# Estimating the Ebro river discharge at 1 km/daily resolution using indirect satellite observations

Victor Pellet, Filipe Aires, Oscar Baez Villa Nueva, Paolo Filippucci

LETTER • OPEN ACCESS

# Estimating the Ebro river discharge at 1 km/daily resolution using indirect satellite observations

To cite this article: Victor Pellet *et al* 2024 *Environ. Res. Commun.* **6** 091014

Environmental Research Communications

LETTER

# Estimating the Ebro river discharge at 1 km/daily resolution using indirect satellite observations

**Victor Pellet**[1,2], **Filipe Aires**[1,2], **Oscar Baez Villa Nueva**[3] and **Paolo Filippucci**[4]

[1] Estellus, Paris, France
[2] The Laboratoire d'Etudes du Rayonnement et de la Matière en Astrophysique et Atmosphères (LERMA), Observatoire de Paris, PSL, 75014 Paris, France
[3] The Hydro-Climate Extremes Lab (H-CEL), Ghent University, Ghent, Belgium
[4] The Research Institute for Geo-Hydrological Protection, National Research Council, Perugia, Italy

**E-mail:** victor.pellet@obspm.fr

**OPEN ACCESS**

**RECEIVED**
10 July 2024

**REVISED**
23 August 2024

**ACCEPTED FOR PUBLICATION**
13 September 2024

**PUBLISHED**
23 September 2024



**Keywords:** river discharge, water cycle, satellite remote sensing, terrestrial hydrology, Ebro river

## Abstract

Estimating river discharge *Q* at global scale from satellite observations is not yet fully satisfactory in part because of limited space/time resolution. Furthermore, on highly anthropized basins, it is essential to anchor the analysis to reliable *Q* measurements. Gauge networks are however very sparse and limited in time, and SWOT (Surface Water Ocean Topography) river discharge estimates at global scale are not yet available. The method proposed here is able to obtain continuous daily *Q* estimates at 1 km/daily resolution, using indirect satellite data and ground-based estimates. We focus here on the Ebro. Over such an anthropized basin (e.g. change of land use, irrigation), the exploitation of 205 available gauges at their nominal resolution (i.e., daily point measurements) is a necessity. The hydrological Continuum model is used to help interpolate spatially and temporally the observations into our optimal interpolation scheme. The proposed *Q*-mapping is similar to an assimilation scheme were Earth observations (precipitation, evapotranspiration and total water storage change) and model simulations are constrained by *in situ* gauge measurements. The *Q* estimates are evaluated using a rigorous leave-one-out experiment, showing a good agreement with the *in situ* data: a correlation of 0.72 (median), and a 75th percentile of Nash-Sutcliffe Efficiency up to 0.62. Our spatio-temporal continuous *Q* estimates at high spatial/temporal resolution can describe complex continental water dynamics, including extreme events. SWOT estimates will soon be available, at the global scale but with irregular space/time sampling: our method should help exploit them to obtain a regular space-temporal description of the water cycle at high resolution.

## 1. Introduction

The Ebro is the longest river (987 km) of Spain, and a major basin of the Mediterranean region. Raising near the Atlantic coast in the Cantabrian Mountains in northern Spain, it drains an area of 86,000 km$^2$ between the Pyrenees and the Iberian mountains. The basin is very heterogeneous, influenced by the Atlantic Ocean on the Northwest and the Mediterranean Sea on the Southeast. The annual mean *P* is unevenly distributed, being higher in the mountainous regions (reaching 1,800 mm/yr), and lower in the central valley (below 500 mm/yr). The basin is highly affected by anthropisation (i.e. irrigation and dam operation) and global warming with an increase of water stress in summer [1, 2]. The management of its water resources is thus a challenge at national level and the Ebro basin is a data rich basin well monitored through national and international campaigns [3, 4]. Its Automatic System of Hydrological Information [4] provides a wealth of real-time and historical river data, which makes it an interesting test-bed for hydrological studies in the context of climate change.

The monitoring of the river discharge in a particular basin such as the Ebro is an efficient way to analyze and even manage such a basin. However, two main difficulties appear: i) A basin like the Ebro is highly anthropized

and it is essential to anchor the water cycle analysis to true *in situ* data. If not, lack of information on human water management can bias what can be obtained with a land or hydrological model combined or not with satellite data. ii) The true understanding and maybe the management of a basin with such complexity necessitates high spatial and temporal resolutions. This is again a true challenge, especially when exploiting satellite observations often given at coarse resolution.

Let us first consider the several ways to estimate river discharge. Obtaining dense regular river discharge $Q$ can be obtained with a Land Surface Model (LSM) that includes a routing module. Such approach, relying on physical equation, struggles in accounting for human activities such as dam operations [5] or any processes for which there is a lack of information or knowledge. For a highly managed basin like the Ebro River, the use of *in situ* observations is a necessity. A traditional method to obtain realistic river discharge estimates that account for human activities involves developing an assimilation scheme, where the model output ($Q$) is constrained/corrected by *in situ* gauge measurements ($q$). However, the development of an assimilation system is not straightforward [6], specially at high resolution (daily and 1 km).

The estimation of river discharge from Earth Observations (EO) observations is generally obtained at point site based on radar nadir altimeter [7–9]. A few attempts have relied on other types of satellite observations such as imaging sensors in Near InfraRed (NIR) [10] or passive microwaves based surface water extent [11]. The idea is to build an 'empirical relationship' representing the low to high flow states, using either i) river height information from altimetry, or ii) proxy of the water surface extent from NIR reflectance or from passive microwave. This is done over a few *in situ* gauges river discharge observations or simulations through hydrological modeling approaches [8]. The difficulty with these empirical relationships is that they can provide river discharge only at point site (e.g. virtual station) and not continuously over the full basin.

Another approach considering water mass conservation was proposed recently to estimate dense regular $Q$ [12]. The proposed methodology uses indirect[5] EOs of Precipitation $P$, Evaporation $E$, and total water storage change $dS$, together with a few available *in situ* gauge measurements. The method relies on a two-steps process. First, the $P$, $E$, and $dS$ products are first corrected at basins scale to better close the water budget, using $q$[6] from the available gauges. Second, the budget is then spatially balanced between the pixels, by using flow direction to consider horizontal water exchange, leading to an estimation $Q$ at pixel scale. This approach in [12] is based solely on satellite products, *in situ* measurements, and river network derived from topography. No dynamical model was used, which can be positive if no good model is available, but it can be detrimental as physical relationships can help perform a better space/time interpolation of the water components.

