On the assimilation of altimetric data in 1D Saint-Venant river flow models

Pierre Brisset, Jerome Monnier, Pierre-André Garambois, Hélène Roux

To cite this version:

Pierre Brisset, Jerome Monnier, Pierre-André Garambois, Hélène Roux. On the assimilation of altimetric data in 1D Saint-Venant river flow models. Advances in Water Resources, 2018, 119, pp.41-59. 10.1016/j.advwatres.2018.06.004 . hal-02044488v3

ON THE ASSIMILATION OF ALTIMETRIC DATA IN 1D SAINT-VENANT RIVER FLOW MODELS

PIERRE BRISSET (1)(2), JÉRÔME MONNIER * (1)(2), PIERRE-ANDRÉ GARAMBOIS (3)(4), HÉLÈNE ROUX (5)(6)

(1) Institut de Mathématiques de Toulouse (IMT), Toulouse, France.
(2) INSA Toulouse, France.
(3) INSA Strasbourg, France.
(4) ICUBE, Strasbourg, France.
(5) INPT Toulouse, France.
(6) Institut de Mécanique des Fluides (IMFT), Toulouse, France.

* Corresponding author: jerome.monnier@insa-toulouse.fr

Abstract. Given altimetry measurements, the identification capability of time varying inflow discharge $Q_{in}(t)$ and the Strickler coefficient K (defined as a power-law in h the water depth) of the 1D river Saint-Venant model is investigated. Various altimetry satellite missions provide water level elevation measurements of wide rivers, in particular the future Surface Water and Ocean Topography (SWOT) mission. An original and synthetic reading of all the available information (data, wave propagation and the Manning-Strickler's law residual) are represented on the so-called *identifiability map.* The latter provides in the space-time plane a comprehensive overview of the inverse problem features. Inferences based on Variational Data Assimilation (VDA) are investigated at the limit of the data-model inversion capability : relatively short river portions, relatively infrequent observations, that is inverse problems presenting a low *identifiability index*. The inflow discharge $Q_{in}(t)$ is infered simultaneously with the varying coefficient $K(h)$. The bed level is either given or infered from a lower complexity model. The experiments and analysis are conducted for different scenarios (SWOT-like or multi-sensors-like). The scenarios differ by the observation frequency and by the identifiability index. Sensitivity analyses with respect to the observation errors and to the first guess values demonstrate the robustness of the VDA inferences. Finally this study aiming at fusing relatively sparse altimetric data and the 1D Saint-Venant river flow model highlights the spatiotemporal resolution lower limit, also the great potential in terms of discharge inference including for a single river reach.

Keywords. River flow, variational data assimilation, altimetry, SWOT, discharge, Saint-Venant, Manning, Strickler.

1. Introduction

While the in situ observation of the continental water cycle, especially river flows, is declining, satellites provide increasingly accurate measurements. The future Surface Water and Ocean Topography (SWOT) mission (CNES-NASA, planned to be launched in 2021) equipped with a swath mapping radar interferometer will provide river surface mapping at a global scale with an unprecedented spatial and temporal resolution - decimetric accuracy on water surface height averaged over $1\,km^2$ [46]. An other highlight of SWOT will be its global coverage and temporal revisits (1 to 4 revisits per 21-days repeat cycle). In complementarity with decades of nadir altimetry on inland waters [7], SWOT should offer the opportunity to increase our knowledge of the spatial and temporal distribution of hydrological fluxes including stream and rivers see e.g. [3, 4]. Thanks to this increased observation of water surfaces worldwide, it will be possible to address a variety of inverse problems in surface hydrology and related fields, see e.g. [43]. Given these surface measurements (elevation, water mask extents), the challenging inverse problems consist to infer the discharge but also the unobservable cross sections, the roughness coefficients and the lateral contributions. These inverse problems are more or less challenging depending on the space-time observations density, the targeted space-time resolution, the potential prior information and the measurements errors.

A relatively recent literature addresses some of these inverse questions including in a pure remote sensing data context potentially sparse both in space and time, see e.g. [4] for a recent review. Few low-complexity methods, based either on steady-state flow models (like the Manning-Strikler's law) or hydraulic geometries (empirical power-laws) have been developed, see [5, 16, 18, 56]. In [15] the performances obtained on 19 rivers with artificially densified daily observables are fluctuating depending on the algorithm tested. In order to better constrain these under-determined inverse problems, prior hydraulic information or empirical laws may be required. It is shown in [18] with a steady model that given one (1) bed level measurement, an effective bathymetry can be infered quite accurately throughout the river reach; see also [21, 22] in a purely academic context. No approach aforementioned does satisfactorily solve the equifinality issue related to the bathymetry and friction. Indeed if infering the triplet formed by the bathymetry, friction and discharge then an equifinality issue is a-priori encountered, see e.g. the discussion led in [18].

In the river hydraulic community, the most employed data assimilation studies are based on sequential algorithms, the Kalman filter and its variants. Let us cite for example [8, 47, 48] who estimate flood hydrographs in the 1D Saint-Venant model from dense water surface width measurements; the bathymetry and roughness are given. [45] considers a diffusive wave model with the bathymetry and friction coefficients given; it corrects the upstream discharge via the assimilation of downstream water depth measurements. The persistence in time of the correction due to the assimilation of synthetic SWOT observations on discharge forecasts of ~ 500 km of the Ohio river is assessed by [1]. [40] shows the benefit of assimilating virtual SWOT observations for optimizing Selingue dam release (lake depth) and river depth in the upper Niger basin. The impact of the hydraulic propagation time (25 days at low flow) compared to synthetic SWOT observation maximum spacing (9 days in this case) on assimilation methods is highlighted through downstream discharge estimates. Most of those twin experiments use temporal observation sampling much greater than the hydrodynamic phenomena time scales, moreover in large river reaches (potentially in network) of several hundreds of km. This ensures multiple measurements of the flow variations. The infered parameters are generally the water depth h *or* a constant Strickler coefficient K but rarely both parameters simultaneously.

Despite the huge improvement of the remotely sensed data (e.g. by satellite altimetry) and the use of data assimilation methods (variational or sequential), the relative sparsity of the acquired data is challenging for river applications. If considering a "small scale" river portion regarding satellite spatio-temporal sampling, typically hundred kilometers long, the hydraulic information propagates faster than the satellites revisit. The model inversions are generally performed at observation times and propagated with a Kalman filter see e.g. [55, 40] and [4] for a review.

The Variational Data Assimilation (VDA) approach based on the optimal control of the dynamics flow model, see [50, 33, 42, 14] and e.g. [6], consists in minimizing a cost function measuring the discrepancy between the model outputs and the observations. This approach aims at optimally combining somehow in the least square sense, the model, the observations and potential prior statistical information This approach is widely used in meteorology and oceanography since it makes possible to "invert" high-dimensional control vectors and models. In some circumstances, it is possible to infer unknown "input parameters" such as the boundary conditions (e.g. inflow discharge), model parameters (e.g. roughness) and/or forcing terms. Among the first VDA studies related to hydraulic models let us cite [44, 11, 49], next [2, 27, 10] which infer the inflow discharge in 2D shallow water river models. Only a few studies tackle the identification by VDA of the complete unknown set that is the inflow discharge, the roughness and the bathymetry. Infering the discharge and hydraulic parameters from water surface measurements is not straightforward and may be even impossible, depending on the flow regime and the adequacy between the observations density and the flow dynamics. The inference of the triplet (inflow discharge, effective bathymetry and friction coefficient) is investigated in [28, 29] from relatively constraining surface Lagrangian observations. Based on a real river dataset (Pearl river in China), the upstream, downstream and few lateral fluxes are identified from water levels measured at in-situ gauging stations in [27]; however the bathymetry and roughness are given. The assimilation of spatially distributed water level observations in a flood plain (a single image acquired by SAR) and a partial in-situ time series (gauging station) are investigated in [31, 30]. In [20, 34] the inference of inflow discharge and lateral fluxes are identified by VDA by superposing a 2D local "zoom model" over the 1D Saint-Venant model. These studies are not conducted in a sparse altimetry measurement context. More recently [19] have investigated discharge identification of the 1D Saint-Venant model by VDA under uncertainties on the bathymetry and the friction coefficient in a purely academic case.

