



**HAL**  
open science

# Discharge estimation in ungauged basins through variational data assimilation: the potential of the SWOT mission

H. Oubanas, I. Gejadze, Pierre-Olivier Malaterre, M. Durand, R. Wei, R. P.  
M. Frasson, A. Domeneghetti

► **To cite this version:**

H. Oubanas, I. Gejadze, Pierre-Olivier Malaterre, M. Durand, R. Wei, et al.. Discharge estimation in ungauged basins through variational data assimilation: the potential of the SWOT mission. *Water Resources Research*, 2018, 54 (3), pp.2405-2423. 10.1002/2017WR021735 . hal-02608723

**HAL Id: hal-02608723**




**<https://hal.inrae.fr/hal-02608723v1>**

Submitted on 16 May 2020

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Discharge estimation in ungauged basins through variational data assimilation: The potential of the SWOT mission

H. Oubanas<sup>1</sup> , I. Gejadze<sup>1</sup>, P-O. Malaterre<sup>1</sup>, M. Durand<sup>2,3</sup> , R. Wei<sup>3</sup>, R.P.M Frasson<sup>3</sup> , A. Domeneghetti<sup>4</sup>

<sup>1</sup>National Research Institute of Science and Technology for Environment and Agriculture, Montpellier, France

<sup>2</sup>School of Earth Sciences, The Ohio State University, Columbus, Ohio, USA

<sup>3</sup>Byrd Polar and Climate Research Center, The Ohio State University, Columbus, Ohio, USA

<sup>4</sup>School of Civil Engineering, Department DICAM, University of Bologna, Bologna, Italy

## Key Points:

- = River discharge is estimated using synthetic remote sensing measurements from the forthcoming SWOT mission without any in-situ observation =
- = 2D SWOT radar measurements are simulated and mapped onto river centerlines, producing realistic errors, temporal and spatial sampling =
- = Discharge accuracy was 12.1% and 11.2% on the Po and Sacramento Rivers, respectively, illustrating potential for ungauged basins=

---

Corresponding author: Hind Oubanas, [hind.oubanas@irstea.fr](mailto:hind.oubanas@irstea.fr)

## Abstract

Space-borne instruments can measure river water surface elevation, slope and width. Remote sensing of river discharge in ungauged basins is far more challenging, however. This work investigates the estimation of river discharge from simulated observations of the forthcoming Surface Water and Ocean Topography (SWOT) satellite mission using a variant of the classical variational data assimilation method “4D-Var”. The variational assimilation scheme simultaneously estimates discharge, river bathymetry and bed roughness in the context of a 1.5D full Saint Venant hydraulic model. Algorithms and procedures are developed to apply the method to fully ungauged basins. The method was tested on the Po and Sacramento Rivers. The SWOT hydrology simulator was used to produce synthetic SWOT observations at each overpass time by simulating the interaction of SWOT radar measurements with the river water surface and nearby land surface topography at a scale of approximately 1 m, thus accounting for layover, thermal noise and other effects. SWOT data products were synthesized by vectorizing the simulated radar returns, leading to height and width estimates at 200 m increments along the river centerlines. The ingestion of simulated SWOT data generally led to local improvements on prior bathymetry and roughness estimates which allowed the prediction of river discharge at the overpass times with relative root-mean-squared errors of 12.1% and 11.2% for the Po and Sacramento rivers respectively. Nevertheless, equifinality issues that arise from the simultaneous estimation of bed elevation and roughness may prevent their use for different applications, other than discharge estimation through the presented framework.

## 1 Introduction

As a primary source of fresh water, rivers are among the most important natural resources and are the nerve system of ecology and human society. Since the earliest civilizations, continental waters have been central to society, featuring in applications such as water supply, irrigation, drainage, navigation, fisheries, flood control, hydropower generation, wastewater treatment, pollution abatement and wildlife protection [Chow *et al.*, 1964; Benke and Cushing, 2011; Tockner *et al.*, 2009].

River discharge represents the flow of water from continental to oceanic environments, and is one of the primary quantities of interest in characterizing fluvial environments. Despite the undeniable importance of global rivers, the availability of the world’s in-situ gauge stations have been steadily declining since the late 1970s due to political, economical and geographical reasons [Fekete and Vörösmarty, 2002; Tourian *et al.*, 2013; Sneeuw *et al.*, 2014]. Moreover, the reluctance from countries to share data in a timely fashion makes the monitoring of discharge and the flood forecasting in international rivers a daunting task [Biancamaria *et al.*, 2011a; Hossain *et al.*, 2014]. Therefore, the need for alternative and/or complementary inland water measuring techniques, such as space-borne sensors, has become a primary concern of the scientific community and space agencies.

In this respect, several nadir altimeters that observe the water height are becoming widely available for the land surface applications, such as JASON-1, JASON-2, ENVISAT, CryoSat-2, SARAL/ALTIKA and the recent near-real time operational missions, JASON-3 and Sentinel-3 [Aldorf and Lettenmaier, 2003; Papa *et al.*, 2006; Crétau and Birkett, 2006; Calmant and Seyler, 2006; Crétau *et al.*, 2011]. Most of them are focused on observing water surface elevation (WSE) in large rivers and lakes [Birkett, 1995, 1998; de Oliveira Campos *et al.*, 2001; Papa *et al.*, 2003; Kouraev *et al.*, 2004; Frappart *et al.*, 2006; Medina *et al.*, 2008; Bartsch *et al.*, 2009].

The forthcoming SWOT mission, scheduled for launch in 2021, will provide a global mapping of the continental water bodies using the wide swath radar interferometer KaRIN (Ka-band Radar Interferometer). SWOT will observe two ground swaths of 50 km separated by a 20 km nadir gap and will provide observations of water surface elevation, slope and river width of rivers wider than 100 m, with a possibility of providing accurate ob-

servations of rivers as narrow as 50 m [Rodriguez, 2016]. The SWOT repeat cycle is 21 days, during which sites will be revisited typically two to four times at irregular intervals. The temporal frequency is dependent on the latitude of the region of interest, with low observation frequency near the equator and higher observation frequency in higher latitude [Biancamaria et al., 2010, 2016].

Several algorithms designed to use future SWOT observations for river discharge estimation have been proposed. Among the proposed methodologies, data assimilation (DA), has become increasingly popular within the hydrological and hydraulic communities. DA methods allow the combination of available observations, prior knowledge and expertise, and natural system dynamics, with a treatment of the associated errors, to provide the best estimates of the unknown system variables and/or parameters. Early DA studies assumed the river bathymetry and the bed roughness to be known a priori [Andreadis et al., 2007; Biancamaria et al., 2011b]. Durand et al. [2008] and Yoon et al. [2012] estimated the river bathymetry, but assumed that the bed roughness was well known. The complexity of implementing DA methods with dynamical models and the difficulty of estimating discharge together with the river bathymetry and the bed roughness have led to the use of simplified models derived from the Saint-Venant equations. Four such algorithms have been presented in the literature, namely Garambois and Monnier (GaMo) [Garambois and Monnier, 2015], Metropolis Manning (MetroMan) [Durand et al., 2014], the Mean-Annual Flow and Geomorphology (MFG) [Durand et al., 2016] and Mean Flow with Constant Roughness (MFCR) [Durand et al., 2016]. These algorithms were assessed in Durand et al. [2016] for 19 rivers with various hydraulic conditions. At least one algorithm (although not always the same one) had a discharge estimate relative root-mean-squared error less than 35%, on the 14 non-braided rivers included in the study.

Among DA methods, variational DA has been the preferable approach in operational geophysical applications, such as numerical weather prediction (NWP) [Gauthier et al., 1999, 2007; Courtier et al., 1998; Rabier et al., 2000; Fischer et al., 2005]. Within the meteorology and oceanography communities, these methods are commonly named in their classical form '3D-Var' and its temporal extension '4D-Var'. The problem is formulated as an optimal control problem and can be considered as a special case of the maximum a-posteriori probability estimator (MAP) in the Bayesian framework. The optimal estimates of the unknown model variables and/or parameters such as initial and/or boundary conditions, source terms (forcing), distributed and/or lumped coefficients (e.g. roughness coefficient, bathymetry, etc.), are obtained via the minimization of a well-defined cost function. Gradient-based optimization methods resolve the minimization problem (e.g. quasi-Newton, conjugate gradient) and require the derivation of the tangent linear and adjoint models associated with the considered direct model. References on variational DA applied to the 1D full Saint-Venant model are rather scarce [Ding and Wang, 2012]. The adjoint model in the latter reference is derived analytically, then implemented numerically; *i.e.* 'optimize-then-discretize' approach, which yields an 'inconsistent' adjoint. The disadvantage of the latter approach is lower accuracy of the gradient and difficulty in applying in the framework of constrained optimization. The consistent adjoint for the full Saint-Venant hydraulic model SIC<sup>2</sup> (Simulation and Integration of Control for Canals) has been reported only recently in Gejadze and Malaterre [2017, 2016] and Oubanas et al. [2015, 2016]. This model has been developed at the National Research Institute of Science and Technology for the Environment and Agriculture (IRSTEA). The difficulty behind the computation of the stable adjoint model is one reason why variational DA has not been very popular within the hydraulic research community. In this case, alternative filtering methods; *i.e.* the Kalman filter and its ensemble extensions, are widely used instead. Note that for two-dimensional (2D) Shallow Water equations based models, variational DA has been reported in Lai and Monnier [2009] and Hostache et al. [2010] for different problem set-ups, *e.g.* single variable estimation, data types, temporal and spatial scales.

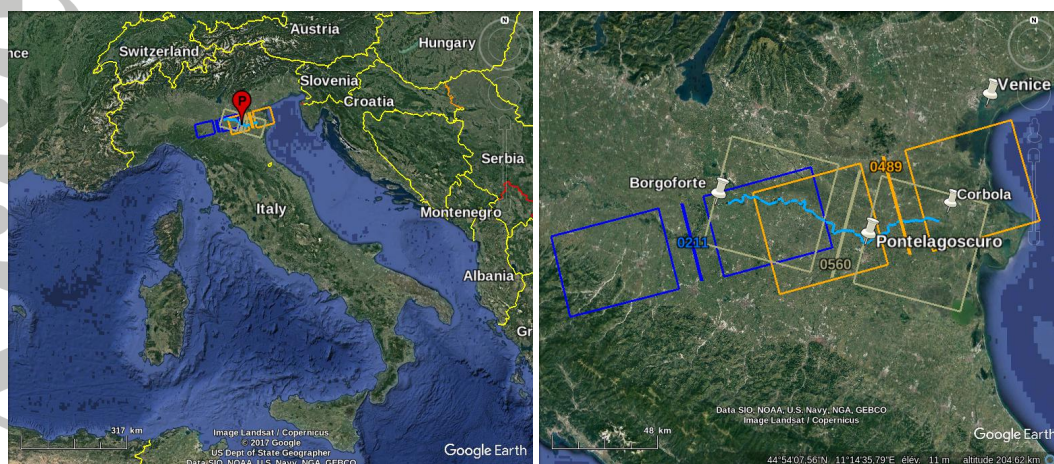
In the present paper, a variant of the classical '4D-Var' method, presented in *Gejadze and Malaterre* [2017, 2016], has been adapted to the general and realistic case of ungauged rivers observed from space. This method, based on the iterative regularization technique, is more suitable for the nonlinear systems where heterogeneous variables, e.g. the state (discharge), the parameters (roughness) and the domain (bathymetry), are estimated simultaneously. Such configuration, involving nonlinear operators, may compromise the robustness of the minimization problem (due to large norm differences among gradient components), which requires appropriate weight assignment to different contributions. Using synthetic SWOT observations produced by the SWOT simulator, we demonstrate the simultaneous estimation of discharge, river bathymetry and roughness for one year over a 133 km section of the Po River and for a 6-month period over a 153 km section of the Sacramento River. Note that the prior knowledge of the variables of interest, needed for the optimization method, is derived from the SWOT observations and globally-available ancillary information only which enables the application of the proposed method for ungauged as well as gauged basins.