Furthermore, the $Q$ mapping in [12] has been obtained at the EOs nominal resolution (monthly, 0.25°). This relatively coarse resolution, however, is a strong limitation for water management applications and for climate-related researches, in particular for a basin like the Ebro. In the framework of the 4D-MED project (https://www.4dmed-hydrology.org/), new downscaled water component estimates have been produced over some basins of the Mediterranean region, at daily and 1 km resolutions. $P$ is derived from a merging/ downscaling process of multi-product [13]. $E$ was obtained through the application of the GLEAM model [14, 15] with high resolution inputs. $dS$ is estimated through a statistical/physical dynamical downscaling that relies on an *a priori* $Q$ estimate derived from a hydrological model [16]. Such higher spatial resolution products for $P$, $E$ and $dS$ can be used to estimate dense $Q$ using available *in situ* $q$ measurements and the methodology presented in [12].

Indeed, despite the existence of these high-resolution datasets, there is still a need for point $q$ measurements either ground-based observations from a relatively dense network of gauges. An alternative will be provided by the recent Surface Water and Ocean Topography (SWOT) NASA-CNES mission [17–19]. SWOT uses a Ka-band radar interferometer, to estimate water surface elevations of rivers globally at 100 m resolution that are used to obtain river discharges at global scale (but with a complex spatio-temporal sampling). Such data can be leveraged to estimate dense regular river discharge estimates, but a convincing approach is required to fully exploit their information at nominal resolution. National and international space agencies such as ESA (European Space Agency) aim to develop a comprehensive view of the interaction between natural phenomena and human activities at global scale and at high spatio-temporal resolution in order to pave the green transition.

This paper aims at demonstrating how to reach high-resolution (daily and 1 km) realistic representation of river discharge ($Q$). The methodological core was previously developed [12] but it was done at a low resolution (i.e., monthly, ~0.25°). Integrating high-resolution EO datasets in order to use *in situ* gauge measurements at their nominal resolution poses a challenge that needs to be addressed. Specifically, the benefits gained from the fusion of multiple sources of hydrological information (model, EO and *in situ* data) needs to be documented.

The datasets used in this study will be presented in section 2. Section 3 presents the closure-based estimate of $Q$. Analysis of the obtained results will be performed in section 4. Finally, section 5 will provide some conclusions and perspectives.

[5] Indirect because not a direct measurement of the river discharge.

[6] Lower-case letter are used for *in situ* point measurements through the manuscript to differ from continuous $Q$ estimate.

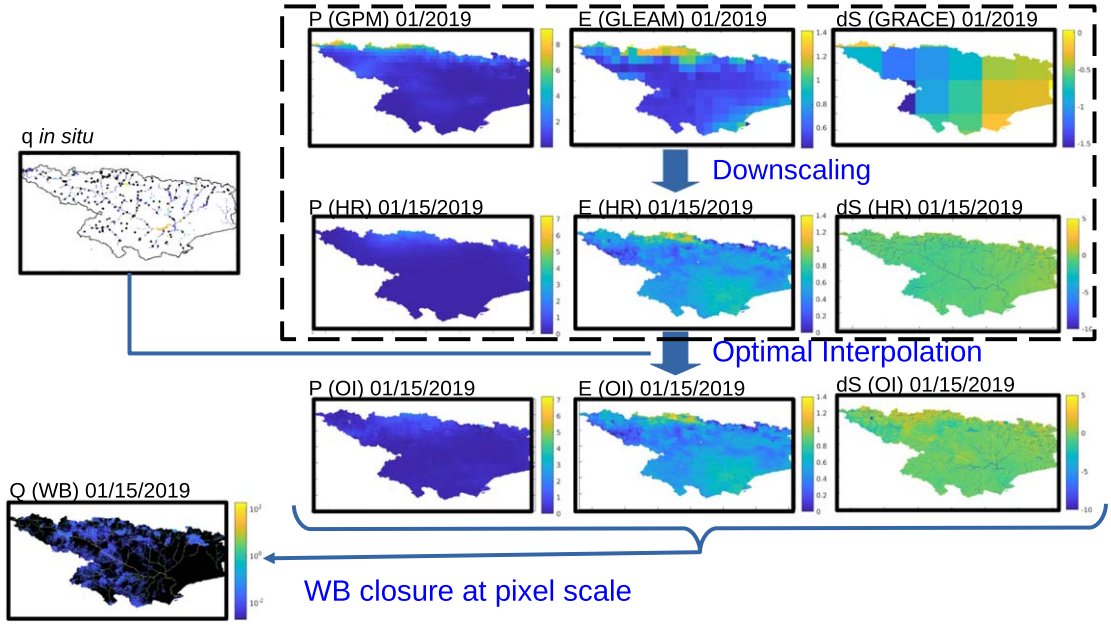


**Figure 1.** Overall integration scheme. First row: Original dataset in January 2019 for *in situ Q* and satellite-based *P*, *E*, and *dS* in *mm/day*. Second row: Downscaled version. Third row: OI balanced version for January 15th 2019. The final continuous water balance *Q* is in the bottom left corner. The dash line highlights the processing done in the 4DMED project and made available for the current study.

## 2. Original datasets and their optimization

### 2.1. Precipitation, *P*

The satellite *P* products developed within 4DMED are based on the integration of multiple *P* datasets to generate a high spatial (1 km) and temporal (daily) resolution over the Mediterranean area. The NOAA Climate Prediction Center (CPC) daily 0.5° precipitation [20, 21] and the Integrated Multi-satellitE Retrievals for Global Precipitation Measurement (GPM-IMERG, Late-run version) 30 mn and 0.1° *P* [22] are merged and downscaled. The downscaling procedure leverages the Climatologies of the Earth's Land Surface Areas (CHELSA), a high resolution (30arcsec, 1 km) global and monthly dataset [23]. It is based on a mechanistical statistical downscaling of global reanalysis data or global circulation model output and it includes climate layers for various time periods and variables. The relative precipitation patterns of CHELSA are exploited to spatialize the coarse resolution information of precipitation products, being based on the modelling of orographic predictors of wind fields, valley exposition and boundary layer height. A paper describing the downscaling procedure is currently under preparation.

Finally, the high spatial (1 km) and temporal (daily) resolution 4DMED precipitation product (https://stac.eurac.edu:8080/collections/rainfall_all_domain) was adopted for this study. This precipitation product is obtained by merging the downscaled GPM-Late run and CPC over the Mediterranean area. The two products are merged leveraging on the results of a triple collocation technique (third product: the reanalysis ERA5 Land precipitation): the obtained signal to noise ratios are used to derive each product weight. The product is available in the period 2000–2022. The original 0.1° GPM monthly value for January 2019 [22] and the 1 km-downscaled product for January 15th are shown in figure 1.

