate a close connection to the associated land cover, following some expected variations, e.g., higher single-scattering albedo over forested areas compared to croplands.

175   The RT1 high-resolution soil moisture product over the Po basin shows an overall good performance compared to ERA5-Land soil moisture, with a median Pearson correlation of 0.55 for croplands and 0.65 over areas primarily covered by natural vegetation. Validation was performed using in situ soil moisture for the Oltrepo station (Bordoni et al., 2019), which resulted in a correlation of 0.58 (raw data) and 0.73 (with a 10-daily rolling mean). These results highlight the potential of Sentinel-1 observations for high-resolution soil moisture retrievals and their use in applied science.

180

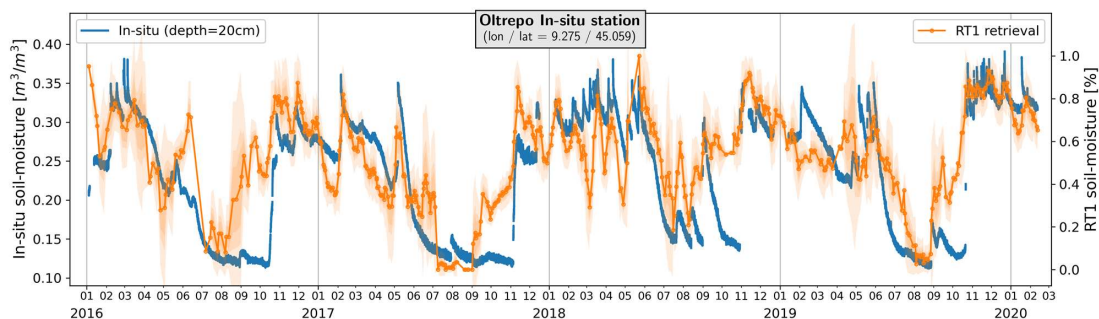



Figure 4: Time series of RT1 Surface Soil Moisture compared to in situ SM in Oltrepo. A 10-daily rolling mean is applied to the RT1 retrievals to reduce noise. The shading indicates the corresponding standard-deviation. Pearson-correlation of 0.58 (raw data) and 0.73 (10-daily rolling mean).

2.3.4 Snow depth

Snow-depth data were obtained from the Sentinel-1-based product proposed by Lievens et al. (2019). The data product has a 1-km spatial resolution and daily granularity, and is available through the public repository of the C-SNOW project (https://ees.kuleuven.be/project/c-snow). The mapping algorithm is based on a change-detection approach and has been validated across the mountain regions in the whole Northern Hemisphere.

For the scope of the present study, C-SNOW data over the period September 2016 - April 2020 were evaluated with 172 ultrasonic snow-depth sensors across the Po river basin (Figure 5a). 77% of the evaluation dataset is located in the range 1000-2500 m above sea level (ASL) (Figure 5b), a frequent condition in the Alps (Avanzi et al., 2021a). Observed snow-depth data were processed and aggregated at daily resolution, and C-SNOW data were extracted for the same location and data range. The evaluation confirmed previous results by Lievens et al. (2019), with C-SNOW successfully reproducing the seasonality and magnitude of snow depth as measured by snow depth sensors (Figure 5c and d). Root mean square errors (RMSE) ranged from less than 20 cm below 1000 m ASL to 60 cm or more above 2000 m ASL, though with no significant trend in the bias versus the elevation.

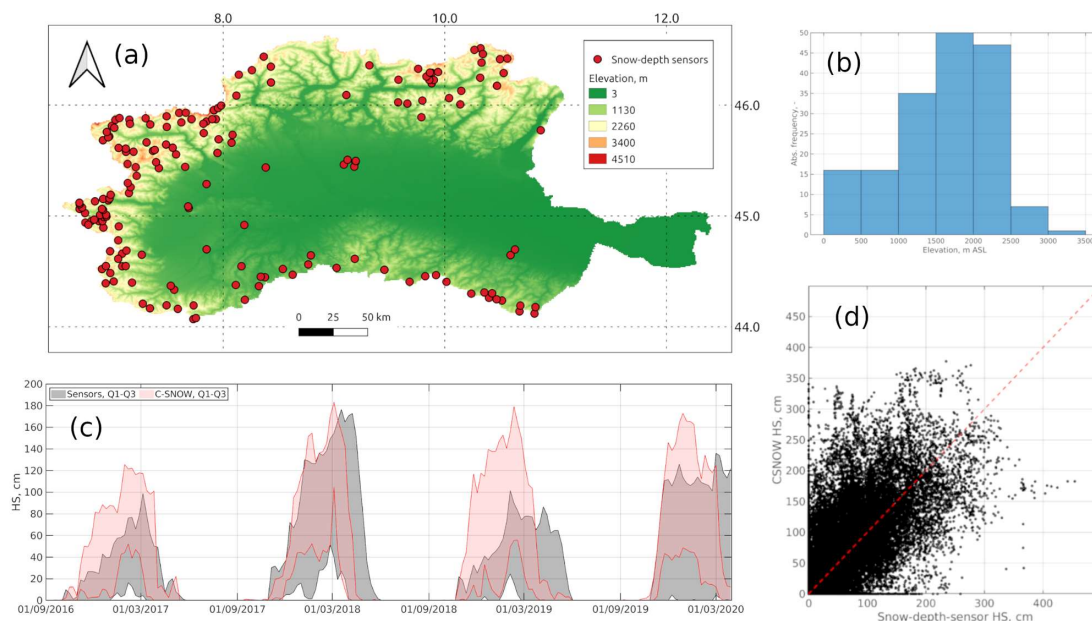



Figure 5: Evaluation of satellite based C-SNOW snow depth estimates. (a) Location of the 172 sensors across the Po river basin, and (b) their elevation distribution. (c) comparison between the interquartile range of C-SNOW and in situ measurements at the 172 sensors. (d) comparison between daily C-SNOW estimates and ground-based snow-depth measurements for all sites.

2.3.5 River discharge

River discharge time series from satellite remote sensing are estimated by integrating data from two sensors: altimeter and multispectral. Traditionally defined as the product of cross-sectional river flow area and velocity, river discharge is calculated by assuming that the satellite sensors measure the two quantities (Tarpanelli et al., 2015). Specifically, once the cross-sectional geometry is known, flow area is calculated as a function of the water height derived from satellite altimetry (Abdalla et al., 2021), while flow velocity, usually measured through in-situ instruments (current meter, acoustic doppler current profiler, velocimeter), is linked to the reflectance measured by the near infrared signal of the multispectral sensor (Tarpanelli et al., 2013), relying on the reflectance ratio between a dry (C) calibration pixel and the corresponding wet (M) measurement pixel.

Multi-mission satellite altimetry data coming from Saral/Altika, Cryosat-2 and Sentinel-3A and 3B are used to derive densified water level time series (Zakharova et al., 2020) at five stations along the main reach of the Po River named Piacenza, Cremona, Borgoforte, Sermide and Pontelagoscuro. At these stations, the multi-mission reflectance was extracted from the MODIS (Aqua and Terra), OLCI (Sentinel-3A) and MSI (Sentinel-2) sensors following the methods shown in

Tarpanelli et al. (2020). Here, river discharge (Q) is estimated as the product of flow velocity (Tarpanelli et al., 2020) and flow area, as a function of altimetry-derived water height (H) (Tarpanelli et al., 2015):

$$Q = \alpha(H)^{\beta}(C/M)^{\gamma} \qquad (3)$$

where the parameters α,β,γ were calibrated using observed discharges at the five stations. The resulting time series for each
220   station are illustrated in Figure 6 against the in situ observations recorded at the gauged stations. Performance metrics (Supplement material, Table S2) show skillful performance of the method in representing the observed daily discharges at the five stations, with average Nash-Sutcliffe (NS) of 0.81, KGE of 0.88 and relative RMSE (rRMSE) of 26%.

