Noname manuscript No.
(will be inserted by the editor)

Hybrid Neural Network - Variational Data Assimilation algorithm to infer river discharges from SWOT-like data

Kevin LARNIER · **Jérôme MONNIER**

Received: date / Accepted: date

Abstract Estimating discharges $Q(x,t)$ from altimetric measurements only, for ungauged rivers (in particular, those with unknown bathymetry b(x)), is an ill-posed inverse problem. We develop here an algorithm to estimate $Q(x,t)$ without prior flow information other than global open datasets. Additionally, the ill-posedness feature of this inverse problem is re-investigated. Inversions based on a Variational Data Assimilation (VDA) approach enable accurate estimation of spatio-temporal variations of the discharge, but with a bias scaling the overall estimate. This key issue, which was already highlighted in our previous studies, is partly solved by considering additional hydrological information (the drainage area, A (km^2)) combined with a Machine Learning (ML) technique. Purely data-driven estimations obtained from an Artificial Neural Network (ANN) provide a reasonably good estimation at a large scale ($\approx 10^3$ m). This first estimation is then employed to define the first guess of an iterative VDA algorithm. The latter relies on the Saint-Venant flow model and aims to compute the complete unknowns (discharge $Q(x,t)$, bathymetry $b(x)$, friction coefficient $K(x,t)$) at a fine scale (approximately 10^2 m). The resulting complete inversion algorithm is called the H2iVDI algorithm for "Hybrid Hierarchical Variational Discharge Inference". Numerical experiments have been analyzed for 29 heterogeneous worldwide river portions. The obtained estimations present an overall bias (less than 30% for rivers with similar characteristics than those used for calibration) smaller than previous results, with accurate spatio-temporal variations of the flow. After a learning period of the observed rivers (e.g. one year), the algorithm provides two complementary estimators: a dynamic flow model enabling estimations at a fine scale and spatio-temporal extrapolations, and a low complexity estimator (based on a dedicated algebraic low Froude flow model). This last estimator provides reasonably accurate estimations (less than 30% for considered rivers) at a large scale from newly acquired WS measurements in real-time, therefore making it a potentially operational algorithm.

Keywords Data assimilation · neural network · inference · rivers · discharge · bathymetry · altimetry · datasets · SWOT mission.

1 Introduction

The estimate of ungauged or poorly gauged river discharges is one of the greatest challenges in hydrology. Numerous satellite missions acquire a huge amount of data of different kinds (altimetry, optical, etc.) every day, which can be useful to set up river flow models (see e.g. [9] and references therein). One of the ultimate goals of river flow models is to estimate the space-time varying discharge $Q(x,t)$. Setting up a river flow model requires knowledge of the bathymetry, an effective friction coefficient, and the (potential) lateral fluxes (see e.g. [10]). The future Surface Water and Ocean Topography (SWOT) mission (NASA-CNES et al.) launched in December 2022 starts to provide unprecedented Water Surface (WS) measurements of rivers wider than $50-100\,m$. The SWOT instrument measures the WS elevation Z (with a decimetric accuracy over 1 km^2) and the WS width W (with a varying uncertainty of a few m, depending on the river planform). This instrument will cover a great majority of the globe with relatively frequent revisits: from 1 to 4 revisits per 21-day repeat cycle (see [48,47]).

Given these WS measurements and in order to set up river flow models, the following inverse problem arises: estimate the discharge $Q(x,t)$, but also the unobserved bathymetry $b(x)$ and a friction law parametrization $K(x,t)$. A few algorithms have been developed to solve related inverse sub-problems (see e.g. [16] and references therein, where five

J. Monnier
INSA & Institut de Mathématiques de Toulouse (IMT), France
E-mail: jerome.monnier@insa-toulouse.fr

K. Larnier
INSA & Institut de Mathématiques de Toulouse (IMT) & CS GROUP - France, Space Business Unit, Toulouse, France

different algorithms are compared on 19 river portions). These algorithms are either based on relatively basic flow models (the algebraic Manning-Strickler's law) or empirical hydraulic and geometric power-laws. No method has proven to be accurate or robust in all the configurations considered. Indeed, all the methods remain sensitive to the introduced priors, e.g. good knowledge of the bathymetry or a mean value of discharge.

It has been mathematically shown in [30] (see also discussions in [6]) that the considered inverse problem is ill-posed if based on flow model equations only (i.e. without any additional prior). This ill-posedness question is addressed in [19] in the simpler case of Manning-Strickler type models. This important result explains the biased estimations obtained in [17] with almost all algorithms tackling this problem. Various data assimilation approaches based on the Kalman filter and its variants have been developed, but initially, none of them considered the complete inverse problem, which aims to infer the complete set of unknowns $(Q(x,t), b(x), K(x,t))$, as discussed in [4] and references therein. However, more recently, [1] addressed the actual complete inverse problem using an Ensemble Kalman Filter (EnKF) with a stochastic description of the discharge. The flow model is a standard backwater equation, which is a scalar first-order ODE, constrained by empirical closing power-laws relating key fields, such as width (WS) W, elevation (WS) H, and discharge Q (classical laws called at-a-station hydraulic geometry laws, [15]). The approach relies on unknown priors (PDFs), estimations of the river's geomorphology obtained by a purely data-driven method, and a classification process. The final discharge estimation [1], which is the optimal solution of the EnKF algorithm called SAD, is encouraging, with typical median NSE of approximately 0.77 and normalized RMSE of approximately 29 obtained on 15 rivers over 19 river portions considered in the called Pepsi dataset [16]. Therefore, it seems that the empirical closure power-laws [15] enable the partial solution of the equifinality issue when combined with a flow model-EnKF scheme. However, the estimations are not yet robust for all tested rivers.

The complete inverse problem has been tackled by a few different Variational Data Assimilation (VDA) algorithms, as discussed in [6,42,41,30]. VDA approaches are based on the optimal control of the considered flow model(s), as discussed in [2]. In [41,42], the triplet of unknowns is accurately inferred from the 1D Saint-Venant flow model. However, the priors are computed from small Gaussian perturbations of the true values of K and $b(x)$ and define a rating curve $Q(Z)$ at downstream (outflow condition). As a consequence, the inversion process easily converges to values corresponding to the rating curve imposed at outflow. If the latter is nearly exact, the obtained estimations are obviously not biased. On the contrary, if it is unknown, as it is in practice, the bias issue remains. In parallel with the present study, [20] has addressed the bias issue by first estimating mean values of the key fields using a Bayesian approach combined with a low Froude flow model, such as in [30,19]. In this Bayesian-type approach, the question of priors is moved to the prior PDFs and to unknown parameters of the discharge likelihood function. Therefore, the ill-posed feature of the problem is somehow transferred to the optimal properties of the Bayesian description. Using this approach, the results presented in [20] appear to be more accurate than those presented in [17]. However, it should be noted that these results are based on different datasets than those compared in [16,17].

[30,6] address the inverse problem for ungauged rivers using the Saint-Venant system flow model, where only the downstream unknown normal depth (Z) is imposed, combined with the low-Froude flow model developed in [6,30]. This model is designed for the scale of satellite measurements (see also [19]). The HiVDI algorithm [30] implements a complete inversion strategy that accurately infers spatio-temporal variations of discharge, but with a potential bias as discussed above. The bias value depends on the first guess values and priors introduced in the minimization formulation. However, it is shown that if the mean value of discharge or bathymetry is known, the bias vanishes [30,6]. Applications of the HiVDI algorithm have been developed in different complex contexts, see e.g. [51,18,45]. Precise comparisons of three "discharge algorithms" based on different concepts (including the HiVDI algorithm above) are presented in [17]. To apply this method to worldwide ungauged rivers, no informed prior other than information available in open global databases can be introduced in the inversions, neither in the direct model nor in the inverse method.

This study proposes to tackle this complex inverse problem by enriching the HiVDI algorithm with additional hydrological information, namely local drainage area $\mathscr{A}$ (km^2), and an ANN-based learning process [31]. The resulting algorithm, called H2iVDI (Hybrid Hierarchical Variational Discharge Inference) as in [31], provides more accurate estimations of discharge for ungauged rivers compared to previous studies [6,30,16,51,17]. This higher accuracy is mainly due to better first guess values for the VDA iterative process than before ([30]); these are derived by an ANN, which provides purely data-driven estimations. The ANN is trained from the WS measurements plus drainage area values $\mathscr{A}$ (km^2), which is an additional physical information compared to the aforementioned studies. It turns out that the ANN-based estimations are already good, at least much better than available global priors such as Water Balance Model (WBM) mean values [11]. Therefore, this purely data-driven estimation naturally defines the first guess value for the iterative VDA process. The solutions obtained from any VDA process depend on the physical model, potentially on the first guess value (it is the case in this ill-posed problem context), and also on the covariance matrices defining the cost function metrics. The present algorithm H2iVDI contains original improvements of the covariance matrices definition. They are non-uniform and therefore somehow physically adaptive, which improves the robustness and accuracy of the VDA-based estimations.

In Section 2, the addressed inverse problem is detailed, and the basic principles of the H2iVDI algorithm to solve this problem are presented. Data types, including altimetry and in-situ measurements, and the three different scales of

data and models are detailed in Section 3. The discharge estimations $Q^{(ANN)}$ obtained by the ANN (which provides purely data-driven estimations) and then re-analyzed through the low Froude flow model are presented in Section 4. In Section 5, the results obtained by the VDA iterative algorithm, which provide the complete unknown parameters set $(Q(x,t),A_0(x),K(x;h(x,t)))$ at a fine scale, are analyzed. Additionally, the ill-posedness feature of the inverse problem is shown when relying on classical flow models only. Section 6 presents how, past the "learning period", the low-complexity algorithm can provide quite accurate discharge estimations at a large scale in real-time. A conclusion is proposed in Section 7. In the appendix, the river geometry model, the dynamic Saint-Venant flow model, the low Froude flow model (low-complexity algebraic system), details on the VDA formulation, and the ANN are presented.

2 Problem statement and H2iVDI algorithm principles

2.1 Problem statement

This study aims at estimating discharges $Q(x,t)$ (m^3/s)), bathymetry profiles $b(x)(m)$ and effective friction parameters K $(m^{1/3}/s)$ of ungauged rivers from WS measurements only. The WS measurements are of SWOT-like type: elevation values $Z(x,t)$ and width values $W(x,t)$, see [47]. These WS measurements are available at two different scales (with different accuracies of course): at "large scale" called reach scale ($\approx$ km and days) and at a smaller scale called node scale ($\approx$ 200m), see Fig. 1. At the reach scale an estimation of the slope $S(x,t)$ (m/m) is also available. Details on the datasets and these two scales are presented in the next section.
Following [6,30] and partly in relation to these two spatio-temporal scales of WS measurements, two hierarchical flow models are considered: the original low-Froude model presented in [30] (see details in Section C) and the classical Saint-Venant flow model (see details in Section B).
In short, the addressed inverse problem is the following.

Problem (P): Given the WS measurements aforementioned at small scale NodeSc and large scale ReachSc, estimate the discharge $Q(x,t)$ (m^3/s), a bathymetry profile $b(x)(m)$ and effective friction parameter values K $(m^{1/3}/s)$.

It is worth noticing that effective friction parameter values K depends in particular on the flow model scale and on the flow regime. In this study, following [6,18], K will be defined as a function depending on the space variable x and the water depth $h(x,t)$ therefore a function of the form $K(x,h(x,t))$. Details are provided in Section 4.2.

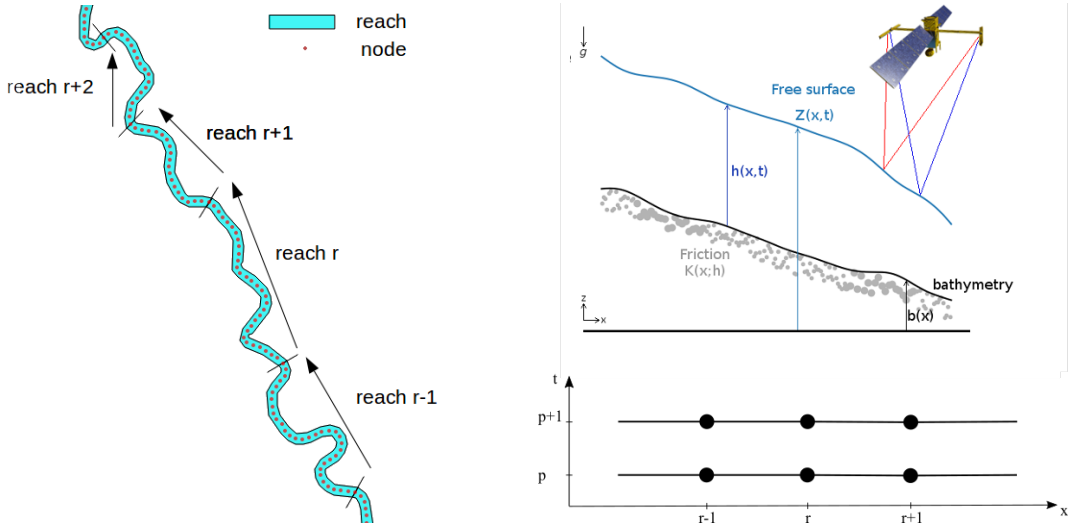

Fig. 1 (Left) Top view of an observed river with the two different scales: reach scale and node scale (denoted by ReachSc and NodeSc respectively).
At each location r (cyan portions) corresponds a set of WS measurements $(Z_{r,p},W_{r,p},S_{r,p})$ (large scale measurements). The algebraic (low complexity) flow model, see Section C, is solved at this large scale.
At each node (red circle, NodeSc) corresponds a presumed true cross-section A_r and another set of WS measurements $(Z_{r,p},W_{r,p})$ (small scale measurements).
At the computational grid scale (denoted by CompGridSc) points ($dx = 100$m, not shown here), no WS measurements are available. The Saint-Venant dynamic flow model, see Section B, is solved at this scale.
(Right)(Top) The inverse problem: infer the flow discharge $Q(x,t)$ (m^3/s), the bathymetry $b(x)$ (m) (equivalently the unmeasured lowest wetted cross-section $A_0(x)$ (m^2), see also Fig. 14) and an effective friction parameter $K(x,t)$ $(m^{1/3}/s)$ from the WS measurements.
(Right)(Bottom) Space-time grid of the measurements: reach number r in x-axis vs satellite overpass instant p re-ordered from lowest to highest flow-line on the y-axis.

2.2 A-priori capabilities and limitations of inversions based on flow models only

Let us recall here a crucial issue when addressing the present inverse problem (P). This result has been first shown in [30] (see also [19] for preliminary discussions).
Let us first consider the (low-complexity) algebraic flow model (5) deriving from the basic Gauckler-Manning -Strickler equilibrium law between gravity and friction forces. The model (5) is based on the Low-Froude assumption $Fr^2 << 1$ (property widely satisfied in the present spatial hydrology context as analyzed in [19,6]). This model enables to determine the ratio $Q/K_{0,r}$, equivalently Q/K, but not the unknowns pair (Q,K). This remarks holds even if the bathymetry b (equivalently A_0) is known. Of course, this remarks holds for the more simple scalar Manning-Strickler's equation too or any other derived version of this basic equilibrium law.

Let us consider now the complete 1D shallow flow model (3) (Saint-Venant's system). The inverse problem aiming at estimating the triplet $(Q_{in}(t),A_0(x),K(h(x,t))$ from this system (3) is also ill-posed. Indeed, it is easy to notice that the solution $(A,Q)(x,t)$ is unchanged if multiplied by an adequate factor. Let us show this. Let $\bar{Q}$ be any scalar value: $\bar{Q}$ may be a mean value of Q or the friction K. We define re-scaled state variables as follows: $(A_*,Q_*) = (A,Q)/\bar{Q}$. The mass equation (3)(a) divided by $\bar{Q}$ is unchanged: $\partial_t(A_*) + \partial_x(Q_*) = 0$. The same mass equation holds, Q and A are simply re-scaled by the same factor.

Next, the re-scaled momentum equation (Eqn (3)(b) divided by $\bar{Q}$) reads:

$$\partial_t(Q_*) + \partial_x\left(\frac{Q_*^2}{A_*}\right) + gA_*\,\partial_x Z = -gA_*S_f \tag{1}$$

with $S_f \equiv S_f(A,Q;h;K) = \frac{1}{K^2}\frac{|Q|Q}{A^2h^{4/3}}$. If defining h as the effective cross-section depth: $h = A/W$, W the WS width, then: $S_f(A,Q,h;K) = S_f(A_*,Q_*,h_*;\bar{Q}^{-2/3}K)$.