### 2.2. Evapotranspiration, *E*

The Global Land Evaporation Amsterdam Model (GLEAM) 1-km *E* has been obtained by running the GLEAM model [14, 24] at 1km using high resolution inputs. GLEAM uses an empirical energy-based equation [25] to calculate a potential evaporation. Afterwards, *E* is estimated per land cover fraction based on an evaporative stress approach. This method estimates separately transpiration, bare-soil evaporation, interception loss, open-water evaporation, and sublimation. Radiation and temperature data were obtained from LSAF (land-saf.eumetsat.int), vegetation fractions from MOD44B v6.0 [26], and soil properties from HiHydroSoil v2.0 [27]. GLEAM was run at 1 km using high-resolution of : i) *P* introduced in section 2.1; ii) VOD and soil moisture based on the Land Parameter Retrieval Model [28] and downscaled microwave observations; and iii) snow water equivalent based on the assimilation of the C-SNOW dataset [29] in the Snow Multidata Mapping and Model [30]. All these data have been produced during the 4DMED project (https://stac.eurac.edu:8080/collections/).

The 0.25° GLEAM monthly value for January 2019 [24] and the 1 km-downscaled product for January 15th 2019 are shown in the second row of figure 1.

### 2.3. Total water storage change, *dS*

The twin satellites of GRACE and GRACE-Follow On missions [31] offer a unique opportunity to monitor the water stored in land (i.e., soil moisture, surface water, snow-pack, glaciers, and groundwater). A state-of-the-art retrieval is performed based on the mass concentration solution known as MASCON (MSC). This technique benefits from gravity field basis functions to better isolate the hydrological (i.e., water storage variation) contributions in the signal from other factors. Typical monthly solution is provided globally on a $0.5^\circ \times 0.5^\circ$ grid. A statistical/physical dynamical downscaling has been proposed to obtain 1 km and daily *dS* estimates, using auxiliary information from the combination of other satellites (i.e. *P*, *E* and river network direction from topography) and modelling data (simulated *Q*) [16]. GRACE information is used to constrain low *dS* space/time variabilities, while high frequencies come from the $P - E - Q$ information. Using *a priori* simulated *Q* at the downscaling level of *dS* allows us to fuse information on *Q* from the hydrological model along with information from GRACE *dS*, *P*, and *E*. This estimate will then be constrained using *in situ* gauge *q* measurements (see section 2.5). This procedure is close to the assimilation framework, where the model is updated using new observations. The original 0.5° GRACE monthly value for January 2019 [32] and the 1 km-downscaled product for January 15th 2019 are shown in figure 1 third row.

### 2.4. *In situ* river discharges *q*

A large dataset of daily river discharge *q* measurements has been obtained for 205 stations over the Ebro basin (top left of figure 1), collected throughout 2016–2019. This extensive *q* record is obtained from the Hydrographic Confederation of the Ebro, through its real-time data portal: Automatic Hydrological Information System (Sistema Automático de Información Hidrológica, SAIH). SAIH provides *q* data from discharge stations (levels and river flows in the basin) and reservoirs (levels, volumes and water filled percentage).

### 2.5. Topography and Continuum hydrological model

The Digital Elevation Model (DEM) of the global USGS (US Geological Survey) Hydrologic Derivatives for Modeling and Analysis [33], upscaled at 1 km, is used for two purposes. First to delineate the catchments of the respective 205 stations to define the water balance in each one of them. Second, the DEM allows to estimate the direction of the horizontal water transport, for each pixel. This gives us the so-called 'flow accumulation' and associated river network.

A simulation from the Continuum distributed hydrological model [34, 35] is used as an *a priori* source of information on the river discharge. It is a dense and continuous *Q* estimate in order to both downscale GRACE *dS* data and later evaluate the Q-mapping. Continuum is a trade-off between an empirical and a physically-based model. Deep flow and water table evolution are modeled with a simple scheme that reproduces the main physical characteristics of the processes and a distributed interaction between water table and soil surface with a low level of parametrization [34]. Continuum model was calibrated for the Ebro River using discharge measurements at 19 locations [36]. The calibration perturbs six scalar parameters of the model all related to hydrological behavior. It is a multi-site calibration procedure that iteratively searches for the model parameterization (values of the six scalars) that best matches the available discharge observations by minimizing the Kling-Gupta efficiency (KGE) used as a cost function. Detailed information regarding the calibration procedure can be found in [36]. Once calibrated, the model does not assimilate any *in situ* when simulating the river discharge.

### 2.6. Optimal interpolation of *P*, *E* and *dS*

As Optimal Interpolation (OI) has been largely presented in previous researches [12, 37, 38], only a short presentation is proposed here. The *P*, *E* and *dS* satellite estimates (from the previous sections 2.1–2.3) are corrected at the sub-basin scale. *In situ* gauges $\{q_j\,;\ j = 1, \cdots, J\}$ (from the section 2.4) are used for this purpose. *P*, *E* and *dS* corrections intend to close the water budget over each one of the 205 drainage areas. At each time step *t*:

$$G \cdot Y_{SAT}(t) = q(t) + \epsilon, \tag{1}$$

where $Y_{SAT}(t)$ is the state vector $(3n \times 1)$ of $P(t)$, $E(t)$ and $dS(t)$ on the *n* pixels of the basin for a given day *t*. Time variable *t* is then omitted in the following for simplicity. *G* is a $(J \times 3n)$ matrix in which $G(j, 3k + 1{:}\ 4k) = [1, -1, -1]$ if the pixel *k* belongs to the drainage area of gauge *j*. In the water budget framework, the water mass conservation of *P*, *E* and *dS* relies on the hypothesis that horizontal exchanges are represented solely by surface water (i.e., discharge *q*).