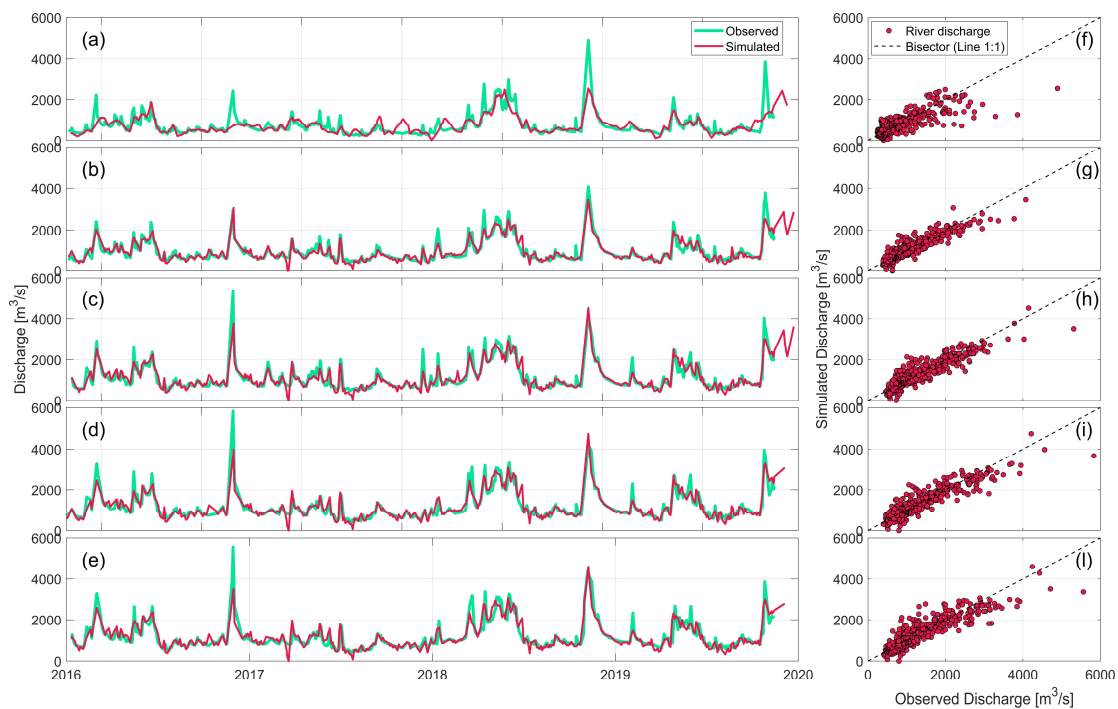



Figure 6: Comparison between discharges simulated by the multi-mission approach versus observations at five gauging stations in the Po
225   River in terms of time series (left column) and scatter plot (right column):(a, f) Piacenza, (b, g) Cremona, (c, h) Borgoforte, (d, i) Sermide and (e, l) Pontelagoscuro.

3 Methods

3.1 Hydrological modeling

Continuum (Silvestro et al., 2013) is a distributed hydrological model relying on a morphological approach based on the
230   identification of the drainage network components (Giannoni et al., 2000). It is a tradeoff between empirical and physically-based models, reproducing all main hydrological processes by relying on parameterization. The physical description of the

hydrological processes is comparatively simple, resulting in high computational efficiency yet generally skillful performance (Silvestro et al., 2013). Continuum reproduces the spatio-temporal evolution of runoff, soil moisture, energy fluxes, surface soil temperature, snow accumulation and melting, and evaporation. Deep flow and water table evolution are based on the

235 Darcy's equation, where each cell drains towards the neighboring cells following the 2D water table gradient and their hydraulic head, while a distributed interaction between water table and soil surface is represented through parameterization. A force-restore equation (Dickinson, 1988) is used to model the surface energy balance and enables the estimation of land surface temperature.

To simulate the cryospheric processes, we used S3M version 5 (Avanzi et al., 2021b), a one-layer snow model accounting

240 for precipitation-phase partitioning, snowpack accumulation and melt, snow rheology and hydraulics, as well as glacier melt (Terzago et al., 2020; Avanzi et al., 2021b). With its hybrid approach to snowmelt, which decouples the radiation- and temperature-driven contributions, S3M combines a parsimonious formulation with a substantial physical realism. For this work, S3M and Continuum were set up and coupled over the entire Po River basin (drainage area of 74,000 km^2), with a constant grid spacing of 1 km and time resolution of 1 hour.

245 **3.2 Model calibration**

To improve the representation of the hydrological states, Continuum was calibrated in the focus region using discharge data as benchmark. We deployed a multi-site calibration procedure that iteratively searches the model parameterization that best matches the available discharge observations over the calibration period at the 22 considered calibration stations. Hydrological simulations run for the model calibration cover the 2 years starting on 2018-01-01, while the calibration period

250 starts on 2018-07-01, leaving out the initial 6 months for model warm-up. The calibration tool perturbs six scalar parameters related to four physical hydrological features: infiltration velocity at saturation (cf), field capacity (ct), Curve Number (CN), and water sources (ws).

While the calibrated value of ws is a constant for the entire region of interest, for ct, cf and CN, the calibration consists in a rescaling of their default maps to the best value, thus preserving their spatial pattern, which depends on geographic spatial

255 datasets of soil characteristics and land cover. The cost function, based on the Kling-Gupta Efficiency (Gupta et al., 2009), computes an error between the duration curves at each percentile, weighted with the logarithm of the upstream area, to give higher weight to the downstream stations without neglecting the contribution of the most upstream ones.

The calibration procedure was performed through the implementation of a parallel search algorithm. The algorithm performs an iterative exploration of the 6-dimensional parameter space; the exploration starts with N=20 initial values sampled with a

260 Gaussian Latin Hypercube approach. For each of these N parameter sets, a hydrological simulation is performed over the calibration period, and the cost function is computed to map the error hypersurface. The point that minimizes J is used as the centre of the following iteration, until the algorithm converges to an optimal solution.

3.3 Data assimilation of satellite snow and soil moisture products

Satellite derived soil moisture from the Sentinel 1 RT1 product was assimilated into the Continuum model through a nudging

265   technique (Stauffer and Seaman, 1990; Lakshmivarahan and Lewis, 2013). The nudging scheme is a computationally inexpensive approach and is particularly suitable for applications in operational frameworks for flood predictions. The update is performed when the satellite data become available, on average once per day for soil moisture, following the equation:

$$X^+_{MOD}(t) = X^-_{MOD}(t) + G[X_{OBS}(t) - X^-_{MOD}(t)] \qquad (4)$$

270   where X^+_{MOD} represents the updated modelled variable , X_{MOD} is the modeled prior value, X_{OBS} is the observation, and G is the Kernel function. Thus, the correction term represents the difference between observed (X_{OBS}) and modelled variable multiplied by a gain (G) that takes into account the uncertainties of both the model and the satellite observations. In this application we used a constant value of G=0.45, following the work of Laiolo et al. (2016).

The assimilation of satellite-derived C-SNOW maps into S3M was performed using the same approach and assuming G=1 to

275   mimic direct insertion.

C-SNOW maps come as snow depths, while S3M supports assimilation in the form of snow water equivalent (SWE), which is a more suitable variable to assimilate to control the water balance. Thus, snow depths from C-SNOW were converted in SWE using simulated snow density values (see Avanzi et al., 2021b). Along with snow depth information, we rely on C-SNOW to determine snow-covered and snow-free areas, and then assimilated this information into S3M to clip modeled

280   snow cover according to the satellite information. More information on the theoretical background of SWE assimilation in S3M can be found in Avanzi et al. (2021b).

4 Results

4.1 Baseline run

The hydrological model Continuum was first calibrated using conventional meteorological data and observed discharges at

285   the 22 calibration stations described above. The calibrated setup was then run over the years 2016-2019 to produce a baseline simulation for 2017-2019, leaving out 2016 as model warm-up. A comparison of simulated versus observed hourly river discharges is shown in Figure 7 for five sample stations, while six performance metrics are shown for all 27 discharge stations in Figure 8 and in Table S1 (see Supplement material). Dimensionless scores, including KGE and its three decomposition terms, i.e., correlation (r), bias rate, and Coefficient of variation rate (CV rate), increase on average with the

290   upstream area. Note that all four scores have optimum value at 1. The mean KGE over all the stations KGE=0.51, rises to 0.63 and 0.70 for basins larger than 1,000 km^2 and 10,000 km^2, respectively. Similar trends versus the same classes of upstream area are found in the mean correlation (0.75, 0.86, 0.88), while bias rate (0.98, 0.99, 0.94) and CV rate (0.89, 0.94, 1.13) are slightly deteriorated for basins larger than 10,000 km^2.

Differences in the mean KGE, r and CV rate between validation and calibration stations are not statistically significant in a

295 two-sample t-test for the mean. Only the mean of the bias rate of the two samples is statistically different at 5% significance level, with validation stations having an average 30% negative bias in comparison to an average 5% positive bias of the calibration stations.

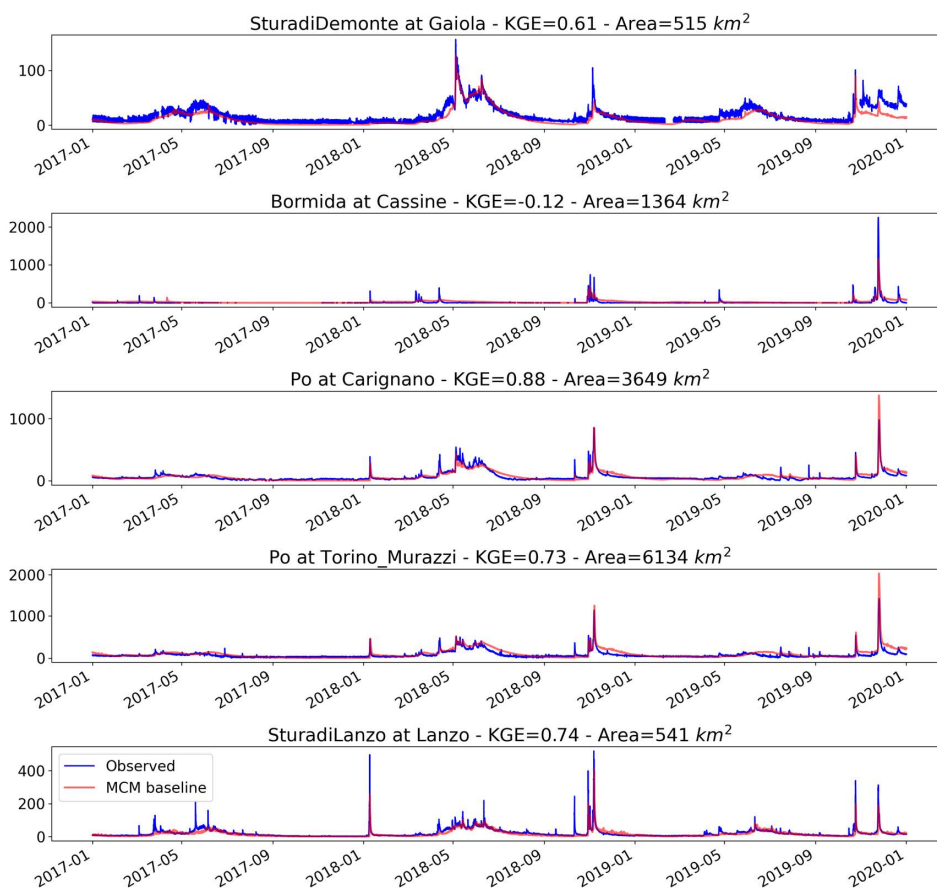



Figure 7: Observed versus simulated (baseline) discharge [m³/s] for the years 2017-2019 at five river gauging stations.