The outline of paper is as follows. In Section 2 we present the SWOT simulator and the study areas and period. The methodology is described in Section 3 introducing the 1.5D full Saint-Venant hydraulic model, the variational data assimilation method with its sequential version, and handling of a priori information. Experimental design is presented in Section 4. Section 5 presents and discusses the results of numerical experiments. Finally, main findings of this study are summarized in the Conclusion.

## 2 Study area, period and simulated data

### 2.1 The Po River

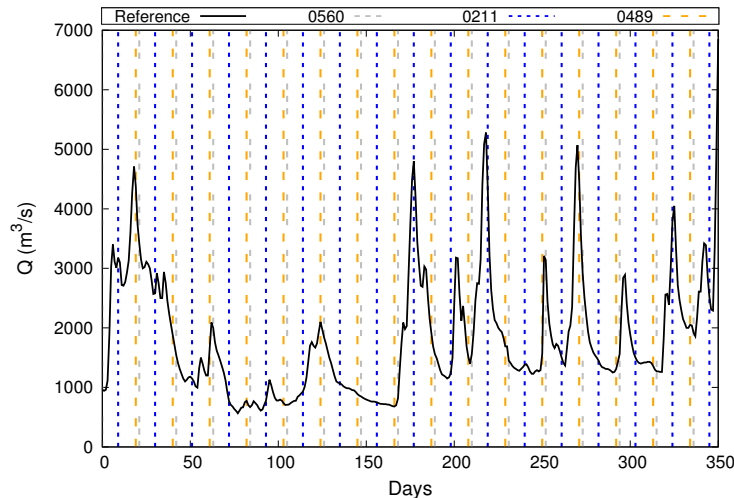
The Po River is the longest river that flows entirely in the Italian peninsula, running from the northeastern Alps to the Adriatic Sea. The mean channel is about 650 km long with over 141 tributaries, draining an area of approximately 71000 km<sup>2</sup>, corresponding to the largest catchment in Italy. The Po River is also the largest Italian river in terms of the streamflow with a maximum historical discharge of 13000 m<sup>3</sup>s<sup>-1</sup> observed at Pontelagoscuro in 1951. The Po Valley has experienced intensive agricultural and industrial development during the 20th century which have led to an increasing vulnerability to hydrological hazards [*Montanari A.*, 2017].



**Figure 1.** The Po River study area (© Google Earth) with the ground track of SWOT overpasses (50 km swaths at each side of the 20 km nadir gap 0560 (grey), 0211 (blue) and 0489 (yellow)).



The modeled stretch of the river is 133 km in length, and is located between the gage of Borgoforte at the upstream boundary and the city of Corbola at the beginning of the Po River delta (Figure 1). It is observed by three passes during each of the 21-day SWOT cycle, for which the corresponding SWOT simulations are available; *i.e.* the left and right swaths of the pass 0560, the left swath of the pass 0211 and the right swath of the pass 0489 (see Figure 1). Note that ascending passes are labeled with pair identification number while odd numbers refer to descending passes. The study period is one year long, ranging from May 2008 to April 2009 during which a hydrodynamic model that solves the Saint-Venant equations was applied to the Po River. It uses the flow hydrograph as an upstream boundary condition to simulate the flow at a daily scale and is conditioned downstream by the observed water surface elevation. The flow hydrograph reported in Figure 2 shows a discharge variability from  $565 \text{ m}^3\text{s}^{-1}$  to  $6850 \text{ m}^3\text{s}^{-1}$ . In this Figure, the SWOT temporal sampling is illustrated by the vertical lines, which represent the time of the SWOT overpasses, with the line color identifying the overpass identity as defined in Figure 1. Note that these lines serve as an indication of the time of the SWOT observations that are assimilated to estimate the upstream discharge at Borgoforte. They are spatially distributed dependent on the corresponding swath coverage.

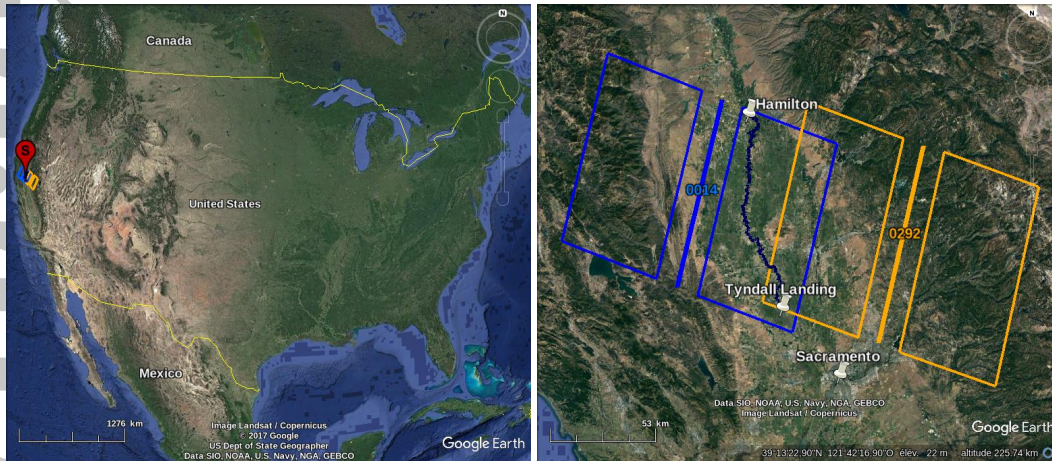


**Figure 2.** Daily discharge data at the Borgoforte streamgage (black) at the Po River during the study period with vertical dashed lines indicating the timing of the SWOT overpasses.

## 2.2 The Sacramento River

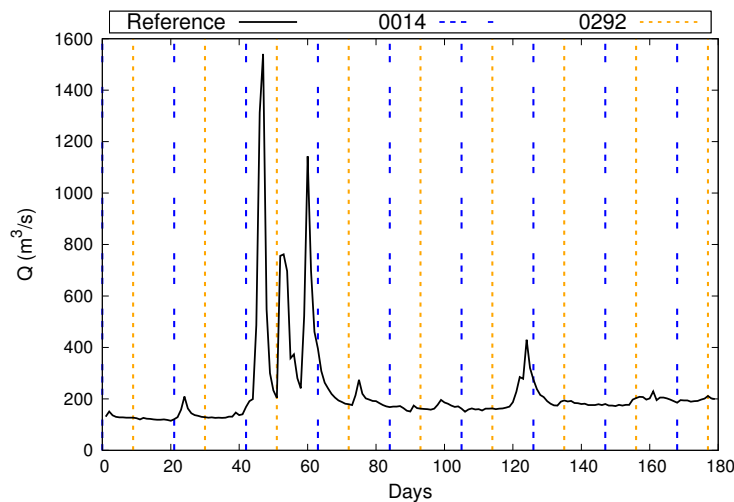
The Sacramento River rises in the Klamath Mountains and drains an area of 69000  $\text{km}^2$  in northern California including the Coast Range, the Sierra Nevada, the Modoc Plateau and the Trinity Mountains. This river is known to be the largest in the state flowing for 640 km long and the major freshwater source for the San Francisco Estuary. Since early 1940s, the flow of the Sacramento River has been regulated by dams to control the magnitude of the flow discharge and the frequency of the flooding events [Buer *et al.*, 1989].

The region of interest is 153 km long and is located between Hamilton and Tyndall Landing cities, upstream of the city of Sacramento. The study area is observed entirely by the right swath of the pass 0014, while the far range portion of the left swath of the pass 0292 intersects a small downstream section of the study area (see Figure 3). The study period is a 6-month simulation from January 2009 to June 2009 conducted with a one-



**Figure 3.** The Sacramento River study area (© Google Earth) with the ground track of SWOT overpasses (50 km swaths at each side of the 20 km nadir gap ) 0014 (blue), 0292 (yellow).

dimensional hydraulic model of the basin released in 2013 by the California Department of Water Resources as part of the Central Valley floodplain evaluation and delineation program [Rogers, 2014]. The model has been simplified to remove diversions, tributaries and storage cells. The hydrograph during this period is presented in Figure 4, covering a range of discharge from  $115.31 \text{ m}^3\text{s}^{-1}$  to  $1541.97 \text{ m}^3\text{s}^{-1}$ . Similarly, the SWOT temporal sampling is illustrated by the vertical lines, with the line color identifying the overpass identity as defined in Figure 3.



**Figure 4.** Daily discharge data at Hamilton streamgage (black) during the study period with vertical dashed lines indicating the timing of the SWOT overpasses.

### 2.3 Deriving 2-D hydraulic properties from 1-D hydraulic simulations

Accurate SWOT simulation requires high-resolution elevation maps, whereas the hydraulic simulations for both the Po and Sacramento were performed in one dimension us-

ing different models. The Po River was simulated by a quasi-two-dimensional hydraulic model built from a combination of a 1 m or 2 m resolution LiDAR Digital Elevation Model (DEM) of the area and boat surveys of the river [Castellarin *et al.*, 2011; Domeneghetti *et al.*, 2014] leading to 111 cross-sections irregularly spaced, with an average distance between the cross-sections of 1.2 km. The quasi-2D scheme depends on the presence of floodplain areas which are protected from frequent inundations by a system of minor embankments and are connected to the main channel by means of lateral structures. The model considers these floodplains as storage areas in which the water level is controlled by a level-volume curve estimated from the DEM and depends on the flow exchange between the main channel and the storage areas (see Castellarin *et al.* [2011] for more details about the model). Similarly, the Sacramento River hydraulic model was built using a combination of airborne LiDAR with land and boat surveys of the area leading to 601 cross-sections, irregularly spaced with an average distance of 258 m. In both cases the estimation of the 2D water surface elevation required by SWOT simulator has been performed interpolating the water elevation associated with all cross-sections to a regular grid. In particular, the interpolation refers to water elevations mapped laterally perpendicular to the river centerline and has been performed with tools suitable for such spatial interpolation (such as e.g. HEC-GeoRAS). These operations resulted in temporally dynamic high-resolution maps of water inundation and elevation, combined with topographic features.