Therefore, given the WS measurements (W,Z), the 1D Saint-Venant equations (3) with parameter K are equivalent to the same equations in the re-scaled variables (A_*,Q_*) but with $(\bar{Q}^{-2/3}K)$ instead of K as Strickler's parameter. Concerning boundary conditions, both upstream and downstream conditions are transparently re-scaled by the factor $\bar{Q}$. It is worth noting that this little calculation and its consequences remain of course true for simplified versions of the Saint-Venant models e.g. like those presented in [3].

As a consequence and as already mentioned in the general introduction, we know how to infer quite accurately space-time variations of the discharge however very likely with a bias, see e.g. [30,18,45,51,17,20]. This bias depends on the priors introduced in the employed numerical methods and the first values of the iterative algorithms. To finish, note that it has been shown in [30] that the knowledge of the mean value (e.g. seasonal or annual) of Q may be enough to remove the bias.

One of the important goal of the present study is to reduce at best this bias in the discharge (and bathymetry) estimations without any additional informed priors other than global open databases.

2.3 Inversion algorithm basic principles

To attempt to solve the bias issue when solving Problem (P) above, the developed algorithm is here the following (see Fig. 2).

- *Step 1) First estimations at large scale (ReachSc)*
  - *Step 1a) Purely-data driven estimation of Q by an ANN*
    An ANN with a classical dense architecture is built up to estimate the discharge value Q at the reach number r and instant p of the observation. The input variables of the ANN are the WS measurements $(Z_{r,p},W_{r,p})$ at large scale (details are presented in the following data description section) plus the river portion drainage area $\mathscr{A}$ (km^2). This last input variable is the only ancillary data employed in this study. It can be derived from HydroSHEDS database (Hydrological data and maps based on SHuttle Elevation Derivatives at multiple Scales) [33] for worldwide rivers. The resulting output of the ANN is the discharge value, estimation denoted by $Q_{r,p}^{(ANN)}$ (r the location index, p the instant index).

  - *Step 1b) Corresponding physically-consistent values of $A_{0,r}$ and $K_{r,p}$*
    Given $Q_{r,p}^{(ANN)}$ and the WS measurements at large scale, estimations of the unobserved wetted cross-sections A_0 and friction (Strickler) parameters K are computed as the solution of the Low Froude flow model (5). This simple flow model has been demonstrated sufficiently accurate at large scale in [30] (see also [19] for a preliminary

version). This step provides an estimation of $A_{0,r}$ at each location r, equivalently an effective bathymetry b_r given the river geometry representation, see Section A and Fig. 14. The friction parameters K are obtained in the form $K_r(h_{r,p})$ by performing the Metropolis–Hastings algorithm (MCMC method).

– *Step 2) Estimation of the complete unknowns $(Q(x,t),b(x),K(x,h(x,t))$ at fine scale (NodeSc)*

A Variational Data Assimilation (VDA) method based on the Saint-Venant flow model has been developed to estimate $Q(x,t)$, $A_0(x)$ and $K(x,h(x,t))$ at fine scale [30]. VDA relies on an iterative algorithm which requires first values of the estimated quantities. The latter are based on the large scale estimations obtained at Step 1): $Q_{r,p}^{(ANN)}$, $A_{0,r}$ and $K_r(h_{r,p})$.

After convergence, estimations of $(Q(x,t),b(x),K(x,h(x,t))$ are obtained at fine scale. Moreover, this step results to a calibrated (spatio-temporal) Saint-Venant model (see Section B) valid at fine scale.

– *Step 3) Low complexity operational model valid at large scale (ReachSc)*
After a complete hydrologic cycle, typically after one year, Saint-Venant flow models have been calibrated for each river; the river portions have been "learned". In particular, effective bathymetries $b(x)$ (equivalently $A_0(x)$) are now available. Given these estimations of bathymetry $b(x)$ and $Q(x,t)$ resulting of Step 2), it is interesting to re-compute an optimal effective large scale friction parameter $K_r(h_{r,p})$ corresponding to the low-complexity flow model. This is done by performing the Metropolis–Hastings algorithm (MCMC method).

Finally, after the learning period of the observed rivers (e.g. one year), the present approach provides three complimentary estimators:
1) the trained ANN (purely data-driven);
2) the calibrated Saint-Venant flow model valid at the finest scale and enabling to extrapolate WS measurements;
3) the algebraic "surrogate" Low Froude flow model valid at larger scale.

The complete algorithm described above is called H2iVDI algorithm (H2iVDI for "Hybrid Hierarchical Variational Discharge Inference"). It is represented in Fig. 2.

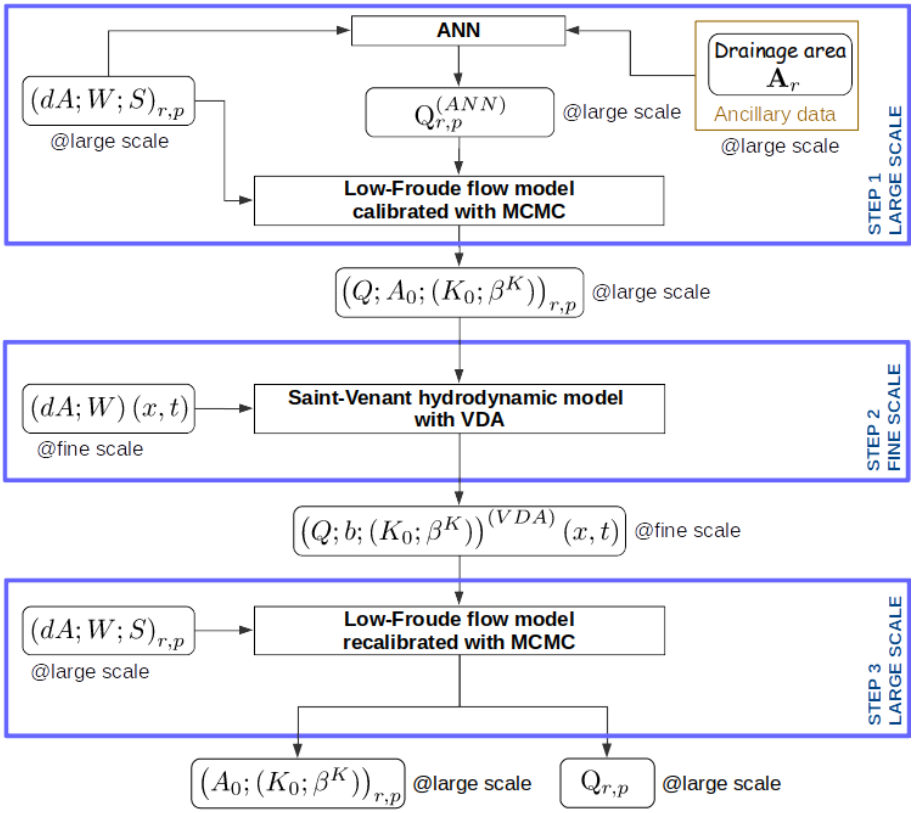

Fig. 2 FlowChart of the inversion algorithm H2iVDI. Input data are the WS measurements (Z,W,S) at large scale ReachSc (at Step 1) and Step 3)) and at smaller scale NodeSc (at Step 2)). Prior of the algorithm is $\mathscr{A}$ (km^2) only (values extracted from e.g. HydroSHEDS database). After a complete representative learning period (typically one year), effective bathymetry $b(x)$ (equivalently $A_0(x)$) and local effective friction laws $K(x;h(x,t))$ are available at both scales.

Operational Near-Real-Time (NRT) estimations Past the learning period, given newly acquired WS measurements, few possibilities will exist to compute discharge estimations in Real-Time (RT) (computations in μs CPU-time).
A first one is as follows. Given the WS measurements including the slopes (therefore at ReachSc), it is possible to generate numerical space rating curves (also called Stage Fall Discharge -SFD- laws) following e.g. [44,37]. SFD laws may also be derived from datasets acquired by multi-missions such as Sentinel-3, Sentinel-6 (ESA), IceSat2 (NASA). Note that in 2022, slope measurements can be expected from the (future) SWOT mission only, moreover at ReachSc only. SFD constitutes efficient NRT estimators. The generation of SFD laws has been implemented in the H2iVDI algorithm; it is however not presented here.
Note that operationally, these estimations will be NRT only because of the delay to produce the inputs of these models from the observations [47]. The low complexity low Froude model aforementioned can provide operational NRT estimations too. This option is investigated in Section 6.1.

3 Data description

3.1 The altimetry and in-situ data

3.1.1 The different scales

Data availability is different depending on the spatial scale. Let us detail the three different scales which are considered in this study, see Fig. 1.

- The large scale is the so-called "reach scale" in the SWOT community, see [47]. It varies between a dozen of km to a few km ($\approx$ 5 km), depending on the river. The reach scale is denoted by ReachSc.
- The SWOT small scale called node scale in the SWOT community. It corresponds to an average distance of 200m. This scale, denoted NodeSc, is defined by "nodes", see Fig. 1. However in the data used in this study, it corresponds to the grid employed in the reference models (e.g. HEC-RAS) to generate the SWOT-like data and presumed true cross-sections $A(x)$. The distance between two nodes here varies between a few km to a few hundreds of meters (generally $\approx$ 200 m) depending on the river.

- The very fine scale called here Computational Grid Scale (CompGridSc). It corresponds to the computational grid of the Saint-Venant dynamics flow model (see Section B). The CompGridSc elements are 100m long.
  Note that a lower complexity flow model (the algebraic model presented in Section C) will be defined at ReachSc only.

In all the sequel, we denote by the integer variables (r,p) the space-time discrete indices of the quantities defined at the reach and node scales e.g. $(A_r, Z_{r,p}, W_{r,p})$, see Fig. 1. We will denote by the real variables (x,t) the space-time dependency of the quantities defined at the Computational Grid Scale (CompGridSc) e.g. $(A(x,t), Q(x,t))$.

3.1.2 Synthetic SWOT-like data

The future SWOT instrument will provide time series ($\sim 4-21$ days frequency depending on the latitude) of WS elevation Z, water extent and therefore the river width W, [47,48]. These measurements will be available at different scales: at NodesSc at the nodes locations and at ReachSc when computed at the reaches locations. The measured WS slopes S will be accurate at large scales only, therefore produced at ReachSc scale.
During the Calibration-Validation (Cal-Val) phase (also referred as the "fast sampling orbit"), the instrument will have a 1-day repeat period. In the present study, the considered data are SWOT-like ones during this Cal-Val phase, [47,48]. These data are of the same nature as the forthcoming nominal SWOT data but with 1-day revisit. Moreover as a first step and following the Pepsi 1 and Pepsi 2 [16,17] benchmarks design (Discharge Algorithm Working Group of the SWOT Science Team), we first focus on the perfect data case to better analyse the capabilities of the complete method. Next, the same experiments but obtained from perturbed data (either with realistic Gaussian noise or statistical model designed from the error budget of the SWOT Science Simulator) are discussed. The latter show the robustness of the approach. The next phase will consist to consider outputs of the SWOT Science simulator or data from AirSWOT (AirBorne) campaigns [47]. This scientific approach enables to rigorously analyse the developed methods capabilities.
The considered data are as follows:

- The complete set of measurements $(Z_{r,p}, W_{r,p}, S_{r,p})$ at ReachSc for each reach r and at each instant p.
- The measurements of $(Z_{r,p}, W_{r,p})$ at NodeSc for each node r and at each instant p.

In the sequel and if ambiguous, it will be clarified at which scale the different fields and data are considered.

The SWOT instrument should provide WS measurements (Z,W) at the "node scale" $200m$ long. This fine scale data is represented by data available in the Pepsi 1 and Pepsi 2 databases at NodeSc.

Each river portion is decomposed into R reaches: $r = 1,..,R$, Fig. 14. It is assumed that $(P+1)$ instants of measurements are available; the corresponding measurements are ordered by flow elevations Z_p; the case $p = 0$ denotes the lowest water level (Z_0) and $p = (P+1)$ denotes the highest (Z_{P+1}).
Given a river portion, the resulting SWOT data set is $\{Z_{r,p}, W_{r,p}\}_{R,P+1}$ plus WS slope $\{S_{r,p}\}_{R,P+1}$ at ReachSc.

Depending on the considered flow model, the r-th "spatial point" denotes either the node or the reach number. More precisely, the node scale is the adequate scale for the Saint-Venant dynamics flow model (3), while the larger reach scale is consistent with the low complexity algebraic model (5), see [19,6] for detailed investigations.
Note that SW elevations Z may be obtained from multiple altimetry missions databases, from e.g. DAHITI database [49], however at heterogeneous accuracy and frequencies. Also, rivers width W may be extracted from e.g. the Global With Database built in [52] or from optical measurements of water extents (e.g. using Sentinel 2 images).

Finally, let us point out that if considering the nominal SWOT orbit (21 days repeat orbit), the present inverse problem remains of same nature as the present one, of course with a time limitation of the discharge estimations validity (see details at the end of the next section).

3.1.3 Synthetic in-situ data

*References data: Pepsi datasets*The synthetic data of discharge and flow observations used in this study are a compilation of the Pepsi databases which have been built up for the Pepsi 1 and 2 challenges, see [16,17]. These databases contain synthetic flow observations generated from outputs of various hydraulic flow models. These models have been calibrated. It is assumed by the Discharge Algorithm Working Group of the SWOT Science Team that these models represent sufficiently well the flow dynamics to constitute references for benchmarking discharge algorithms, [16,17]. The present SWOT like observations have been generated from these flow models outputs at daily sampling (corresponding to the CalVal orbit phase), both at NodeSc and at ReachSc. For the aforementioned reasons, here no errors have been added to the models outputs.

The considered datasets contain numerous river portions with various hydro-geomorphological properties and various regimes.
Other more sophisticated datasets exist: SWOT Instrument Simulator and AirSWOT campaigns data, see e.g. [51]. However these real-like datasets are restricted to very few river portions only. These datasets are in particular not well suited for ML experiments.
The number of days, nodes and reaches varies from one river portion to another. The number of days varies from 12 days to a full year. The number of nodes by river portion varies from 21 to 3189; the number of reaches varies from 4 to 16. Some of the river portions in this dataset were outside the range of SWOT visibility since the width was less than 50m; they were then removed from the dataset. Similarly river portions with less than 100 days of observations were removed. Finally, a total number of 29 river portions were selected which represents a total count of 145 reaches and (time multiplied by space) of 55 525 observations of any variable at ReachSc. At NodeSc, values of (Z,W) and (Q,A) are available. At ReachSc, values of (Z,W,S) and (Q,A) are available.

*Ancillary data*To train the ANN, we will use local drainage area values $\mathscr{A}$ (km^2) as an input variable. In the present altimetry context, this variable may be considered as an ancillary data. The knowledge of $\mathscr{A}$ will be the only additional prior of the inversion algorithm. The employed values of $\mathscr{A}$ are extracted from HydroSHEDS (Hydrological data and maps based on SHuttle Elevation Derivatives at multiple Scales), [33]. This database is a collection of geo-referenced datasets (vector and raster) at various scales (from 3 arc seconds to 30 arc seconds). It includes river networks, void filled DEM, watershed boundaries, drainage directions and flow accumulations. As the flow accumulation in HydroSHEDS is expressed in number of cells, a dedicated script to compute drainage area (flow accumulation in kilometers square) from the drainage directions has been developed. Then the drainage area $\mathscr{A}$ at every reach r of the PEPSI database has been computed using the geo-location of every river portions.

3.2 Statistic description of the datasets

Data representing the important features of the considered rivers portions are presented in Fig. 3. More precisely for each river portion are presented the mean value, quantiles (and outliers) for the discharge Q and width W, Fig. 3 (Top). The drainage area $\mathscr{A}$ (km^2) related to the considered river portion is also plotted, Fig. 3 (Bottom). This variable is not present in the flow models however it is an important additional information. This variable enables to better estimate discharges using ANNs (see next Section).

Three rivers (Jamuna, Mississipi downstream and Padma) present particularly high values of discharges and widths (as well as high values of drainage area $\mathscr{A}$). It can also be noted that the Missouri river portion presents high values of drainage area.

This preliminary statistic analysis will help to define learning classes for the ANN. Moreover it will help to set up a-priori PDFs and covariance kernels to solve the algebraic flow model (Section C) and the VDA optimization problem (Section D.2).