300

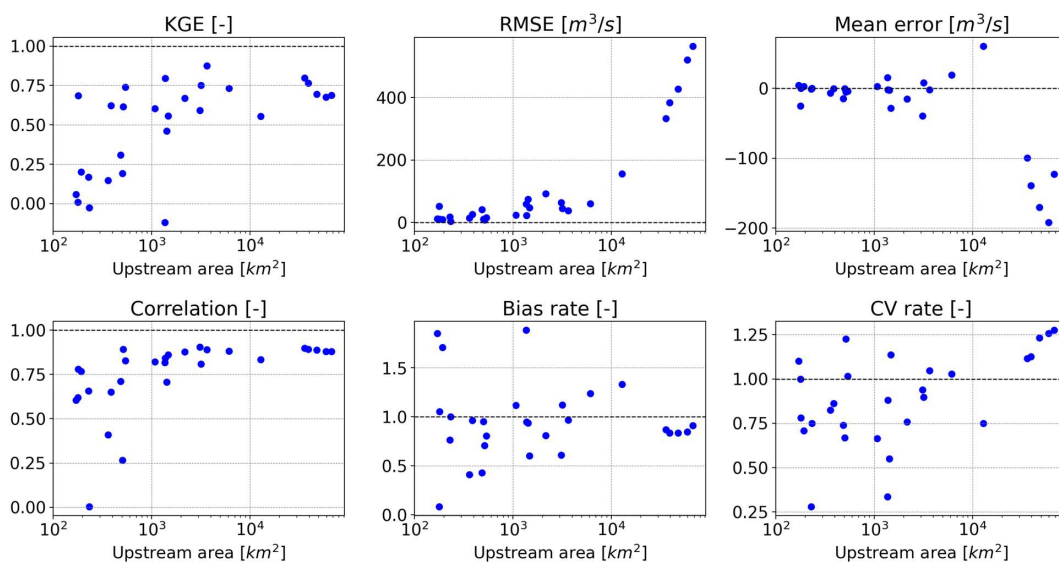



Figure 8: Skills of the baseline run versus upstream area at the 27 measurement stations. Dashed lines indicate the optimum value of each score.

4.2 Model runs with satellite input

In a second phase, we performed four hydrological simulations, each of them based on the configuration and input data of the baseline run and by replacing in turn one input dataset with one of four satellite products described in Sect. 2.3: 1) precipitation from SM2RAIN, 2) evaporation from GLEAM, data assimilation of 3) soil moisture from RT1, and of 4) snow depths from C-SNOW. Two additional configurations were run including multiple satellite-based data sources: 5) all four satellite Earth observation datasets, hereafter referred to as EO, and 6) a combination of the satellite precipitation and evaporation, referred to as SM2RAIN+GLEAM. The spatial distribution of the performance of the six model simulations at the 27 river gauges is shown as maps of KGE (Figure 9) and its three decomposition terms (see Supplement material). Further, boxplots of the KGE of the six experiments and comparison with the baseline run are shown in Figure 10.

Results denote a generally skillful reconstruction of river discharges for all experiments, with mean KGE at the 27 stations ranging between 0.13 (SM2RAIN+GLEAM) and 0.53 (C-SNOW), all well above the no-skill threshold of $KGE_0 = 1-2^{1/2} \cong -0.41$ (see Knoben et al., 2019). Simulations including C-SNOW and GLEAM perform on average better than the baseline run, with mean improvements in KGE of 0.02 and 0.01 (+4% and +2%), respectively. Largest differences in the overall performances are due to the wide range of the mean bias across the six simulations, with the largest bias rates for SM2RAIN+GLEAM (1.58) and EO (0.69), and the lowest bias rate for GLEAM (1.02) and C-SNOW (0.97), both improving that of the baseline run (0.95). On the other hand, average correlations across the six experiments fall in a much narrower

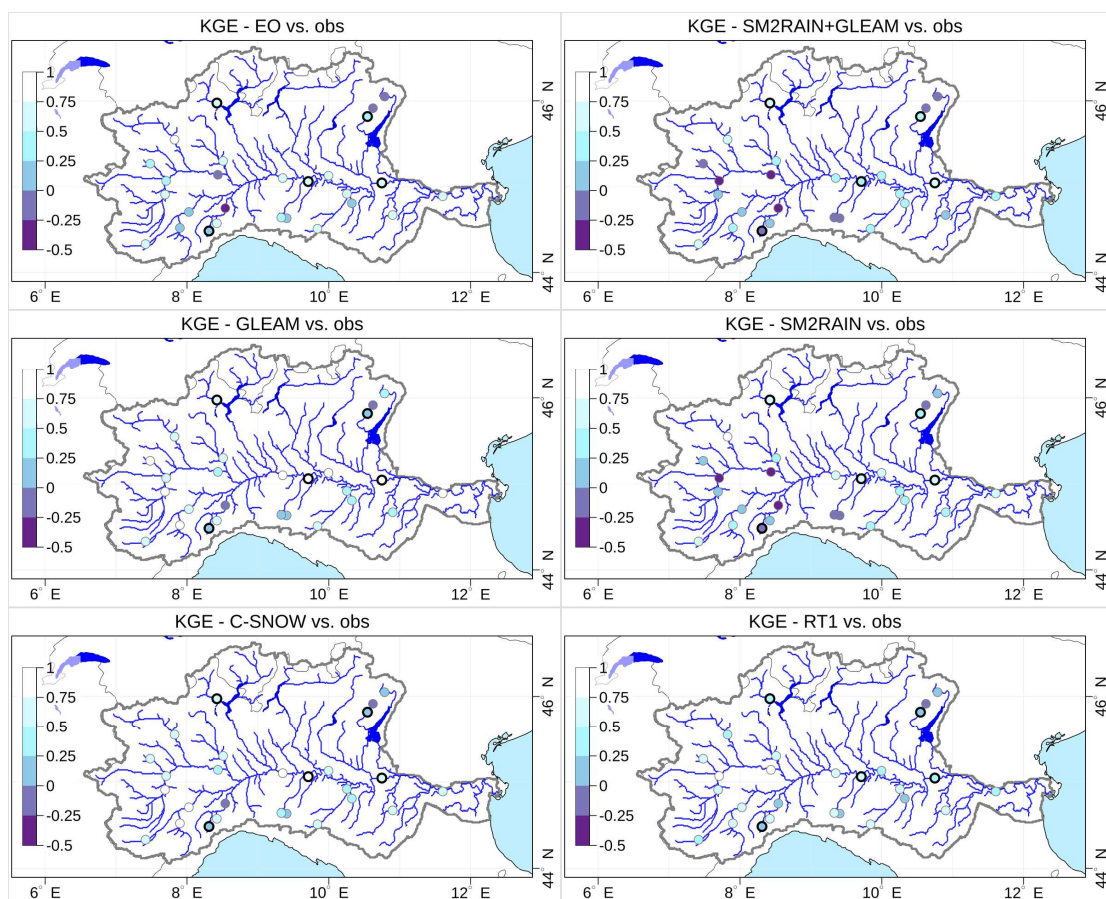
interval, ranging between 0.61 for EO and 0.75 for both C-SNOW and the baseline run. Running the model with all Earth observations produces on average a 28% deterioration of the mean performance (KGE=0.37), though it surprisingly generates the best performance at the five validation stations (KGE=0.54) among all simulations (Figure 10).

325

Figure 9: Spatial distribution of the Kling-Gupta Efficiency (KGE) of the six model runs driven by the four input satellite products versus observed discharges at the measurement stations. Validation stations are marked with a bold circle. Multi-product experiments are in the first row, while single-product experiments in row 2 (forcing input) and 3 (data assimilation input).