## 2.4 SWOT Simulator

The SWOT simulator designed by the Jet Propulsion Laboratory (JPL) is used for the first time in this work for DA purposes to simulate the expected performance of the KaRIn instrument on-board the SWOT satellite. It was used for the first time by our group in this paper for DA purposes. The SWOT simulator takes as inputs the SWOT orbit and radar parameters, the expected water-land radar contrast, the DEM of the study area including water and terrain, and a water mask, *i.e.* a two dimensional map that distinguishes between inundated areas and dry land. The water DEM, *i.e.* a digital elevation model that contains only the elevation of the water surface, and the water mask are built based on the results of the dynamic hydraulic simulations of the rivers at the time of the satellite overpass. As a first step, the SWOT simulator builds synthetic radar interferograms of the scene, containing no noise and only subjected to terrain layover errors, *i.e.* errors that happen when the radar return from two or more distinct targets, generally water and surrounding terrain at higher elevations, reach the satellite at the same time, leading to overestimation of the water surface elevation (for more information of terrain layover, refer to Fjortoft *et al.* [2014]). Subsequently, the simulator adds noise to the synthetic interferograms, which are modeled as correlated circular Gaussian noise. The noise added to the target depends on the surface type, *i.e.* land or water, the position in the swath, and radar parameters. As the next step, the simulator processes the noisy interferograms to create pixel cloud, which is composed of a pixel cloud with target class, elevation, and area associated with the pixel.

The processing of the noisy interferograms entails the following steps: a smoothing procedure called multi-looking, target classification, and geolocation. The multi-looking step is a procedure that averages the returned power from consecutive pixels in the along-track and cross-track dimensions, effectively decreasing noise at the cost of reduced spatial resolution [Cuchi, 1986; Ulaby *et al.*, 2014]. The number of looks used in the present study was 4, which is currently envisioned as the lowest level of smoothing for SWOT data products. Next, the simulator classifies the target based on the returned power. The classification relies on the expectancy that for SWOT band and antenna characteristics, water will appear brighter than land targets [Alsdorf and Lettenmaier, 2003; Fjortoft *et al.*, 2014]. Ulaby and Dobson [1989] provides a list of expected backscatter coefficient for different targets when viewed by a range of viewing angles. Given SWOT's viewing angle, the contrast between water and land is conservatively on the order of 10 dB [Bian-



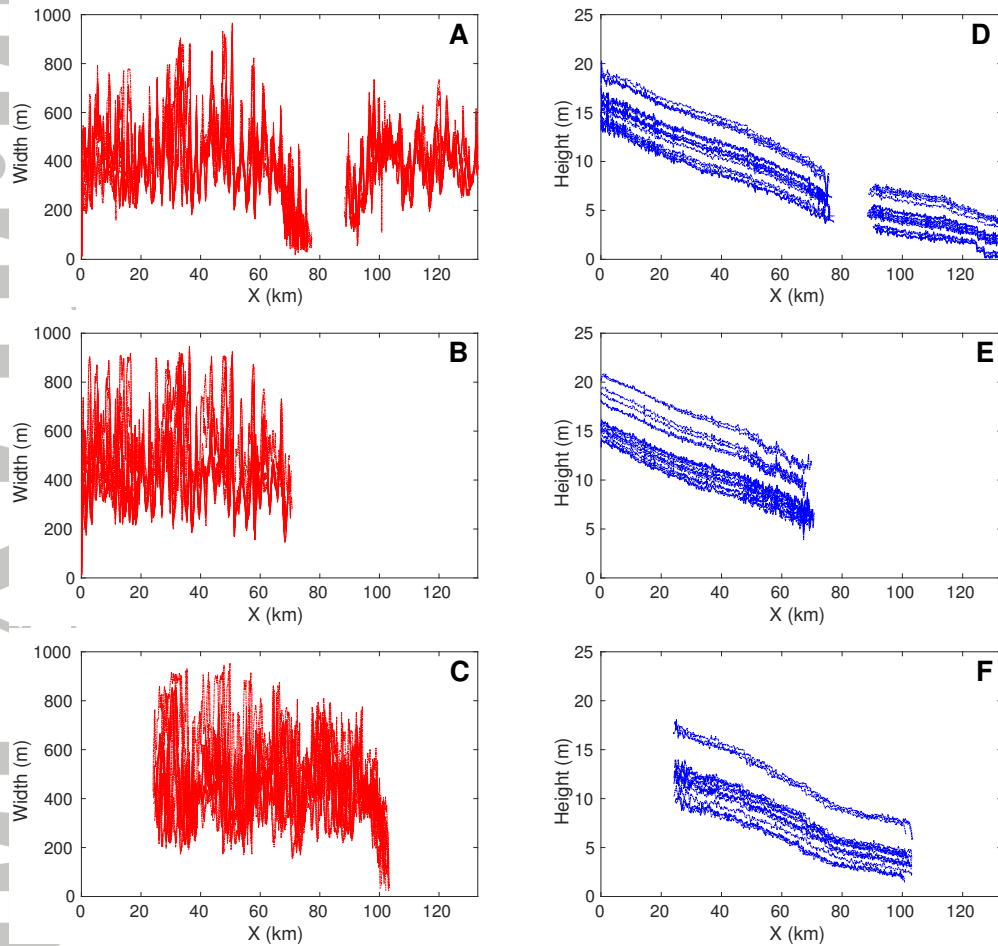
*camaria et al.*, 2016], which would allow differentiation based on thresholding. However, sharp corners in buildings may appear as bright as water, as seen the images published by *Fjortoft et al.* [2014]. Filtering of false detections can be done by searching for nearby bodies of water or by implementing more advanced classification techniques. Finally, the simulator geolocates the targets, translating the pixels from radar coordinates into elevation, latitude and longitude (for more information of terrain layover, refer to *Fjortoft et al.* [2014]).

## 2.5 Deriving 1-D simulated observations from 2-D SWOT simulations

The pixel clouds were processed with the RiverObs package developed at JPL. RiverObs aggregates the 2D pixel clouds produced by the SWOT simulator into regularly spaced points, called nodes, located at the river centerline and estimates node-averaged height, width, and associated observational uncertainties. RiverObs produces node statistics by assigning pixels located within a user-defined search window to the nearest river node. The search window is a polygon with outer boundaries running parallel to the river centerline. In the present work; we utilized a search window of 1200 m for the Sacramento River which corresponds 6-12 times the averaged width of the river. For the Po River, we used a 1600 m-wide search window for all but the three highest flow overpasses, for which we increased the width to 5000 m to account for overbank flow. This represents 4 times the averaged width during the low flow and 9 times during high flow [*Frasson et al.*, 2017]. The node height assumes the value of the average of all water pixel heights associated with that node whereas the node width is estimated by dividing the inundated area associated with the node by the node spacing. We use the synthetic SWOT node-averaged products within our data assimilation framework to estimate river discharge as described in sections 3.2 and 3.3. The Po River is observed 52 times during the one-year study period while the Sacramento study area is observed by 18 overpasses during the 6-month study period. The simulated node-averaged water surface elevation and the top river width, along the Po and Sacramento Rivers are presented in Figure 5 and 6. In the case of the Sacramento River, the width variations are smaller after 90 km due to the presence of levees which confine the river during our simulations. In the first 90 km, the river is wider which allows to decrease the random error in the WSE measurements. The errors are more important in the downstream part because the river is narrower. Moreover, the presence of the levees increases the layover errors.

The simulated observations are subjected to biases, although not all the sources of systematic errors are taken into account. The bias depends on the orientation of the river with respect to the satellite swath; it is more important when the river is parallel to the swath (*e.g.* case of the Sacramento River) and is negligible when the river is perpendicular to the swath (*e.g.* case of the Po River). For more information about the characteristics of the SWOT data, one can refer to *Fernandez et al.* [2013]; *Frasson et al.* [2017]. It should be mentioned that the SWOT measurements are also subject to systematic and spatially correlated errors introduced by different factors (*e.g.* satellite roll, wet troposphere, etc.). These types of errors are not addressed in the present paper and will be subject to future investigation.

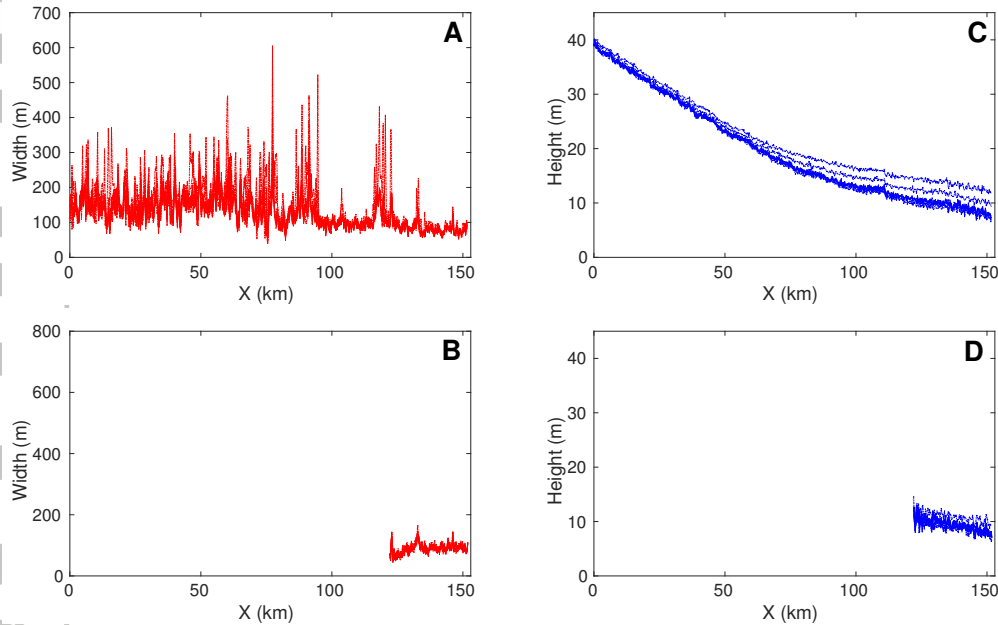
RiverObs provides the standard deviation of pixel heights associated to each node, at each SWOT overpass. No information about correlation between noise in different pixels is currently available, thus we cannot properly define the nodes-associated observation error covariance matrix, which is actually required for a classical DA algorithm. This is why a simplified representation of the observation error covariance matrix must be used. Let us note that in our DA scheme, presented in Section 3.2, it is sufficient to know the standard deviation at nodes up to a scaling factor. That is why we can rely on the standard deviation of the pixel. Moreover, if it is assumed a constant for all nodes, its value does not affect the result of DA. When more detailed information (correlation between pixels/nodes)