Finally it is worth to mention that the VDA approach provides instructive local analysis sensitivity maps, making possible to better understand the flow and the model, in particular the influence of the bathymetry and local friction coefficient values, see e.g. [37].

The present study investigates the capabilities of accurate, repetitive but relatively sparse altimetry dataset (SWOT like) to infer time varying river discharges. To do so, firstly the inverse problem is simply represented by the so-called *identifiability map*. This map represents all the available information in the (x, t) plane, that is the observations (the observed "space-time windows"), the hydrodynamic waves propagation (1D Saint-Venant model) and the misfit to the "local equilibrium" (more precisely the local misfit with the steady state uniform flow represented by the Manning-Strikler law). This preliminary analysis makes possible to roughly estimate the time-windows which can be quantified by VDA since the inflow discharge values arise from these observed "space-time windows". This original reading of the hydraulic inverse problem is qualitative only but fully instructive. Indeed this makes possible to roughly estimate whether the sought information has been observed or not, in particular in terms of frequency (providing orders of magnitude). Next the inference of the inflow discharge $Q_{in}(t)$ and the Strickler coefficient K, with K depending on the water depth h (that is a power-law depending on the state of the system) is analyses into details. This analysis provides an answer to the temporal variability identifiable given a spatio-temporal distribution of water surface observations.

The numerical results are presented first on a so-called "academic" case with synthetic data, making possible to focus on the computational method (based on the classically called twin experiments) without the specific real data difficulties (difficulties due to potential difference of scales, measurement errors, un-modeled subscale phenomena etc). This case presents a relatively low *identifiability index*, that is a quite high frequency hydrograph variations compared to the observation frequency. A basic guideline to estimate the a-priori minimal identifiable frequency is provided. Next a river portion (74 km long) of the Garonne river (France) [51, 32] is considered with few scenarios of observation frequency: from the SWOT like data (21 days period with 1 to 4 passes at mid-latitudes) to a multiple-sensor scenario (or SWOT Cal-Val orbit, ∼ 1 day period). The bathymetry is either provided or estimated from one in-situ measurement following [21, 18]. The computational code developed for the present inverse analyses is part of the computational software DassFlow [36].

The outline of the article is as follows. In Section 2, the 1D Saint-Venant forward model and the inverse method based on VDA are presented, along with the academic test case and the Garonne river case. In Section 3 the identifiability maps are presented and analyses. Next based on the VDA process, the discharge identification is discussed for various observation samplings. In Section 4, numerical experiments are conducted to infer by VDA the pair $(Q_{in}(t),\, K(h))$; the bed level is either given or estimated from one (1) in-situ value and a low complexity model. Sensitivities of the infered quantities are analyses with respect to the first guess and the observation errors. In Section 5, the Garonne test case is investigated for two scenarios: the real SWOT temporal sampling (∼21 days revisiting period) and a data sampling densified by a factor 100 . A conclusion and perspectives are proposed in Section 6. The two appendices present details of the numerical scheme in the present context of altimetry measurements.

2. FORWARD-INVERSE MODELS AND TEST CASES

In this part, the forward model (1D Saint-Venant equations) and the inverse model, Variational Data Assimilation (VDA), are described. In particular the model geometry (effective river bathymetry), the observation operator and the minimized cost function are detailed.

2.1. **Forward model.** Open channel flows are commonly described with the 1D Saint Venant equations in (S, Q) variables [12, 9]. The model based on the depth-integrated variables is valid under the long-wave assumption (shallow-water). The equations read :

$$(2.1)\qquad \left\{\begin{array}{lcll} \dfrac{\partial S}{\partial t} + \dfrac{\partial Q}{\partial x} & = & 0 & (2.1.1)\\ \dfrac{\partial Q}{\partial t} + \dfrac{\partial}{\partial x}\left(\dfrac{Q^2}{S} + P\right) & = & g\displaystyle\int_0^h (h-z)\frac{\partial \tilde{w}}{\partial x}dz - gS[\frac{\partial z_b}{\partial x} + S_f] & (2.1.2) \end{array}\right.$$

where S is the wet-cross section (m^2), Q is the discharge $(m^3.s^{-1})$, $P = g\int_0^h (h-z)\tilde{w}dz$ is the pressure term as proposed in [54], $\tilde{w}$ is the water surface top width (m), g is the gravity magnitude $(m.s^{-2})$, H is the water surface elevation (m), $H = (z_b + h)$ where z_b is the lowest bed level (m) and h is the water depth (m). S_f denotes the basal friction slope (dimensionless) and $S_f = \frac{|Q|Q}{K^2 S^2 R_h^{4/3}}$ (classical Manning-Strikler parameterization) with K the Strickler coefficient $(m^{1/3}.s^{-1})$ and R_h the hydraulic radius (m). The discharge Q is related to the average cross sectional velocity u $(m.s^{-1})$ by: $Q = uS$. The left-hand side of the momentum equation is written in its conservative form (hyperbolic part of the model) while the right-hand is a source term. This source term can viewed as pulling the model to the basic equilibrium: the gravitational force vs the friction forces. This classical model is considered with a specific bathymetry geometry built from the water surface observables. The discrete cross sections are asymmetrical trapezium layers; each layer is defined by one triplet (H_i,w_i,Y_i) corresponding respectively to the water elevation, the water surface width associated to H_i and a centering parameter. In a SWOT context, each layer corresponds to a satellite pass.

Remark 1. If the Froude number, $Fr = \frac{u}{c}$, tends towards 0 then the 1D St Venant model can be written as a depth-averaged scalar equation: the diffusive wave model, see e.g. [12, 39]. In the case of a wide channel (the hydraulic radius $R_h \approx h$), the advective term of the equation corresponds to the velocity $\frac{5}{3}u$. In the identifiability maps presented in a next section, this wave velocity $\frac{5}{3}u$ is plotted simultaneously with the Saint-Venant model wave velocities, that is $(u - c)$ and $(u + c)$ (gravity waves model).

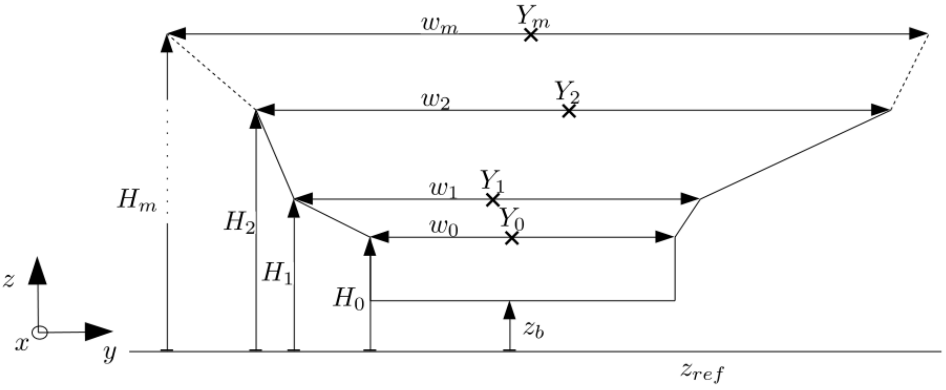

FIGURE 2.1. Effective geometry considered for each cross section: superimposition of m trapeziums (yz-view).