The Pearson correlation coefficient R^2 has been computed between numerous variables: Z, W, elevation variations dZ, wetted crossed-sections variations dA (both being computed between two ordered overpasses), also soil composition (percentage of clay, sand and silt), mean annual rain, mean annual temperature and land use. The numerical results are not detailed here; the only high correlations are between (Q,dZ), (Q,dA) and somehow (dA,W).
The Pearson correlation coefficient R^2 has been computed between Q and numerous variables: Z, W, elevation variations, dZ (above lowest observed elevation Z_0), wetted crossed-sections variations dA (above Z_0), also soil composition (percentage of clay, sand and silt), mean annual rain, mean annual temperature, land use and drainage area. The numerical results are not detailed here. The only high correlations are (in decreasing order) (Q,dA), (Q,W) and somehow $(Q,\mathscr{A})$.

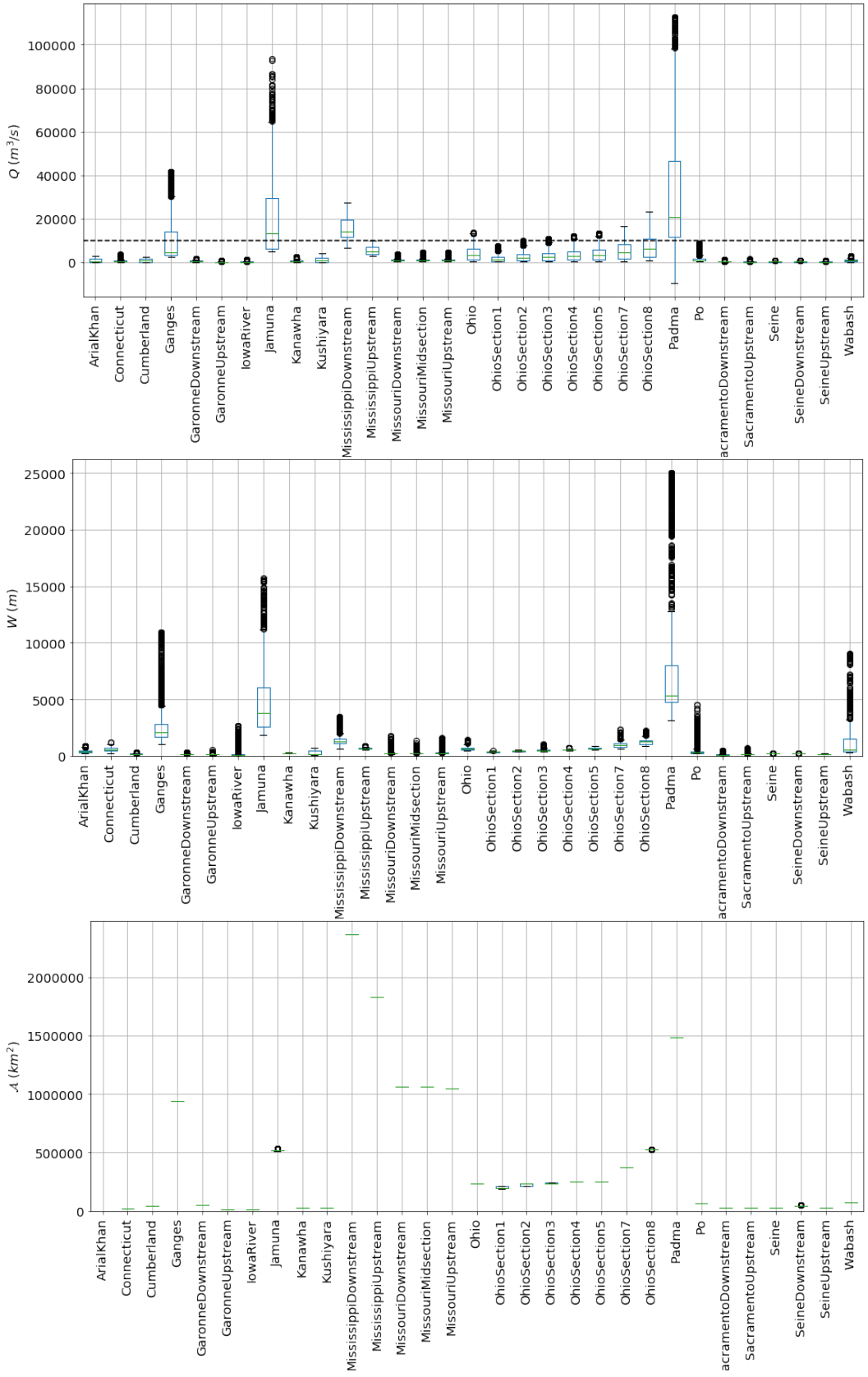

Fig. 3 Hydraulic features of the considered river portions. The green bar indicates the mean value, boxes indicate ±25% quartiles, circles are outliers (Python boxplot command). (Top) Discharge Q (m^3/s). (Middle) Width W (m). (Bottom) Drainage area $\mathscr{A}$ (km^2).

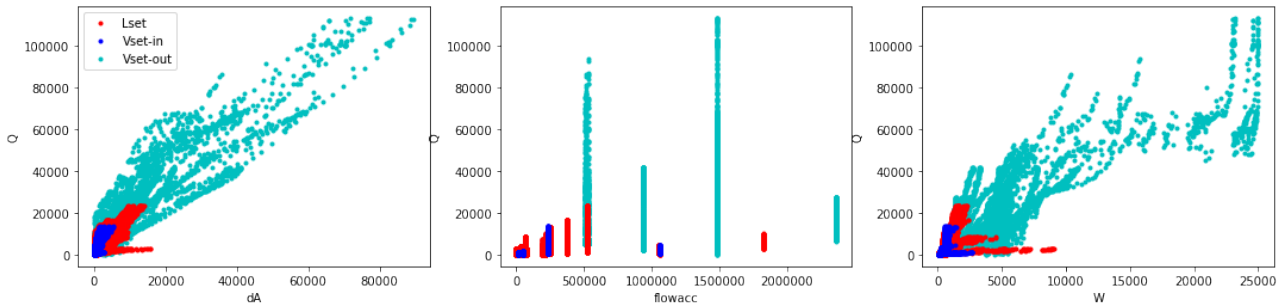

Fig. 4 Distribution of the data in Q-Lset (red), Q-Vset-in (blue) and Q-Vset-out (cyan) in planes (dA, Q), ($\mathscr{A}$, Q) and (W,Q).

3.3 Effective low Froude Strickler coefficient

Given the datasets previously presented, the friction coefficient K is computed by solving the algebraic flow model (5). Since this low complexity flow model relies on the low Froude assumption, this provides the effective low Froude Strickler coefficient. The obtained estimations, see Fig. 5, highlight the large range of the effective low Froude Strickler coefficient. Moreover it confirms a-posteriori physically-consistent values of K resulting from the measurements.

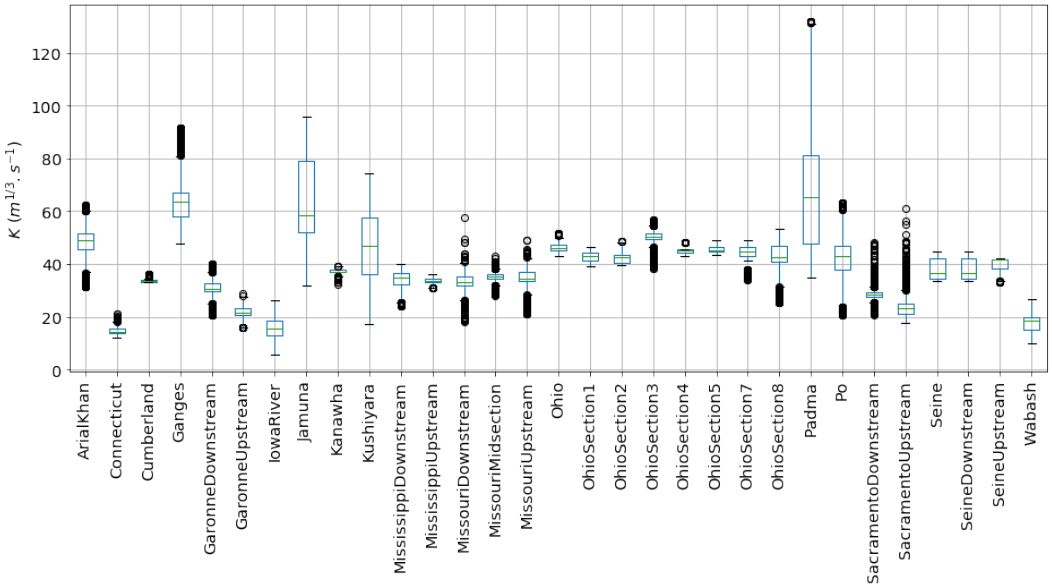

Fig. 5 Effective Strickler friction coefficient K computed by solving the low Froude (algebraic) flow model (5), given data of the considered river portions.

4 Numerical results of Step 1): First estimation at large scale

4.1 Step 1a) Estimations of Q by ANN

A first estimation of discharge is obtained by performing the ANN detailed in Section E. The obtained estimation is denoted by $Q^{(ANN)}$. This ANN is first trained on the learning set denoted by Q-Lset. Next, the trained ANN accuracy is analyzed on two different validation datasets denoted by Q-Vset-in and Q-Vset-out respectively. These three distinct datasets are detailed below.

4.1.1 Learning set Q-Lset and validation sets (Q-Vset-in, Q-Vset-out)

ANN constitutes excellent interpolators (if well trained); on the contrary, they are poor extrapolators. Then, it is necessary to train an ANN such as those employed here (see Section E for details) on well-designed data classes. Here, data can be easily classified into 2 rough classes. The threshold to define these 2 classes follows from the statistical analysis presented in figures 3 and 4: one considers rivers presenting a discharge value lower or greater than 10 000 $m3/s$. This threshold value is obviously not the only possible, however similar results would be obtained with a different threshold value if

respecting a reasonably good splitting of the dataset. The number of 2 classes is here chosen to illustrate the concept.
The learning dataset denoted by Q-Lset is constituted by river portions presenting mean discharge value lower than 10 000m^3/s, see Fig. 3 (Top Left). All river portions satisfying this criteria are incorporated into Q-Lset excepted 4 of them. Three (3) rivers randomly chosen have not been included into Q-Lset: Garonne downstream, Missouri mid-section and Ohio. An additional fourth river has been selected since partly inside the learning dataset only: the Iowa river. Indeed, this river presents discharge values Q, drainage area $\mathscr{A}$ and lowest observed cross-section A_0 at the very lower limit of the dataset values and with values of width W and A_0 outside the other dataset values see Fig. 4.1.1. As a consequence, this river will represent a partly learned case only for the supervised ML process presented in next section.
All these 4 rivers will be used to evaluate the prediction capabilities of the trained ANN. They constitute the validation dataset denoted by Q-Vset-in.

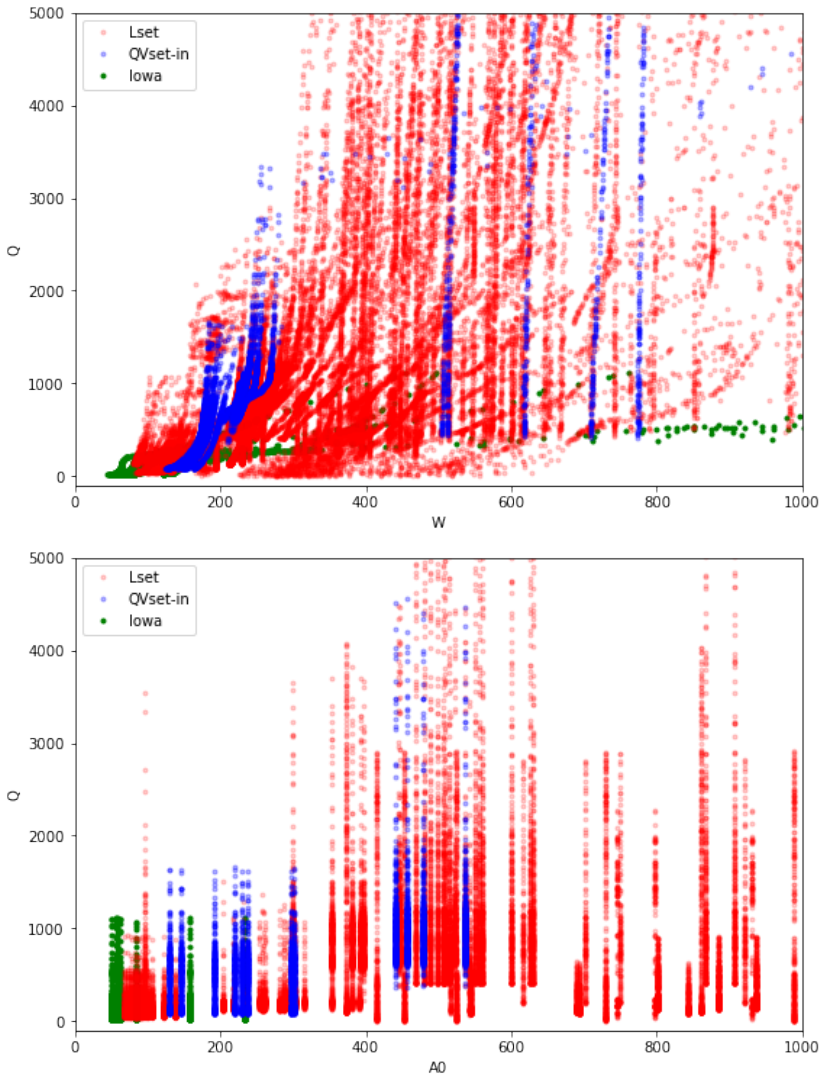

Fig. 6 Distribution of the data in (red) Q-Lset, (blue) Q-Vset-in (except the Iowa river) and the (green) Iowa river. (Top) in (W,Q) plane. (Bottom) in (A_0,Q) plane

Finally, the learning dataset Q-Lset contains data related to $(24-4)=20$ river portions. This corresponds to a total of 41 747 training "samples" (ML jargon). At each observation location r and each overpass instant p corresponds a sample. Each sample contains the 4 predictor variables $(dA, W, S, \mathscr{A})$ and the single target variable Q.
In the presented numerical results, the 4 non-considered rivers in Q-Lset are: Garonne downstream, Missouri mid-section, Ohio and Iowa. Recall that the latter is an extreme case in terms of values range, see figures 3 and 4.

The rivers portions presenting discharge values greater than 10 000 m^3/s, that is Jamuna, Mississippi downstream and Padma, see Fig. 3 (Top Left), are gathered in the dataset denoted by Q-Vset-out. These data represent less than 10% of

Criteria	nRMSE	NSE	R^2
Mean value for the 20 rivers	12.85 %	0.95	0.98

Table 1 Accuracy of the trained ANN for the 20 rivers of Q-Lset: obtained mean value of criteria. (Recall that these preliminary estimations are next improved by the algebraic flow model).

the total samples. Q-Vset-out will be used to evaluate the estimation capabilities of the trained ANN for rivers presenting values of Q far outside the learning values range.

The ANN is trained on Q-Lset (supervised learning) that is rivers presenting discharges lower than the threshold, like rivers contained in Q-Vset-in. On the contrary, rivers contained in Q-Vset-out present discharges greater than this threshold. As a consequence, one can hope relatively good estimations (predictions) for rivers belonging to Q-Vset-in (if the ANN is well-trained), however bad estimations for rivers belonging Q-Vset-out. Also, it is expected that the ANN provides less accurate estimations for the partly learned Iowa case compared to the 3 other rivers of Q-V-set-in, since the Iowa represents a partly learned case.

Finally, let us remark that classifying rivers within a few rough classes (like the present two here) is realistic. Indeed, rough classifications can be done e.g. from the GRDC database [8] or from the GRADES database [34].

4.1.2 Assessment of the trained ANN

Performance criteria Few criteria are used to measure the estimation accuracy: the normalized RMSE (*nRMSE*), the Nash–Sutcliffe Efficiency coefficient (*NSE*) and the Pearson correlation coefficient (R^2). Applied to the variable Q, these criteria read:

- $nRMSE(Q)$=$RMSE(Q)/\bar{Q}^{obs}$ with $RMSE(Q) = \left(\frac{1}{n}\sum_{i=1}^{n}(Q_i^{est} - Q_i^{obs})^2\right)^{1/2}$, Q_i^{est} (resp. Q_i^{obs}) is the estimated (resp. observed) i-th discharge value.

- NSE criteria reads: $NSE = 1 - \frac{\sum_{i=1}^{n}(Q_i^{est}-Q_i^{obs})^2}{\sum_{i=1}^{n}(Q_i^{obs}-\bar{Q}^{obs})^2}$. NSE value range within $[-\infty, 1]$.