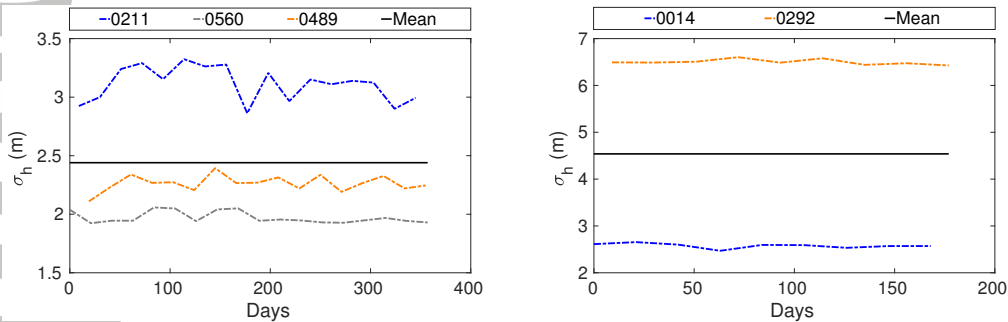
**Figure 5.** Panels A, B, and C show the river width measured at the nodes for the Po River using the simulated SWOT passes 0560 (18 superimposed revisits), 0211 (17 revisits), and 0489 (17 revisits) respectively. Panels D, E, and F show the water surface elevation at nodes produced from the passes 0560, 0211, and 0489 respectively.

is available, we will introduce the observation covariance matrix that handles the temporal and spatial variability of the error.

Figure 7 presents an average in space for each overpass during the study period, for the Po and the Sacramento Rivers. The averaged standard deviation in space and time is about  $\sigma_z = 2.5$  m in the case of the Po River, while the averaged uncertainty is higher in the case of the Sacramento River,  $\sigma = 4.5$  m. In fact, the observations of the pass 0292 are located in the far range of the swath and have higher errors due to the low signal return. The corresponding errors are not representative of the observation uncertainty, therefore, we only consider the average standard deviation associated with the observations of the pass 0014, which is about  $\sigma_z = 2.5$  m.



**Figure 6.** Panels A and B show the river width measured at the nodes for the Sacramento River using the simulated SWOT passes 0014 (9 revisits) and 0292 (9 revisits) respectively. Panels C and D show the water surface elevation at nodes produced from the passes 0014 and 0292 respectively.



**Figure 7.** The standard deviation associated to the simulated WSE averaged over the RiverObs nodes for each pass, and the mean of all the passes, for the Po (left) and the Sacramento (right) Rivers.

### 3 Methodology

#### 3.1 Hydraulic model SIC<sup>2</sup>

The 1.5D hydraulic model SIC<sup>2</sup>, based on the full Saint-Venant equations, has been under development at IRSTEA for about 30 years, succeeding the former CEMAGREF hydraulic models (Talweg-Fluvia-Sirene) [<http://sic.g-eau.net/>]. The longitudinal and transversal hydraulic effects are described at a finite number of cross-sections along the reach using the compound channel bathymetry representation. A simplified representation of the out-of-bank flow involves connected storage areas. For each cross-section, the hydraulic variables such as the wetted area  $A(z, p_g)$ , the wetted perimeter  $P(z, p_g)$ , the hydraulic radius  $R(z, p_g)$  and the top width  $L(z, p_g)$  are computed, for a given water surface elevation  $z$ . The  $p_g(x)$  refers to the parameters which define the geometry of the corresponding computational cross-section, *i.e.* bottom width  $l$ , bank slope  $b$  and the bed elevation  $Z_b$ .

with respect to a chosen reference level, in the case of trapezoidal approximation. Note that SIC<sup>2</sup> can handle any type of cross-sections including irregular ones. Later on, when talking about the bathymetry estimation we mean the estimation of the distributed vector  $Z_b(x)$ .

The flow dynamics in the longitudinal direction  $x$  at time  $t$  are described by the Saint-Venant equations:

$$\frac{\partial A}{\partial t} + \frac{\partial q}{\partial x} = Q_L, \quad (1)$$

$$\frac{\partial q}{\partial t} + \frac{\partial q^2/A}{\partial x} + gA \frac{\partial z}{\partial x} = -gAS_f + C_L Q_L v, \quad (2)$$

where  $q(x, t)$  is the local discharge,  $z(x, t)$  is the WSE,  $Q_L(x, t)$  is the lateral discharge,  $C_L(x) \in [0, 1]$  is the lateral discharge coefficient,  $v(x, t) = q/A$  is the mean longitudinal velocity and  $S_f$  is the friction slope defined as

$$S_f = \frac{q^2}{K_s^2 A^2 R^{4/3}} \quad (3)$$

which is related to the Strickler coefficient  $K_s(x)$  (inverse of the Manning coefficient) by equation 3. The initial state  $(z_0, q_0)$  and the boundary conditions (*i.e.* the inflow discharge  $Q(t)$  at the upstream node and a rating curve, defined by the rating curve parameters  $p_{rc}$ , at the downstream node) are needed to solve the equations (1)-(3). Note that the SIC<sup>2</sup> model supports different types of boundary conditions; here we present only those relevant to the chosen test cases. The four-point implicit finite-difference method, called the Preissmann scheme [Cunge *et al.*, 1980; Novak *et al.*, 2010], is utilized for discretizing the problem (1)-(3). The fixed-point iterations are used to resolve nonlinearity. For more details on the solver, refer to Malaterre *et al.* [2014].

In order to use the hydraulic model SIC<sup>2</sup> in a chosen river system, some inputs need to be provided. These include the cross-sectional geometrical parameters, roughness coefficients, upstream and downstream boundary conditions. In the framework of ungauged basins, we propose a methodology, presented in Section 3.3, to generate initial approximations of these inputs from SWOT observations (WSE, width and slope) and globally available ancillary information. From these inputs, the model SIC<sup>2</sup> simulates the WSE and the local discharge fields. Next, we solve the corresponding inverse problem using variational DA to improve approximations of the inputs of interest.

### 3.2 Variational data assimilation

Dynamical systems can be described by a numerical model, here denoted  $\mathcal{M} : \mathcal{U} \rightarrow \mathcal{X}$ , which maps the model inputs  $U \in \mathcal{U}$ , also called the “control vector”, into the model state  $X \in \mathcal{X}$ :

$$\mathcal{M}(U) = X. \quad (4)$$

The model state  $X$  is assumed to be related to observations via an observation operator  $\mathcal{H} : \mathcal{X} \rightarrow \mathcal{Y}$ , where  $\mathcal{Y}$  is the observation space:

$$Y = \mathcal{H}(X) \in \mathcal{Y}. \quad (5)$$

Therefore, the control-to-observation nonlinear mapping  $\mathcal{G} : \mathcal{U} \rightarrow \mathcal{Y}$  can be defined as:

$$\mathcal{H}(X) = \mathcal{H}(\mathcal{M}(U)) := \mathcal{G}(U) = Y. \quad (6)$$

Both observations  $Y$  and model inputs  $U$  embody uncertainties  $\xi_o$  and  $\xi_b$  that arise from various sources (e.g. instrumental noise, parametrization, discretization, etc.) and should be taken into account:

$$Y^* = Y^t + \xi_o, \quad (7)$$



where  $Y^t = \mathcal{G}(U^t)$  is the 'true' observation vector. And:

$$U_b = U^t + \xi_b, \quad (8)$$

where  $U^t$  is the true model input vector and  $U_b$  is its best available first guess, known as a 'background' in variational DA or as a 'prior' in Bayesian statistics.

The full input vector of the hydraulic model presented above consists of the following variables:

$$U = (z_0, q_0, Q, p_{rc}, Q_L, C_L, K_s, Z_b, p_g, p_{nm})^T, \quad (9)$$

where  $p_{nm}$  are the numerical scheme parameters. For a given  $U$ , we obtain the flow fields by solving the model equations (1)-(3) such that:

$$(z, q) = \{(z(x_i, t), q(x_i, t)), i = 1, \dots, N\}, t \in [0, T]. \quad (10)$$

In certain components of the input vector  $V \subset U$  the uncertainty is significant and strongly affects the model predictions. The aim of data assimilation is to estimate  $V$  using observations  $Y$  (i.e., to improve the first guess  $V_b$ ), whereas the remaining components  $U^0 = U \setminus V$  are fixed at their background value  $U_b^0$ .

In the present study, the inputs of interest are:

$$V = (Q(t), Z_b(x), K_S(x)), \quad (11)$$

and the assimilated observations are:

$$Y = \{z(x_i, t_j), i = 1, \dots, N_N, j = 1, \dots, N_P\}, \quad (12)$$

where  $N_N$  and  $N_P$  are respectively the numbers of RiverObs nodes and SWOT overpasses. In this case, the dynamical model  $\mathcal{M}$  refers to the code of the hydraulic model SIC<sup>2</sup>, while  $\mathcal{H}$  is a additional module representing the observation operator.

The variational DA method gives the best estimate of the control vector by minimizing a cost-function. Assuming that errors associated with the observation and the background information are Gaussian, i.e.  $\xi_b \sim N(0, B)$  and  $\xi_o \sim N(0, R)$ , where  $B$  and  $R$  are the corresponding covariance matrices, the conventional formulation of the DA problem is as follows:

$$\hat{V} = \underset{V}{\operatorname{argmin}} J(V), \quad (13)$$

where

$$J(V) = \frac{1}{2} \|R^{-1/2}(\mathcal{G}(V, U_b^0) - Y^*)\|^2 + \frac{1}{2} \|B^{-1/2}(V - V_b)\|^2. \quad (14)$$

A variant of the conventional variational DA method useful for the estimation of the model variables and parameters affected by uncertainties in hydraulic applications is described in detail in *Gejadze and Malaterre* [2017]. First, instead of (14) we consider a modified cost-function:

$$J(V, \alpha) = \frac{1}{2} \|R^{-1/2}(\mathcal{G}(V, U_b^0) - Y^*)\|^2 + \frac{\alpha}{2} \|B^{-1/2}(V - V_b)\|^2, \quad (15)$$

where  $\alpha > 0$  is a regularization parameter. This is done to reduce the impact of possible errors in assigning  $B$ . The cost function (15) is used in the Tikhonov regularization method [Tikhonov et al., 1977]. Second, we consider the change of variables  $V = V_b + B^{1/2}W$ , in which case the above cost-function takes the form:

$$J(W, \alpha) = \frac{1}{2} \|R^{-1/2}(\mathcal{G}(V_b + B^{1/2}W, U_b^0) - Y^*)\|^2 + \frac{\alpha}{2} \|W\|^2. \quad (16)$$

If the control vector  $V$  is composed of heterogeneous components (i.e. flow variables, physical parameters and parameters describing the domain geometry), then the corresponding parts of the gradient  $J'_V$  could have a very different norm. Since  $\mathcal{G}$  is a non-linear operator, this may compromise the robustness of the minimization process. The

change of variables helps to assign the appropriate weights to the gradient, see *Gejadze and Malaterre* [2017]. This is the advantage of the latter formulation as compared to the classical one where at first iterations the contribution of the background term is negligible (zero at first iteration). Let us mention that this change of variables serves as a preconditioning in the framework of the incremental approach of variational DA. To implement (16) one needs  $B^{1/2}$  instead of  $B^{-1}$  (or  $B^{-1/2}$ ) in the original formulation (14). This is obtained using the approach presented in *Gejadze and Malaterre* [2017], based on the assumption that the control variable is a distributed function of space or time which belongs to the Sobolev space of the second order. The square-root of the matrix  $B$  is obtained using Cholesky decomposition. This approach can be performed using modest computational resources in the case of one-dimensional hydraulic problems.