Equation (1) can be inverted, based on an *a priori* error covariance matrix $B_{SAT}$ describing our uncertainly on $Y_{SAT}$ [37, 39]. The definition of $B_{SAT}$ is shown in the appendix.

$$Y_a = Y_{SAT} + [B_{SAT} \cdot G_o^T (G_o \cdot B_{SAT} \cdot G_o^T)] \cdot [q_o - G_o \cdot Y_{SAT}], \quad (2)$$

where *a* represents the 'analysis' of the previous solution $Y_{SAT}$. $Y_a$ closes the water budget by minimizing errors $\epsilon$ in equation (1) over the catchments, while being as close as possible to $Y_{SAT}$. OI aims here only at optimizing *P*, *E*, and *dS*, as *q* is considered to be more reliable since it is an *in situ* observation. The gauge measurements represent a reference on which the whole OI is built. Equation (1) can be read as an update in a assimilation cycle where *a priori* $Y_{SAT}$ is corrected based on the observation of *q* to give the analysis $Y_a$. The OI procedure is a closed-form analytical solution that directly used *in situ q* measurements at each time step based on their availability.

In this optimization problem, the state vector $Y_{SAT}$ has a dimension=380220 representing 126 740 pixels×3 (for *P*, *E*, *dS*). It then involves the inversion of a very large matrix of dimension (380220 × 380220).

The OI-optimized *P*, *E* and *dS* estimates are shown for January 15th in the third row of figure 1. The OI only slightly corrects the water components, but it reduces substantially the water budget imbalance [38, 40]. The obtained water components are more coherent and in better agreement with the measured *q*. This should help the spatial 'interpolation' of the gauge measurements in the following step.

## 3. Closure-based river discharge Q estimation

The OI-optimized *P*, *E*, and *dS* are used here to estimate daily *Q* at a spatial resolution of 1 km, over the whole Ebro domain.

### 3.1. River discharge estimation using budget closure

The river discharge $Q(i)$ at a location *i* is estimated from the optimized satellite observations $Y_a = [P_a, E_a, dS_a]$ over its upstream area:

$$Q(i) = UP(i, :) \times Y_a. \quad (3)$$

*UP* is a $n \times 3n$ matrix in which $UP(i, 3k + 1{:}\,4k) = [1 - 1 - 1]$ if pixel *k* belongs to the drainage area of pixel *i*, and zeros elsewhere (this is actually the extension of *G* to all pixels in the Ebro basin). Equation (3) allows estimating pixel-wise *Q* by closing the water budget upstream of each pixel. The upstream area, for any pixel 'i' in *UP*, is given by the flow direction derived from topography. This river network constrains the inter-pixel exchanges to'interpolate' the *in situ q*. The resulting *Q* can be interpreted as an interpolation of the river gauge measurements *q* through the use of the spatial patterns $Y_a$ of the indirect observations, using the two-step process of equations (1) and (3). The obtained *Q* estimate are represented in the bottom left corner of figure 1.

### 3.2. Evaluation of the closure-based river discharges

To test the robustness of the estimation, a Leave-One-Out (LOO) cross validation technique is applied: one gauge is withdrawn from the ensemble that is used for the *Q* mapping. The *Q* estimate in this location is then compared to the gauge measurement. This operation is repeated for all the 205 available gauges. This experiment evaluates the true ability of our approach to estimate *Q* over an ungauged pixel. Table 1 summarizes the performance of the estimations over gauges that were not considered in the OI process data for six metrics: the Kling-Gupta Efficiency (KGE), Nash-Sutcliffe Efficiency (NSE), correlation (CORR), the Root Mean Square Error (RMSE), Mean Error (ME), and the STandard Deviation of Error (STDE). For all the metrics, the 75th, 50th (median), and 25th percentiles are shown.

The results of our method in terms of RMSE in table 1 can be analyzed from the perspective of closing the water budget. Indeed, accurately estimating *q* on a particular testing gauge using the water budget equation ($\hat{q} = P - E - dS$) and obtaining $\hat{q}$ close to the observed *in situ q* during testing is by means minimizing the water budget imbalance: $imb = \hat{q} - q$.

The proposed approach appears to be of good quality for all metrics. Median correlation is 0.8, and median KGE = 0.13. The variability in NSE metrics comes from higher errors over small rivers.

The method is evaluated along with the Continuum model simulation. Even though the model is calibrated, it does not use explicitly *in situ* measurements in its simulations. The gauge data are used to 'calibrate' the model in the sense that the parameters of the model are optimized so that its RD are as close as possible to the *in situ* data. Once the model is calibrate (once and for all) then its simulations do not the *in situ* data. In order to make the model simulations closer to the *in situ* data, an assimilation scheme would be required. However, leveraging on the *in situ* gauge data to better constrain EO has significant advantages, in particular anchoring the analysis on true data, which is essential in highly anthropized basins such as this one. In contrast, our OI analysis fuse the

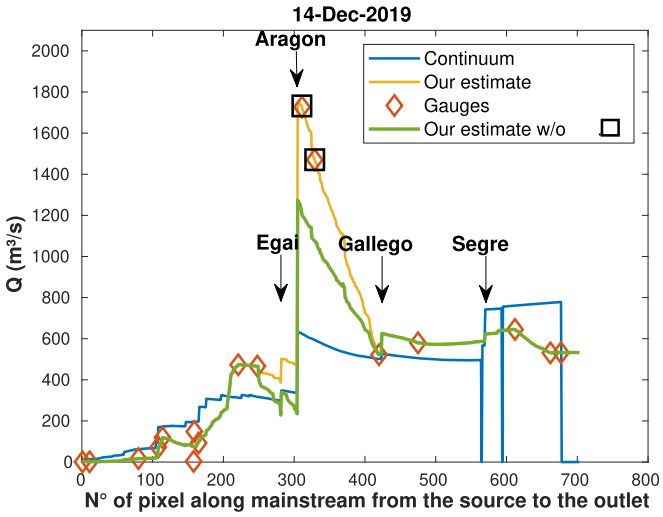


**Figure 2.** Longitudinal profile of the Ebro river discharge in December 14th 2019 for: the Continuum model (blue), our estimate (yellow), and the *in situ* measurements (red diamonds). For evaluation purposes, the closure-based estimate calibrated without the Aragon confluence gauge (black square) nor the closer gauge on mainstream is also shown in green. Main Ebro tributaries are indicated.

**Table 1.** Evaluation of the *Q* estimates using the testing *in situ* data (not used during the optimisation) for: our estimates (left) and for the Continuum model (right). See text for definition of the metrics.

| | Our estimate | | | Continuum model | | |
|---|---|---|---|---|---|---|
| Metrics | 75th | 50th | 25th | 75th | 50th | 25th |
| KGE | 0.61 | 0.13 | −0.84 | 0.34 | −0.03 | -0.49 |
| CORR | 0.88 | 0.72 | 0.41 | 0.69 | 0.59 | 0.22 |
| NSE | 0.62 | 0.07 | −3.6 | 0.30 | −0.02 | -0.83 |
| RMSE | 11.6 | 4.1 | 2.0 | 12 | 4.6 | 1.5 |
| ME | 1.9 | 0.2 | −0.5 | 1.0 | 0.2 | -2 |
| STDE | 10.6 | 4.0 | 1.8 | 11.5 | 3.9 | 1.4 |

information of EOs, modeling *a priori* and *in situ* data, in a similar way than what assimilation does. Table 1 shows the benefits of exploiting the *in situ* data to obtain a more accurate estimation of the river discharge.