- R^2 criteria reads: $R^2(Q) = \frac{\sum_{i=1}^{n}(Q_i^{est}-\bar{Q}^{est})(Q_i^{obs}-\bar{Q}^{obs})}{\left(\sum_{i=1}^{n}(Q_i^{est}-\bar{Q}^{est})^2\right)^{1/2}\left(\sum_{i=1}^{n}(Q_i^{obs}-\bar{Q}^{obs})^2\right)^{1/2}}$.

Learning phase After optimization (learning stage), the loss function value (15) is low: the mean values of the misfit equals 189 (m^3/s). The mean *nRMSE* and R^2 over the 20 learned rivers are excellent, see Tab. 1. After training, the ANN approximates very accurately the learned rivers. The estimated discharges for the 4 first rivers in alphabetical order of Q-Lset are presented in Fig. 7. This figure aims at verifying the good learning process only.

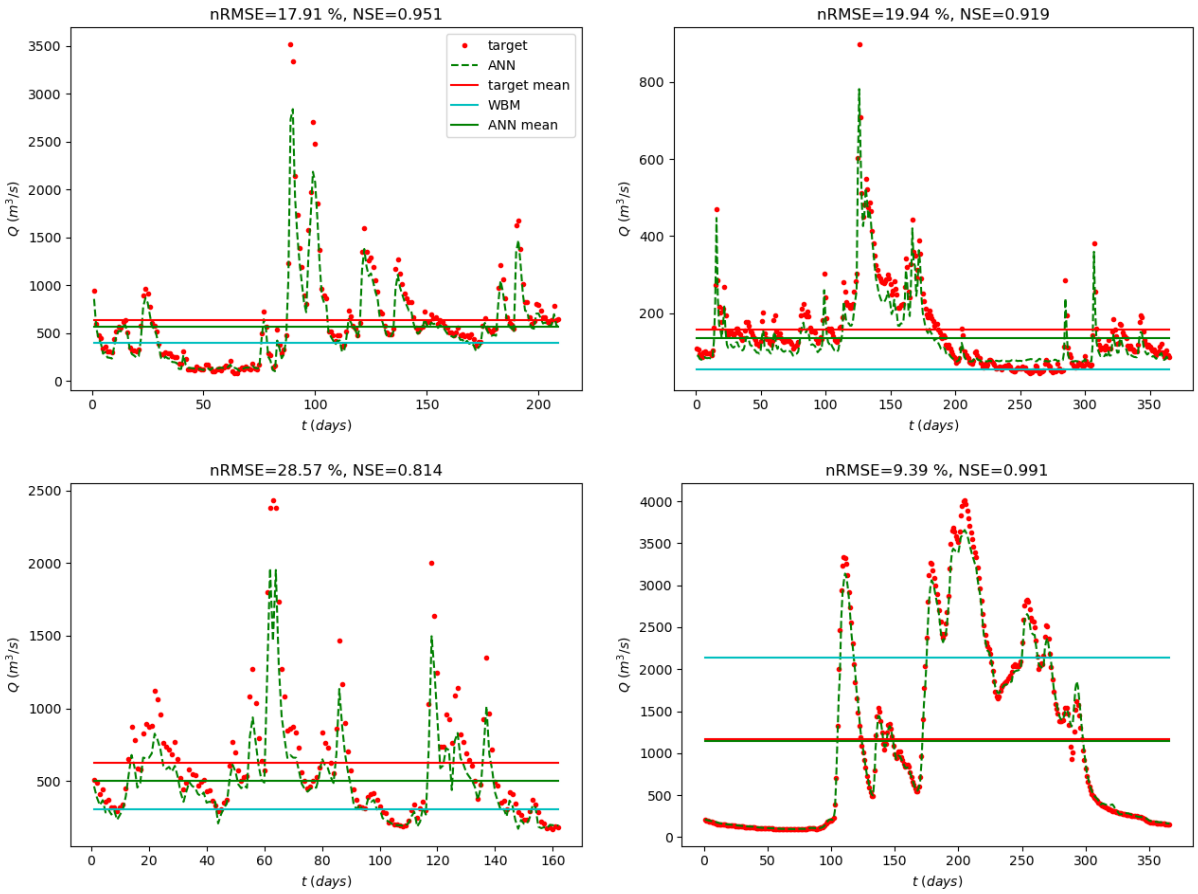

Fig. 7 Discharge values estimated by the trained ANN for the 4 first rivers in alphabetical order of the learning set Q-Lset. (Top Left) Connecticut. (Top Right) Garonne upstream. (Bottom Left) Kanawha. (Bottom Right) Kushiyara.

4.1.3 Estimations

Estimations for rivers belonging to Q-Vset-in Below are presented the results obtained for the 4 river portions belonging to Q-Vset-in: these rivers do not belong to the learning set Q-Lset. This experiment participates to the estimator assessment. Q-Vset-in: each river present mean discharge values lower than 10 000 m^3/s like those belonging to Q-Lset, see Fig. 3(Top) and Fig.4. Recall that these 3 of these 4 river portions have been randomly chosen. The fourth one (Iowa river) has been selected for its extreme features (it represents a partly learned case only). The results show that the ANN globally provides good estimations, see Tab.2 and the hydrographs in Fig. 8.

The same experiment has been performed for all possible combinations of rivers that is a complete K-fold cross-validation, the results are very similar to the present one (presented for the rivers portions Garonne downstream, Iowa, Missouri mid-section and Ohio only).

The obtained results are as follows. Garonne downstream, Missouri mid-section and Ohio hydrographs are very well estimated: the nRMSE are lower than 30%. If removing the bias value for each case, the accuracy becomes excellent, Tab.2. Local peaks only are not well captured. Note that in the Ohio case, the peak values are greater than 10000 m^3/s that is outside the learning values range.
As previously mentioned, the case of Iowa river is different. This river presents values partly outside the learning range, in particular for the width W and for the lowest observed wetted cross-section A_0. Then, the estimated hydrograph is accurate excepted for the lowest values: the very low discharges are over-estimated by the ANN. This result is consistent with the string feature of ANNs: potential excellent interpolators however poor extrapolators. This over-estimation of the lowest values is the reason why the nRMSE criteria is quite high, and even if one removes the bias, Tab.2.
This experiments shows that the ANN is a good predictor for rivers clearly inside the learning set.

Let us recall that uncertainty error on discharge measurements may be considered as $\approx 30\%$ (see e.g. [22] and references therein) that is higher than the obtained nRMSE on the estimations (Tab. 2 without the Iowa case).

Rivers	nRMSE	NSE	nRMSE with bias removed
Garonne downstream	26.6 %	0.81	7.7 %
Iowa	44.4 %	0.85	44.5 %
Missouri mid-section	18.7 %	0.85	7.0 %
Ohio	18.9 %	0.94	18.9 %

Table 2 Accuracy of the trained ANN for the 4 rivers of the validation set Q-Vset-in. (Recall that these preliminary estimations are next improved by the algebraic flow model)

In terms of mean value of the estimations, the ANN mean value is better than the WBM values (horizontal dotted lines in Fig. 8), excepted in the Garonne downstream case (Fig. 8 (Top)(L)) where all estimations are already good.

In conclusion, in terms of mean value (therefore in a bias point of view), the present ML approach enables to decrease the bias compared if starting from the WBM values. Moreover, the ANN provides already very good estimations of discharge (including the space-time variations). This is true for rivers presenting values clearly in the same range as those in the learning set. On the contrary, as soon as the values of input parameters are slightly outside the learning value ranges, the prediction becomes much less accurate as shows the Iowa case. For the Iowa case, the ANN is unable to reproduce some of the lowest flows (see 8 for days 0-70 and 225-360). Remember the Iowa case has lower values of W than the learning set Q-Lset (see 4.1.1), which explains why the prediction of the ANN are not accurate at low flows. This result is consistent with the widely observed property of dense ANNs as the present one.

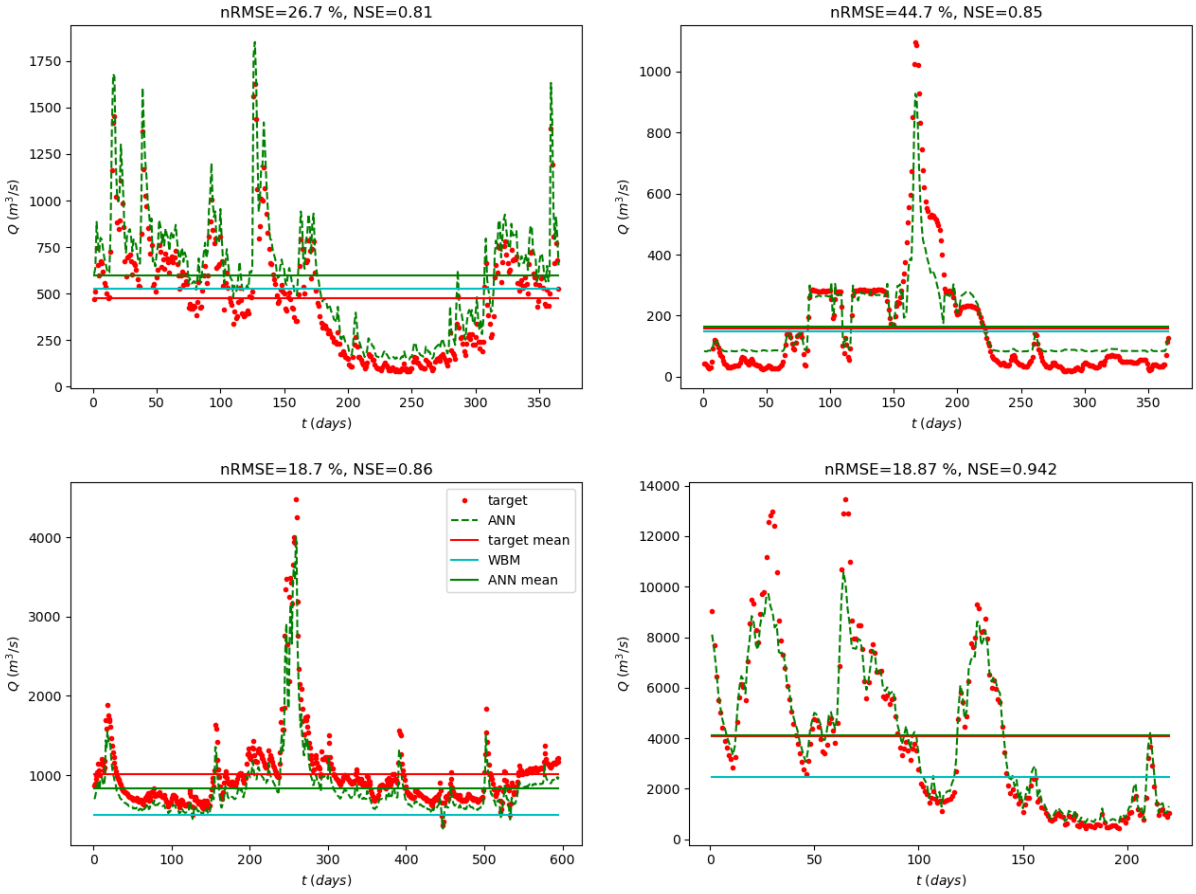

Fig. 8 Discharge values estimated by the trained ANN for the river portions belonging to Q-Vset-in. (Top Left) Garonne Downstream. (Top Right) Iowa. (Bottom Left) Missouri mid-section. (Bottom Right) Ohio.

Estimations for rivers belonging to Q-Vset-out Despite the well-known poor property of extrapolation of ANNs, we briefly present here results obtained for the 3 rivers belonging to Q-Vset-out, that is rivers presenting discharge values much greater than 10 000 m^3/s, see figures 3 (Top) and 4. The obtained performance criteria are indicated in Tab. 3 and the results for two rivers are plotted in Fig. 9. As expected, the estimated values by the ANN are lower than the true ones.

Rivers	nRMSE	NSE	nRMSE with bias removed
Jamuna	73.3 %	0.28	52.5 %
Mississipi downstream	43.6 %	-0.56	16.4 %
Padma	109.4 %	-0.67	71.4 %

Table 3 Accuracy of the trained ANN for the 3 rivers of Q-Vset-out. (Recall that these first purely data-driven estimations are next improved by the algebraic flow model).

Variations of the hydrographs are roughly recovered; however peaks are greatly smoothed (the obtained hydrographs are not presented here).
These numerical results confirm that an ANN does not provide good estimations for input values outside the learning ranges. This experiment highlights the importance to first make a classification of rivers (see Fig. 4), next to apply a correctly trained ANN.

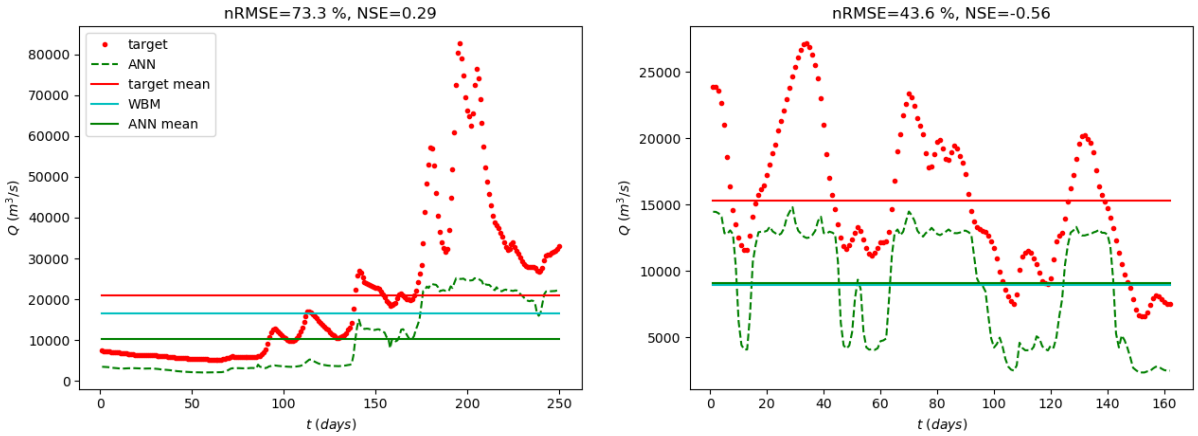

Fig. 9 Discharge values estimated by the trained ANN for two river portions belonging to Q-Vset-out. (Left) Jamuna. (Right) Mississippi Downstream.

4.1.4 On the sensitivity of the estimations with respect to error measurements and data frequency

As already mentioned, as a first step and following the Pepsi 1 and Pepsi 2 [16,17] benchmarks design (from the Discharge Algorithm Working Group of the SWOT Science Team), algorithms evaluations are performed on data with no noise and 1 day sampling. Recall that the first SWOT observations will have a 1 day repeat period during the Cal/Val phase. Numerous experiments with realistic noise have been performed to be confident in the method robustness and its relative insensitivity with respect to a few input data properties. The results are here briefly presented.

With perturbed WS measurements (Z,W) We have tested the ANN estimation if considering perturbed measurements (Z,W), respecting the expected instrument accuracy. Data have been perturbed using a statistical model developed from the error budget of the SWOT Science Simulator, see [17]. The obtained results are as follows: NRMSE equals 18.1% and NSE equals 0.91. These results have to be compared with those indicated in Table 1, that is: 18.1% vs 12.8% and 0.91 vs 0.98. These results show that the ANN estimations remain robust to the inaccuracy of the WS measurements.

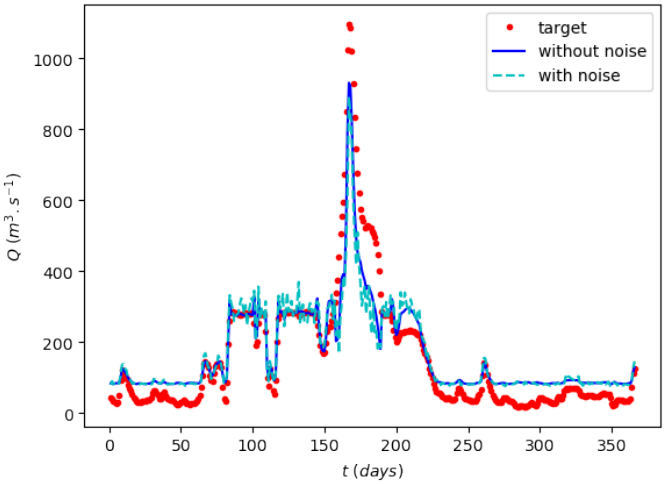

Fig. 10 Discharge values estimated by the trained ANN for the Iowa River using perfect (blue) and noisy (cyan) WS measurements.