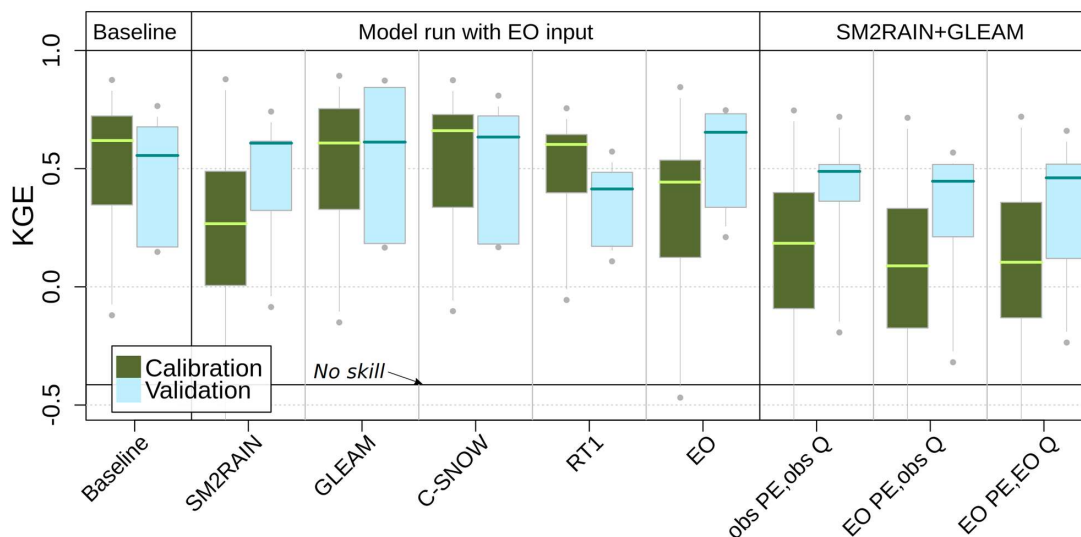



Figure 10: Box-plots comparing the Kling-Gupta Efficiency (KGE) of simulated river discharges for all the considered experiments versus observations at the calibration and at the validation stations. The no-skill line at $1-2^{1/2}$ is indicated with a solid horizontal line.

The six simulations forced by satellite products were then compared to the baseline run, to detect similarities and deviations in the entire simulation domain, including where no observations are available. To reduce the correlation effects along the river network, we consider only one value per simulated river reach, located just upstream each confluence. Using RT1 and GLEAM does not result in significant spatial differences with respect to the baseline (Figure 11). As expected, the use of C-SNOW results in differences mainly in alpine areas, especially in Ticino (Switzerland), where the MCM dataset used in the baseline run is known to underestimate precipitation rates, due to the lack of ground measurements outside the Italian territory. Larger deviations are visible in the runs including SM2RAIN, particularly in the upper Po basin in the west and in the upper Adda River in the north, confirming the stronger sensitivity of river discharge to precipitation dynamics.

Figure 12 shows a comparison of the six simulations forced by satellite products, the baseline run, and observed discharges at two validation stations, for a series of moderate to high intensity events which hit a large portion of the Po River basin in Fall 2019. The second of the three main events, in the second half of November, caused the exceedance of the maximum alert level and widespread flooding in several river sections in the main reach of the Po River across the Lombardia and Emilia Romagna regions, including the area of Piacenza (Figure 12, bottom). In Piacenza, all model simulations performed reasonably well, with maximum error on the peak discharge below 20%. The best performances over the three months are found in the baseline run and in the two runs with data assimilation (RT1 and C-SNOW), all three with KGE=0.89. Lower performances are produced by the three runs forced by SM2RAIN, mainly due to an overestimation of the first event in late October 2019. At the Candoglia station, results show an opposite pattern, with best performance by SM2RAIN and

SM2RAIN+GLEAM both with KGE=0.74 over the three months, mildly improving upon the performance of the baseline

350   run (KGE=0.71).

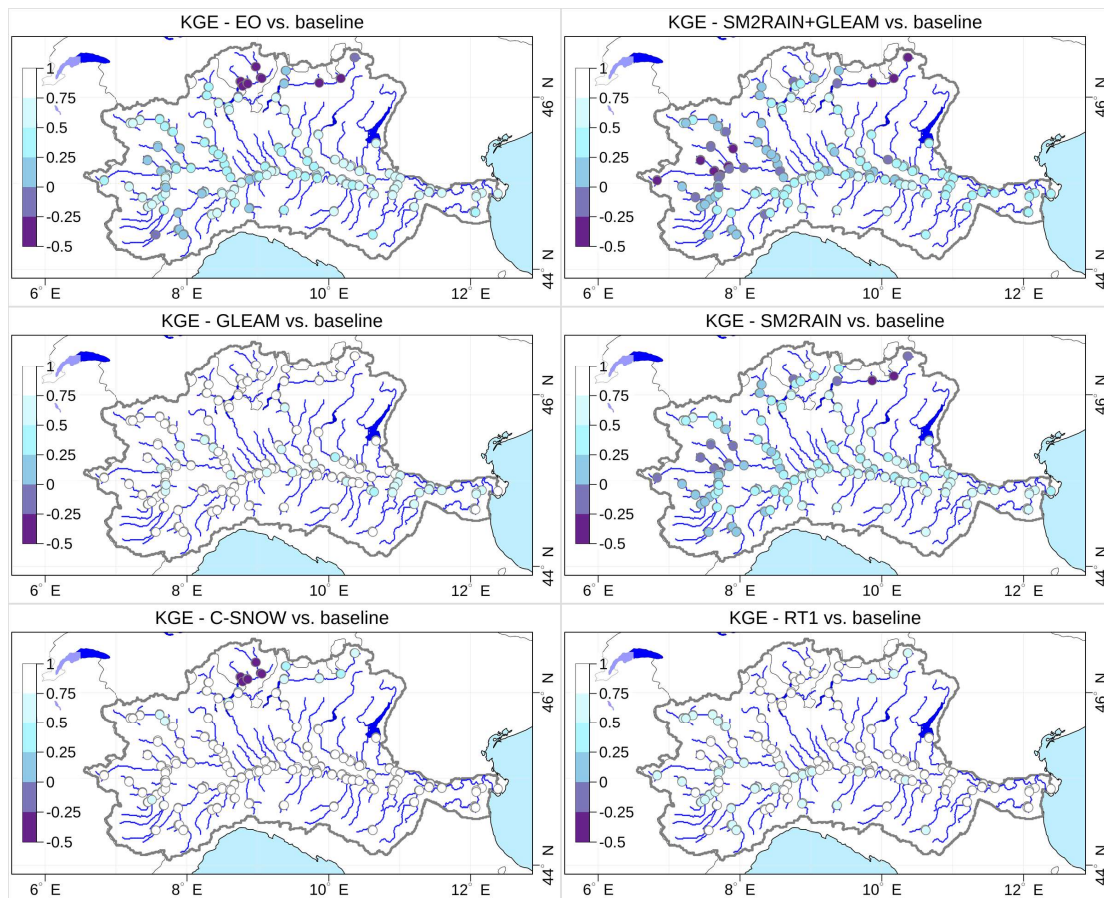



Figure 11: Like Figure 9, but using the baseline run as reference discharge. Results are shown for each modeled river reach.

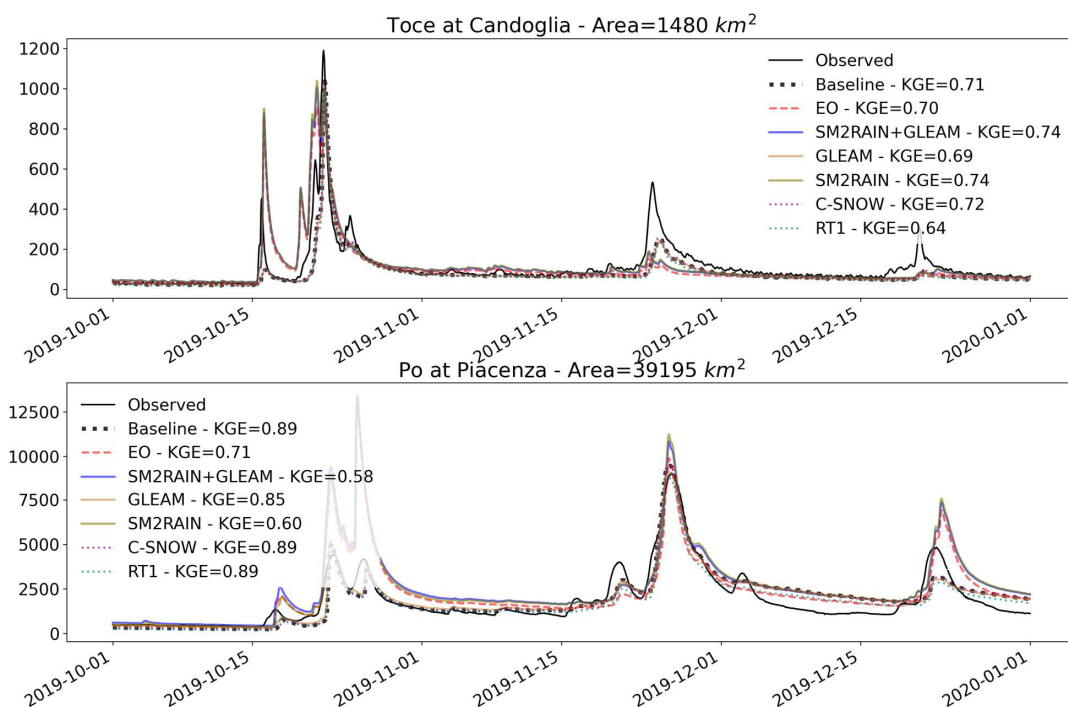



Figure 12: Comparison of observed and simulated hydrographs [m³/s] for the events of October-December 2019 at two validation stations: Candoglia (top) and Piacenza (bottom), together with KGE calculated versus the observed discharges for the same three months.