## 4. Analysis of the obtained river discharges

### 4.1. Along-the-Ebro analysis

Figure 2 shows an instantaneous view of the mainstream longitudinal *Q* profile for one day. *In situ* measurements are shown in diamond, Continuum model *Q* estimates in blue, and closure-based *Q* in yellow. By definition, the closure-based estimates match very well the *in situ* data, while giving a continuous estimate along the river. The method does not only interpolate between two gauges: By exploiting the spatial patterns of the indirect observations on *P*, *E* and *dS*, it is possible to infer complex behaviours between two gauges such as the discontinuous rise in *Q* due to a tributary contribution (e.g. Ega, Aragon, or Gallego). In contrast, the Continuum model is not able to represent adequately the high flow in the river reach between the 300th and the 400th pixels.

An experiment was conducted (in green), for which the reconstruction is done without the Aragon confluence gauge nor the closer gauge on mainstream: Despite the relatively high underestimation of *Q*, the Aragon peak is however well recovered, without a direct constrain on it, demonstrating how EO data can be leveraged to substantially improve the representation of *Q* patterns along the river.

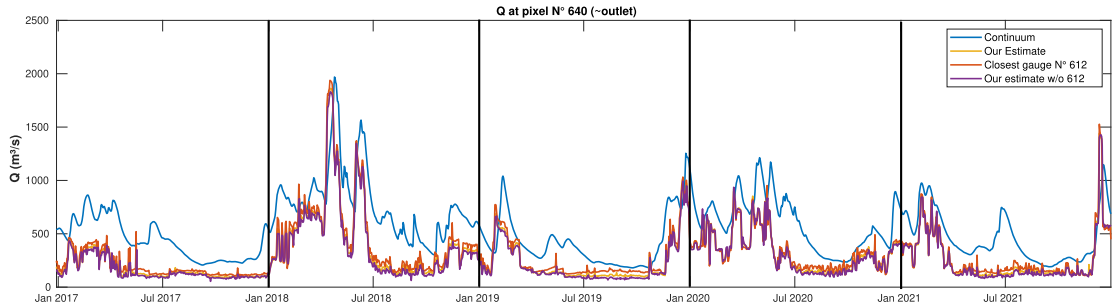


**Figure 3.** *Q* time series close to the Ebro outlet for: our estimate (red); Continuum simulation (blue); nearest *in situ* gauge (red).

### 4.2. Temporal analysis at the outlet

Figure 3 shows in yellow the time series of the *Q* estimates close to the Ebro outlet, together with the closest gauge in red about 50 km upstream (612th pixel in figure 2). Correlation is equal to 0.98 between them. Again, when this gauge is not used to constrain the solution, the result (purple) is barely degraded. At the outlet of the basin, even when the closest stations are removed, our approach yield substantially better results compared to the hydrologic modelling exercise.

The Continuum estimate is shown in blue. A scaling error of Continuum seems to be present in the simulation. E.g., for particularly dry year, a bias at low flow can be observed in the simulation (e.g. 2019). This could be related to the existence of bias in the runoff used as input by the Continuum model and/or soil water content issue in the model. Furthermore, the modelling seems unable to represent high frequency signal contrary to the closure-based *Q* estimate that includes a larger range of variability. In terms of peak, a one-day shift is observed in the simulation of the main flooding event in 2018, probably related to celerity parameters. Overall, figure 3 shows well the advantage of observation-based *Q* estimates, and a possible opportunity to improve model simulations.

### 4.3. Spatio-temporal analysis

In the two previous sections, the *Q*-estimates are analyzed in space and time separately. Figure 4 represents the space-time Continuum of the river flow estimation. Time is represented from January 2016 to December 2021 (y-axis) and the mainstream is represented along its 703 pixels from the upstream (pixel 1) to the outlet (pixel 715).

The discontinuities along the mainstream (x-axis) highlight the main confluent rivers flowing into the mainstream. The seasonal pattern is clear, with an increase from January to March (y-axis) and from upstream to downstream (in *x*-axis). Among the six available yearly patterns, 2017 shows the lowest dynamic with a very narrow period of high-flow (in time). The cumulative *P* from Oct. 1st 2016 to Sept. 30st is down 12%, making 2017 the third driest year, behind 1981 and 2005. For some years (2019, 2021), *Q* is higher in the reach from the 400th to 600th pixels than downstream, reflecting some dam operations along the mainstream. Indeed, our method accounts implicitly for dam operations relying on *q* observations which integrate regulation operations from dam. 2019 showcases a two-peaks high-flow season (the first peak centered around February 2019 while the second is centered in December 2019). Finally, the spatio-temporal graph stresses the flood event that took place in April 2018 for the second part of the mainstream (pixels 350-715). The analysis of extreme events benefits here from the spatially continuous aspect of the *Q* estimate. Droughts (e.g. 2017, (A) in figure 4) and floods (e.g. April 2018, (B) in figure 4) are monitored throughout the entire mainstream. Floods can be caused by one particular confluent (the northern ones).

This figure shows that the the implementation of such approach using near real time datasets could open avenues for a continuous monitoring of continental water dynamics including the occurrence of extreme events from space.