The Strickler coefficient K is defined as a power law in the water depth h:

$$K(h) = \alpha \, h^{\beta} \tag{2.2}$$

where α and β are two constants to be determined. This a-priori law makes possible to set the roughness in function of the flow regime. This power-law is richer than a constant uniform value as it is often set in the literature. Also such a power-law can be defined by sections or reaches.

The discharge at upstream boundary $Q_{in}(t)$ will be considered as an unknown variable of the model (it will be a control parameter of the model). It will be defined by one of these two methods:

IDbasic.: At each identification time t_j, $t_j \in [t_1..t_p]$, a value of $Q_{in}(t_j)$ is computed by the VDA process. Next the identified inflow discharge is continuously constructed by simple linear interpolation.

IDFourier.: The inflow discharge is defined as Fourier series:

$$Q_{in}(t) = \frac{a_0}{2} + \sum_{n=1}^{N_{FS}} \left(a_n \cos(nt\frac{2\pi}{T}) + b_n \sin(nt\frac{2\pi}{T}) \right) \tag{2.3}$$

where $\{a_0; a_n, b_n\}$, $n \in [1..N_{FS}]$, are the Fourier coefficients and T is the total simulation time. The lower frequency represented by the Fourier series is $1/T$ and the highest one is N_{FS}/T. Then this way to identify $Q_{in}(t)$ is global in time (on the contrary to punctual basic approach above). Obviously, the hydrograph must be periodic. However this is not an issue since the hydrograph can be extended to make a T-periodic function (T denoting the final simulation time).

The numerical scheme used is the classical finite volume scheme HLL [25] with Euler integration in time. This numerical scheme with the specificities due to the particular geometrical transformations are presented in Appendix 7.1 and Appendix 7.2. The equations above have been implemented into the computational code DassFlow [36]. Note that few numerical schemes are possible: the classical implicit Preissmann's scheme, the HLL finite volume scheme and also an original semi-implicit multi-regime scheme.

2.2. **Inverse problem: Variational Data Assimilation (VDA) formulation .** The inference of the unknown parameters are performed by the VDA approach. It consists in minimizing a cost function $J(\mathbf{k})$ measuring the discrepancy between the model output (state variables) and the available measurements (which are sparse and uncertain): $\min_{\mathbf{k}} J(\mathbf{k})$. Since J depends on k through the model solution (S, Q), it is an optimal control problem. It is classically solved by introducing the adjoint model and by computing iteratively a "better" control vector $\mathbf{k}$. The latter contains the inflow discharge $Q_{in}(t)$ and the coefficient $K(h)$ defined by (2.2). In the case the unknown parameters are computed at given times $[t_1..t_p]$ (it is the identification time grid), $\mathbf{k}$ is defined by:

$$\mathbf{k} = (Q_{in,1}, \, ..., \, Q_{in,p}, \, \alpha, \, \beta)^T$$

In the case the inflow discharge is decomposed as a Fourier series, see 2.3, $\mathbf{k}$ is defined by:

$$\mathbf{k} = (a_0, \, a_1, \, b_1 ..., \, a_{N_{FS}}, \, b_{N_{FS}}, \, ..., \, \alpha, \, \beta)^T$$

The VDA process requires the computation of the gradient of the cost function ∇J with respect to $\mathbf{k}$. The computation of ∇J is done with DassFlow software which has been originally designed to generate automatically the discrete adjoint model using the source to source differentiation tool Tapenade [26]. The cost function expression J depends on the observations; the latter are presented below while the expression of J is detailed in Section 2.5.

The employed optimization algorithm is a the L-BFGS algorithm (here the M1QN3 routine [23]). Details on the basis of VDA can be found e.g. in [35]. Given a first guess on parameters $\mathbf{k}_0$, the iterates $\mathbf{k}_i$ are searched with the descent algorithm such as the cost function J decreases. For each iteration of the minimization:

1. The cost function $J(\mathbf{k}_i)$ and its gradient $\nabla J(\mathbf{k}_i)$ are computed by performing the forward model (from 0 to T) and its adjoint (from T to 0).
2. Given $\mathbf{k}_i$, $J(\mathbf{k}_i)$ and $\nabla J(\mathbf{k}_i)$, the M1QN3 routine is invoked to compute a new iterate such that: $J(\mathbf{k}_{i+1}) < J(\mathbf{k}_i)$.
3. The few convergence criteria are tested: either $|J| \leq 10^{-7}$, or $|J(\mathbf{k}_{i+1}) - J(\mathbf{k}_i)| \leq 10^{-5}$ or $i > 100$.

In order to measure the accuracy of the identified discharge $Q_{in}^{\text{ident}} = (Q_{in,1}^{\text{ident}}, Q_{in,2}^{\text{ident}}, ..., Q_{in,p}^{\text{ident}})^T$, the classical Nash-Sutcliffe criteria E is considered, [41]:

$$E(Q_{in}^{\text{ident}}) = 1 - \frac{\sum_{i=1}^{p} \left(Q_{in,i}^{\text{real}} - Q_{in,i}^{\text{ident}}\right)^2}{\sum_{i=1}^{p} \left(Q_{in,i}^{\text{real}} - \bar{Q}_{in}^{\text{real}}\right)^2}, \text{ with } \bar{Q}_{in}^{\text{real}} = \sum_{i=1}^{p} \frac{Q_{in,i}^{\text{real}}}{p} \tag{2.4}$$

The vector $Q_{in}^{\text{real}} = (Q_{in,1}^{\text{real}}, Q_{in,2}^{\text{real}}, ..., Q_{in,p}^{\text{real}})^T$ contains the true values. The Nash-Sutcliffe value E is close to 1 for values of Q_{in}^{ident} close to Q_{in}^{real}; it is close to 0 for values of Q_{in}^{ident} close to $\bar{Q}_{in}^{\text{real}}$; finally it is close to $-\infty$ for values of Q_{in}^{ident} not correlated to the true value Q_{in}^{real}.

For a given quantity u (it will be Q_{in}, α or β), $e_2(u)$ denotes the 2-norm relative error:

$$e_2(u) = \frac{\|u^{\text{ident}} - u^{\text{real}}\|_2}{\|u^{\text{real}}\|_2} \tag{2.5}$$

2.3. **Design of the inversion experiments.**

The identifiability of the river flow model parameters from water surface observables is studied on a so-called academic test case before being studied on a real data set (a portion of the Garonne river, France). Analyzing an "academic" case first is important to properly analyse the numerical inversions. Indeed, the academic test case makes possible to focus on the computational method (based on the classically called twin experiments) without the specific real data difficulties (difficulties due to potential difference of scales, measurement errors, un-modeled subscale phenomena etc). Then so-called twin experiments are considered. It consist to set the inverse problem as follows:

- Realistic true values of the parameters (roughness uniform in space and discharge hydrographs) are fixed. Then the forward model is run, which allows to compute the SWOT like data (that is water elevation H and WS width w at the reach scale -see details in next section-).
- Given the perturbed synthetic data, the parameter identifiability is investigated for various temporal samplings of observations. The input "parameters", inflow discharge $Q_{in}(t)$ and coefficient power-law $K(h)$, are computed by VDA. The inflow discharge may be sought in a reduced Fourier basis; the latter being defined from a-priori fixed frequency. In the first numerical experiments, the bathymetry is given. This makes possible to focus the investigation on the identifiability of the inflow discharge in terms of frequency ratio between the observation and the minimal identified frequency. In the last experiment (Garonne river), the considered bed level can be given or estimated from one in-situ value and following the method presented in [18].