The optimal choice of the regularization parameter  $\alpha$  is a key issue in the Tikhonov regularization method. For example,  $\alpha$  can be chosen from the residual principle as follows:

$$J(\hat{W}) \sim \chi^2(M, \delta) \quad (17)$$

where  $M$  and  $\delta$  are the observation space dimension and the confidence level, respectively. This implies solving the minimization problem for different values of  $\alpha$  to satisfy the condition (17). This could be computationally expensive. An alternative is to use the iterative regularization method [*Kaltenbacher et al.*, 2008], in which the cost function to be minimized is reduced to its residual term:

$$J(W) = \frac{1}{2} \|R^{-1/2}(\mathcal{G}(V_b + B^{1/2}W, U_b^0) - Y^*)\|^2. \quad (18)$$

This should be combined with a stopping criterion in order to obtain a regularized solution by early termination of the iterations. We employ the residual principle (17) to stop the iterations in the iterative form:

$$J(\hat{W}_i) \sim \chi^2(M, \delta) \quad (19)$$

where  $i$  is the iteration number. The equivalence of these two approaches has been proved in *Kaltenbacher et al.* [2008] for a class of the so-called “regular” iterative methods which include the steepest descent, conjugate gradient, BFGS or L-BFGS. Let us underline that the equivalence is only valid for the cost function in the form (16). Note that the residual principle can only be used in case the statistical properties of the observation error are well known. Otherwise, one should use different approaches such as L-curve [*Hansen and O’Leary*, 1993], cross-validation [*Golub et al.*, 1979], etc.

The update step of the L-BFGS algorithm reads as follows:

$$W_{i+1} = W_i + \beta_i \tilde{H}_i^{-1} B^{1/2} J'(V_i), \quad W_0 = 0, \quad (20)$$

$$V_{i+1} = V_b + B^{1/2} W_{i+1}, \quad V_0 = V_b, \quad (21)$$

where  $\tilde{H}_i^{-1}$  is the approximated inverse of the Hessian built by the algorithm. The gradient of the cost-function is given by:

$$J'(V_i) = (\mathcal{G}'(V_i))^* \mathcal{R}^{-1}(\mathcal{G}(V_i, U_b^0) - Y) \quad (22)$$

where,  $(\mathcal{G}'(V))$  and  $(\mathcal{G}'(V))^*$  are respectively the tangent linear and adjoint counterparts of the nonlinear operator  $(\mathcal{G}(V, U_b^0))$  given by the following Gateaux derivative:

$$\mathcal{G}'(V)w = \lim_{t \rightarrow 0} \frac{\mathcal{G}(V + tw, U_b^0) - \mathcal{G}(V, U_b^0)}{t}, \quad (23)$$

$$\langle w, (\mathcal{G}'(V))^* w^* \rangle_{\mathcal{U}} = \langle \mathcal{G}'(V)w, w^* \rangle_{\mathcal{Y}}, \quad \forall w, w^*. \quad (24)$$

The tangent linear and adjoint operators associated with the SIC<sup>2</sup> model have been produced using the automatic differentiation tool TAPENADE developed at INRIA [*Hascoët and Pascual*, 2004].

The gradient-based minimization involves the use of the direct and adjoint models. While the direct model represents the forward time integration, the adjoint model is regarded as the corresponding backward time integration. For the latter, the system trajectory needs to be saved at each forward integration time step. Thus, for a long assimilation window (e.g. a few months) and a time step consistent with the Preissmann numerical scheme (e.g. 20min), the process memory requirements become prohibitively expensive. To overcome this limitation, we consider a sequential implementation of the variational DA approach which operates with assimilation sub-windows.

Let us introduce a time lag  $\delta T$  consistent with the characteristic time of the dynamical system. At the start of the DA process, i.e. for the first assimilation sub-window, the initial condition is given by the steady-state flow solution consistent with the initial guess, i.e.  $Q(t) = Q_b(t_0) = \text{const}$ . For any subsequent sub-window, the initial condition is given by the estimated state at time  $T - \delta T$  from the previous assimilation sub-window, i.e.  $z_0(x) = \hat{z}(x, T - \delta T)$  and  $q_0(x) = \hat{q}(x, T - \delta T)$ , whereas the background is given by the estimated discharge at time  $T - \delta T$ , i.e.  $Q_b(t) = \hat{Q}(T - \delta T) = \text{const}$ . This sequential implementation is an additional option to the method presented in *Gejadze and Malaterre [2017]*. The conducted study is only possible using the sequential version of DA, when using modest computational resources, i.e. a laptop with 16-32GB of RAM. Note that all operational variational DA systems are based on sequential (cyclic) technologies.

### 3.3 Computing first guess model inputs

The optimization method requires a first guess on the variables of interest  $Q(t)$ ,  $Z_b(x)$  and  $K_S(x)$ . Therefore, we suggest a way of generating this initial background using the SWOT observations only and globally-available background information applicable to ungauged basins. Note that the background values of  $Q$ ,  $Z_b$  and  $K_S$ , generated using this method, are only introduced during the first DA sub-window then will be later improved during the data assimilation process.

This method is based on a prior knowledge of discharge  $Q_b$ , which for the current test cases was taken as an output of the global Water Balance Model (WBM), that uses monthly atmospheric forcing data (mean air temperature and precipitation), obtained from the Climatic Research Unit (CRU, <http://www.cru.uea.ac.uk/>) [*Wisser et al., 2010*], and SWOT observations of the water surface elevation  $z$ , the top width  $L$  and the slope  $S = \frac{\partial z}{\partial x}$ . The first guesses on  $A$  and  $K_S$  are calculated using the Manning's equation, assuming trapezoidal cross-sections at each river node. The WBM-based prior discharge  $Q_b$  is  $841.8 \text{ m}^3\text{s}^{-1}$  throughout the study area for the Po River and  $377 \text{ m}^3\text{s}^{-1}$  for the Sacramento River. Note that any available first guess on discharge, although not reliable, can be used as long as the resulting dynamics are supported by SIC<sup>2</sup> model. As we aim to apply the presented methodology to ungauged basins, no in-situ gauge data were considered.

An approximate bathymetry is then built by estimating the bottom width  $l$ , the depth  $h$  and the bank slope  $b$ , assuming trapezoidal cross-sections. To do so, let us consider the discharge formula derived from the Manning equation  $Q_M = K_S R^{\frac{2}{3}} S^{\frac{1}{2}} A$ , where the index  $M$  refers to the 'Manning', the wetted perimeter  $P_M$  can be then expressed as :

$$P_M = A \left( \frac{1}{K_S} Q_b S^{-\frac{1}{2}} A^{-1} \right)^{-\frac{3}{2}} \quad (25)$$

Therefore, the top width  $L$ , the area  $A$  and the wetted perimeter  $P$  are given by:

$$L = l + 2bh \quad (26)$$

$$A = \frac{h}{2} (L + l) \quad (27)$$

$$P = l + 2h\sqrt{1 + b^2} \quad (28)$$

From the above formulas,  $P$  can be written as a function of  $l$  as:

$$P(l) = l + 2\sqrt{\left(\frac{2A}{l+L}\right)^2 + \left(\frac{L-l}{2}\right)^2} \quad (29)$$

The solution to this problem was computed using an iterative numerical method to find the estimates of  $l$ ,  $h$  and  $b$ . Finally, the bed elevation is retrieved by subtracting the depth  $h$  from the  $h$  from the WSE  $z$ ;  $Z_b = z - h$ .

#### 4 Experimental design

The challenge of this study stems from different factors such as low data accuracy and temporal frequency, unknown river bathymetry and bed roughness, the need for estimating a heterogeneous control vector, etc. Therefore, our aim is to demonstrate the performance and robustness of the proposed variational DA method for discharge estimation under uncertainties from the simulated SWOT observations, with no in-situ information. Since the long term hydraulic behavior is governed by the boundary conditions and the lateral inflow, we focus on estimating the upstream inflow discharge hydrograph  $Q(t)$ . Note that lateral inflow discharge  $Q_L(t)$  is not considered in this study. Future work will investigate more complex hydraulic configurations involving tributaries and their interactions with the main channel; a configuration that is supported by the hydraulic model SIC<sup>2</sup>.

We perform simultaneous estimation of the inflow discharge  $Q(t)$ , the bed level  $Z_b(x)$  and the Strickler coefficient  $K_S(x)$  at the nodes scale, *i.e.* at every 200 m. The background information is subjected to bias, which is the systematic part of the original uncertainty in the model inputs. Hence, the corresponding discharge estimation error can be arbitrarily large. It is the purpose of DA to remove/reduce this uncertainty. This is exactly the reason why  $Z_b$  and  $K_S$  are included into the control vector  $V$ . Removing the bias in observations is more tricky and requires a special analysis of the residuals and their derivatives in a series of independent experiments (reanalysis) [Dee and Da Silva, 1999; Desroziers *et al.*, 2005]. This is not implemented in the current version of the algorithm.

In order to deal with long study periods, the sequential version of the variational DA method presented in section 3 is considered. The size of the sub-window can be arbitrary as long as the required computational resources are available (long sub-window needs more memory). The priors of discharge, river bathymetry and roughness coefficient, during the first DA sub-window, are taken as described in section 3.3. For subsequent sub-windows, the prior of discharge is taken as the final time minus time-lag discharge estimate from the previous sub-window according to the sequential approach described in section 3.2. In order to verify the usefulness of the combination of the estimates of  $Z_b(x)$  and  $K_S(x)$ , these variables are simultaneously estimated with  $Q(t)$  during the first sub-window only then are fixed at their estimated value for subsequent sub-windows.