With less frequent WS measurements In the present ANN, the concept of spatial correlation or time correlation between the samples does not exist. Indeed, each sample corresponds to a set of $(4+1)$ values which are correlated neither in time nor in space; they are simple point-wise snapshots. As a consequence, the ANN does not "see" any space or time structure in the datasets. After optimization (training), an optimal ANN model has been built. The latter enables to reproduce invariants between the four input variables and the output variable Q.

Following the argument above, the present ANN should provide a similar accuracy if considering much less frequent observations (of course with same volume and quality of data). This assertion has been numerically verified (for example for a frequency of 5 days).

Recall that the discharge estimations are valid a time approximately equal to the wave travelling through the river portion that is roughly a few hours (and potentially up to a day depending on rivers), see e.g. [43,6,30]. If considering the nominal SWOT orbit (21 days repeat orbit) with perturbed data, the inverse problem remains of same nature as the present one, with the estimations valid a few hours around the observations instant only.

4.2 Step 1b) Deductions of A_0 and K from the Low-Froude model

Given the WS measurements and $Q^{(ANN)}$ from Step 1a), estimations of $(A_0(x), K(x;h(x,t)))$ are computed as the solution of the algebraic Low Froude flow model (5). As a consequence, these estimations of A_0 and K are valid at large scale only, see Section A. Observe that estimating the unobserved wetted cross-section $A_{0,r}$ is here equivalent to estimate the bathymetry b_r, see Section A and Fig. 14.

4.2.1 Parametrization of K

The Strickler coefficient K depends on the location r but also on the water depth value $h(x,t)$: it is a space-time dependent parameter of the flow model. Different parametrizations for K exist in the literature like the Einstein Formula or the Debord formula, see e.g. [7]. Here to reduce this space-time dependent model parameter, K is simply defined as local power-laws as it is proposed in [18,6]:

$$K_{r,p} \equiv K((K_{0,r}, \beta_r^K); h_{r,p}) = K_{0,r}\,(h_{r,p})^{\beta_r^K} \quad \forall r\, \forall p \tag{2}$$

In the following, $K_{r,p}$ refers to its parametrization defined by (2).

4.2.2 Numerical results

The algebraic system (5) is solved by using the Metropolis-Hasting algorithm (MCMC method implemented in the Python package PyMC3). In the Metropolis-Hasting algorithm, the a-priori PDF are set as follows: $\mathscr{U}(10,100)$ for $K_{0,r}$, $\mathscr{N}(0,0.3)$ for β_r^K and $\mathscr{N}(\mu_{A0/\bar{A}}, \sigma_{A0/\bar{A}})$ for $(A_0/\overline{A})_r$. Following the statistics obtained from the HydroSWOT and Pepsi databases, one has: $\mu_{A0/\bar{A}} = 0.73$, $\sigma_{A0/\bar{A}} = 0.21$.

Given $A_{0,r}^{(0)}$ and the measurements $(Z_{r,0}, W_{r,0})$, the corresponding bathymetry profile $b_r^{(0)}$ is explicitly obtained, see Section

3.1. These estimations are obtained at ReachSc only; see the bathymetry values in Fig. 12 (”prior VDA” values, black solid lines).
The target bathymetry values are those employed in the various calibrated reference flow models (HEC-Ras, LisFlood, DassFlow-2D) which have been performed to obtain the synthetic data available in the Pepsi 1 and Pepsi 2 datasets, see [16, 17] and references therein. In Fig. 12, the ”true” (target) values are represented by the red dots. The latter are computed from the effective rectangular values of the unobserved lowest cross-section $A_0(W = W_0, H_0 = Z_0 - b)$. These “true” values of bathymetry b and A_0 are available at node scale (NodeSc) (see Section 3 and red dots in Fig. 12).

The bathymetry obtained by solving the algebraic system (5) (”prior (VDA)” lines in Fig. 12) may be biased or not, depending on the rivers or the locations. In the Garonne case, the large scale pattern is quite good and relatively unbiased. On the contrary in the Ohio case, the bias is much more important (≈ 10 m).

5 Numerical results of Step 2): Estimations at fine scale based on the Saint-Venant model and VDA

The VDA algorithm aims at estimating $Q(x,t)$, $A_0(x)$ and $K(x,h(x,t))$ at fine scale (ComputGridSc). VDA relies on an iterative algorithm which requires first values of the inferred quantities. The latter are set to the first estimations of $Q_{r,p}^{(ANN)}$, $A_{0,r}$ and $K_r(h_{r,p})$ obtained at Step 1). The VDA algorithm aims at estimating ”input parameters” of the Saint-Venant flow model (3) (see details in Appendix B): the time-dependent discharge at inflow $Q_{in}(t)$, the bathymetry $b(x)$ (equivalently the unobserved wetted cross-section $A_0(x)$, see Fig. 14) and an effective friction law $K(x,h(x,t))$.
Let us denote by c the complete set of parameters: $c(x,t) = (Q_{in}(t), A_0(x), K(x,h(x,t)))$. VDA consists to minimize a cost function $j(c)$, see (10), which measures the discrepancy between data and flow model outputs. The present formulation of VDA is quite sophisticated; it is detailed in Section D. Like it is classical in non-linear geophysics problems, the cost function $j(c)$ is likely not convex therefore presenting different local minima. Moreover, it is likely ill-conditioned: the sensitivity of $j(c)$ with respect to local variations of c near a local optimal value is low. As a consequence, the definition of the first guess value $c^{(0)} = (Q_{in}^{(0)}(t), A_0^{(0)}(x), K^{(0)}(x,h(x,t))))$ is crucial.

5.1 First guess values

The first guess values are set from the values obtained at Step 1). First guesses are denoted by an superscript $^{(0)}$. The first guesses $A_{0,r}^{(0)}$ and $K_r(h_{r,p})^{(0)}$ are directly the values obtained at Step 1b). The first guess value $Q_{in,p}^{(0)}$ is not the value $Q_{in,p}^{(ANN)}$: it is the solution of the Low Froude model (5) given $A_{0,r}^{(0)}$ and $K_r(h_{r,p})^{(0)}$. Indeed, this low Froude estimation $Q_{in,p}^{(0)}$ better catches the variations of the true values than $Q^{(ANN)}$, see e.g. Fig. 12 (“prior” curve). This estimation $Q_{in,p}^{(0)}$ may be viewed as a physically-consistent correction of the purely data driven estimation $Q^{(ANN)}$.

Remark. If a mean value of $Q^{(true)}$ is known for a given period (e.g. a week, a month), then one can make fit this information with $Q_{in,p}^{(0)}$ therefore providing a less biased first estimation. However, in ungauged rivers, such mean value is unavailable.

5.2 On the VDA algorithm convergence

VDA aims at solving the minimization problem (10) by an iterative descent algorithm. It is important to first analyse the algorithm convergence. The employed minimization algorithm (see Section D for details) generally converges in less than 100 iterations. In a very few cases, the convergence is reached after ≈ 150 iterations only, see e.g. Fig. 11. After convergence, the misfit values on WS elevation, see (8), is always excellent: standard deviation $\sigma_{misfit} \approx 10$ cm, see Fig. 11. This value of σ_{misfit} is lower than the expected value for the SWOT instrument ($\sigma_{SWOT} = 25$ cm, see [47]).

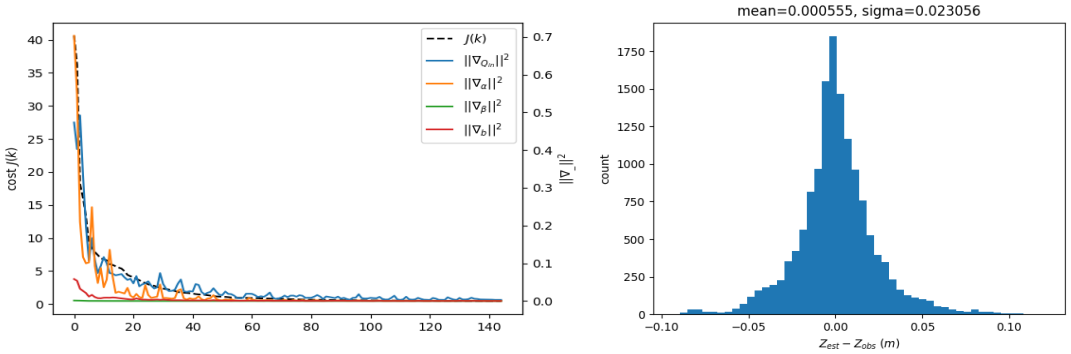

Fig. 11 VDA algorithm convergence: (Left) A typical convergence curve: cost function $J(k)$ (dashed black line) and gradient components (colored solid lines) vs iterations (Garonne Downstream case). 'BCOO1', 'KPAR01', 'KPAR02', 'BATHY' corresponds to the gradient component wrt to $Qin(t)$, $K_0(x)$, $\beta^K(x)$, $b(x)$ respectively (all in norms 2). (Right) Misfit values $|Z_{r,p} - Z^{obs}_{r,p}|\ \forall r \forall p$ in meters (see (8)) after convergence.

5.3 The bias issue in a VDA context

The equifinality issue pointed out in Section 2.2 does not depend on the method to solve the inverse problem (P) e.g. VDA or sequential filters. It is intrinsic to the use of the classical flow equations only. Let us detail consequences of the bias issue when solving Problem (P) by VDA. At each minimization iteration, the "model constraint" (3) (see Section 5 for details) is satisfied by an infinity of flow states values (A,Q) characterized by the friction parameter K. In other words, the flow model (3) constrains the inverse problem solution $(Q_{in}(t), A_0(x), K(h))$ up a to a multiplicative factor only. This feature is (of course) retrieved in the numerical results: the space-time variations of the infered discharge values are reasonably accurate however up to a bias. This bias depends to the first guess and to the priors introduced in the VDA formulation (the covariance matrix parameters), see Section 2.2 for details.
Note that if the bathymetry $b(x)$ is given (therefore $A_0(x)$), the re-scaled unknown (A^*, Q^*) introduced in Section 2.2 does not satisfy the flow model anymore. As a consequence in this case, Problem (P) based on Saint-Venant like equations may be well-posed. In other respect, recall that a single measurement of bathymetry enables the bathymetry estimation along a relatively long river portion, see [19,6].

In the case a mean value of Q is known (e.g. seasonal or annual value) then this enables to fix the bias issue too. The numerical results below confirm this statement: given an accurate mean value of Q, estimations of $Q(x,t)$ are very accurate, without bias.
As a consequence, the VDA solution (solution of the minimization problem (10) under the flow model constraint) depends on priors: the first guess but also the covariance matrix B, see Section D.2. The covariance matrix B is necessary to ensure the robustness of the algorithm convergence. However, B is a prior probabilistic model (defined from the prior parameters $\sigma_\square$)[1] and the computed optimal solution depends these prior parameters $\sigma_\square$. This feature is classical and well known in VDA communities, see e.g. [36,24] and references therein. The reader may refer e.g. to [2,38] for a formal proof showing equivalences between VDA covariance based solutions and Bayesian estimations.

5.4 Results for rivers belonging to Q-Vset-in

Given the first guess values $(Q^{(0)}_{in,p}, A_{0,r}, (K_0, \beta^K)_r)^{(0)}$ computed as previously described, the VDA based on the Saint-Venant flow model provides $(Q(x,t), A_0(x), K(x;h))$ at the finest scale (CompGridSc).
Recall that the first guess $(Q_{in}(t), A_0(x), K)^{(0)}$ is a physically-consistent solution, however potentially presenting a bias. The bias is here a-priori smaller than if using WBM values but still, see Section 4.1.3.
The VDA algorithm explores solutions in a "vicinity" of this first guess $(Q_{in}(t), A_0(x), K)^{(0)}$. As already mentioned in Section 3, the target bathymetries are not true ones; they are those employed in the reference numerical flow models which have been performed to generate the synthetic data available in the Pepsi 1 and Pepsi 2 datasets [16,17]. In Fig. 12, these "true" (target) values are represented by the red dots. The latter are computed from the effective rectangular values of the unobserved flow area $A_0(W = W_0, H_0 = (Z_0 - b))$ (see Section A). These values of bathymetry b and A_0 are available at node scale (NodeSc) only.

[1] For sake of simplicity, the subscript $\square$ denotes here any index of the variable σ

Numerical results for the three rivers randomly chosen in Q-Vset-in plus the partly learned case (Iowa river) are presented in Fig. 12. Performance criteria are indicated in Tab. 4. The estimations are plotted in Fig. 12 ("prior (VDA)" denotes $Q_{in,p}^{(0)}$). For the three rivers aleatory chosen (Garonne downstream, Missouri mid-section and Ohio), the three estimations of Q (ANN, first guess and VDA solution) are accurate. The nRMSE are lower than the standard error made on discharge measurements, see e.g. [12,28]. In terms of nRMSE and NSE criteria, Q^{VDA} values are similar to Q^{ANN} ones. However, the VDA estimation Q^{VDA} captures better the variations than the purely data-driven estimation Q^{ANN}, Fig. 12, since physically-based.
If removing the bias value for each case, the accuracy is excellent again, see Tab.4.

Concerning the Iowa river, again, the estimated hydrograph is accurate for low values and much less for high ones and the peak, see Fig. 12 b). The peak corresponds to a 200-years return period. The cross-section shape presents an important discontinuity of $W(Z)$ between the minor bed and the major bed (as no other case presents in the Pepsi-2 dataset). Then, the power-law model of $K(Z)$ prevents to well represent this sharp change when flooding. A better representation of K would be the use of the Einstein or the Debord formula, see [7]. However, these formulas require an accurate detection of the minor bed boundaries, which is a task far to be trivial in all cases using SWOT observables only.

Finally, this second set of estimations confirm that the ML phase (Step 1) must be performed on datasets well inside the learning values ranges. Moreover, the VDA process acts as a physical filter (both in space and time) of the uncorrelated purely data-driven estimations. However, this step does not significantly reduce a potential bias introduced in the first guess.

Concerning the bathymetry, the VDA improves its estimation in the Garonne case, in particular in pools (low values of $b(x)$), see Fig. 12 (Down)(L). In the Ohio case, the bathymetry estimation remains relatively inaccurate despite the excellent discharge estimation. In this case, the bathymetry error is balanced by the adjustment of K.

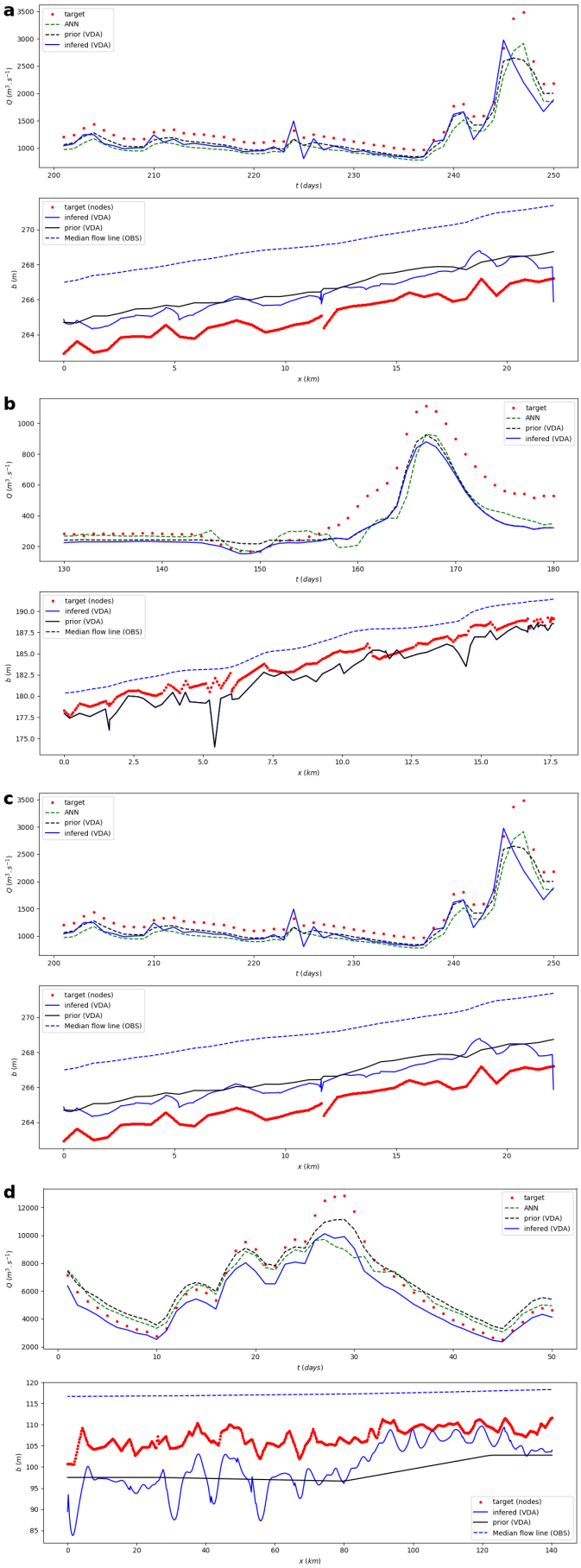

Fig. 12 Estimations for the rivers in Q-Vset-in. (a) Garonne downstream. (b) Iowa. (c) Missouri mid-section. (d) Ohio.
On each figure: (top) discharge values $Q_in(t)$ vs time during the assimilation window only, (bottom) bathymetry values $b(x)$ and the observed median flow line $Z_{med}^{(obs)}(x)$.