2.3.1. *Academic test case.*

The aim of this test case is to investigate the identifiability of several discharge hydrographs and roughness on a fully controlled and low CPU time test case Its geometry consists in a 1000 m length channel. Each cross-section is defined as a superposition of 5 trapeziums. The river bed elevation z_b and water surface width w are not constant; they are defined as follows: $z_b(x) = z_\Delta(x) + z_\delta(x)$, with mean slopes defined by:

$$z_\Delta(x) = \begin{cases} 10 - 0.001x & \text{if } 0 \leq x \leq 300 \\ 9.7 - 0.004(x - 300) & \text{if } 300 < x \leq 700 \\ 8.1 - 0.002(x - 700) & \text{else} \end{cases}$$

and local bed level oscillations as follows: $z_\delta(x) = \sum_{i=0}^{4} c_n \sin(d_n(x - 50)\dfrac{2\pi}{T})$if $50 \leq x \leq 950$ and equal to 0 otherwise.

with $c_n = \{0.01, 0.01, 0.015, 0.02, 0.02\}$ and $d_n = \{1, 2, 4, 8, 16\}$. The triplets $(H_{i,j}, w_{i,j}, Y_{i,j})$ for cross section j as defined in Section 2.1 with i being a vertical index read: $H_{i,j} = H_i' + z_b(x_j)$ with $H_i' = \{1, 2, 3, 4, 5\}$, $Y_{i,j} = \{0, 0, 0, 0, 0\}$ and:

$$w_{i,j} = \begin{cases} w'_{i,j} + \sin\left(\dfrac{\pi(x_j-50)}{900}\right) & \text{if } 50 \le x \le 950 \\ w'_{i,j} & \text{else} \end{cases} \quad \text{with } w'_{i,j} = \{3, 4.9, 5.1, 6.4, 7.3\} \tag{2.6}$$

The coefficient K equals 25 $m^{1/3}.s^{-1}$ ($\alpha = 25$ and $\beta = 0$ in Eq. (2.2)). The considered inflow discharge respecting realistic discharge magnitudes and time scales creates a comparable flooding than those considered in the considered real case (Garonne river). The hydraulic propagation time T_{wave} over the whole river domain equals $\sim$ 160s for a wave velocity $(u + c)$ and the total simulation time is 1000s (cf.Table 1. Recall that T_{wave} is of great interst when using observations of water surface features within a river domain for identifying an inflow discharge (in $x = 0$). The steady-state backwater curve, velocities and local Froude number values ($Fr = \frac{u}{\sqrt{g\frac{A}{W}}}$, with $u = Q/A$ the mean cross sectional velocity) are presented on Fig. 2.2 for $Q_{in} = 10\, m^3.s^{-1}$. The downstream boundary condition is a power law rating curve defined by: $h_{out}(Q) = 0.45\ Q_{out}^{0.6}$ (m).

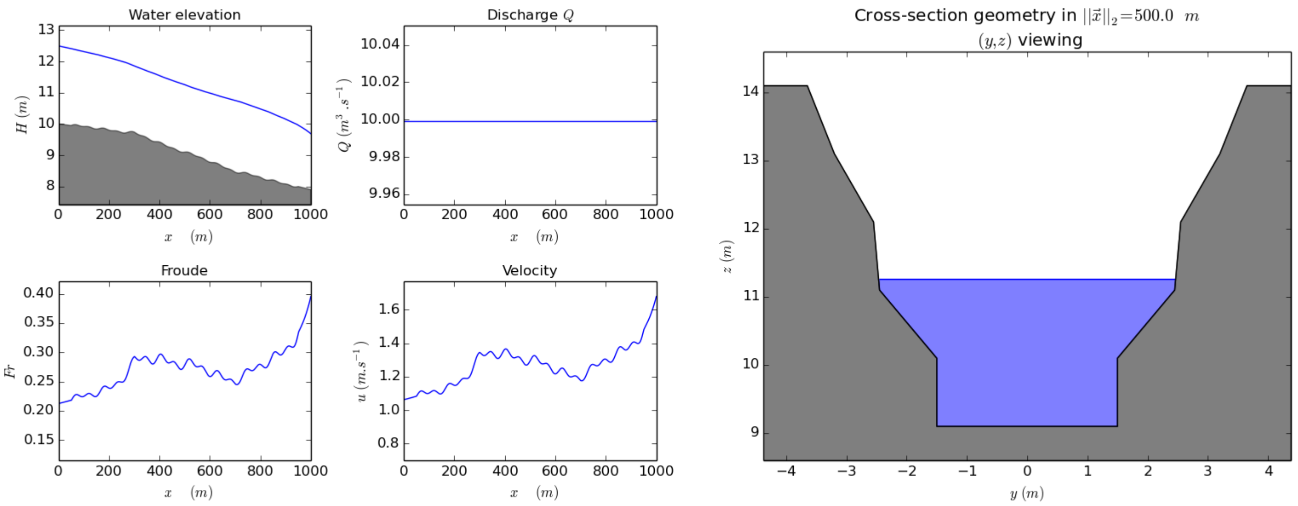

FIGURE 2.2. Academic test case. (Left) Steady state flow for $Q_{in} = 10m^3.s^{-1}$ (quite a low value with respect to the considered hydrograph in the forthcoming experiments): (Left, top) Water elevation H (Right, top) Discharge Q. (Left, Bottom) Froude F and (Right, Bottom) Velocity U vs river curvilinear abscissa. (Right) Cross-section example (for $x = 500\, m$).

	Academic test case		Garonne river	
	mean μ $[m/s]$	standard deviation σ $[m/s]$	mean μ	standard deviation σ
$\min(\lvert u+c\rvert)$	6.3027	0.6063	5.4502	0.6805
$\text{mean}(\lvert u+c\rvert)$	6.3521	0.6224	6.0739	0.6561
$\max(\lvert u+c\rvert)$	6.4028	0.6394	6.6827	0.8052
$\min(\lvert u\rvert)$	1.3018	0.2916	0.7198	0.2507
$\text{mean}(\lvert u\rvert)$	1.376	0.2972	1.1023	0.1968
$\max(\lvert u\rvert)$	1.4361	0.3069	1.391	0.2607
Hydraulic propagation time $T_{wave} = \frac{L}{mean(\lvert u+c\rvert)}$	$\sim$ 160 sec		$\sim$ 3.5 hours	

TABLE 1. Statistics on the wave velocity $(u+c)$, velocity u and the hydraulic propagation time T_{wave} both for the academic and the Garonne test case.

2.3.2. *Garonne river test case.* The 1D Garonne dataset contains a DEM of the river bathymetry between Toulouse and Malause (South West of France, [51, 32, 18]) defined as follows:

- 173 cross sections measurements from the field, distant of 56 to 2200 meters with a median value of 438 m,
- a mesh containing 1158 cross sections; they result of linear interpolations of the original 173 cross sections,
- the cross sections are merged into lidar data of banks and floodplain elevations (5 m horizontal accuracy).

The mean slope of this 74 km portion of the Garonne River is $-0.0866\,\%$ (86.6 cm/km). The reference bathymetry is the effective one respecting the trapezium superimposition structure as described in the academic case and preserving the wetted areas, see figures 2.1 and 2.3. The considered bathymetry can be the reference one or the so-called "low-Froude bathymetry" estimated from one (1) in-situ measurement and the method proposed in [18]. On the present case it is those at the location $x = 40$ km (the reference point indicated in Fig. 2.1).