The estimation quality is measured based on different error metrics :

- Root-Mean-Square-Error (RMSE) defined by:

$$RMSE[Q] = \sqrt{\frac{1}{T} \sum_t (\hat{Q}(t) - Q^r(t))^2}, \quad (30)$$

- relative Root-Mean-Square-Error (rRMSE):

$$rRMSE[Q] = \sqrt{\frac{1}{T} \sum_t \left(\frac{\hat{Q}(t) - Q^r(t)}{Q^r(t)}\right)^2}, \quad (31)$$



- Normalized-Root-Mean-Square-Error (NRMSE):

$$NRMSE[Q] = \frac{1}{\bar{Q}^t} \sqrt{\frac{1}{T} \sum_t (\hat{Q}(t) - Q^t(t))^2}, \quad (32)$$

- Nash-Sutcliffe Efficiency (NSE):

$$NSE[Q] = 1 - \frac{\sum_t (\hat{Q}(t) - Q^t(t))^2}{\sum_t (\bar{Q}^t - Q^t(t))^2}, \quad (33)$$

- Volumetric Efficiency (VE):

$$VE[Q] = 1 - \frac{\sum_t |\hat{Q}(t) - Q^t(t)|}{\sum_t Q^t(t)}, \quad (34)$$

where  $Q^t(t)$  and  $\bar{Q}^t$  are the 'true' (reference) inflow discharge and its time-average value, respectively,  $\hat{Q}(t)$  is the inflow discharge estimate and  $T$  is the full time-length of the experiment.

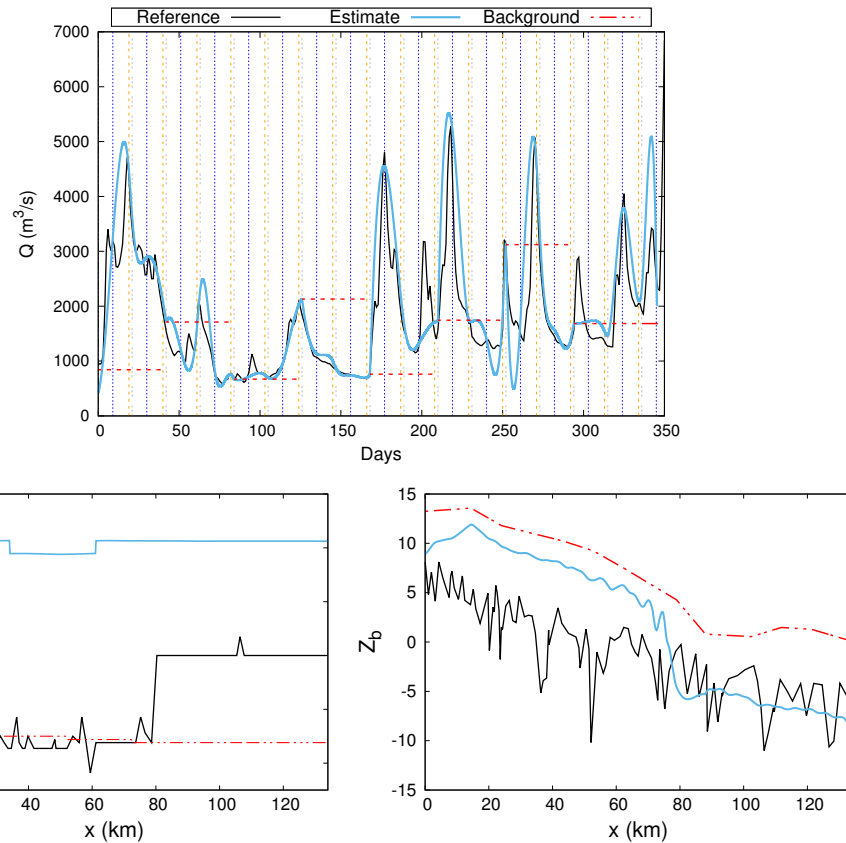
## 5 Results & Discussion

Despite the highly uncertain first guess on the Po River discharge, with a relative root mean square error of  $rRMSE[Q_b] = 81.1\%$ , simultaneous estimation of the upstream discharge  $Q(t)$ , the bed level  $Z_b(x)$  and the Strickler coefficient  $K_S(x)$  was successfully performed using eight sub-windows 42 days wide. The comparison between the daily discharge estimated with our method and the assumed true discharge from the Po River hydraulic model are presented in Figure 8. Summary statistics for the estimated daily discharge and the estimated discharge strictly during the SWOT overpasses are presented in Table 1, using the error metrics introduced in section 4. Discharge errors were lower at the time of the SWOT observations, amounting to a relative root mean square error (rRMSE) of 12.1%. The estimate of discharge at the daily scale increases rRMSE to 29.8% due to unobserved dynamics between the SWOT overpasses.

**Table 1.** Error metrics of discharge estimation for the Po River during the full assimilation windows, for daily discretization and irregular observation sampling.

Computational $\Delta T$	RMSE ( $\text{m}^3 \text{s}^{-1}$ )	rRMSE (%)	NRMSE (%)	NSE	VE
Daily	657.9	29.8	36.5	0.54	0.77
Irregular $\Delta T_{obs}$	221.4	12.6	12.1	0.95	0.91

Similarly, the first guess on the Sacramento River discharge had high uncertainty, amounting to a relative error of  $rRMSE[Q_b] = 84.4\%$ . Here, two assimilation sub-windows of 63 and 105 days are considered for DA. Figure 9 shows the estimates of  $Q(t)$ ,  $Z_b(x)$  and  $K_S(x)$  over the Sacramento study area, during the 6-month study period. The flow dynamics at the Sacramento River change are significantly faster than at the Po River, which exemplifies how the limited SWOT temporal resolution might affect the understanding of narrower rivers. As the Sacramento River study area is only revisited twice per SWOT cycle and due to the faster discharge dynamics, the flood peaks were not sampled, with one of the most significant flood events having no observations. For example, the maximum discharge  $1541.97 \text{ m}^3 \text{ s}^{-1}$  was recorded on February 17<sup>th</sup>, after the February 13<sup>th</sup> overpass. By the time SWOT returned, on February 22<sup>nd</sup>, the flood wave had already



**Figure 8.** The Po River discharge estimation at Borgoforte station (upper) together with the Strickler coefficient (lower left) and bed level (lower right) along the study area.

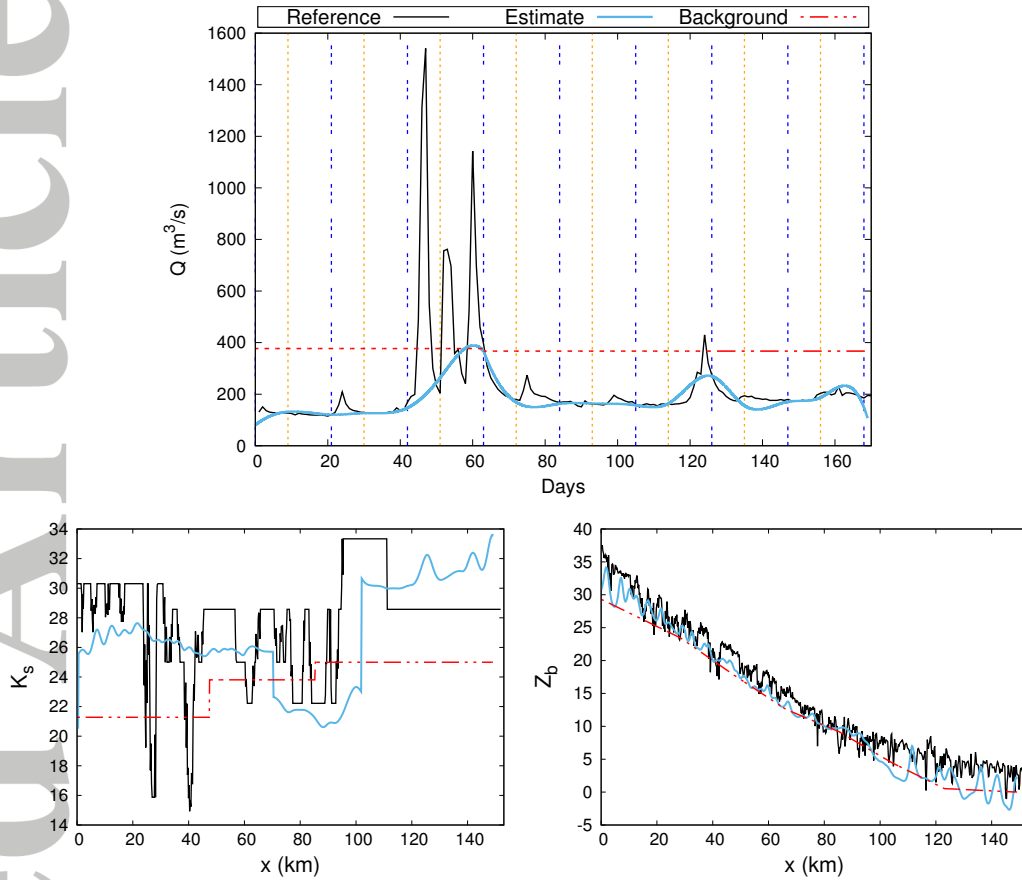
left the area. Nevertheless, discharge was successfully estimated, with rRMSE computed solely during SWOT overpasses being as low as 11.2%, whereas daily discharge rRMSE reached a higher value of 21.2%. The increased rRMSE were driven by the unobserved higher discharge events, however, the model showed considerable skill estimating the pre-flood flows (before day of year 40), despite the errors in the discharge background.

**Table 2.** Error metrics of discharge estimation for the Sacramento River during the full assimilation windows, for daily discretization and irregular observation sampling.