The most important hydrological event of the period in the Ebro is the flooding episode that hit the Zaragoza region in April 2018. It is represented in figure 5. The 2018 winter was one of the rainiest since records are kept. In late February and in March, there were considerable episodes of rain throughout the Iberian Peninsula especially in the upper and middle sections, while snow-melt occurred in the Pyrenees mountains at the headwaters of the rivers, which caused a severe flooding event. The flooding started on April 12th and affected upstream areas such as Tudela that could have seen the peak as soon as the next day, before the peak of the flood was observed to occur in Zaragoza city on April 15th. Severe damages and at least one casualty were reported. The closure-based estimate represents well this extreme event with the maximum discharge estimated for

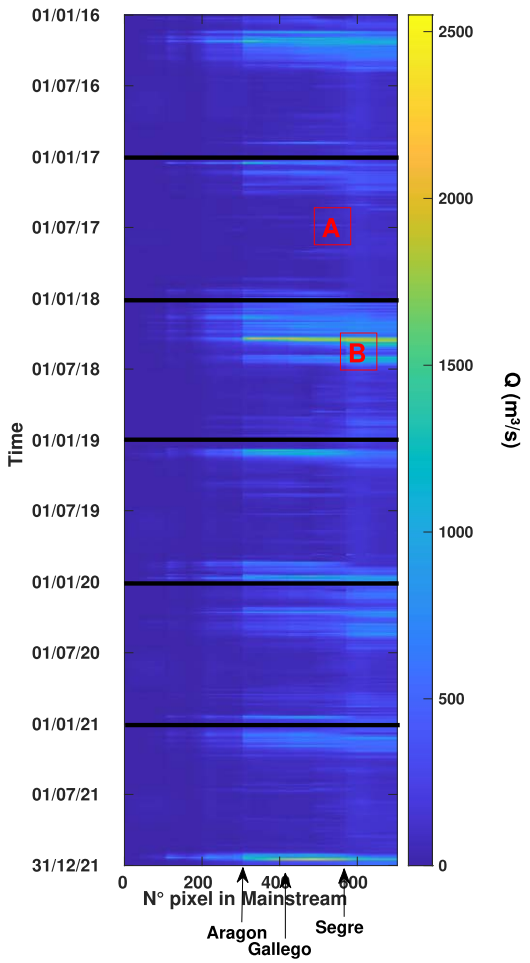


**Figure 4.** Spatio-temporal graph of the river discharge along the Ebro mainstream (x-axis) and over time (2016-2021, *y*-axis). Main tributaries rivers along the mainstream (x-axis) are given in the *x*-axis. Two particular events are indicated: (A) for the drought of 2017 and (B) for the flood in April 2018.

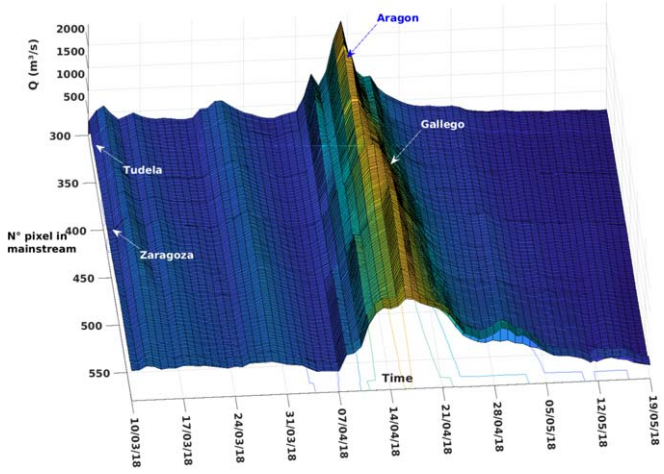


**Figure 5.** Spatio-temporal evolution of the Ebro river discharge along the mainstream (x-axis) and over time (2016-2021, *y*-axis) centered around the flooding event between Tudela and Zaragoza in April 2018.

April 14th around Zaragoza, but on the 13th before Tudela. The shape of the floods also change in space and time: narrower upstream to wider downstream. It can be seen that a smaller event impacts the mainstream around April 28th for pixels above 530th, stressing the impact from the confluent downstream.

## 5. Conclusion

The monitoring of the water cycle in highly anthropized basins such as the Ebro is a true challenge when using either satellite data or a surface/hydrological model. Since exhaustive and precise information on dams management, canals, and irrigation is not available, only *in situ* observations can constrain the analysis towards realistic conditions. For that purpose, the water cycle analysis needs to be performed at high spatial/temporal resolutions, to exploit as much as possible the native resolution of the *in situ* data. Our previous analysis of the water cycle using optimal interpolation fused Earth Observation (EO), *in situ* gauge measurements, and horizontal exchange direction information from topography constraints, but only at a coarse resolution (monthly, 0.25°). New developments were required to obtain results at much higher space/time resolutions. This has been performed here by integrating more information in our framework, namely from the dynamical hydrological model Continuum.

Our new framework to estimate a continuous river discharge is getting closer to an assimilation scheme, but at a much lower computational and development cost. Spatial patterns of the EOs are constrained towards the set of sparse *in situ* gauge measurements. This is done thanks to the *a priori* dynamical and spatial information of the Continuum model, resulting in a water cycle monitoring that is hydrologically more coherent. The benefits gained from this fusion have been demonstrated in terms of spatial/temporal resolutions. Thanks to their proximity to the *in situ* measurements, the river discharge estimates are more realistic because they account for human activity such as dams operations, canals and irrigation that are implicitly recorded in the measurements.

Our new integrated database will be used to better calibrate and validate models such as Continuum. Applications on water management could be investigated, such as the impact of channelisation and irrigation. In this area, water demand for agriculture needs is high [41]. Irrigation impacts the water balance of the river basin, with enhanced evaporation and this must be analyzed too [2]. The possibility to monitor from space the impact of human activity on the water cycle will be investigated. For that purpose, more information needs to be introduced in our analysis (e.g., irrigation or even aqueduct), in a similar way that gauge measurements were introduced. OI appears to be a very good and practical tool that allows for the exploitation of real *in situ* data of diverse nature. This might be easier to develop than a full assimilation scheme.

Machine-learning approaches for estimating river discharge from satellite data have received a lot of attention recently [11, 42, 43] thanks to the development of large river discharge databases [44, 45] at the global scale. However, there are still difficulties and issues to extrapolate the empirical relationships from the very sparse gauges network to the global scale. The combination of such empirical approaches with our OI framework including physical constraint (e.g. water mass conservation) should be beneficial and needs to be investigated [46]. New dedicated and hand-tailored artificial intelligence model architectures might be necessary to achieve such tasks.

The SWOT mission [17, 19, 47] launched in 2023 should provide high-resolution river discharges at the global scale , a true revolution in hydrology. However, SWOT will give those with a complex time/space sampling [19]. Combining SWOT data with more classical EO datasets (regularly sampled in time and space and at different resolutions) will be a challenge, that needs to be overcome for the full exploitation of the SWOT data. Our integration scheme is able to combine such disparate data (e.g. GRACE has an original 200km resolution) and should therefore benefit the SWOT mission by helping the exploitation of its data. Furthermore, the new SWOT data could complement the current gauge network in our analysis, which should help us extend it at the global scale. In addition, when performing this analysis during the SWOT period, it could help us extend our analysis before SWOT. This could be a way to extend the SWOT data back in time, for the GRACE and GRACE-FO period (2002-2023).