The effective SWOT like bathymetry (superimposition of trapeziums respecting the true wetted section values) is compared to the Low Froude bathymetry (same trapeziums but not the same z_b) in Fig 2.3. The difference $\frac{1}{N}\sum_{n=1}^{N}|Z_b^{true} - Z_b^{LF}|$ equals 38 cm.

The final mesh size, i.e. the spacing between interpolated cross sections extended on banks, is between $37.26\,m$ and $70.0\,m$ at maximum (the average spacing being $63.96m$). The friction coefficient may be variable, depending on the water depth. Its value is detailed in the identification experiment section.

The considered hydrograph is those measured at Toulouse during a 80 days period in 2010, see e.g. Fig. 5.2. In terms of wave propagation, basic statistics are indicated in Table 1 and the hydraulic propagation time within the whole river portion equals $T_{wave} \sim 3.5$ hours.

All the forthcoming numerical inversions can be performed from either the effective true value of bed level or from the low-Froude one. Indeed the obtained results in terms of infered discharge and roughness coefficient are similar. The assimilation of partial in-situ data in addition of the altimetry measurements is addressed into details in a forthcoming study. Then the error pourcentages on the estimated discharge values in next sections are those obtained from the effective true value.

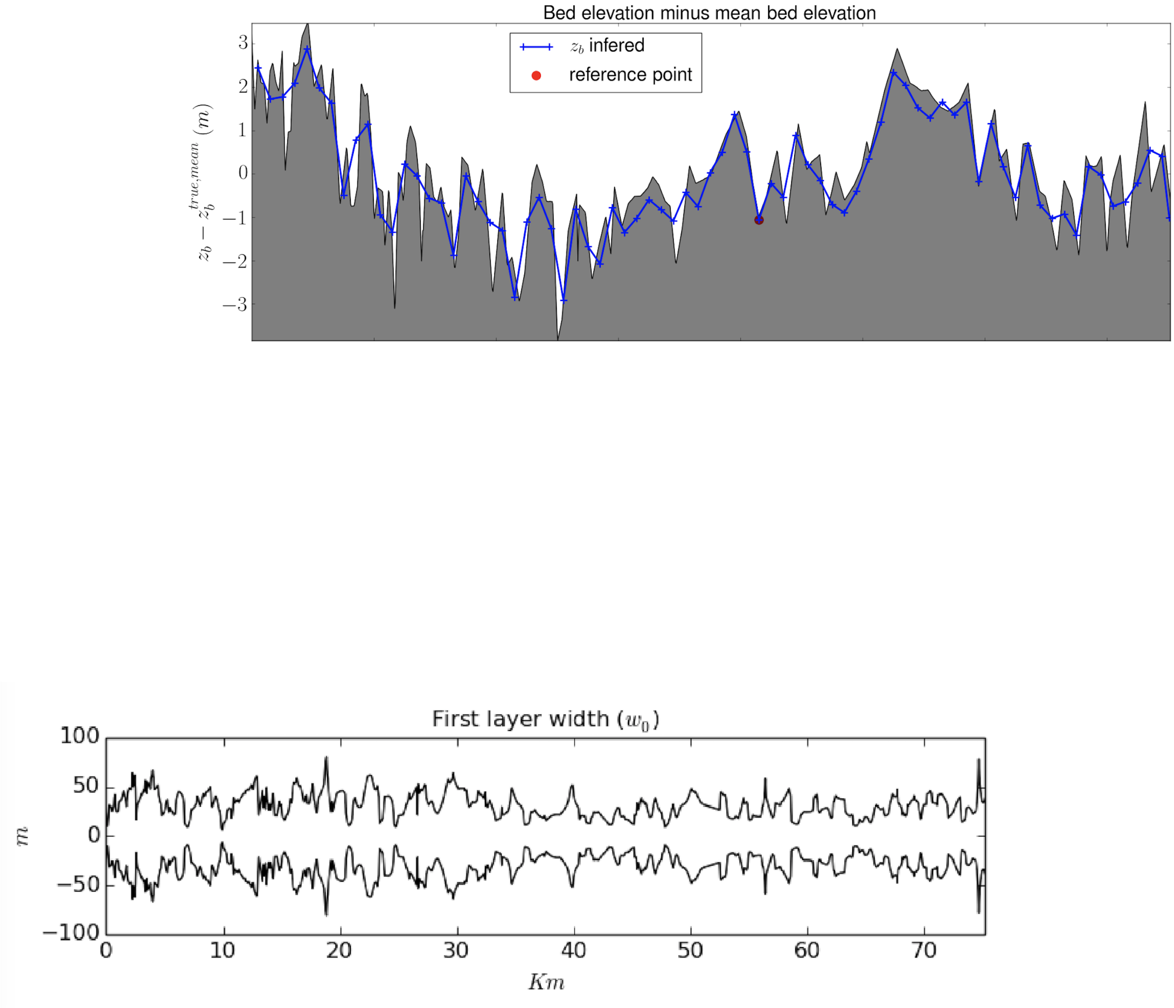

FIGURE 2.3. Garonne data (Top) Effective bed elevation (z_b−mean slope) : the effective true value (gray) and the low-Froude estimated value (blue). (Bottom) First cross sectional layer width w_0 on 151 vertical layers, see Fig.

Remark 2. Concerning the unusual definition of K, an uniform power-law, see Eq. (2.2), it is worth to notice that the forthcoming inversions performed by a VDA approach could have been done with a locally defined power-law $K_r(h)$ with r the "reach" number. However since the main goal of the present study is to focus on the identifiability of the inflow discharge, in particular in terms of frequency flow variations, an uniform power law $K(h)$ has been considered. Moreover as it has been already mentioned, such a power-law gives already more degree of freedom than a mean uniform constant value $\bar{K}$ as it is almost always considered in the literature.

2.4. **The (SWOT-like) altimetric data .** The identifiability capability of the present inverse method depends on the spatial and temporal density of the water surface measurements. Synthetic SWOT observations are generated over the studied domain (Fig. 2.4) from the expected SWOT ground tracks representing three temporal revisits over the domain during a 21 days cycle. Then each swath (50 km wide) defined by the SWOT ascending and descending tracks are split into 1 km stripes. These stripes define the so-called reaches; these splitting lenghts may related with the physical flow features, see e.g. [17]. Only 25 stripes contain the considered Garonne river portions. These 25 observed reaches can be classified in 3 groups observed at different times δT^i, $i = 1..3$ within a 21 days satellite period (see Eqn (2.10) and Fig. 2.4).

1D forward model outputs are averaged in space at each observation time ($\bar{H}_r(t)$ and $\bar{w}_r(t)$) in order to reproduce SWOT like observations at the reach scale; next a random noise is added in order to be representative of SWOT observation errors averaged this reach scale.