Computational $\Delta T$	RMSE ( $\text{m}^3 \text{s}^{-1}$ )	rRMSE (%)	NRMSE (%)	NSE	VE
Daily	167.2	21.2	75.0	0.18	0.75
Irregular $\Delta T_{obs}$	22.6	11.2	12.0	0.87	0.91

The Sacramento River example illustrates the fact that the characteristic time of the hydraulic system, *i.e.* the dynamical transition time between two possible equilibrium states, should be consistent with the temporal frequency of observations in order to ensure that the river dynamics between satellite overpasses can be accurately recovered. The characteristic time of the Po and Sacramento study Rivers under consideration are 6 and 5 days, respectively. Therefore, longer study areas may be more suitable for this type

Accepted Article

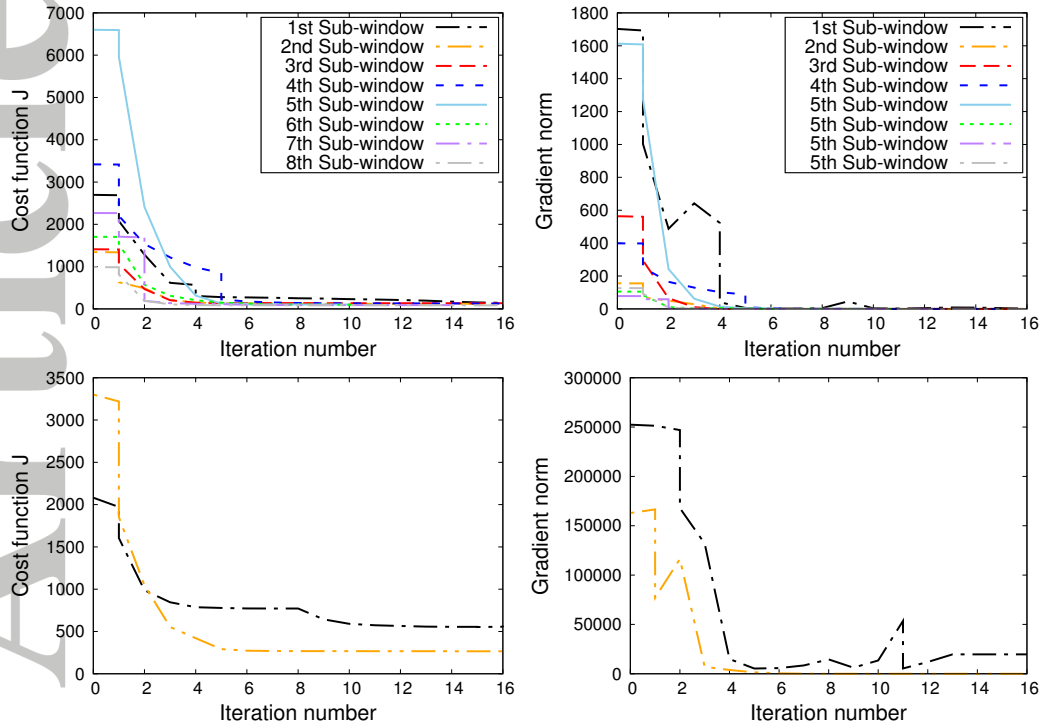


**Figure 9.** The Sacramento River discharge estimation at Hamilton station (upper) together with the Strickler coefficient (lower left) and bed level (lower right) along the study area.

of sparse temporal sampling. Moreover, the river orientation with respect to the satellite ground tracks has a key role in the temporal and spatial observation sampling, *i.e.* across-track rivers have a higher temporal frequency (The Po River) while the along-track rivers are observed with better spatial distribution (The Sacramento River). The temporal and spatial gaps between the SWOT observations can be completed by supplementary sources of data such as the virtual stations of the other existing satellite missions.

The Strickler coefficient  $K_S$  and the bed level  $Z_b$  may be locally improved during the minimization process (Figures 8 and 9). However, this combination of optimal estimates ( $Z_b, K_S$ ) only allow accurate estimation of discharge  $Q$  in both study areas despite the use of highly uncertain background information. This behavior is typical of ill-posed estimation problems (the so-called 'equifinality issue') [Gejadze and Malaterre, 2016]. Thus, the estimates of  $Z_b$  and  $K_S$  may not provide a reliable global information of river bathymetry and bed roughness.

The performance of the variational DA method is illustrated in Figure 10. The cost function  $J$  and the norm of its gradient  $\|\nabla J\|$  are presented with respect to the number of iterations at each sub-window, for the Po and the Sacramento Rivers. The minimization algorithm has converged within less than 10 iterations; *i.e.* 5 iterations for the Po River, while few more iterations, 6-8, were required for the Sacramento River case. Note that in the first sub-window, the three variables  $Q, Z_b$  and  $K_S$  are estimated simultaneously while



**Figure 10.** The minimization process for the Po River (upper panels) and the Sacramento River (lower panels) study cases. Left: the cost function  $J$ , right: the norm of the gradient  $||\nabla J||$ .

during subsequent sub-windows, only  $Q$  is estimated which explains why fewer iterations can be required for the convergence.

In case the statistical properties of the observation error are not very well known, the iterations are stopped using the criterion  $J_{i+1} - J_i < \epsilon$ , where,  $\epsilon$  is a threshold equal to  $10^{-4}$ , which is a simplified implementation of the L-curve criterion. Moreover, Figure 10 indicates the actual level of the observation noise standard deviation, which is significantly lower than the one predicted by RiverObs. This level can be roughly assessed from the residual principle :

$$J_\epsilon \approx m \tilde{\sigma}_z^2 / \sigma_z^2, \quad (35)$$

where  $m$  is the total number of observation points and  $\sigma_z$  and  $\tilde{\sigma}_z$  are the estimated standard deviations from RiverObs and simplified L-curve, respectively. For example, the first sub-window of data assimilation has a cost function equal to  $J_\epsilon = 122.3$  at the end of the iterations, in the case of the Po River. The total number of observations during the sub-window is  $m = 2938$ , while the initial estimation of the standard deviations is  $\sigma_z = 2.5$  m. Therefore, following the residual principle (35), the estimated actual error standard deviation is about  $\tilde{\sigma}_z = 0.5$  m. Future work will include simultaneous estimation of the model inputs together with parameters of the observation noise.

## 6 Conclusions

Our work presented a methodology suitable for the estimation of discharge on ungauged basins using solely remote sensing information from platforms such as the upcoming SWOT satellite. The present article describes necessary modifications to the classical variational data assimilation method applied to the full Saint-Venant-based hydraulic model SIC<sup>2</sup>. Moreover, the method offers the flexibility of also assimilating in-situ data



in addition to satellite observations when those are available. The implemented sequential version enables better use of computer resources, overcoming memory limitations, which hindered the applicability of the classical data assimilation methods to long-time periods.

We demonstrated the applicability of the proposed methodology using synthetic SWOT overpasses generated with the SWOT simulator developed at the JPL over the Sacramento and the Po Rivers. In our two study cases, discharge was successfully estimated at the time of the overpasses, with rRMSE of 12.1% and 11.2% for the Po and the Sacramento Rivers, respectively. The estimated river discharge at a daily scale increases the rRMSE to 29.8% and 21.2%, respectively. The higher degradation of the model skill for the Sacramento River when estimating daily discharge happened mostly due to poor temporal sampling during the observed flood waves, showing the importance of having temporal sampling that is compatible with the characteristic time of the hydraulic system.

The presented method demonstrates good accuracy and robustness, while requiring modest computational resources. In both cases, the method converged after less than 10 iterations when estimating  $Q$ ,  $Z_b$  and  $K_S$  simultaneously during the first sub-window, while fewer iterations (2 to 5) were required to estimate discharge during the subsequent sub-windows. Future work will report multi-mission variational data assimilation including the existing satellites missions such as JASON, ENVISAT and Sentinel in addition to the future SWOT platform.

### Acknowledgments

The presented work was undertaken with the financial support of IRSTEA and Collecte Localisation Satellite (CLS) and takes part of the PhD of Hind Oubanas. The authors would like to thank Dr. Franck Mercier, CLS, for his implication in the project. The experimental data used in this study are attached as supporting information files “SI.zip”. Complementary data can be requested from Dr. Alessio Domeneghetti ([alessio.domeneghetti@unibo.it](mailto:alessio.domeneghetti@unibo.it)), and Rui Wei ([wei.263@osu.edu](mailto:wei.263@osu.edu)).

### References

- Alsdorf, D. E., and D. P. Lettenmaier (2003), Tracking fresh water from space, *Science*, 301(5639), 1491–1494.
- Andreadis, K. M., E. A. Clark, D. P. Lettenmaier, and D. E. Alsdorf (2007), Prospects for river discharge and depth estimation through assimilation of swath-altimetry into a raster-based hydrodynamics model, *Geophysical Research Letters*, 34(10).
- Bartsch, A., W. Wagner, K. Scipal, C. Pathe, D. Sabel, and P. Wolski (2009), Global monitoring of wetlands—the value of envisat asar global mode, *Journal of environmental management*, 90(7), 2226–2233.
- Benke, A. C., and C. E. Cushing (2011), *Rivers of North America*, Academic Press.
- Biancamaria, S., K. M. Andreadis, M. Durand, E. A. Clark, E. Rodriguez, N. M. Mognard, D. E. Alsdorf, D. P. Lettenmaier, and Y. Oudin (2010), Preliminary characterization of swot hydrology error budget and global capabilities, *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 3(1), 6–19.
- Biancamaria, S., F. Hossain, and D. Lettenmaier (2011a), Forecasting transboundary flood with satellites, *Geophysical Research Letters*, 38, L11,401.
- Biancamaria, S., M. Durand, K. Andreadis, P. Bates, A. Boone, N. Mognard, E. Rodriguez, D. Alsdorf, D. Lettenmaier, and E. Clark (2011b), Assimilation of virtual wide swath altimetry to improve arctic river modeling, *Remote Sensing of Environment*, 115(2), 373–381.
- Biancamaria, S., D. P. Lettenmaier, and T. M. Pavelsky (2016), The swot mission and its capabilities for land hydrology, *Surveys in Geophysics*, 37(2), 307–337.