## Acknowledgments

This research has been funded by the European Space Agency through contract N°4000136272/21/I-EF. The authors would like to thank D Miralles and P Quintana-Segui for their interesting discussion, the European Space Agency, and in particular Volden Espen, as well as Christian Massari for monitoring the ESA-4DMED project.

## Data availability statement

The data that support the findings of this study are openly available at the following URL https://www.4dmed-hydrology.org/

## A priori covariance matrix in Equation (2)

The *a priori* error covariance matrix $B_{SAT}$ in Equation (2) is a $(3n \times 3n)$ block diagonal matrix defined as:

$$B_{SAT} = \begin{bmatrix} B_P & 0 & 0 \\ 0 & B_E & 0 \\ 0 & 0 & B_{dS} \end{bmatrix} \tag{A.1}$$

where $B_P$ (resp. $B_E$ and $B_{dS}$) is the covariance matrix of $P$ (resp. $E$ and $dS$). Since equation (2) considers all the pixels at a time, the error correlation of the observations need to be considered, following [12], $B_{SAT}$ contains off-diagonal terms have been added to the covariance error matrix $B$. A simple assumption is taken: error correlations decay exponentially in space:

$$B_P(i, j) = [\epsilon_P \cdot P(i)] \cdot [\epsilon_P \cdot P(j)] \cdot e^{\frac{d(i,j)}{D}}, \tag{A.2}$$

where $d(i, j)$ is the distance between pixel $i$ and $j$; $D = 5$ is the e-folding distance (in pixel). This expression extends easily to $B_E$ and $B_{dS}$. The error base estimates is: $\epsilon_P = 10\%$, $\epsilon_E = 10\%$ and $\epsilon_{dS} = 20\%$. No covariance terms have been considered between two different variables (i.e. $P$ vs $E$, $P$ vs $dS$, and $E$ vs $dS$).

## ORCID iDs

Victor Pellet https://orcid.org/0000-0001-6996-0032

## References


[1] López-Moreno J I, Vicente-Serrano S M, Moran-Tejeda E, Zabalza J, Lorenzo-Lacruz J and García-Ruiz J M 2011 Impact of climate evolution and land use changes on water yield in the ebro basin *Hydrol. Earth Syst. Sci.* **15** 311–22

[2] Quiroga S, Garrote L, Iglesias A, Fernández-Haddad Z, Schlickenrieder J, De Lama B, Mosso C and Sánchez-Arcilla A 2011 The economic value of drought information for water management under climate change: a case study in the Ebro basin *Natural Hazards and Earth System Sciences* **11** 643–57

[3] Boone A *et al* 2022 Land surface Interactions with the Atmosphere over the Iberian Semi-arid Environment (LIAISE) Project: overview of the Field Campaign intense phase *EGU22* (https://doi.org/10.5194/egusphere-egu22-8028)

[4] Conde C d, Vallvé J L G and Gutiérrez S C 2019 Los sistemas automáticos de información hidrológica (SAIH) una innovación que se exporta *Revista Digital del Cedex* **193** 101–6 https://ingenieriacivil.cedex.es/index.php/ingenieria-civil/article/view/2382

[5] Sadki M, Munier S, Boone A and Ricci S 2023 Implementation and sensitivity analysis of the Dam-Reservoir OPeration model (DROP v1.0) over Spain *Geosci. Model Dev.* **16** 427–48

[6] Wang F, Polcher J, Peylin P and Bastrikov V 2018 Assimilation of river discharge in a land surface model to improve estimates of the continental water cycles *Hydrol. Earth Syst. Sci.* **22** 3863–82

[7] Tarpanelli A, Barbetta S, Brocca L and Moramarco T 2013 Remote sensing river discharge estimation by using altimetry data and simplified flood routing modeling *Remote Sens* **5** 4145–62

[8] Paris A, Dias de Paiva R, Santos da Silva J, Medeiros Moreira D, Calmant S, Garambois P A, Collischonn W, Bonnet M P and Seyler F 2016 Stage-discharge rating curves based on satellite altimetry and modeled discharge in the Amazon basin *Water Resour. Res.* **52** 3787–814

[9] Tourian M J, Schwatke C and Sneeuw N 2017 River discharge estimation at daily resolution from satellite altimetry over an entire river basin *J. Hydrol.* **546** 230–47

[10] Tarpanelli A, Amarnath G, Brocca L, Massari C and Moramarco T 2017 Discharge estimation and forecasting by MODIS and altimetry data in Niger-Benue River *Remote Sens. Environ.* **195** 96–106

[11] Anh D T L and Aires F 2019 River Discharge Estimation based on Satellite Water Extent and Topography: An Application over the Amazon *Journal of Hydrometeorology* **20** 1851–66

[12] Pellet V, Aires F, Yamazaki D, Zhou X and Paris A 2022 A first continuous and distributed satellite-based mapping of river discharge over the Amazon *J. Hydrol.* **614** 128481

[13] Filippucci P, Mosaffa H, Brocca L and Massari C 2023 High spatial and temporal resolution precipitation over Mediterranean basin for Digital Twin Earth Hydrology and 4dMED projects *EGU* (https://doi.org/10.5194/egusphere-egu23-15320)

[14] Miralles D G, De Jeu R A M, Gash J H, Holmes T R H and Dolman A J 2011 Magnitude and variability of land evaporation and its components at the global scale *Hydrol. Earth Syst. Sci.* **15** 967–81

[15] Martens B, de Jeu R A, Verhoest N E, Schuurmans H, Kleijer J and Miralles D G 2018 Towards estimating land evaporation at field scales using GLEAM *Remote Sensing* **10** 1720

[16] Pellet V, Aires F, Alfieri L and Bruno G 2024 A physical/statistical data-fusion for the dynamical downscaling of GRACE data at daily and 1 km resolution *J. Hydrol.* **628** 130565

[17] Alsdorf D E, Rodríguez E and Lettenmaier D P 2007 Measuring surface water from space *Rev. Geophys.* **45** 2002

[18] Durand M, Fu L L, Lettenmaier D P, Alsdorf D E, Rodriguez E and Esteban-Fernandez D 2010 The surface water and ocean topography mission: Observing terrestrial surface water and oceanic submesoscale eddies *Proc. IEEE* **98** 766–79

[19] Biancamaria S, Lettenmaier D P and Pavelsky T M 2016 The SWOT mission and its capabilities for land hydrology *Surv. Geophys.* **37** 307–37