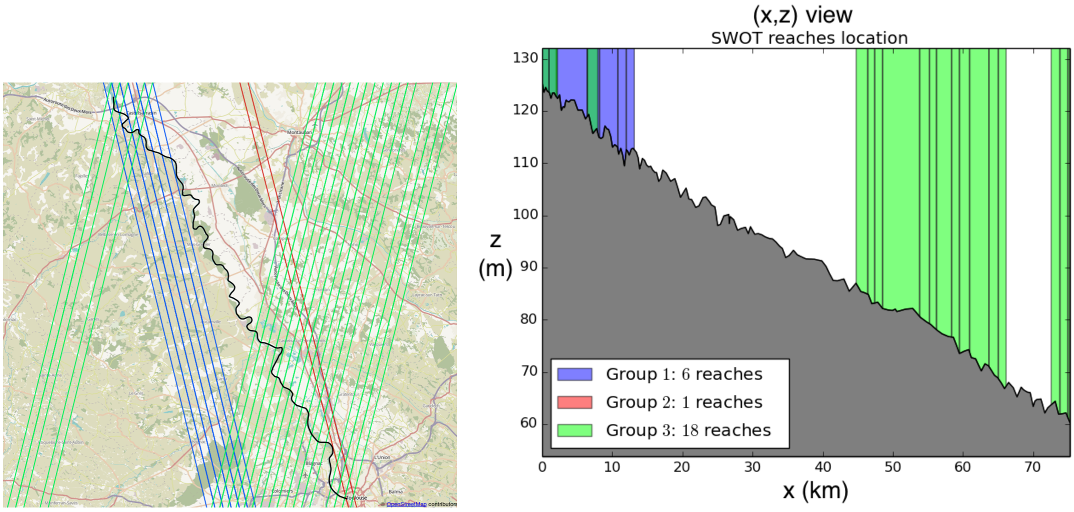

FIGURE 2.4. Location of SWOT reaches on the Garonne river.
(Left) Aerial view from OpenStreetMap of Garonne River (black line) and SWOT reaches location with ascending (green) and descending (blue when the river is seen and red otherwise) tracks. (Right) Longitudinal river profile with the three groups respectively observed at $\delta T^1 = 12.58$ days, $\delta T^2 = 14.11$ days and $\delta T^3 = 1.51$ days

2.5. **Cost function.** The cost function J to be minimized is defined from the available measurements as follows:

$$J(\mathbf{k}) = j^{\text{obs}}(\mathbf{k}) + \gamma\, j^{\text{reg}}(\mathbf{k}) \tag{2.7}$$

where $j^{\text{reg}}(\mathbf{k})$ is a regularization term defined later, and $j^{\text{obs}}(\mathbf{k})$ is defined by:

$$j^{\text{obs}}(\mathbf{k}) = \frac{1}{2}\int_0^T ||\bar{H}^{\mathbf{k}}(t) - H^{\text{obs}}(t)||_W^2 dt \tag{2.8}$$

where $\bar{H}^{\mathbf{k}}(t)$ and $H^{\text{obs}}(t)$ are defined by:

- $\bar{H}^{\mathbf{k}}(t) = \left(\bar{H}_0^{\mathbf{k}}(t),\, \bar{H}_1^{\mathbf{k}}(t),\, \bar{H}_2^{\mathbf{k}}(t),\, ...,\, \bar{H}_{N_r-2}^{\mathbf{k}}(t),\, \bar{H}_{N_r-1}^{\mathbf{k}}(t)\right)^T$
- $H^{\text{obs}}(t) = \left(H_0^{\text{obs}}(t),\, H_1^{\text{obs}}(t),\, H_2^{\text{obs}}(t),\, ...,\, H_{N_r-2}^{\text{obs}}(t),\, H_{N_r-1}^{\text{obs}}(t)\right)^T$

W is a symmetrical positive semi-define matrix $N_r \times N_r$, N_r the number of observed reaches, and it defines an error covariance matrix. Its extra diagonal terms $w_{i,j}$, $i \neq j$, represent the correlation of error observations between reach i and reach j; its diagonal terms $w_{i,i}$ are the a-priori confidence on the observation of reach i. In a real measurement context, reaches close to the satellite nadir would be observed with lower errors. Hence, the diagonal coefficient values should depend on the distance between the reach r and the nadir. Extra-diagonal terms are difficult to estimate and considered to be null here. In all the following, the matrix W is the identity matrix of $\mathbb{R}^{Nr}$ (same confidence on all observations).

The regularization term $j^{\text{reg}}(\mathbf{k})$ is defined by:

$$j^{\text{reg}}(\mathbf{k}) = j_Q^{\text{reg}}(\mathbf{k}) + \gamma j_K^{\text{reg}}(\mathbf{k})$$

where $j_Q^{\text{reg}}(\mathbf{k})$ (respectively $j_K^{\text{reg}}(\mathbf{k})$) is the regularization term on the discharge (respectively the Strickler coefficient).

The balance coefficient γ between $j^{\text{reg}}(\mathbf{k})$ and $j^{\text{obs}}(\mathbf{k})$ can be classically set following the empirical Morozov's discrepancy principle and/or the classical L-curve strategy [38]. It will be observed in the numerical experiments that no prior regularization needs to be considered on the friction term parameterized with a power law with constant coefficients in space. Moreover, in the SWOT context, given the low frequency and high sparsity of the observations, it is difficult (and numerically unnecessary) to define such a regularization term on $Q_{in}(t)$. This regularization may be done by defining $Q_{in}(t)$ in the Fourier basis with frequencies a priori defined from the observation frequency, see the discussion in the next section.

Let $N_{t,r}$ denote the number of SWOT observation of the reach r. Then the discrete form of the cost function J reads:

$$J(\mathbf{k}) = \frac{1}{2} \sum_{r=1,\mathrm{N_r}} \sum_{j=1,\mathrm{N_{t,r}}} \left(\bar{H}^{k}_{r,j} \; - \; \bar{H}^{obs}_{r,j}\right)^2 \tag{2.9}$$

with $\bar{H}^{\mathbf{k}}_{r,j} = \frac{1}{\Omega_r} \sum_{i=1,\mathrm{N_r}} H^{\mathbf{k}}_{i,j} dx$. With Ω_r the curvilinear length of reach r.

Let us remark that in an altimetry context, the i^{th} observation time of reach group g, t^g_i satisfies:

$$t^g_i = i\Delta T + \delta T^g \tag{2.10}$$

where ΔT is the satellite period and δT^g is the time lap of the first observation of the reach group g. Thus if a river is observed by 3 satellite passes during 1 repeat period (like it is the case for the Garonne river, see Fig. 2.4), then there are 3 different δT^g (i.e. $g = 1, 2$ or 3).

All the equations and algorithms previously described have been implemented into the computational code DassFlow [36]. It contains the 1D shallow water model dedicated to the altimetric data (effective cross section geometries) with all required boundary conditions, a Strickler coefficient $K(h)$ depending on the water depth plus a complete VDA process. The adjoint equations are obtained by automatic differentiation [26] and the minimizer is a BFGS algorithm. Note that few numerical schemes are possible: the classical implicit Preissmann's scheme, the classical explicit HLL finite volume scheme and also an original semi-implicit multi-regime scheme.

3. Discharge identification on the academic test case

This section aims at analyzing the inference capability of the 1D river Saint-Venant model from the water surface observables described previously. As a first step, the unknown parameter is the inflow discharge $Q_{in}(t)$ only on the academic channel described previously (Section 2.3.1). From the available observation distribution, (x,t)-identifiability maps are calculated. They provide an overview of the inference capability of the forthcoming VDA process. These maps are analyses in three contexts depending on three scenario of observation sparsity (see Fig. 3.1):

OD1: (Observation Distribution #1), the whole domain is observed (10 reaches),

OD2: the observations are available at upstream and downstream only (2x3 reaches), Fig. 3.1 (middle).

OD3: the observations are available in the middle only (4 reaches), Fig. 3.1(right).

Then the inference of $Q_{in}(t)$ is performed either classically by identifying its values on a fixed identification grid (IDbasic case, with dt_a the constant assimilation time step), or by computing $Q_{in}(t)$ as a Fourier series (IDFourier case). IDFourier case leads to a "global" computation of $Q_{in}(t)$ (on the contrary to the IDbasic). In both cases, an analysis of the influence of the identification time grid is done.