4.3 Sensitivity of satellite data to three model parameterizations

A subsequent experiment investigated the performance of the hydrological model in reproducing discharges at the 27 river gauges, by forcing it with the satellite datasets SM2RAIN and GLEAM. In details, we compared the results of three model runs over 2017-2019, using three different model parameterizations obtained through dedicated calibrations (over 2018-2019), derived by applying the steps described in Sect. 3.2 to different configurations of input and benchmark discharges:

1. The first is the simulation SM2RAIN+GLEAM described in Sect. 4.2, i.e., run with the model parameters obtained by calibrating with conventional ground observations (interpolated measurements and MCM precipitation) and optimizing the objective function using observed discharge at the 22 calibration stations as benchmark (obs PE, obs Q in Figure 10).

2. The simulation SM2RAIN+GLEAM run on a model calibration forced by the same satellite datasets SM2RAIN and GLEAM as input and optimizing the objective function using observed discharge at the 22 calibration stations as benchmark (EO PE, obs Q in Figure 10).

3. The simulation SM2RAIN+GLEAM run on a model calibration forced by the satellite datasets SM2RAIN and
370 GLEAM as input and optimizing the objective function using satellite-derived discharge estimates at the 5 virtual
stations (see Sect. 2.3.5) as benchmark (EO PE, EO Q in Figure 10).

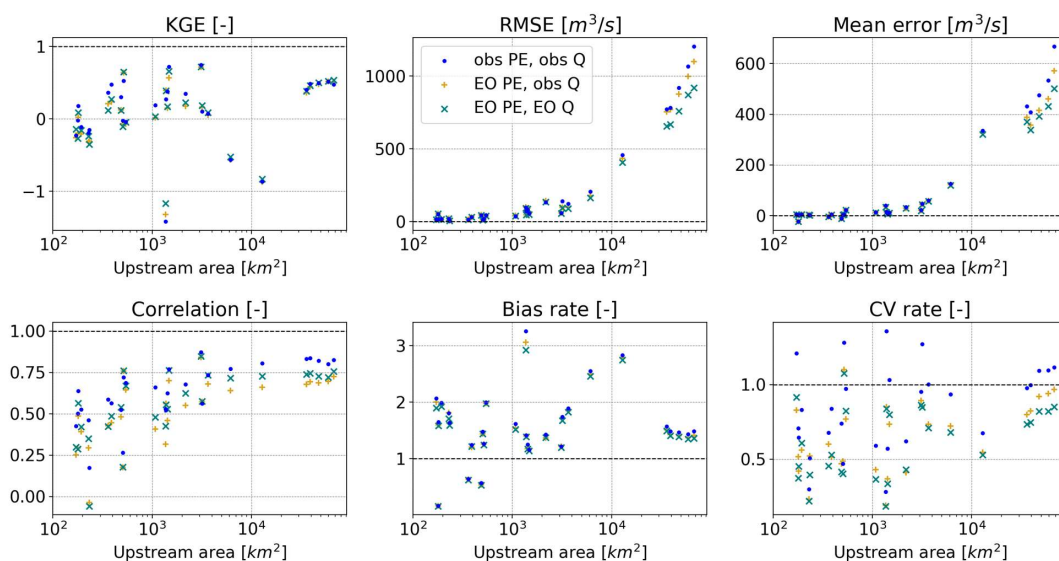
Surprisingly, simulation #2 forced by the same SM2RAIN and GLEAM used in the calibration shows the lowest
performance among the three (mean KGE=0.07 over all 27 stations). Simulation #3 (EO PE, EO Q) gives satisfactory
performance (mean KGE=0.10), relatively close to #1 (mean KGE=0.13), despite relying largely on satellite data.
375 Interestingly, the five validation stations on average outperform the set of calibration stations, with average KGE of 0.38,
0.30 and 0.29 for the three experiments. Performance of the three model runs versus the upstream area at the 27 stations
(Figure 13) shows a general improvement in the correlation with the upstream area, while for the other metrics trends are
less clear. Simulation #3 shows reduced variability (CV rate) yet smaller absolute errors (RMSE and mean error in Figure
13), also thanks to a calibration focused on the downstream virtual stations.

380

Figure 13: Skills of the run forced by satellite precipitation and evaporation (PE) versus upstream area at the 27 measurement stations. The three markers denote three calibrated parameter sets, obtained with different configurations of PE input and of benchmark discharge (Q). Conventional observational datasets are indicated with "obs", while "EO" are the satellite-derived datasets. Dashed lines indicate the optimum value of each score.

385 **5 Discussions**

A critical evaluation of the results of the experiments performed can help identify strengths and weaknesses, as well as
directions to take to maximize the benefits of satellite observations in Earth system modeling. Overall, hydrological

simulations driven by satellite datasets produced encouraging results, with 95% of the KGE of the station-experiment combinations above the no-skill threshold (versus 100% for the baseline run). The remaining 5% of combinations with KGE

390   below the no-skill threshold occur in just 3 stations out of 27 and only in model configurations including SM2RAIN. Generally, the precipitation dataset is found to have the largest weight on the resulting model performance, with standard deviation of changes in KGE versus the baseline simulation $SD_{\Delta KGE,SM2RAIN}=0.37$ being more than twice that of all the other satellite driven configurations ($SD_{\Delta KGE,RT1}=0.16$, $SD_{\Delta KGE,GLEAM}=0.09$, $SD_{\Delta KGE,C\text{-}SNOW}=0.06$). In other words, the simulation performance shows strongest sensitivity to the precipitation forcing, which in fact leads to the largest deteriorations

395   compared to the baseline run, as well as some of the largest improvements in KGE, up to $\Delta KGE_{MAX,SM2RAIN}=0.29$, well above all the improvements produced by GLEAM ($\Delta KGE_{MAX,GLEAM}=0.17$) and C-SNOW ($\Delta KGE_{MAX,C\text{-}SNOW}=0.12$) at any single station. This result is largely in agreement with previous findings (e.g., Jones et al., 2006; Sperna Weiland et al., 2015) and highlights the importance of advances in satellite precipitation estimation for hydrological applications. Qi et al. (2016) showed that model performance can also be impacted by model-precipitation product interactions, though this can partly be

400   mitigated by dedicated model calibrations for each combination of input products. The high resolution version of SM2RAIN used in this work leads to comparable hydrological performance to that of the best non-gauge-corrected satellite products found in the literature (Camici et al., 2018; Amorim et al., 2020), and local results are better than those obtained with previous coarser resolution versions (see e.g., Beck et al., 2017; Tang et al., 2020). These works also show that satellite precipitation datasets bias-corrected with ground observations further improve the overall quality, including the performance

405   in hydrological modelling.

With regard to the precipitation forcing, one must also note that the MCM dataset used in the baseline represents a particularly difficult benchmark to overcome. The high station density and the merging with the Italian radar composite make MCM a high-quality and detailed product both spatially and temporally. Yet, only few world areas can rely on seamless and nearly unbiased gauge-radar products, while satellite datasets remain prime candidates in ungauged regions,

410   especially for real time applications, thanks to key features such as extended coverage, high resolution, short latency, and spatial consistency. In addition, satellite datasets are unaffected by country borders, which make them suitable for applications in transboundary river basins, especially in countries where data sharing agreements are not easily implemented. In contrast, GLEAM and C-SNOW consistently produced moderate improvements, though on a larger number of river sections, with only a minority of stations where skills deteriorated in comparison to the baseline run. Finally, the assimilation

415   of RT1 soil moisture shows contrasting behaviour. On the one hand, it deteriorated KGE values throughout most of the stations in the main reach of the Po river, due to a general increasing negative bias. On the other hand, it shows general benefits in small-size upstream catchments and notably the best improvement in KGE ($\Delta KGE_{MAX,RT1}=0.41$) among all 216 station-experiment combinations, for the Trebbia river at Valsigiara.

A final comment goes to the surprisingly high skills of hydrological simulations at the 5 validation stations, which on

420   average exceed those at the calibration stations in 5 out of 9 experiments (see Figure 10). Validation stations were chosen to represent different areas of the Po basin, including a mix of small and large sub-catchments with varying influence of lakes

and reservoirs. The multi-site calibration strategy is designed to find an optimal parameter set for the entire domain, thus reducing the effect of highly variable model performance typical of cascading calibrations (e.g., Alfieri et al., 2020). All results are then compared at the calibration and validation stations for the same period 2017-2019, which is twice the

425 duration of the calibration period, implicitly adding a validation component also at the calibration stations. Higher performance at the validation stations seem to be particularly evident in simulations forced by SM2RAIN, though a connection between these facts is not known and it may simply be related to spatial differences in the skills of the satellite precipitation forcing in the sub-catchments where validation stations are located. A noteworthy case is that of the validation station of the Toce River at Candoglia, in the north-western part of the Po basin. It is influenced by a large number of

430 reservoirs upstream and the Lake Maggiore located just downstream hugely smoothens its runoff characteristics from the rest of the river network. This makes the sub basin almost disconnected by the rest of the Po basin. Notwithstanding, simulation performance at Candoglia are higher than those of the calibration stations in all experiments but one (RT1), with the case of SM2RAIN scoring a KGE=0.74, hence 0.22 points higher than the average calibration KGE among all stations.