[20] Xie P, Yatagai A, Chen M, Hayasaka T, Fukushima Y, Liu C and Yang S 2007 A Gauge-Based Analysis of Daily Precipitation over East Asia *Journal of Hydrometeorology* **8** 607–26

[21] Chen M, Shi W, Xie P, Silva V B, Kousky V E, Higgins R W and Janowiak J E 2008 Assessing objective techniques for gauge-based analyses of global daily precipitation *Journal of Geophysical Research Atmospheres* **113**

[22] Huffman G J, Bolvin D T, Nelkin E J and Tan J 2019 *Integrated Multi-satellitE Retrievals for GPM (IMERG) Technical Documentation* NASA Goddard Space Flight Center

[23] Karger D N, Lange S, Hari C, Reyer C P O and Zimmermann N E 2022 HELSA-W5E5 v1.0: W5E5 v1.0 downscaled with CHELSA v2.0 *ISIMIP Repository* (https://doi.org/10.48364/ISIMIP.836809.3)
[24] Martens B, Miralles D G, Lievens H, van der Schalie R, de Jeu R A M, Férnandez-Prieto D, Beck H E, Dorigo W A and Verhoest N E C 2016 GLEAM v3: satellite-based land evaporation and root-zone soil moisture *Geoscientific Model Development* **10** 1903–25
[25] Priestley C H B and Taylor R J 1972 On the assessment of surface heat flux and evaporation using large-scale parameters *Mon. Weather Rev.* **100** 81–92
[26] DiMiceli C, Carroll M, Sohlberg R, Kim D, Kelly M and Townshend J 2015 MOD44B MODIS/Terra Vegetation Continuous Fields Yearly L3 Global 250m SIN Grid V006 [Data set] *NASA EOSDIS Land Processes DAAC* (http://doi.org/10.5067/MODIS/MOD44B.006)
[27] Simons G, Koster R and Droogers P 2020 *HiHydroSoil v2.0-high resolution soil Maps of global hydraulic properties* 213 FutureWater Report 1–18
[28] Van der Schalie R, De Jeu R, Parinussa R, Rodríguez-Fernández N, Kerr Y, Al-Yaari A, Wigneron J-P and Drusch M 2018 The effect of three different data fusion approaches on the quality of soil moisture retrievals from multiple passive microwave sensors *Remote Sens.* **10** 107
[29] Lievens H *et al* 2019 Snow depth variability in the Northern Hemisphere mountains observed from space *Nat. Commun.* **10** 4629
[30] Avanzi F, Gabellani S, Delogu F, Silvestro F, Cremonese E, Morra di Cella U, Ratto S and Stevenin H 2022 Snow Multidata Mapping and Modeling (S3M) 5.1: a distributed cryospheric model with dry and wet snow, data assimilation, glacier mass balance, and debris-driven melt *Geosci. Model Dev.* **15** 4853–79
[31] Tapley B D, Bettadpur S, Watkins M and Reigber C 2004 The gravity recovery and climate experiment: mission overview and early results *Geophys. Res. Lett.* **31**
[32] Watkins M and Yuan D-N 2014 GRACE gravity recovery and climate experiment JPL Level-2 processing standards document for level-2 product release 05.1 *Technical Document* JPL https://icgem.gfz-potsdam.de/docs/L2-JPL_ProcStds_v5.1.pdf
[33] Verdin K L 2017 Hydrologic derivatives for modeling and analysis? new global high-resolution database *Data Series* 1053 USGS Reston, VA (https://doi.org/10.3133/ds1053)
[34] Silvestro F, Gabellani S, Delogu F, Rudari R and Boni G 2013 Exploiting remote sensing land surface temperature in distributed hydrological modelling: The example of the Continuum model *Hydrol. Earth Syst. Sci.* **17** 39–62
[35] Giannoni F, Roth G and Rudari R 2000 A semi-distributed rainfall-runoff model based on a geomorphologic approach *Phys. Chem. Earth Part B* **25** 665–71
[36] Alfieri L *et al* 2022 High-resolution satellite products improve hydrological modeling in northern Italy *Hydrol. Earth Syst. Sci.* **26** 3921–3939
[37] Aires F 2014 Combining datasets of satellite-retrieved products. Part I: methodology and water budget closure *Journal of Hydrometeorology* **15** 1677–91
[38] Pellet V, Aires F, Munier S, Fernández Prieto D, Jordá G, Arnoud Dorigo W, Polcher J and Brocca L 2019 Integrating multiple satellite observations into a coherent dataset to monitor the full water cycle—application to the Mediterranean region *Hydrol. Earth Syst. Sci.* **23**
[39] Pan M and Wood E F 2006 Data assimilation for estimating the terrestrial water budget using a constrained ensemble kalman filter *Journal of Hydrometeorology* **7** 534–47
[40] Munier S, Aires F, Schlaffer S, Prigent C, Papa F, Maisongrande P and Pan M 2014 Combining data sets of satellite-retrieved products for basin-scale water balance study: 2. Evaluation on the Mississippi Basin and closure correction model *Journal of Geophysical Research-Atmospheres* **119** 12
[41] Dari J *et al* 2023 Regional data sets of high-resolution (1 and 6 km) irrigation estimates from space *Earth System Science Data* **15** 1555–75
[42] Nearing G S *et al* 2021 What role does hydrological science play in the age of machine learning? *Water Resour. Res.* **57** e2020WR028091
[43] Kratzert F, Klotz D, Brenner C, Schulz K and Herrnegger M 2018 Rainfall-runoff modelling using Long Short-Term Memory (LSTM) networks *Hydrol. Earth Syst. Sci.* **22** 6005–22
[44] Do H X, Gudmundsson L, Leonard M and Westra S 2018 The Global Streamflow Indices and Metadata Archive (GSIM)-Part 1: The production of a daily streamflow archive and metadata *Earth System Science Data* **10** 765–85
[45] Kratzert F *et al* 2023 Caravan—A global community dataset for large-sample hydrology *Scientific Data* **10** 1–11
[46] Aires F and Pellet V Synergizing AI and physical expertise to close the water budget from satellite data *Journal of Hydrometeorology* (https://doi.org/10.1175/JHM-D-23-0001.1)
[47] Prigent C, Lettenmaier D P, Aires F and Papa F 2016 Toward a high-resolution monitoring of continental surface water extent and dynamics, at global scale: from GIEMS (Global Inundation Extent from Multi-Satellites) to SWOT (surface water ocean topography) *Surv. Geophys.* **37** 339–55