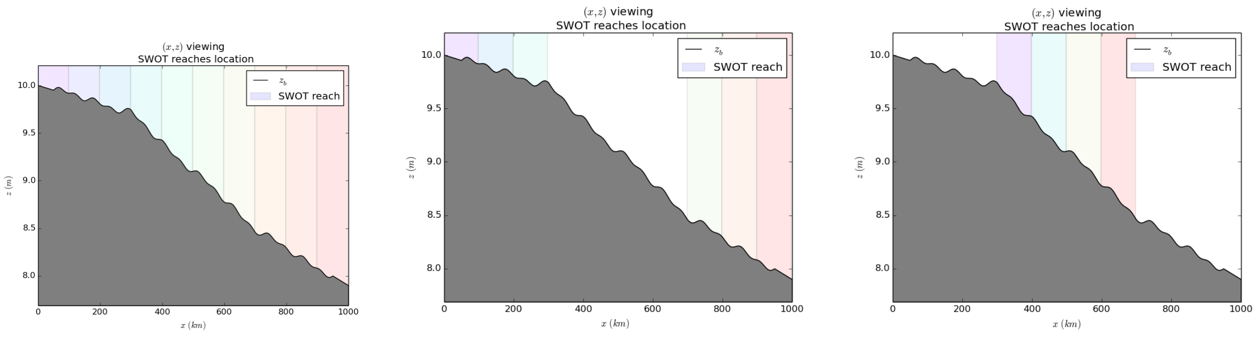

Figure 3.1. Location of the observation reaches. (Left) Case OD1: the whole domain is observed (10 reaches). (Middle) Case OD2: observations are located at upstream and downstream (6 reaches). (Right) Case OD3: observations are located in the middle (4 reaches).

3.1. The identifiability map.

This subsection introduces the identifiability maps. This map represents the complete information in the (x,t) plane: the observations (the observed "space-time windows"), the hydrodynamic waves propagation (1D Saint-Venant model) and the misfit to the "local equilibrium" (more precisely the local misfit with the steady state uniform flow represented by the Manning-Strikler law). This preliminary analysis makes possible to roughly estimate the time-windows which can be quantified by VDA since the inflow discharge values arise from these observed "space-time windows". This original reading of the hydraulic inverse problem is qualitative only but fully instructive. Indeed this will make possible in the next section to roughly estimate whether the sought information has been observed or not, in particular in terms of frequency (orders of magnitude).

The observations are generated from the hydrograph $Q_{in}^{\text{real}}(t)$ shown in Fig. 3.3 Top-Left. Recall that the so-called hydraulic propagation time $T_{wave} \sim 160$ s (estimation based on the mean wave velocity $(u + c)$). From the hydraulic propagation time and the observation time step dt_{obs} (time between satellite overpasses), the identifiability index is defined as follows:

$$I_{ident} = \frac{Twave}{dt_{obs}} \tag{3.1}$$

In the present case, $dt_{obs} = 100\ s$, hence lower than the hydraulic propagation time; the identifiability index I_{ident} ~ 1.6. This means that at least the low frequency variations are observed.

An instructive analysis of the inverse problem consists to plot the so-called identifiability map in the plane (x, t).

Since the inflow discharge (that is $Q(t)$ defined at $x = 0$) is the central sought "parameter", the important wave velocity is the positive one i.e. $(u + c)$ in considering the Saint-Venant system waves. Indeed recall that without the source terms (i.e. gravity waves model), the 1D Saint-Venant model wave velocities are $(u - c)$ and $(u + c)$. Moreoiver if considering the diffusive wave model, that is including the RHS of the Saint-Venant system, the wave velocity equals $\frac{5}{3}u$ in the case of a wide channel (see e.g. [12, 39, 52]and 1.

For each reach r ($N_r = 10$ in the OD1 case) and for each observation time t_i^r (11 in the OD1 case), the velocity waves of the 1D Saint-Venant model (and the diffusive wave model) are plotted, see Fig. 3.2. To do so, $\bar{u}_i^r$ and $\bar{c}_i^r$ corresponding to the reach r at time i are approximated; $\bar{u}$ denotes the mean velocity value and $\bar{c} = (g\bar{h})^{1/2}$ (assuming a rectangular cross section) with $\bar{h}$ the mean water depth. Let us point out that in the present twin experiments, $\bar{u}$ is known. In a realistic context, $\bar{u}$ can be estimated from a low complexity 0.5D model (Manning-Strikler's equation applied at each reach). Such estimations are sufficiently accurate to make the present analysis.

The (r, i) observation time interval is defined as follows: $T_{r,i}^w = [t_i^r - L_r/(\bar{u} + \bar{c})_i^r, t_i^r]$ with L_r the reach length and t_i^r the observation time. Each observation space time window $T_{r,i}^w$ is plotted (in color) in Fig. 3.2. Each rectangle diagonal corresponds to the local $(\bar{u} + \bar{c})$ line; indeed the height of the rectangle $T_{r,i}^w$ corresponds to $(\bar{u}_i^r + \bar{c}_i^r) \times L_r$. It can be noticed that the space-time variation of $(\bar{u} + \bar{c})$ is not significant, see the rectangle height variations and Table 1.

In the present case, the whole domain is observed at $t = 0$ hence the wave velocity $(\bar{u} + \bar{c})$ at $t = 0$ can be estimated accurately.

The identifiability map in (x, t) is plotted for the three cases depending on the observation sparsity: cases OD1, OD2 and OD3, see Fig. 3.2.

The rectangle colors represent the misfit to the steady uniform flow (in norm 1). It is the right-hand side (the source term) in norm 1 of the momentum equation, see (2.1):

$$\text{"Steady uniform flow misfit"} = Norm_1[\ g\int_0^h (h - z)\frac{\partial \tilde{w}}{\partial x}dz - gS(\frac{\partial z_b}{\partial x} + S_f)\ ] \tag{3.2}$$

If this source term vanishes (blue colors in Fig. 3.2), it means that locally in space and time the flow variables satisfy the steady state uniform flow equation (here the Manning-Strikler equation). On the contrary, if the misfit term becomes important (e.g. orange - red colors) then the hyperbolic feature of the model is important.

In terms of energy, this non-conservative source term contains the dissipative friction term S_f; while the left-hand side of the 1D Saint-Venant model is conservative, see (2.1). Therefore Fig. 3.2 provide a rough estimation of the propagation features of the flow model including advection diffusion phenomena.

Typically, the peak time at inflow is represented by the rectangle $(r, i) = (1, 6)$. The corresponding wave velocity $(\bar{u} + \bar{c})$ is faster than those arising from the middle of the domain for example, see rectangle $(6, 6)$.

To illustrate differently the advective-diffusion phenomena corresponding to Fig. 3.2, the discharge throughout the domain is plotted at the three observations times 400s, 500s (peak time at inflow) and 600s in Fig. 3.2 Top Right.

All these information represented in the (x, t) plane constitute the so-called *identifiability map*. Its analysis provides a comprehensive overview of the inversion capability, in particular with respect to the inflow discharge $Q_{in}(t)$.

If a characteristic $(\bar{u} + \bar{c})$ line crosses one or more observed reaches (the colored rectangles on Fig. 3.2), the identifiability of discharge is ensured at the time corresponding approximately to the intersection between the characteristic and the vertical axis. In other words, for a reach observed at time t_i^r and abscissa rL_r, the inflow discharge should be identifiable at time $\sim (t_i^r - rL_r/(\bar{u} + \bar{c}))$. Typically, in the present case, the identifiability maps show that any change on Q_{in} is observed at least few times, and this is true for the three scenarios OD1, OD2 and OD3. In other words, there is no blind time-space window; the same hydraulic information may be observed even few times. Then the forthcoming identification computations based on VDA will be robust and accurate for the complete simulation time range $[0, T]$.