River name	Prior (ANN) nRMSE	Prior (ANN) NSE	Infered by VDA nRMSE	Infered by VDA NSE	Infered by VDA nRMSE with bias removed
Garonne downstream	22.0 %	0.76	18.4 %	0.83	7.7 %
Iowa	28.3 %	0.84	29.8 %	0.82	20.6 %
Missouri mid-section	15.2 %	0.85	21.2 %	0.71	15.3 %
Ohio	12.0 %	0.93	14.5 %	0.90	10.0 %

Table 4 Performance obtained on the discharge estimation $Q^{(VDA)}$ for the rivers in Q-Vset-in.

5.5 On the robustness of the estimations in presence of noise

Let us recall that the accuracy of Q^{ANN} is affected slightly only if considering perturbed SW measurements (with realistic Gaussian noises), see Section 4.1.4. As a consequence and following our previous studies [6,51,30,17], one can expect that the VDA process propagates these errors but with the same order of magnitude i.e. without amplification. In [6,51, 30,17], sensitivity analyses of the VDA algorithm (and of the algebraic flow model) have demonstrated a good robustness with respect to the WS measurement errors.
In summary, the main source of the estimation errors is from far the bias potentially present in the first guess.
The VDA results obtained from the first guess that would be computed using the ANN with perturbed WS measurements (Fig. 10) are not presented here. Indeed, they are similar to those with noise within the range of variation already observed in [30,17].

Finally, recall that the accuracy of $Q^{(ANN)}$ is not affected if having less frequent data e.g. 10 days period (with similar amount and quality of data), since the input data of the ANN are non correlated. As a consequence, the remarks above made for the Cal-Val orbit case remain true for the nominal orbit case.

Remark 1 What about the rivers belonging to Q-Vset-out? As expected, the results obtained for rivers belonging to Q-Vset-out (see Section 4.1.1), that is rivers presenting values outside the learning ranges, are much less accurate. For these rivers, the VDA estimations are better than the first guess values: the VDA estimation captures relatively well the time discharge variations, on contrary to the purely ANN-based ones. However, a relatively large bias remains: the true values are under-estimated like the ANN does in this case. These results are here not analysed more in details.

In the end, these experiments confirm that for fully ungauged rivers (in particular without any prior rough mean value), the present VDA formulation based on first guesses derived from the ANN - Low Froude model strategy, enables to accurately capture the space-time discharge variations at small scale (CompGridSc). However, as expected following the mathematical analysis presented in [30] and Section 5.3, the VDA estimation does not make vanish the potential bias of the first guess.
Recall that if any mean value of Q is known (e.g. the annual value) then the estimations become accurate with nRMSE much less than the standard error made on discharge measurements ([12,28]).

6 Step 3) Low complexity operational model

Past the assimilation period (learning period), the river has been "learned" and the H2iVDI algorithm (Fig. 2) provides a calibrated dynamic flow model valid at fine scale (based the dynamics Saint-Venant flow model (3)). This model provides estimations of discharge at $\approx 14-22\%$ error (rRMSE, see Tab. 4) at fine scale for rivers respecting the validity range of the ANN (the Garonne down stream, Missouri mid-section, Ohio rivers in the present case). This accuracy is better than the discharge measurements one, see e.g. [12,28].
Moreover, the estimation of an effective bathymetry $b(x)$ (equivalently $A_0(x)$) is available at fine scale (therefore at large scale too). However it may be biased as previously discussed. The effective friction parameter $K(x;h(x,t))$ is scale dependent and model dependent. Its adjustment absorbs modeling errors and the potential bias made on b.

The algebraic low Froude flow model (5) can be performed in Real-Time (extremely short CPU-time): it can be used in an operational objective as a surrogate model given newly acquired data.
Note that while the dynamics flow model (3) enables space-time extrapolation of discharge values (typically outside the measurements locations), this is not possible if with the algebraic flow model (5).

6.1 Real-Time estimations given newly acquired WS measurements: the RT algorithm

Given newly acquired data, an algorithm to estimate $Q_{r,p}$ in real computational time (μ-sec CPU-time) can be as follows.

Step 1) Re-calibration of the friction coefficient K. Given $(Q_{r,p}^{(VDA)}, A_{0,r}^{(VDA)})$ obtained after the VDA process, the algebraic Low Froude (LF) model (5) is solved to obtain $(K_{0,r}, \beta_r^K)^{(LF)}$. This algebraic system is solved using the Metropolis-Hasting algorithm. This provides effective LF friction parameters $K_{r,p}^{(LF)}$; $K_{r,p}^{(LF)} = K((K_{0,r}, \beta_r^K)^{(LF)}, A_{0,r}; Z_{r,p})$, see (4). This stage can be done "offline" at any moment past the calibration period.

Step 2) RT estimations given newly acquired WS measurements. Given $(A_{0,r}^{(VDA)}, (K_{0,r}, \beta_r^K)^{(LF)})$ $\forall r$, given new WS measurements $(Z_{r,p}, W_{r,p}, S_{r,p})_{R,P+1}$, the coefficients $(c_{r,p}^{(k)})$, for $k = 1,2,3$, and $c_r^{(4)}$ in (5), can be evaluated.
Next, $Q_{r,p}^{(RT)}, \forall r\ \forall p$, can be explicitly obtained from (5), therefore computed in Real Time (RT).

In the sequel, the algorithm based on these two steps is called the RT algorithm.

Recall that the algebraic Low Froude model is valid at the Reach Scale and at the measurement "instants" only, see [6, 30].

6.2 Numerical results

In this section, the VDA estimations have been obtained from relatively short time periods compared to a complete year, see Tab. 5. However these learning periods are relatively representative of the potential annual variations. Outside these assimilation periods, the WS measurements are considered as newly acquired: the estimation denoted by $Q^{(RT)}$ is obtained following the RT algorithm above.
$Q^{(RT)}$ is computed for the assimilation period too. This enables to compare $Q^{(RT)}$ with $Q^{(target)}$ for the two time periods, see Fig. 13.
During the assimilation period, the estimation $Q^{(RT)}$ differs from $Q_{in,p}^{(0)}$ ("prior VDA") because of the bathymetry has been re-evaluated after the VDA phase.
Results are presented for the three rivers fully belonging to Q-Vset-in (the partly learned case is here skipped). Performance criteria are indicated in Tab. 5.
For the rivers Garonne downstream and Ohio, the prior estimation $Q^{(ANN)}$ is already excellent, see Section 4.1.3. Then, the present RT-estimations given newly acquired WS measurements are accurate ($\approx 17-30\%$ nRMSE, see Tab. 5).

River name	Assimilation period (days)	Complete estimation period (days)	nRMSE	NSE
Garonne downstream	[120-170]	[1-365]	28.6 %	0.78
Missouri mid-section	[201-250]	[1-595]	29.9 %	0.64
Ohio	[1-50]	[1-220]	17.3 %	0.95

Table 5 The assimilation periods are those considered for the VDA processes. The performance scores are those obtained for $Q^{(RT)}$ during the complete period.

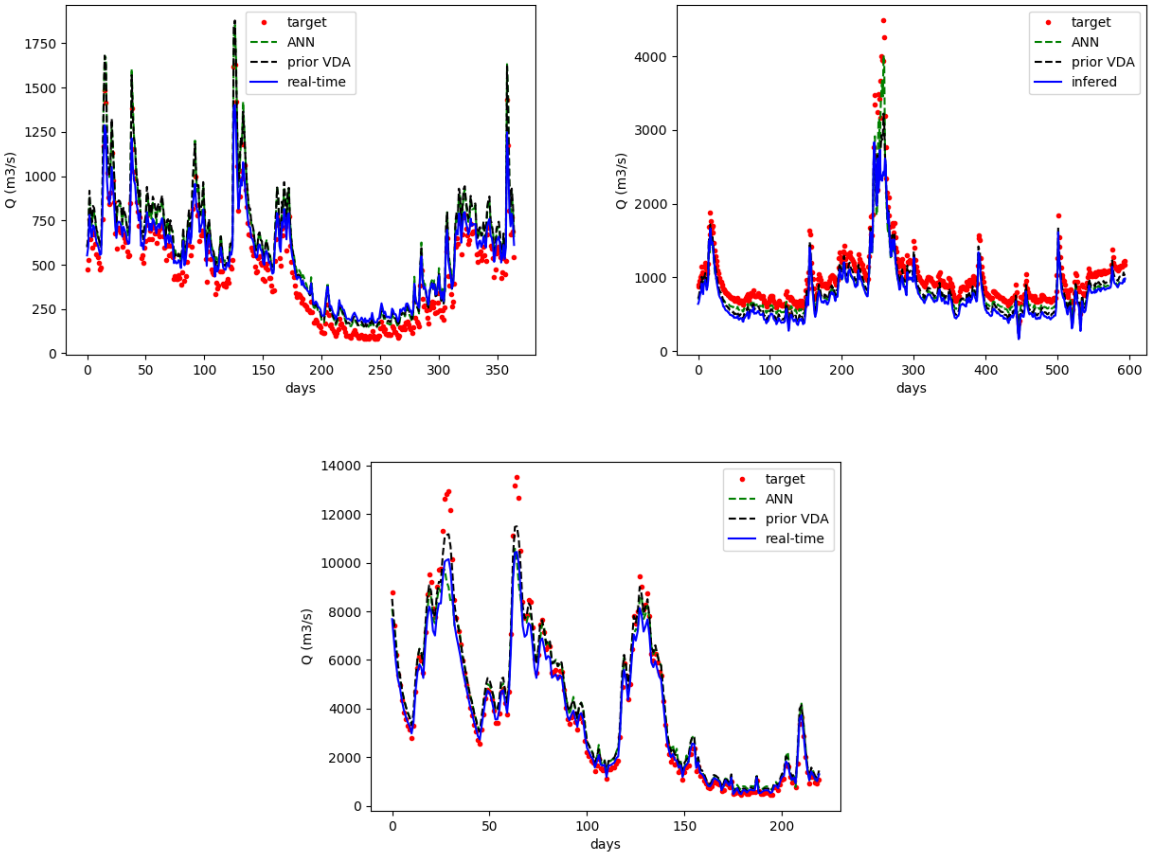

Fig. 13 Real-Time (RT) discharge estimations $Q(t)$ vs time during the complete time period by solving the algebraic flow model (5). "Prior VDA" corresponds to the first guess $Q^{(0)}_{in,p}$ presented in Section 5.1. (Top-left) Garonne Downstream, (Top-Right) Missouri mid-section, (Bottom) Ohio.

Remark 2 On the possibility to compute uncertainty envelopes. In the RT algorithm (Section 6.1), one can easily produce uncertainty envelops on the final estimation $Q^{(RT)}$ as follows.
At Step 1), one can introduce an uncertainty model on $Q^{(VDA)}$ by considering it as a random variable e.g. $Q^{(VDA)} \sim \mathcal{N}(\bar{Q}^{(VDA)}, \sigma_Q)$. Next when performing the Metropolis-Hasting algorithm to compute the effective Low-Froude values $(K_0, \beta^K)^{(LF)}_r$, one obtains $K_{r,p}$ as a random variable with a corresponding standard deviation σ_K.
At Step 2), $Q^{(RT)}_{r,p}$ (the explicit solution of (5)) becomes a random variable with a corresponding standard deviation σ_{final}.

In summary, by setting a-priori uncertainty through the PDFs of $Q^{(VDA)}$, $K(K_{0,r}, \beta^K_r)$, $A_{0,r}$ and the WS measurements, one obtains the posterior PDF of $Q^{(RT)}$ (therefore the uncertainties).

Like in any Bayesian approach, the prior PDFs are arbitrary (here those on $Q^{(VDA)}$ in particular). That is the reason why we prefer here to do not present such uncertainty envelopes which fully depends on the priors which are unknown.

7 Conclusion

This study proposes a hybrid Machine Learning - Physically Informed Data Assimilation approach (an algorithm combining ANN and VDA) to infer rivers discharge for ungauged rivers from altimetry measurements. The developed algorithm,

named H2iVDI for Hierarchical Hybrid Variational Discharge Inference, constitutes an important improvement at different levels of the algorithm HiVDI presented in [30,51,17]. The challenging goal is to diminish the bias obtained in the previous studies, see e.g. [16,17]. The bias is the result of the equifinality issue mathematically demonstrated in [30] (see also [19] for preliminary discussions) and re-discussed here in detail.
The H2iVDI algorithm is based on three steps. At Step 1), a deep ANN provides a first discharge estimation Q^{ANN}. Given Q^{ANN}, first estimations of the bathymetry and the parameters of the friction empirical law are derived from the low Froude flow model first developed in [30,19]. Step 1) provides already quite good estimations of the complete unknowns set $(Q(x,t),b(x),K(x,h(x,t))$ at large scale ($\approx km$). The input variables of the ANN (purely data-driven estimations first introduced in [31]) are the SWOT-like WS measurements plus an ancillary data: the drainage area $\mathscr{A}$ (km^2) here taken from the HydroSHEDS database [33]. This additional physical information improves the ANN accuracy; this combination somehow playing the role of a closure law. This purely data-driven estimation is next used to define the first guess of an advanced formulation of VDA at Step 2). The VDA algorithm enables to accurately capture the space-time variations of the discharge, provides a bathymetry and effective friction laws at small scale ($\approx 10\ m$). The potential bias present in the first guess is not strongly diminished by the VDA iterative algorithm, in accordance with the mathematical analysis. In the end, the accuracy using the complete hybrid ANN-VDA based algorithm is better than previous inter-comparison studies [16,17], moreover with a bias greatly diminished.

Past the assimilation period, the rivers have been "learned", the H2iVDI algorithm provides two calibrated flow models: the dynamics Saint-Venant flow model valid at fine scale and the low complexity algebraic low Froude flow model valid at large scale. Given newly acquired WS measurements, the low complexity algebraic low Froude flow model constitutes a possible operational estimator by providing estimation of Q (at large scale) in real computational time (μ-sec CPU-time). Moreover, the resulting dynamic Saint-Venant flow model enables to generate numerical rating curves at virtual stations (Stage Fall Discharge laws) like in e.g. [44,37]. Such simple and operational laws may be derived from datasets acquired by multi-missions such as Sentinel-3, Sentinel-6 (ESA), IceSat2 (NASA) too.

The numerical tests and performance criteria are here analyzed for synthetic SWOT-like data generated from the outputs of various calibrated flow models for 29 rivers portions available in the reference Pepsi 1 and 2 datasets, [16,17]). The obtained flow models / estimators provide estimations of discharge within the expected 30% range error (rRMSE), that is within the range of uncertainty on discharge measurements (see e.g. [12,28]) for all cases presenting input data within the learning ranges.
Since the estimators at Step 1) are neither spatially nor temporally correlated, their results quality is not affected when considering more or less frequent data e.g. with a few days frequency instead of 1 day (with estimations valid a time approximately equal to the wave travelling through the river portion that is between approximately between a hour and a day depending on rivers portions). The results show that the ANN estimations remain robust in presence of realistic noise on the WS measurements. The VDA-based estimations (Step 2) with respect to realistic noises are robust too, confirming previous studies [30,51,6]. All these numerical experiments ensure the robustness and demonstrate a reasonable accuracy of the presently developed algorithm.

This version of the algorithm is implemented in the Confluence framework that will be in charge of the testing and production of the discharge parameters for the SWOT mission. Thus this version is ready for operational ingestion of real SWOT WS measurements as soon as they are available, and particularly for the first phase of the mission (Cal/Val phase with 1 day repeat orbit).

The present H2iVDI algorithm is implemented in the open-source computational software DassFlow [39]. Based on the recently released multi-dimensional 2D-1D software version [46], the estimation of discharge and bathymetry from satellite WS measurements (SWOT mission in particular) could be used to calibrate this complete surface flows numerical model which enable to simulate complex 1D networks, including braided portions [45] and local flood plains, see e.g. [46].