6 Conclusions

435 This research explored the impact of five high resolution satellite products in distributed hydrological modelling. In a set of experiments we tested the use of satellite precipitation and evaporation as forcing input, data assimilation of satellite soil moisture and snow depth, and satellite river discharge estimates as benchmark for model calibration. We found skillful performance for all simulations including satellite derived products, with GLEAM evaporation and C-SNOW snow depth yielding an average 2% and 4% improvements over a baseline run driven by high-quality ground-based datasets. The skills

440 of model runs including earth observation data showed considerable variability in space and time. In addition, we found skillful results in a model calibration heavily relying on satellite products, both with regard to forcing input and to benchmark discharge. This heralds the use of hydrological models fully relying on satellite data as an appealing solution for large scale applications and for regions where ground-based observations are not available, particularly in near-real time.

445 *Code and data availability.* Satellite products developed for this work can be requested from the authors. S3M and Continuum are open source models and their code is available at https://github.com/c-hydro.

Author contributions. LA, LB, SG conceptualized the work; FA, FD, CM, LB, AT, DR, DGM, RQ, MV prepared and validated the satellite products; FA, FD, LC, AL developed model code for the analysis; LA, GB and AL prepared the model input data; LA performed the formal analysis, data visualization and wrote the draft article with the contribution of all the

450 authors.

Competing interests. The authors declare no conflict of interest.

Acknowledgements. The research leading to these results has received funding from the European Space Agency (ESA) contract n. 4000129870/20/I-NB (CCN. N.1) "Digital Twin Earth (DTE) - Hydrology". Observed snow depths were provided by the Italian Civil Protection Agency and the administrative regions of Valle d'Aosta, Piemonte, Lombardia,

455   Liguria, Veneto, and Emilia Romagna. DGM acknowledges support from the European Research Council (ERC, grant no. 715254 DRY–2–DRY).

References

Abdalla, S., Abdeh Kolahchi, A., Ablain, M., Adusumilli, S., Aich Bhowmick, S., Alou-Font, E., Amarouche, L., Andersen, O. B., Antich, H., Aouf, L., Arbic, B., Armitage, T., Arnault, S., Artana, C., Aulicino, G., Ayoub, N., Badulin, S., Baker, S.,

460   Banks, C., Bao, L., Barbetta, S., Barceló-Llull, B., Barlier, F., Basu, S., Bauer-Gottwein, P., Becker, M., Beckley, B., Bellefond, N., Belonenko, T., Benkiran, M., Benkouider, T., Bennartz, R., Benveniste, J., Bercher, N., Berge-Nguyen, M., Bettencourt, J., Blarel, F., Blazquez, A., Blumstein, D., Bonnefond, P., Borde, F., Bouffard, J., Boy, F., Boy, J.-P., Brachet, C., Brasseur, P., Braun, A., Brocca, L., Brockley, D., Brodeau, L., Brown, S., Bruinsma, S., Bulczak, A., Buzzard, S., Cahill, M., Calmant, S., Calzas, M., Camici, S., Cancet, M., Capdeville, H., Carabajal, C. C., Carrere, L., Cazenave, A., Chassignet,

465   E. P., Chauhan, P., Cherchali, S., Chereskin, T., Cheymol, C., Ciani, D., Cipollini, P., Cirillo, F., Cosme, E., Coss, S., Cotroneo, Y., Cotton, D., Couhert, A., Coutin-Faye, S., Crétaux, J.-F., Cyr, F., d'Ovidio, F., Darrozes, J., David, C., Dayoub, N., De Staerke, D., Deng, X., Desai, S., Desjonqueres, J.-D., Dettmering, D., Di Bella, A., Díaz-Barroso, L., Dibarboure, G., Dieng, H. B., Dinardo, S., Dobslaw, H., Dodet, G., Doglioli, A., Domeneghetti, A., Donahue, D., Dong, S., et al.: Altimetry for the future: Building on 25 years of progress, Advances in Space Research, 68, 319–363,

470   https://doi.org/10.1016/j.asr.2021.01.022, 2021.

Alfieri, L., Cohen, S., Galantowicz, J., Schumann, G. J.-P., Trigg, M. A., Zsoter, E., Prudhomme, C., Kruczkiewicz, A., Coughlan de Perez, E., Flamig, Z., Rudari, R., Wu, H., Adler, R. F., Brakenridge, R. G., Kettner, A., Weerts, A., Matgen, P., Islam, S. A. K. M., de Groeve, T., and Salamon, P.: A global network for operational flood risk reduction, Environmental Science & Policy, 84, 149–158, https://doi.org/10.1016/j.envsci.2018.03.014, 2018.

475   Alfieri, L., Lorini, V., Hirpa, F. A., Harrigan, S., Zsoter, E., Prudhomme, C., and Salamon, P.: A global streamflow reanalysis for 1980–2018, Journal of Hydrology X, 6, 100049, https://doi.org/10.1016/j.hydroa.2019.100049, 2020.

Amorim, J. da S., Viola, M. R., Junqueira, R., Oliveira, V. A. de, and Mello, C. R. de: Evaluation of Satellite Precipitation Products for Hydrological Modeling in the Brazilian Cerrado Biome, 12, 2571, https://doi.org/10.3390/w12092571, 2020.

Avanzi, F., Ercolani, G., Gabellani, S., Cremonese, E., Pogliotti, P., Filippa, G., Morra di Cella, U., Ratto, S., Stevenin, H.,

480   Cauduro, M., and Juglair, S.: Learning about precipitation lapse rates from snow course data improves water balance modeling, 25, 2109–2131, https://doi.org/10.5194/hess-25-2109-2021, 2021a.

Avanzi, F., Gabellani, S., Delogu, F., Silvestro, F., Cremonese, E., Morra di Cella, U., Ratto, S., and Stevenin, H.: S3M 5.1: a distributed cryospheric model with dry and wet snow, data assimilation, glacier mass balance, and debris-driven melt, 1–50, https://doi.org/10.5194/gmd-2021-92, 2021b.

485   Bauer-Marschallinger, B., Freeman, V., Cao, S., Paulik, C., Schaufler, S., Stachl, T., Modanesi, S., Massari, C., Ciabatta, L., Brocca, L., and Wagner, W.: Toward Global Soil Moisture Monitoring With Sentinel-1: Harnessing Assets and Overcoming Obstacles, 57, 520–539, https://doi.org/10.1109/TGRS.2018.2858004, 2019.

Beck, H. E., Vergopolan, N., Pan, M., Levizzani, V., Van Dijk, A. I. J. M., Weedon, G. P., Brocca, L., Pappenberger, F., Huffman, G. J., and Wood, E. F.: Global-scale evaluation of 22 precipitation datasets using gauge observations and
490   hydrological modeling, 21, 6201–6217, https://doi.org/10.5194/hess-21-6201-2017, 2017.

Bordoni, M., Corradini, B., Lucchelli, L., Valentino, R., Bittelli, M., Vivaldi, V., and Meisina, C.: Empirical and Physically Based Thresholds for the Occurrence of Shallow Landslides in a Prone Area of Northern Italian Apennines, 11, 2653, https://doi.org/10.3390/w11122653, 2019.

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

Bruno, G., Pignone, F., Silvestro, F., Gabellani, S., Schiavi, F., Rebora, N., Giordano, P., and Falzacappa, M.: Performing Hydrological Monitoring at a National Scale by Exploiting Rain-Gauge and Radar Networks: The Italian Case, 12, 771, https://doi.org/10.3390/atmos12060771, 2021.

500   Camici, S., Ciabatta, L., Massari, C., and Brocca, L.: How reliable are satellite precipitation estimates for driving hydrological models: A verification study over the Mediterranean area, Journal of Hydrology, 563, 950–961, https://doi.org/10.1016/j.jhydrol.2018.06.067, 2018.

Chen, L. and Wang, L.: Recent advance in earth observation big data for hydrology, 2, 86–107, https://doi.org/10.1080/20964471.2018.1435072, 2018.

505   Dembélé, M., Ceperley, N., Zwart, S. J., Salvadore, E., Mariethoz, G., and Schaefli, B.: Potential of satellite and reanalysis evaporation datasets for hydrological modelling under various model calibration strategies, Advances in Water Resources, 143, 103667, https://doi.org/10.1016/j.advwatres.2020.103667, 2020a.

Dembélé, M., Schaefli, B., van de Giesen, N., and Mariéthoz, G.: Suitability of 17 gridded rainfall and temperature datasets for large-scale hydrological modelling in West Africa, 24, 5379–5406, https://doi.org/10.5194/hess-24-5379-2020, 2020b.

510   Dhote, P. R., Thakur, P. K., Domeneghetti, A., Chouksey, A., Garg, V., Aggarwal, S. P., and Chauhan, P.: The use of SARAL/AltiKa altimeter measurements for multi-site hydrodynamic model validation and rating curves estimation: An application to Brahmaputra River, Advances in Space Research, 68, 691–702, https://doi.org/10.1016/j.asr.2020.05.012, 2021.

Dickinson, R. E.: The Force-Restore Model for Surface Temperatures and Its Generalizations, 1, 1086–1097, 1988.

515 Domeneghetti, A., Molari, G., Tourian, M. J., Tarpanelli, A., Behnia, S., Moramarco, T., Sneeuw, N., and Brath, A.: Testing the use of single- and multi-mission satellite altimetry for the calibration of hydraulic models, Advances in Water Resources, 151, 103887, https://doi.org/10.1016/j.advwatres.2021.103887, 2021.