- Birkett, C. (1995), The contribution of topex/poseidon to the global monitoring of climatically sensitive lakes, *Journal of Geophysical Research: Oceans*, 100(C12), 25,179–25,204.
- Birkett, C. M. (1998), Contribution of the topex nasa radar altimeter to the global monitoring of large rivers and wetlands, *Water Resources Research*, 34(5), 1223–1239.
- Buer, K., D. Forwalter, M. Kissel, and B. Stohler (1989), The middle sacramento river: human impacts on physical and ecological processes along a meandering river.
- Calmant, S., and F. Seyler (2006), Continental surface waters from satellite altimetry, *Comptes Rendus Geoscience*, 338(14), 1113–1122.
- Castellarin, A., A. Domeneghetti, and A. Brath (2011), Identifying robust large-scale flood risk mitigation strategies: a quasi-2d hydraulic model as a tool for the po river, *Physics and Chemistry of the Earth, Parts A/B/C*, 36(7), 299–308.
- Chow, V. T., et al. (1964), Handbook of applied hydrology.
- Courtier, P., E. Andersson, W. Heckley, D. Vasiljevic, M. Hamrud, A. Hollingsworth, F. Rabier, M. Fisher, and J. Pailleux (1998), The ecmwf implementation of three-dimensional variational assimilation (3d-var). i: Formulation, *Quarterly Journal of the Royal Meteorological Society*, 124(550), 1783–1807.
- Crétau, J.-F., and C. Birkett (2006), Lake studies from satellite radar altimetry, *Comptes Rendus Geoscience*, 338(14), 1098–1112.
- Crétau, J.-F., M. Berge-Nguyen, M. Leblanc, R. A. Del Rio, F. Delclaux, N. Mognard, C. Lion, R. K. Pandey, S. Tweed, S. Calmant, et al. (2011), Flood mapping inferred from remote sensing data, *Int. Water Technol. J.*, 1, 48–62.
- Cuchi, K. (1986), Multi-look processing of synthetic aperture radar data from dynamic ocean surfaces, *Pattern recognition letters*, 4(4), 305–314.
- Cunge, J. A., F. M. Holly, and A. Verwey (1980), Practical aspects of computational river hydraulics.
- de Oliveira Campos, I., F. Mercier, C. Maheu, G. Cochonneau, P. Kosuth, D. Blitzkow, and A. Cazenave (2001), Temporal variations of river basin waters from topex/poseidon satellite altimetry. application to the amazon basin, *Comptes Rendus de l'Académie des Sciences-Series IIA-Earth and Planetary Science*, 333(10), 633–643.
- Dee, D. P., and A. M. Da Silva (1999), Maximum-likelihood estimation of forecast and observation error covariance parameters. part i: Methodology, *Monthly Weather Review*, 127(8), 1822–1834.
- Desroziers, G., L. Berre, B. Chapnik, and P. Poli (2005), Diagnosis of observation, background and analysis-error statistics in observation space, *Quarterly Journal of the Royal Meteorological Society*, 131(613), 3385–3396.
- Ding, Y., and S. S. Wang (2012), Optimal control of flood diversion in watershed using nonlinear optimization, *Advances in Water Resources*, 44, 30–48.
- Domeneghetti, A., A. Tarpanelli, L. Brocca, S. Barbetta, T. Moramarco, A. Castellarin, and A. Brath (2014), The use of remote sensing-derived water surface data for hydraulic model calibration, *Remote sensing of environment*, 149, 130–141.
- Durand, M., K. M. Andreadis, D. E. Alsdorf, D. P. Lettenmaier, D. Moller, and M. Wilson (2008), Estimation of bathymetric depth and slope from data assimilation of swath altimetry into a hydrodynamic model, *Geophysical Research Letters*, 35(20).
- Durand, M., J. Neal, E. Rodríguez, K. M. Andreadis, L. C. Smith, and Y. Yoon (2014), Estimating reach-averaged discharge for the river severn from measurements of river water surface elevation and slope, *Journal of Hydrology*, 511, 92–104.
- Durand, M., C. Gleason, P.-A. Garambois, D. Bjerklie, L. Smith, H. Roux, E. Rodriguez, P. Bates, T. Pavelsky, J. Monnier, et al. (2016), An intercomparison of remote sensing river discharge estimation algorithms from measurements of river height, width, and slope, *Water Resources Research*, 52(6), 4527–4549.
- Fekete, B. M., and C. J. Vörösmarty (2002), The current status of global river discharge monitoring and potential new technologies complementing traditional discharge measurements, in *Predictions in Ungauged Basins: PUB kick-off (Proceedings of the PUB*

- kick-off meeting held in Brasilia, 20–22 November 2002). *IAHS Publication*, vol. 349.
- Fernandez, D. E., B. Pollard, and P. Vaze (2013), Swot mission performance error budget, revision a, *Tech. rep.*, JPL document D-79804.
- Fischer, C., T. Montmerle, L. Berre, L. Auger, and S. E. ȘTEFĂNESCU (2005), An overview of the variational assimilation in the aladin/france numerical weather-prediction system, *Quarterly Journal of the Royal Meteorological Society*, 131(613), 3477–3492.
- Fjortoft, R., J.-M. Gaudin, N. Pourthie, J.-C. Lalaurie, A. Mallet, J.-F. Nouvel, J. Martinot-Lagarde, H. Oriot, P. Borderies, C. Ruiz, et al. (2014), Karin on swot: characteristics of near-nadir ka-band interferometric sar imagery, *IEEE Transactions on Geoscience and Remote Sensing*, 52(4), 2172–2185.
- Frappart, F., S. Calmant, M. Cauhopé, F. Seyler, and A. Cazenave (2006), Preliminary results of envisat ra-2-derived water levels validation over the amazon basin, *Remote sensing of Environment*, 100(2), 252–264.
- Frasson, R. P. d. M., R. Wei, M. Durand, J. T. Minear, A. Domeneghetti, G. Schumann, B. A. Williams, E. Rodriguez, C. Picamilh, C. Lion, et al. (2017), Automated river reach definition strategies: Applications for the surface water and ocean topography mission, *Water Resources Research*.
- Garambois, P.-A., and J. Monnier (2015), Inference of effective river properties from remotely sensed observations of water surface, *Advances in Water Resources*, 79, 103–120.
- Gauthier, P., C. Charette, L. Fillion, P. Koclas, and S. Laroche (1999), Implementation of a 3d variational data assimilation system at the canadian meteorological centre. part i: The global analysis, *Atmosphere-Ocean*, 37(2), 103–156.
- Gauthier, P., M. Tanguay, S. Laroche, S. Pellerin, and J. Morneau (2007), Extension of 3dvar to 4dvar: Implementation of 4dvar at the meteorological service of canada, *Monthly weather review*, 135(6), 2339–2354.
- Gejadze, I., and P.-O. Malaterre (2017), Discharge estimation under uncertainty using variational methods with application to the full saint-venant hydraulic network model, *International Journal for Numerical Methods in Fluids*, 83(5), 405–430.
- Gejadze, I. Y., and P.-O. Malaterre (2016), Design of the control set in the framework of variational data assimilation, *Journal of Computational Physics*, 325, 358–379.
- Golub, G. H., M. Heath, and G. Wahba (1979), Generalized cross-validation as a method for choosing a good ridge parameter, *Technometrics*, 21(2), 215–223.
- Hansen, P. C., and D. P. O’Leary (1993), The use of the l-curve in the regularization of discrete ill-posed problems, *SIAM Journal on Scientific Computing*, 14(6), 1487–1503.
- Hascoët, L., and V. Pascual (2004), Tapenade 2.1 user’s guide, *Tech. rep.*, INRIA.
- Hossain, F., A. Siddique-E-Akbor, L. C. Mazumder, S. M. ShahNewaz, S. Biancamaria, H. Lee, and C. Shum (2014), Proof of concept of an altimeter-based river forecasting system for transboundary flow inside bangladesh, *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 7(2), 587–601.
- Hostache, R., X. Lai, J. Monnier, and C. Puech (2010), Assimilation of spatially distributed water levels into a shallow-water flood model. part ii: Use of a remote sensing image of mosel river, *Journal of hydrology*, 390(3), 257–268.
- Kaltenbacher, B., A. Neubauer, and O. Scherzer (2008), *Iterative regularization methods for nonlinear ill-posed problems*, vol. 6, Walter de Gruyter.
- Kouraev, A. V., E. A. Zakharova, O. Samain, N. M. Mognard, and A. Cazenave (2004), Ob’river discharge from topex/poseidon satellite altimetry (1992–2002), *Remote sensing of environment*, 93(1), 238–245.
- Lai, X., and J. Monnier (2009), Assimilation of spatially distributed water levels into a shallow-water flood model. part i: Mathematical method and test case, *Journal of Hydrology*, 377(1), 1–11.
- Malaterre, P., J. Baume, and D. Dorchies (2014), Simulation and integration of control for canals software (sic 2), for the design and verification of manual or automatic controllers for irrigation canals, in *USCID Conference on Planning, Operation and Automa-*

- tion of Irrigation Delivery Systems, pp. 377–382.
- Medina, C. E., J. Gomez-Enri, J. J. Alonso, and P. Villares (2008), Water level fluctuations derived from envisat radar altimeter (ra-2) and in-situ measurements in a subtropical waterbody: Lake izabal (guatemala), *Remote Sensing of Environment*, 112(9), 3604–3617.
- Montanari A., B. E. D. A. B. A., Ceola S. (2017), *Handbook of Applied Hydrology, Second Edition*, McGraw-Hill Education.
- Novak, P., V. Guinot, A. Jeffrey, and D. E. Reeve (2010), *Hydraulic modelling—an introduction: principles, methods and applications*, CRC Press.
- Oubanas, H., I. Gejadze, P. Malaterre, and F. Mercier (2015), Estimation of river discharge from in-situ and remote sensing data, using variational data assimilation and a full saint-venant hydraulic model, in *3rd Space for Hydrology Workshop, Frascati (Rome), Italy, September*, pp. 15–17.
- Oubanas, H., I. Gejadze, P. Malaterre, M. Durand, A. Domeneghetti, and F. Mercier (2016), Simultaneous estimation of inflow discharge, river bathymetry and friction from synthetic swot data using variational data assimilation, in *American Geophysical Union, San Francisco, California, United States, December*.
- Papa, F., B. Legrésy, and F. Rémy (2003), Use of the topex–poseidon dual-frequency radar altimeter over land surfaces, *Remote sensing of Environment*, 87(2), 136–147.
- Papa, F., C. Prigent, F. Durand, and W. B. Rossow (2006), Wetland dynamics using a suite of satellite observations: A case study of application and evaluation for the indian sub-continent, *Geophysical Research Letters*, 33(8).
- Rabier, F., H. Järvinen, E. Klinker, J.-F. Mahfouf, and A. Simmons (2000), The ecmwf operational implementation of four-dimensional variational assimilation. i: Experimental results with simplified physics, *Quarterly Journal of the Royal Meteorological Society*, 126(564), 1143–1170.
- Rodriguez, E. (2016), Swot science requirements document, jpl document d-61923.
- Rogers, W. (2014), Central valley floodplain evaluation and delineation.
- Sneeuw, N., C. Lorenz, B. Devaraju, M. J. Tourian, J. Riegger, H. Kunstmann, and A. Bárdossy (2014), Estimating runoff using hydro-geodetic approaches, *Surveys in Geophysics*, 35(6), 1333–1359.
- Tikhonov, A. N., V. I. Arsenin, and F. John (1977), *Solutions of ill-posed problems*, vol. 14, Winston Washington, DC.
- Tockner, K., U. Uehlinger, and C. T. Robinson (2009), *Rivers of Europe*, Academic Press.
- Tourian, M., N. Sneeuw, and A. Bárdossy (2013), A quantile function approach to discharge estimation from satellite altimetry (envisat), *Water Resources Research*, 49(7), 4174–4186.
- Ulaby, F. T., and M. C. Dobson (1989), *Handbook of radar scattering statistics for terrain (artech house remote sensing library)*, Norwood, MA, USA: Artech House.
- Ulaby, F. T., D. G. Long, W. J. Blackwell, C. Elachi, A. K. Fung, C. Ruf, K. Sarabandi, H. A. Zebker, and J. Van Zyl (2014), *Microwave radar and radiometric remote sensing*, vol. 4, University of Michigan Press Ann Arbor.
- Wisser, D., B. Fekete, C. Vörösmarty, and A. Schumann (2010), Reconstructing 20th century global hydrography: a contribution to the global terrestrial network-hydrology (gtnh), *Hydrology and Earth System Sciences*, 14(1), 1–24.
- Yoon, Y., M. Durand, C. J. Merry, E. A. Clark, K. M. Andreadis, and D. E. Alsdorf (2012), Estimating river bathymetry from data assimilation of synthetic swot measurements, *Journal of hydrology*, 464, 363–375.