Declaration of Conflict of Interests / Competing Interests

The authors declare no conflict of interests / competing interests, despite the existence of other teams working on developing other algorithms to estimate discharge from SWOT data. The source code for DassFlow is fully open and has been distributed to many institutions worldwide including e.g. INRAE Montpellier (DassFlow V1, Malaterre's group). In addition, our co-authors of joint comparison articles (who are also our friends) are regularly invited to our laboratory and to serve on our PhD defense committees.

Acknowledgments

K. Larnier, research engineer at CS GROUP - France corp., has been funded by CNES. The authors would like to acknowledge Renato Frasson (NASA/JPL, USA) and Mike Durand (Univ. Ohio, USA) for sharing Pepsi-2 dataset.
They would like to acknowledge two anonymous reviewers for their comments, which helped to improve the clarity of the manuscript.

References

1. KM Andreadis, CB Brinkerhoff, and CJ Gleason. Constraining the assimilation of swot observations with hydraulic geometry relations. *Water Resources Research*, 56(5):e2019WR026611, 2020.
2. Mark Asch, Marc Bocquet, and Maëlle Nodet. *Data assimilation: methods, algorithms, and applications*, volume 11. SIAM, 2016.
3. Paul D Bates, Matthew S Horritt, and Timothy J Fewtrell. A simple inertial formulation of the shallow water equations for efficient two-dimensional flood inundation modelling. *Journal of hydrology*, 387(1-2):33–45, 2010.
4. S. Biancamaria, D. P. Lettenmaier, and T. M. Pavelsky. The swot mission and its capabilities for land hydrology. *Surveys in Geophysics*, 37(2):307–337, 2016.
5. F Bouttier and P Courtier. Data assimilation concepts and methods march 1999. *Meteorological training course lecture series. ECMWF*, page 59, 2002.
6. P. Brisset, J. Monnier, P.-A. Garambois, and H. Roux. On the assimilation of altimetric data in 1d saint-venant river flow models. *Adv. Water Ress.*, 119:41–59, 2018.
7. M. Carlier. *Hydraulique générale et appliquée*. Centre de Recherche et d'Essais Chatou: Collection du. Eyrolles, 1972.
8. Global Runoff Data Centre. 56068 koblenz, germany.
9. Lajiao Chen and Lizhe Wang. Recent advance in earth observation big data for hydrology. *Big Earth Data*, 2(1):86–107, 2018.
10. V.T. Chow. Handbook of applied hydrology. *McGraw-Hill Book Co., New-York, 1467 pages*, 1964.
11. Sagy Cohen, Albert J. Kettner, and James P.M. Syvitski. Global suspended sediment and water discharge dynamics between 1960 and 2010: Continental trends and intra-basin sensitivity. *Global and Planetary Change*, 115:44–58, 2014.
12. G. Coxon, J. Freer, I. K. Westerberg, T. Wagener, R. Woods, and P. J. Smith. A novel framework for discharge uncertainty quantification applied to 500 uk gauging stations. *Water Resources Research*, 51(7):5531–5546, 2015.
13. J. Cunge. Practical aspects of computational river hydraulics. *Pitman Publishing Ltd. London,(17 CUN), 1980, 420*, 1980.
14. L. Dalcín, R. Paz, and M. Storti. Mpi for python. *Journal of Parallel and Distributed Computing*, 65(9):1108–1115, 2005.
15. S Lawrence Dingman. Analytical derivation of at-a-station hydraulic–geometry relations. *Journal of Hydrology*, 334(1-2):17–27, 2007.
16. M. Durand, CJ. Gleason, P-A. Garambois, D. Bjerklie, LC. Smith, H. Roux, E. Rodriguez, PD. Bates, TM. Pavelsky, J Monnier, et al. An intercomparison of remote sensing river discharge estimation algorithms from measurements of river height, width, and slope. *Water Resources Research*, 2016.
17. Renato Prata de Moraes Frasson, Michael T Durand, Kevin Larnier, Colin Gleason, Konstantinos M Andreadis, Mark Hagemann, Robert Dudley, David Bjerklie, Hind Oubanas, Pierre-André Garambois, et al. Exploring the factors controlling the error characteristics of the surface water and ocean topography mission discharge estimates. *Water Resources Research*, 57(6):e2020WR028519, 2021.
18. P-A. Garambois, K. Larnier, J. Monnier, P. Finaud-Guyot, J. Verley, A. Montazem, and S. Calmant. Variational inference of effective channel and ungauged anabranching river discharge from multi-satellite water heights of different spatial sparsity. *Journal of Hydrology*, 581:124409, 2020.
19. P-A. Garambois and J. Monnier. Inference of effective river properties from remotely sensed observations of water surface. *Advances in Water Resources*, 79:103–120, 2015.
20. I Gejadze, P-O Malaterre, H Oubanas, and V Shutyaev. A new robust discharge estimation method applied in the context of swot satellite data processing. *Journal of Hydrology*, page 127909, 2022.
21. X. Glorot, A. Bordes, and Y. Bengio. Deep sparse rectifier neural networks. In *Proceedings of the fourteenth international conference on artificial intelligence and statistics*, pages 315–323, 2011.
22. James A Gore and James Banning. Discharge measurements and streamflow analysis. In *Methods in Stream Ecology, Volume 1*, pages 49–70. Elsevier, 2017.
23. S. Haben, A. Lawless, and N. Nichols. Conditioning and preconditioning of the variational data assimilation problem. *Computers & Fluids*, 46(1):252–256, 2011.
24. S. Haben, A. Lawless, and N. Nichols. Conditioning of incremental variational data assimilation, with application to the met office system. *Tellus A*, 63(4):782–792, 2011.
25. L. Hascoët and V. Pascual. The Tapenade Automatic Differentiation tool: Principles, Model, and Specification. *ACM Transactions On Mathematical Software*, 39(3), 2013.
26. B. Kaltenbacher, A. Neubauer, and O. Scherzer. *Iterative regularization methods for nonlinear ill-posed problems*, volume 6. Walter de Gruyter, 2008.
27. M. Kanevski, A. Pozdnoukhov, A. Pozdnukhov, and V. Timonin. *Machine learning for spatial environmental data: theory, applications, and software*. EPFL press, 2009.
28. Julie E. Kiang, Chris Gazoorian, Hilary McMillan, Gemma Coxon, Jérôme Le Coz, Ida K. Westerberg, Arnaud Belleville, Damien Sevrez, Anna E. Sikorska, Asgeir Petersen-Øverleir, Trond Reitan, Jim Freer, Benjamin Renard, Valentin Mansanarez, and Robert Mason. A comparison of methods for streamflow uncertainty estimation. *Water Resources Research*, 54(10):7149–7176, 2018.
29. D. Kingma and J. Ba. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014.
30. K. Larnier, J. Monnier, P. A. Garambois, and J. Verley. River discharge and bathymetry estimation from swot altimetry measurements. *Inverse Problems in Sciences and Engineering*, 29(6):759–789, 2020.
31. Kevin Larnier and Jerome Monnier. Hybrid neural network–variational data assimilation algorithm to infer river discharges from swot-like data. *Nonlinear Processes in Geophysics Discussions*, pages 1–30, 2020.
32. Y. LeCun, Y. Bengio, and G. Hinton. Deep learning. *nature*, 521(7553):436, 2015.
33. Bernhard Lehner, Kristine Verdin, and Andy Jarvis. New global hydrography derived from spaceborne elevation data. *Eos, Transactions American Geophysical Union*, 89(10):93–94, 2008.
34. Peirong Lin, Ming Pan, Hylke E Beck, Yuan Yang, Dai Yamazaki, Renato Frasson, Cédric H David, Michael Durand, Tamlin M Pavelsky, George H Allen, et al. Global reconstruction of naturalized river flows at 2.94 million reaches. *Water resources research*, 55(8):6499–6516, 2019.

35. A. Lorenc, S. Ballard, R. Bell, N. Ingleby, P. Andrews, D. Barker, J. Bray, A. Clayton, T. Dalby, D. Li, et al. The met. office global three-dimensional variational data assimilation scheme. *Quarterly Journal of the Royal Meteorological Society*, 126(570):2991–3012, 2000.
36. A. C. Lorenc. Optimal nonlinear objective analysis. *Quarterly Journal of the Royal Meteorological Society*, 114(479):205–240, 1988.
37. Thibault Malou, P-A Garambois, Adrien Paris, Jerome Monnier, and Kevin Larnier. Generation and analysis of stage-fall-discharge laws from coupled hydrological-hydraulic river network model integrating sparse multi-satellite data. *Journal of Hydrology*, 603:126993, 2021.
38. J. Monnier. *Data assimilation, Optimal Control and Model Learning*. Open Online Course, INSA Toulouse, France, 2021.
39. J. Monnier, K. Larnier, and . others. Data assimilation for free surface flows. Technical report, Mathematics Institute of Toulouse - INSA - CNES - CNRS - CS group - INRAE, 2020.
40. J. Monnier and J. Zhu. Inference of the bottom topography in anisothermal mildly-sheared shallow ice flows. *Computer Methods in Applied Mechanics and Engineering*, 348:954–977, 2019.
41. H. Oubanas, I. Gejadze, P.-O. Malaterre, M. Durand, R. Wei, R. P. M. Frasson, and A. Domeneghetti. Discharge estimation in ungauged basins through variational data assimilation: The potential of the swot mission. *Water Resources Research*, 54(3):2405–2423, 2018.
42. H Oubanas, I Gejadze, P-O Malaterre, and F Mercier. River discharge estimation from synthetic swot-type observations using variational data assimilation and the full saint-venant hydraulic model. *Journal of Hydrology*, 559:638–647, 2018.
43. Rodrigo CD Paiva, Michael T Durand, and Faisal Hossain. Spatiotemporal interpolation of discharge across a river network by using synthetic swot satellite data. *Water Resources Research*, 51(1):430–449, 2015.
44. Adrien Paris, Rodrigo Dias de Paiva, Joecila Santos da Silva, Daniel Medeiros Moreira, Stéphane Calmant, Pierre-André Garambois, Walter Collischonn, Marie-Paule Bonnet, and Frederique Seyler. Stage-discharge rating curves based on satellite altimetry and modeled discharge in the amazon basin. *Water Resources Research*, 52(5):3787–3814, 2016.
45. J. Pujol, P-A. Garambois, P. Finaud-Guyaot, J. Monnier, K. Larnier, R. Mose, S. Biancamaria, H. Yesou, D. Moreira, A. Paris, and S. Calmant. Estimation of multiple inflows and effective channel by assimilation of multi-satellite hydraulic signatures: The ungauged anabranching negro river. *Journal of Hydrology*, 2020.
46. Léo Pujol, Pierre-André Garambois, and Jérôme Monnier. Multi-dimensional hydrological-hydraulic model with variational data assimilation for river networks and floodplains. *EGUsphere*, pages 1–44, 2022.
47. E. Rodriguez and . others. Swot science requirements document. *JPL document, JPL*, 2012.
48. Ernesto Rodriguez and Daniel Esteban-Fernandez. The surface water and ocean topography mission (swot): The ka-band radar interferometer (karin) for water level measurements at all scales. In *Sensors, Systems, and Next-Generation Satellites XIV*, volume 7826, page 782614. International Society for Optics and Photonics, 2010.
49. Christian Schwatke, Denise Dettmering, Wolfgang Bosch, and Florian Seitz. Dahiti–an innovative approach for estimating water level time series over inland waters using multi-mission satellite altimetry. *Hydrology and Earth System Sciences 19 (10): 4345-4364*, 2015.
50. A. Tarantola. *Inverse problem theory and methods for model parameter estimation*, volume 89. SIAM, 2005.
51. S. Tuozzolo, G. Lind, B. Overstreet, J. Mangano, M. Fonstad, M. Hagemann, R. Frasson, K. Larnier, P-A. Garambois, J. Monnier, and M. Durand. Estimating river discharge with swath altimetry: A proof of concept using airswot observations. *Geophysical Research Letters*, 46(3):1459–1466, 2019.
52. Dai Yamazaki, Fiachra O'Loughlin, Mark A Trigg, Zachary F Miller, Tamlin M Pavelsky, and Paul D Bates. Development of the global width database for large rivers. *Water Resources Research*, 50(4):3467–3480, 2014.

A River geometries

Recall that the SWOT-like measurements consist in sets $(Z_{r,p}, W_{r,p})_{R,P+1}$. Moreover at ReachSc, WS slopes values $S_{r,p}$ are available and taken into account into the algebraic flow model (see next section). The values $S_{r,p}$ are either deduced from the elevation values Z or estimated by an internal instrument process.

The considered river geometries are derived from the SWOT-like datasets $(Z_{r,p}, W_{r,p})_{R,P+1}$ following the suggestion made in [19]. The cross-sectional geometry consists in discrete cross sections formed by asymmetrical trapezium layers $(Z_{r,p}, W_{r,p})$, see Fig. 14. The cross-sectional areas $A_{r,p}$ satisfy: $A_{r,p} = A_{r,0} + \delta A_{r,p} = A_{r,0} + \int_{Z_{r,0}}^{Z_{r,p}} W_r(h)dh \;\; \forall r \, \forall p \geq 1$.

The variations $\delta A_{r,p}$ are approximated by the trapeziums: $\delta A_{r,p} \approx \sum_{q=1}^{p} \frac{1}{2}(W_r^q + W_r^{q-1})(h_r^q - h_r^{q-1})$.

The lowest cross-sectional areas denoted by $A_{r,0}$ $(r = 1, \cdots, R)$ are unobserved; they are key unknowns of the flow models. They can be represented by rectangles or any other fixed shape (e.g. a parabola); all the other cross-sectional areas are trapezoidal. Next, for simplicity and regularization purposes, the shape is approximated at a cubic spline curve in the least square sense (green curve), see Fig. 14.

For all considered rivers we have the hydraulic radius R^h which satisfies: $R^h_{r,p} \approx h_{r,p}$. Also since $W >> h$, it follows the effective depth expression: $h_{r,p} \approx (A_{r,0} + \delta A_{r,p})(W_{r,0} + W_{r,p})^{-1}$.

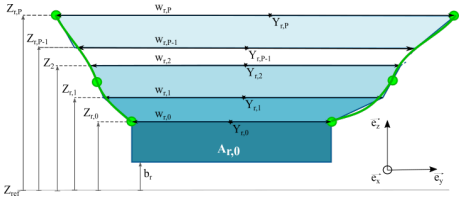

Fig. 14 A cross section A_r is the superimposition of the observed trapeziums $A_{r,p}$ defined from the p-ordered SWOT measurements $(Z_{r,p}, W_{r,p})$ + the unobserved lowest cross-section A_0. Next, the shape is approximated by a cubic spline (green curve).

B The dynamic flow model: Saint-Venant's equations

The considered dynamic flow model is the 1D Saint-Venant equations in their non conservative form in (A, Q) variables; A the wetted-cross section $\left[m^2\right]$, Q the discharge $\left[m^3.s^{-1}\right]$. The equations read as follows, see e.g. [10]

$$\begin{cases} \partial_t A + \partial_x Q & = 0 \\ \partial_t Q + \partial_x \left(\frac{Q^2}{A}\right) + gA\,\partial_x Z & = -gA\,S_f(A,Q;K) \end{cases} \tag{3}$$

where g is the gravity magnitude $\left[m.s^{-2}\right]$, Z is the WS elevation $[m]$, $Z=(b+h)$ where b is the lowest rectangular cross-section (bed) level $[m]$ and h is the water depth $[m]$.

At inflow (upstream), the discharge $Q_{in}(t)$ is imposed.

At outflow (downstream), if known the WS elevation Z_{out} is imposed. If unknown, the normal depth (based on the Manning-Strickler equilibrium equation) is imposed. Recall that the normal depth depends on the prior values (K,A_0) at outflow.

The RHS term S_f is the classical Manning-Strickler friction term: $S_f(A,Q;K)=\frac{|Q|Q}{K^2A^2R_h^{4/3}}$ with K the Strickler roughness coefficient $\left[m^{1/3}.s^{-1}\right]$ with $R_h \backsimeq h\ [m]$. K is defined following the local power-law (4). The discharge Q is related to the water velocity u $\left[m.s^{-1}\right]$ by the relation: $Q=uA$.

This 1D Saint-Venant model is discretized using the classical implicit Preissmann scheme (see e.g. [13]) with a space cell length $\Delta x = 200$m and time step $\Delta t = 1$h.

The numerical model has been implemented in the computational software DassFlow [39].