Dorigo, W., Wagner, W., Albergel, C., Albrecht, F., Balsamo, G., Brocca, L., Chung, D., Ertl, M., Forkel, M., Gruber, A., Haas, E., Hamer, P. D., Hirschi, M., Ikonen, J., de Jeu, R., Kidd, R., Lahoz, W., Liu, Y. Y., Miralles, D., Mistelbauer, T.,
520 Nicolai-Shaw, N., Parinussa, R., Pratola, C., Reimer, C., van der Schalie, R., Seneviratne, S. I., Smolander, T., and Lecomte, P.: ESA CCI Soil Moisture for improved Earth system understanding: State-of-the art and future directions, Remote Sensing of Environment, 203, 185–215, https://doi.org/10.1016/j.rse.2017.07.001, 2017.

ESA: Land Cover CCI Product User Guide Version 2. Tech. Rep., 2017.

Faroux, S., Kaptué Tchuenté, A. T., Roujean, J.-L., Masson, V., Martin, E., and Le Moigne, P.: ECOCLIMAP-II/Europe: a
525 twofold database of ecosystems and surface parameters at 1 km resolution based on satellite information for use in land surface, meteorological and climate models, 6, 563–582, https://doi.org/10.5194/gmd-6-563-2013, 2013.

Getirana, A. C. V., Boone, A., Yamazaki, D., and Mognard, N.: Automatic parameterization of a flow routing scheme driven by radar altimetry data: Evaluation in the Amazon basin, 49, 614–629, https://doi.org/10.1002/wrcr.20077, 2013.

Giannoni, F., Roth, G., and Rudari, R.: A semi-distributed rainfall-runoff model based on a geomorphologic approach, 25,
530 665–671, https://doi.org/10.1016/s1464-1909(00)00082-4, 2000.

Gruber, A., Dorigo, W. A., Crow, W., and Wagner, W.: Triple Collocation-Based Merging of Satellite Soil Moisture Retrievals, 55, 6780–6792, https://doi.org/10.1109/TGRS.2017.2734070, 2017.

Gupta, H. V., Kling, H., Yilmaz, K. K., and Martinez, G. F.: Decomposition of the mean squared error and NSE performance criteria: Implications for improving hydrological modelling, 377, 80–91, https://doi.org/10.1016/j.jhydrol.2009.08.003, 2009.

535 Hartanto, I. M., van der Kwast, J., Alexandridis, T. K., Almeida, W., Song, Y., van Andel, S. J., and Solomatine, D. P.: Data assimilation of satellite-based actual evapotranspiration in a distributed hydrological model of a controlled water system, International Journal of Applied Earth Observation and Geoinformation, 57, 123–135, https://doi.org/10.1016/j.jag.2016.12.015, 2017.

Hengl, T., Jesus, J. M. de, Heuvelink, G. B. M., Gonzalez, M. R., Kilibarda, M., Blagotić, A., Shangguan, W., Wright, M.
540 N., Geng, X., Bauer-Marschallinger, B., Guevara, M. A., Vargas, R., MacMillan, R. A., Batjes, N. H., Leenaars, J. G. B., Ribeiro, E., Wheeler, I., Mantel, S., and Kempen, B.: SoilGrids250m: Global gridded soil information based on machine learning, PLOS ONE, 12, e0169748, https://doi.org/10.1371/journal.pone.0169748, 2017.

Hersbach, H., Bell, B., Berrisford, P., Hirahara, S., Horányi, A., Muñoz-Sabater, J., Nicolas, J., Peubey, C., Radu, R., Schepers, D., Simmons, A., Soci, C., Abdalla, S., Abellan, X., Balsamo, G., Bechtold, P., Biavati, G., Bidlot, J., Bonavita,

545   M., De Chiara, G., Dahlgren, P., Dee, D., Diamantakis, M., Dragani, R., Flemming, J., Forbes, R., Fuentes, M., Geer, A., Haimberger, L., Healy, S., Hogan, R. J., Hólm, E., Janisková, M., Keeley, S., Laloyaux, P., Lopez, P., Lupu, C., Radnoti, G., de Rosnay, P., Rozum, I., Vamborg, F., Villaume, S., and Thépaut, J.-N.: The ERA5 global reanalysis, 146, 1999–2049, https://doi.org/10.1002/qj.3803, 2020.

Hostache, R., Chini, M., Giustarini, L., Neal, J., Kavetski, D., Wood, M., Corato, G., Pelich, R.-M., and Matgen, P.: Near-
550   Real-Time Assimilation of SAR-Derived Flood Maps for Improving Flood Forecasts, 54, 5516–5535, https://doi.org/10.1029/2017WR022205, 2018.

Huffman, G. J., Bolvin, D. T., Braithwaite, D., Hsu, K., Joyce, R., Xie, P., and Yoo, S.-H.: NASA global precipitation measurement (GPM) integrated multi-satellite retrievals for GPM (IMERG), 4, 26, 2015.

Ishitsuka, Y., Gleason, C. J., Hagemann, M. W., Beighley, E., Allen, G. H., Feng, D., Lin, P., Pan, M., Andreadis, K., and
555   Pavelsky, T. M.: Combining Optical Remote Sensing, McFLI Discharge Estimation, Global Hydrologic Modeling, and Data Assimilation to Improve Daily Discharge Estimates Across an Entire Large Watershed, 57, https://doi.org/10.1029/2020WR027794, 2021.

Jones, R. N., Chiew, F. H. S., Boughton, W. C., and Zhang, L.: Estimating the sensitivity of mean annual runoff to climate change using selected hydrological models, Advances in Water Resources, 29, 1419–1429,
560   https://doi.org/10.1016/j.advwatres.2005.11.001, 2006.

Knoben, W. J. M., Freer, J. E., and Woods, R. A.: Technical note: Inherent benchmark or not? Comparing Nash-Sutcliffe and Kling-Gupta efficiency scores, 1–7, https://doi.org/10.5194/hess-2019-327, 2019.

Laiolo, P., Gabellani, S., Campo, L., Silvestro, F., Delogu, F., Rudari, R., Pulvirenti, L., Boni, G., Fascetti, F., Pierdicca, N., Crapolicchio, R., Hasenauer, S., and Puca, S.: Impact of different satellite soil moisture products on the predictions of a
565   continuous distributed hydrological model, International Journal of Applied Earth Observation and Geoinformation, 48, 131–145, https://doi.org/10.1016/j.jag.2015.06.002, 2016.

Lakshmivarahan, S. and Lewis, J. M.: Nudging Methods: A Critical Overview, in: Data Assimilation for Atmospheric, Oceanic and Hydrologic Applications (Vol. II), edited by: Park, S. K. and Xu, L., Springer, Berlin, Heidelberg, 27–57, https://doi.org/10.1007/978-3-642-35088-7_2, 2013.

570   Lehner, B., Liermann, C. R., Revenga, C., Vörösmarty, C., Fekete, B., Crouzet, P., Döll, P., Endejan, M., Frenken, K., Magome, J., Nilsson, C., Robertson, J. C., Rödel, R., Sindorf, N., and Wisser, D.: High-resolution mapping of the world's reservoirs and dams for sustainable river-flow management, 9, 494–502, https://doi.org/10.1890/100125, 2011.

Li, B., Rodell, M., Zaitchik, B. F., Reichle, R. H., Koster, R. D., and van Dam, T. M.: Assimilation of GRACE terrestrial water storage into a land surface model: Evaluation and potential value for drought monitoring in western and central
575   Europe, Journal of Hydrology, 446–447, 103–115, https://doi.org/10.1016/j.jhydrol.2012.04.035, 2012.

Lievens, H., Demuzere, M., Marshall, H.-P., Reichle, R. H., Brucker, L., Brangers, I., de Rosnay, P., Dumont, M., Girotto, M., Immerzeel, W. W., Jonas, T., Kim, E. J., Koch, I., Marty, C., Saloranta, T., Schöber, J., and De Lannoy, G. J. M.: Snow depth variability in the Northern Hemisphere mountains observed from space, Nat Commun, 10, 4629, https://doi.org/10.1038/s41467-019-12566-y, 2019.

580 Martens, B., De Jeu, R. A. M., Verhoest, N. E. C., Schuurmans, H., Kleijer, J., and Miralles, D. G.: Towards Estimating Land Evaporation at Field Scales Using GLEAM, 10, 1720, https://doi.org/10.3390/rs10111720, 2018.

Massari, C., Brocca, L., Ciabatta, L., Moramarco, T., Gabellani, S., Albergel, C., De Rosnay, P., Puca, S., and Wagner, W.: The Use of H-SAF Soil Moisture Products for Operational Hydrology: Flood Modelling over Italy, 2, 2–22, https://doi.org/10.3390/hydrology2010002, 2015.

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

Massari, C., Modanesi, S., Dari, J., Gruber, A., De Lannoy, G. J. M., Girotto, M., Quintana-Seguí, P., Le Page, M., Jarlan, 590 L., Zribi, M., Ouaadi, N., Vreugdenhil, M., Zappa, L., Dorigo, W., Wagner, W., Brombacher, J., Pelgrum, H., Jaquot, P., Freeman, V., Volden, E., Fernandez Prieto, D., Tarpanelli, A., Barbetta, S., and Brocca, L.: A Review of Irrigation Information Retrievals from Space and Their Utility for Users, 13, 4112, https://doi.org/10.3390/rs13204112, 2021.

McCabe, M. F., Rodell, M., Alsdorf, D. E., Miralles, D. G., Uijlenhoet, R., Wagner, W., Lucieer, A., Houborg, R., Verhoest, N. E. C., Franz, T. E., Shi, J., Gao, H., and Wood, E. F.: The future of Earth observation in hydrology, 21, 3879–3914, 595 https://doi.org/10.5194/hess-21-3879-2017, 2017.

Miralles, D. G., Gash, J. H., Holmes, T. R. H., de Jeu, R. A. M., and Dolman, A. J.: Global canopy interception from satellite observations, 115, https://doi.org/10.1029/2009JD013530, 2010.

Miralles, D. G., Holmes, T. R. H., De Jeu, R. a. M., Gash, J. H., Meesters, A. G. C. A., and Dolman, A. J.: Global land-surface evaporation estimated from satellite-based observations, 15, 453–469, https://doi.org/10.5194/hess-15-453-2011, 600 2011.

Mosaffa, H., Sadeghi, M., Mallakpour, I., Naghdyzadegan Jahromi, M., and Pourghasemi, H. R.: Chapter 43 - Application of machine learning algorithms in hydrology, in: Computers in Earth and Environmental Sciences, edited by: Pourghasemi, H. R., Elsevier, 585–591, https://doi.org/10.1016/B978-0-323-89861-4.00027-0, 2022.

Muñoz-Sabater, J., Dutra, E., Agustí-Panareda, A., Albergel, C., Arduini, G., Balsamo, G., Boussetta, S., Choulga, M., 605 Harrigan, S., Hersbach, H., Martens, B., Miralles, D. G., Piles, M., Rodríguez-Fernández, N. J., Zsoter, E., Buontempo, C.,

and Thépaut, J.-N.: ERA5-Land: a state-of-the-art global reanalysis dataset for land applications, 13, 4349–4383, https://doi.org/10.5194/essd-13-4349-2021, 2021.

Paiva, R. C. D., Collischonn, W., Bonnet, M.-P., de Gonçalves, L. G. G., Calmant, S., Getirana, A., and Santos da Silva, J.: Assimilating in situ and radar altimetry data into a large-scale hydrologic-hydrodynamic model for streamflow forecast in 610 the Amazon, Hydrol. Earth Syst. Sci., 17, 2929–2946, https://doi.org/10.5194/hess-17-2929-2013, 2013.

Pechlivanidis, I. G., Jackson, B. M., Mcintyre, N. R., and Wheater, H. S.: Catchment scale hydrological modelling: a review of model types, calibration approaches and uncertainty analysis methods in the context of recent developments in technology and applications, 13, 193–214, 2011.

Priestley, C. H. B. and Taylor, R. J.: On the Assessment of Surface Heat Flux and Evaporation Using Large-Scale 615 Parameters, 100, 81–92, https://doi.org/10.1175/1520-0493(1972)100<0081:OTAOSH>2.3.CO;2, 1972.

Qi, W., Zhang, C., Fu, G., Sweetapple, C., and Zhou, H.: Evaluation of global fine-resolution precipitation products and their uncertainty quantification in ensemble discharge simulations, 20, 903–920, https://doi.org/10.5194/hess-20-903-2016, 2016.

Quast, R. and Wagner, W.: Analytical solution for first-order scattering in bistatic radiative transfer interaction problems of layered media, Appl Opt, 55, 5379–5386, https://doi.org/10.1364/AO.55.005379, 2016.

620 Quast, R., Albergel, C., Calvet, J.-C., and Wagner, W.: A Generic First-Order Radiative Transfer Modelling Approach for the Inversion of Soil and Vegetation Parameters from Scatterometer Observations, 11, 285, https://doi.org/10.3390/rs11030285, 2019.

Raup, B., Racoviteanu, A., Khalsa, S. J. S., Helm, C., Armstrong, R., and Arnaud, Y.: The GLIMS geospatial glacier database: A new tool for studying glacier change, Global and Planetary Change, 56, 101–110, 625 https://doi.org/10.1016/j.gloplacha.2006.07.018, 2007.

Reichle, R. H.: Data assimilation methods in the Earth sciences, Advances in Water Resources, 31, 1411–1418, https://doi.org/10.1016/j.advwatres.2008.01.001, 2008.

Revilla-Romero, B., Wanders, N., Burek, P., Salamon, P., and de, R.: Integrating remotely sensed surface water extent into continental scale hydrology, 543, 659–670, https://doi.org/10.1016/j.jhydrol.2016.10.041, 2016.

630 Ross, C. W., PRIHODKO, L., ANCHANG, J., KUMAR, S., JI, W., and HANAN, N. P.: Global Hydrologic Soil Groups (HYSOGs250m) for Curve Number-Based Runoff Modeling, ORNL Distributed Active Archive Center, https://doi.org/10.3334/ORNLDAAC/1566, 2018.

Silvestro, F., Gabellani, S., Delogu, F., Rudari, R., and Boni, G.: Exploiting remote sensing land surface temperature in distributed hydrological modelling: the example of the Continuum model, 17, 39–62, https://doi.org/10.5194/hess-17-39-635 2013, 2013.

Sinclair, S. and Pegram, G.: Combining radar and rain gauge rainfall estimates using conditional merging, 6, 19–22, 2005.

Spaaks, J. H. and Bouten, W.: Resolving structural errors in a spatially distributed hydrologic model using ensemble Kalman filter state updates, 17, 3455–3472, https://doi.org/10.5194/hess-17-3455-2013, 2013.

Sperna Weiland, F. C., Vrugt, J. A., van Beek, R. (L.) P. H., Weerts, A. H., and Bierkens, M. F. P.: Significant uncertainty in global scale hydrological modeling from precipitation data errors, Journal of Hydrology, 529, 1095–1115, https://doi.org/10.1016/j.jhydrol.2015.08.061, 2015.

Stauffer, D. R. and Seaman, N. L.: Use of four-dimensional data assimilation in a limited-area mesoscale model. Part I: experiments with synoptic-scale data, 118, 1250–1277, https://doi.org/10.1175/1520-0493(1990)118<1250:UOFDDA>2.0.CO;2, 1990.

Tang, G., Clark, M. P., Papalexiou, S. M., Ma, Z., and Hong, Y.: Have satellite precipitation products improved over last two decades? A comprehensive comparison of GPM IMERG with nine satellite and reanalysis datasets, Remote Sensing of Environment, 240, 111697, https://doi.org/10.1016/j.rse.2020.111697, 2020.

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

Tarpanelli, A., Brocca, L., Barbetta, S., Faruolo, M., Lacava, T., and Moramarco, T.: Coupling MODIS and Radar Altimetry Data for Discharge Estimation in Poorly Gauged River Basins, 8, 141–148, https://doi.org/10.1109/JSTARS.2014.2320582, 2015.

Tarpanelli, A., Iodice, F., Brocca, L., Restano, M., and Benveniste, J.: River Flow Monitoring by Sentinel-3 OLCI and MODIS: Comparison and Combination, 12, 3867, https://doi.org/10.3390/rs12233867, 2020.

Terzago, S., Andreoli, V., Arduini, G., Balsamo, G., Campo, L., Cassardo, C., Cremonese, E., Dolia, D., Gabellani, S., von Hardenberg, J., Morra di Cella, U., Palazzi, E., Piazzi, G., Pogliotti, P., and Provenzale, A.: Sensitivity of snow models to the accuracy of meteorological forcings in mountain environments, 24, 4061–4090, https://doi.org/10.5194/hess-24-4061-2020, 2020.

Thirel, G., Salamon, P., Burek, P., and Kalas, M.: Assimilation of MODIS Snow Cover Area Data in a Distributed Hydrological Model Using the Particle Filter, 5, 5825–5850, https://doi.org/10.3390/rs5115825, 2013.

Verdin, K. L.: Hydrologic Derivatives for Modeling and Analysis—A new global high-resolution database, Hydrologic Derivatives for Modeling and Analysis—A new global high-resolution database, U.S. Geological Survey, Reston, VA, https://doi.org/10.3133/ds1053, 2017.

665   Wanders, N., Karssenberg, D., De Roo, A., De Jong, S. M., and Bierkens, M. F. P.: The suitability of remotely sensed soil moisture for improving operational flood forecasting, 18, 2343–2357, https://doi.org/10.5194/hess-18-2343-2014, 2014.

Wongchuig-Correa, S., Cauduro Dias de Paiva, R., Biancamaria, S., and Collischonn, W.: Assimilation of future SWOT-based river elevations, surface extent observations and discharge estimations into uncertain global hydrological models, 590, https://doi.org/10.1016/j.jhydrol.2020.125473, 2020.

670   Wu, H., Adler, R. F., Tian, Y., Gu, G., and Huffman, G. J.: Evaluation of Quantitative Precipitation Estimations through Hydrological Modeling in IFloodS River Basins, 18, 529–553, https://doi.org/10.1175/JHM-D-15-0149.1, 2017.

Zakharova, E., Nielsen, K., Kamenev, G., and Kouraev, A.: River discharge estimation from radar altimetry: Assessment of satellite performance, river scales and methods, Journal of Hydrology, 583, 124561, https://doi.org/10.1016/j.jhydrol.2020.124561, 2020.

675