C The algebraic low-Froude flow model

In this section, the so-called low Froude flow model dedicated is presented. This model has been developed especially for the scale of satellite measurements. The low-Froude validity has been investigated in detail in [19]; next the model has been improved and re-evaluated in [6,30]. Mathematically speaking it is an algebraic system depending on the three variables $(Q_{r,p},K_{r,p},A_{0,r})$ containing equations similar to the Manning-Strickler law at (r,p) given. The Strickler friction coefficient $K_{r,p}$ depends on space (index r) and time (index p). To reduce its complexity, K is represented as a power-law in function of the water depth h.

Reduced parametrization of K As already mentioned, the Strickler friction coefficient K is defined as local power-laws at the large scale (ReachSc) as defined in (2). As a consequence, given $R\times(P+1)$ measurements $Z_{r,p}$, the friction parameter $K_{r,p}$ is represented by $2R$ parameters only: $(K_{0,r},\beta_r^K)_{1\le r\le R}$. This reduced parametrization provides a local effective power-law in h. The law reads in function of the WS measurements as:

$$K_{r,p} \equiv K((K_{0,r},\beta_r^K);A_{0,r},W_{r,0},Z_{r,p}) = K_{0,r}\left(Z_{r,p}-Z_{r,0}+\frac{1}{W_{r,0}}A_{0,r}\right)^{\beta_r^K} \quad \forall r\ \forall p \tag{4}$$

The algebraic flow model While deriving the flow equations (mass and momentum conservation laws), the Low Froude assumption ($Fr^2 << 1$) is applied. The resulting model is an algebraic system of R equations (one equation per reach r); each equation is similar to the Manning-Strickler law, see [30,6]. Since this "Low Froude" flow model is algebraic, its complexity is low. Using the present reduced parametrization (4), this system reads as follows:

$$Q_{r,p}^{\frac{3}{5}} = K_{0,r}^{3/5}\ \left(c_{r,p}^{(1)}A_{0,r}+c_{r,p}^{(2)}\right)\left(c_r^{(4)}A_{0,r}+c_{r,p}^{(3)}\right)^{3/5\beta_r^K} \quad 1\le r\le R, \quad 0\le p\le P \tag{5}$$

The coefficients $c_{r,p}^{(k)}$, $k=1,\cdots,3$, and $c_r^{(4)}$ can be evaluated from the altimetry measurements. Their expressions are:

$$c_{r,p}^{(1)} = W_{r,p}^{\frac{-2}{5}}\ S_{r,p}^{3/10}, \quad c_{r,p}^{(2)} = c_{r,p}^{(1)}\,\delta A_{r,p}, \quad c_{r,p}^{(3)} = (Z_{r,p}-Z_{r,0}), \quad c_r^{(4)} = \frac{1}{W_{r,0}} \tag{6}$$

System (5) constitutes the so-called algebraic flow model. It contains $R(P+1)$ equations.

If considering the full set of unknowns $((K_{0,r},\beta_r^K),A_{0,r},Q_{r,p})$ i.e. $R(3+(P+1))$ unknowns, it is an underdetermined system therefore admitting an infinity of solutions.

If the discharge values $Q_{r,p}$ are given, the system admits an unique solution for the two other variables $((K_{0,r},\beta_r^K),A_{0,r})$ ($2R$ unknowns). This is the way the first guesses $(K_{r,p},A_{0,r})^{(0)}$ are computed given $Q_{r,p}^{ANN}$, see Section 5.1.

Moreover this system is employed to compute real-time estimations of Q, see Section 6.

Finally it is worth to notice that if $A_{0,r}$ is given $\forall r$ (therefore all wetted areas $A_{r,p}=A_{r,0}+\delta A_{r,p}\ \forall r\forall p$ are given) then by solving the algebraic flow model (5) the inference of the *ratio* $(Q/K)_{r,p}$ is possible but not the sough variables $(Q_{r,p},K_{r,p})$. (Of course, this remark applies to the classical scalar Manning-Strickler law too).

D The Variational Data Assimilation (VDA) method

In this section we detail the developed VDA method to infer the "input parameters" of the Saint-Venant flow model (3). The estimated parameters are: the time-dependent discharge at inflow $Q_{in}(t)$, the bathymetry $b(x)$ (equivalently $A_0(x)$) and the friction coefficient K. K is parametrized as indicated in (4). VDA consists to minimize a cost function which measures the discrepancy between data and the flow model outputs.

D.1 The optimization formulation

The employed VDA formulation is the one developed in [30] with a few improvements detailed later. At the observational scale, the discrete unknown "parameter" of the dynamic flow model (Saint-Venant's equations) reads:

$$c = \left(Q_{in,0},\, ...,\, Q_{in,P};\, b_1,\, ...,\, b_R;(K_{0,1},\beta_1^K),\ldots,(K_{0,R},\beta_{0,R}^K)\right)^T \tag{7}$$

The subscript p denotes the instant, $p \in [0..P]$, r denotes the reach number, $r \in [1..R]$, see Fig. 14. The parameters used to impose a normal depth at downstream, see Section B, are considered as unknown parameters too (otherwise the flow would be controlled by the imposed outflow condition).

The cost function aims at measuring the misfit between data (therefore at the observations scale) and the Saint-Venant (fine scale) flow model output. It is defined as:

$$j(c) \equiv j_{obs}(c) = \frac{1}{2}\sum_{p=0}^{P}\sum_{r=1}^{R}\left(Z_{r,p}(c) - Z_{r,p}^{obs}\right)^2 \tag{8}$$

This cost function j has to be minimized, starting from a first guess value (prior) $c^{(0)}$. However following [35,30], the following change of variable is applied:

$$k = B^{-1/2}(\mathrm{c} - \mathrm{c}^{prior}) \tag{9}$$

with B a covariance (symmetric definite positive) matrix, $B = B^{1/2}B^{1/2}$.

Then by setting $\mathrm{J}(k) = j(c)$, the considered optimization problem reads:

$$\min_k \mathrm{J}(k) \tag{10}$$

The first order optimality condition of this optimization problem reads: $B^{1/2}\nabla j(c) = 0$. The change of variable based on the covariance matrix B acts as a preconditioner for the optimization problem, see e.g. [23,24] for related analysis.

Recall that in the linear-quadratic case (the model is linear, the functional is quadratic), one can show the equivalence between the VDA solution of (10) (considering (9)) and Bayesian estimations based on B, see e.g. [38]. The VDA algorith is implemented in the DassFlow computational code [39]; it employs the automatic differentiation tool Tapenade [25].

It is necessary to add a regularization ("convexifying") term to the cost function $j(c)$ to define a better conditioned optimization problem, see e.g. [5]. The classical way to do it is to define j as follows: $j(c) = j_{obs}(c) + j_{reg}(c)$ with j_{reg} a Tikhonov regularization term. Here the regularization term reads as:
$j_{reg}(c) = \frac{1}{2}\left(\gamma_b\sum_{r=1}^{R}|\partial_r b_r(c)|^2 + \gamma_{K_0}\sum_{r=1}^{R}|\partial_r K_{0,r}(c)|^2\right)$.
The regularization term weight coefficients $\gamma_\square$ are empirically set (such that at initial iteration the regularization terms $\approx 10\%$ of j_{obs}). Following an adaptive regularization strategy, see e.g. [26], the weight coefficients are divided by 2 every 10 iterations.

Moreover thanks to the formulation (9), a regularization term is also implicitly introduced through the covariance matrix B. Indeed one can show the equivalence between the chosen covariance kernel B (e.g. as the second order auto-regressive kernel like those employed below, see (12)) and a regularization functional. The reader may refer to e.g. [50,40] for detailed examples. The definition of B is detailed in the next subsection.

Remark. Compared to the H2iVDI algorithm presented in [30] (and implemented in [39]), a technical but important improvement have been introduced. The vertical discretization of the river geometry (superimposition of the measured trapeziums, see Section A) is now represented by a smooth curve parametrized by a very low number of points e.g. 5. These points are optimal in the sense they minimize the R^2 (Pierson) criteria. Defining a regularized vertical geometry is important since it is differentiated in the reverse code. Indeed, the adjoint method (implemented here using automatic differentiation) computes the differential of the geometry function. Therefore if this function presents numerous stiff local gradients, this may affect the algorithm convergence robustness. The present regularized geometry provides more robust convergence of the optimizer while it remains physically-consistent.

D.2 The covariance matrix B in the VDA formulation

The choice of B greatly determines the computed solution of the inverse problem; this "prior model B" constitutes an important feature of the VDA formulation. In the present study, these covariances are defined from classical operators but with non constant coefficients therefore defining somehow physically-adaptive regularizations.

D.2.1 Expression of B

Here the three unknown parameters $(Q_{in}(t), b(x), K(x))$ are supposed to be independent variables. This assumption is a-priori incorrect but one don't known a-priori universal covariances between these variables. As a consequence B is defined as a block diagonal matrix:

$$B = blockdiag(B_{Qin}, B_b, B_K) \tag{11}$$

Each block matrix $B_\square$ is defined as a covariance matrix (symmetric positive definite matrix). The matrices B_Q and B_b are set as the classical second order auto-regressive correlation matrices:

$$(B_{Qin})_{i,j} = \sigma_{Qin}(t_i)\sigma_{Qin}(t_j)\,\exp\left(-\frac{|t_j - t_i|}{T_{Qin}}\right) \text{ and } (B_b)_{i,j} = \sigma_b(x_i)\sigma_b(x_j)\,\exp\left(-\frac{|x_j - x_i|}{L_b}\right) \tag{12}$$

The matrix B_K is set as $B_K = blockdiag(B_{K_0}, B_{\beta^K})$ with:

$$(B_{K_0})_{i,j} = \sigma_{K_0}^2\,\exp\left(-\frac{|x_j - x_i|}{L_K}\right) \text{and } B_{\beta^K} = diag(\sigma_{\beta^K}^2(x)) \tag{13}$$

The parameters T_{Qin} and (L_b, L_K) act as correlation lengths.

D.2.2 Setting of the parameters $\sigma_\square$ and $(T_{Qin}, L_\square)$

These parameters are important prior information of the inversions.They are set from the first guesses values $(Q_{in,p}, A_{0,r}, (K_0, \beta^K)_r)^{(0)}$.

Recall that the observation frequency is 24h. The measurements spacing varies from a few dozen meters to a few hundreds of meters. Local Froude numbers range in great majority within $\approx [0.05 - 0.3]$, with some very local maximum values up to ≈ 0.5.

The discharge parameters are set as follows. $T_{Qin} = 24$ h. The normalization coefficient σ_{Qin} is time-dependent: $\sigma_{Qin}(t)$ equals 30% of the mean value of $Q^{(0)}(t)$. (Recall that uncertainty error on discharge measurements may be considered as $\approx 30\%$, see e.g. [22] and references therein).

Concerning the bathymetry, σ_b is space dependent: $\sigma_b(x)$ is set such that it corresponds to 50% of the mean value of $A_0^{(0)}(x)$. Recall that the bathymetry values $b^{(0)}(x)$ are deduced from the unobserved flow area values $A_0^{(0)}(x)$.

Concerning the correlation length, we set: $L_b = 1$ km. However if this last parameter is too large, the matrix B_b may be not positive. In such a case, the characteristic length L_b is adaptively decreased until the matrix becomes positive. This has happened in a few cases, then the minimal resulting value was $L_b = 500$ m.

The normalization coefficients related to the friction are constant: $\sigma_{K_0} = 10$ and $\sigma_{\beta^K} = 0.3$. These values have been chosen following statistical analysis made on the databases and by analyzes on the gradient components. We set $L_{K_0} = dx = 100$m (dx is the computation grid spacing).

As an illustration, some covariance matrices $B_\square$ are plotted in Fig. 15.

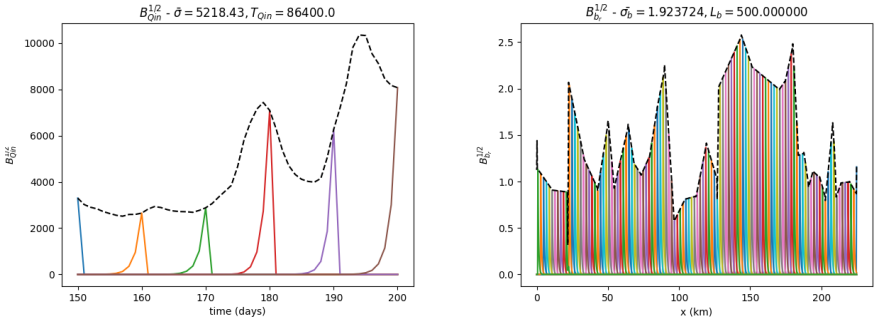

Fig. 15 The covariance matrices $B_\square^{1/2}$ in the Jamuna river case. (L) $B_{Qin}^{1/2}$ (with $T_{Q_{in}} = 24h$). A few covariance values only are plotted for sake of readability. (R) $B_b^{1/2}$ (with $L_b = 500$m). Note that the scaling factor of $B_\square^{-1/2}$ is $\sigma_\square^{-1/2}$ and not $\sigma_\square$.

E The Artificial Neural Network (ANN)

An ANN is built up to estimate the discharge value Q at a given location r and a given instant p, from the WS observations plus the ancillary data $\mathscr{A}$, see Section 3. This ANN is designed as follows.

The training dataset $\mathscr{D}$ contains N_{lp} learning pairs (samples) (I_i, Q_i), $i = 1, \cdots, N_{lp}$. The i-th input is $I_i = (dA, W, S, \mathscr{A})_i$ where i denotes the i-th value at the considered location and day. Recall that dA (m^2) denotes the variations of the wetted crossed-sections above the unobserved A_0; it is straightforwardly computed from the variations of Z (m) and W (m). The slopes values S are extracted from the Pepsi databases (model outputs) at ReachSc. Values of $\mathscr{A}$ (km^2) are extracted from HydroSHEDS database.

Measurements are daily sampled. This frequency corresponds to the important Cal-Val phase of the forthcoming SWOT instrument. However, note that the data could be less frequent (provided eg. every few days) without altering the ANN accuracy; this point is discussed at the end of the section.

The i-th output is the discharge value Q_i at the same location and same instant (day).

The parameters of the neural network are denoted by W_k, $k = 1, \cdots, N_{hl}$; N_{hl} being the number of hidden layers.

Each layer contains N_{nn} neurons. Since neurons are connected to each other, the size of each parameter W_k equals $N_{nn} \times N_{nn}$. The input variables are re-scaled by removing the mean and scaling to unit variance.

Numerous numerical experiments based on numerous different network architectures have been tested. We have observed that fairly deep networks improve the estimation capabilities (ability to find quite correctly nonlinear trends between data); From our experiments, the set $N_{hl} = 64$ and $N_{nn} = 64$ has proven the best precision w.r.t. performance.

Therefore W_1 contains $4 \times N_{nn} = 256$ parameters, each W_j, $j = 2, \cdots, (N_{hl} - 1)$, contains $N_{nn}^2 = 4096$ parameters, while $W_{N_{hl}}$ contains $N_{nn} \times 1 = 64$ parameters.

Training an ANN consists to solve the following optimization problem:

$$W^* = arg \min_W l_Q(W) \tag{14}$$

with the loss function (misfit-cost function) l_Q classically set as

$$l_Q(W) = \frac{1}{N_{ls}} \sum_{i=1}^{N_{ls}} \Big(Q_i(W) - Q^{obs}(I_i) \Big)^2 = \| Q(W) - Q^{obs}(I) \|_{2,N_{ls}}^2 \tag{15}$$

(We may also denote: $Q_i^{obs} = Q^{obs}(I_i)$). The resulting estimator is:

$$Q^{(ANN)} = Q(W^*; I) \tag{16}$$

The activation function of the ANN is the usual rectified linear unit (ReLU) function, see e.g. [21, 32] for details. The ANN have been coded in Python using Keras and Mpi4Py libraries [14]. The minimization of $l_Q(W)$ is performed using the classical Adam method [29], a first-order gradient-based stochastic optimization. The learning rate (the gradient descent step size) is classically adjusted during the optimization procedure. As usual, the

hyper-parameters of the algorithm (learning rate, decay rate, dropout probability) are experimentally chosen; the selected values are those providing the minimal value of l_Q. The reader may refer e.g. to [27] for more details and know-hows on ANN algorithms.

Remark. The drainage area $\mathscr{A}$ (km^2) is not represented in the hydrodynamics models, at least neither (5) nor (3). This additional physical is connected to the un-modeled infiltration fluxes; it somehow constitutes here a closure law. If training the same ANN as described above but based on the SWOT-like observations only, the numerical experiments have shown that the ANN does not provide informative enough results.