This a-priori analysis is confirmed by the VDA experiments presented in next paragraphs.

Remark 3. The present source term estimation provides an a-posteriori model error if employing the usual Manning-Strikler's law to model the flow.

Remark 4. It can be noticed that since the wave velocity $\frac{5}{3}\bar{u}$ is slower than $(\bar{u}+\bar{c})$, see Fig. 3.2. Following the same analysis, it shows that the inflow discharge identifiability in the diffusive wave model would be higher than in the present 1D Saint-Venant model. However if considering a wave velocity value or another, the present analysis remains qualitatively the same; while the quantitive conclusions would differ slightly. The present identifiability map has been arbitrarily plotted using the Saint-Venant waves velocities (recall, values valid if not considering the RHS). For a comparison between the diffusive wave model and the present Saint-Venant model, see e.g. [39] and references therein.

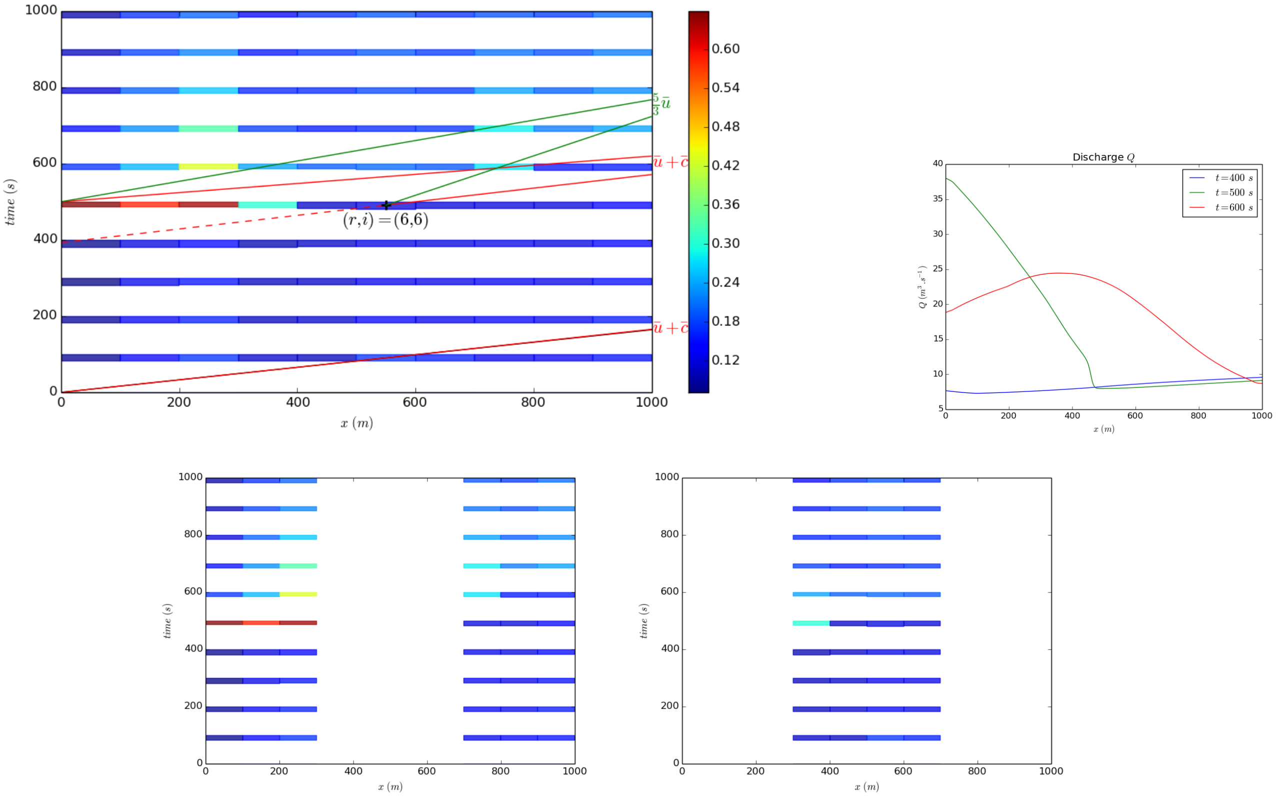

FIGURE 3.2. The identifiability maps in (x,t) in the case: (Top, Left) OD1 (full observations); (Bottom, Left) OD2; (Bottom, Right) OD3.
The estimated wave velocities are plotted in red (continuous line) for the 1D St-Venant model $(\overline{u+c})$, and in green for the diffusive wave model $(\frac{5}{3}\bar{u})$. The red dotted line represents outgoing wave velocity $(\overline{u-c})$ (on the Fig., from the reach r=6 at observation instant i=6). The "steady uniform flow misfit" defined by (3.2) is represented in each rectangle by the colors.
(Top, Right). Discharge $Q(x,\cdot)$ vs x at three observations times: 400s, 500s (=the peak time at inflow), 600s.

3.2. **Identification for various dt_a.** In this section, assuming that both the river bathymetry and the friction law are given, few identifications of inflow discharge are performed with :

- a fixed observation time step $dt_{obs} = 100\,s$ and various assimilation time steps dt_a, ranging from 1/10 to 1 dt_{obs} - IDBasic case in Section 2.1. The parameter vector is $\mathbf{k} = (Q_1, \ldots, Q_p)^T$ with $dt_a = (t_{i+1} - t_i) = \forall i \in [1..p-1]$.
- $Q_{in}(t)$ represented in a reduced basis (Fourier series, see (2.3) - IDFourier case in Section 2.1. The parameter vector reads: $\mathbf{k} = (a_0, a_1, b_1 \ldots, a_{N_{FS}}, b_{N_{FS}})^T$ and the identification with VDA of $N_{FS} = 7$ and $N_{FS} = 25$ Fourrier coefficients is tested.

The inflow discharge and the gradient value are plotted in Fig. 3.3 Top, for IDBasic case with $dt_a = dt_{obs}/10$ and $dt_a = dt_{obs}$. In the case $dt_a = dt_{obs}/10$ the result is excellent, and if $dt_a = dt_{obs}$ the accuracy remains good (excepted at peak time, $t \approx 500$s) - convergence reached in 45 and 17 iterations respectively.The errors on the identified inflow discharge are plotted in Fig. 3.3. Both the 2-norm error and $(1-E)$, with E the Nash–Sutcliffe criteria, are the lowest for dt_a between $20\,s$ and $50s = dt_{obs}/2$ (with $(1-E) \sim 0.0077$). Roughly, the error is improved if $dt_a < dt_{obs}$ but dt_a not too small. Indeed, for $dt_a << dt_{obs}$, typically $dt_a = dt_{obs}/10$, over and under estimations of the discharge appear. Indeed, if dt_a is small, any change of Q_{in} between two identification times, is not observed, Fig. 3.3 Left-top. The value obtained for $dt_a = dt_{obs}/2$ is almost the optimal value. As a practical guide, a simple rule would be to set $dt_a = \frac{dt_{obs}}{2}$ i.e. consider one intermediate point only between two observations.

For IDFourier case roughly the same accuracy and behaviors as in the previous case are obtained, see Fig. 3.3 Bottom. Indeed the minimal error, $(1-E) \sim 0.005$, is obtained with $T/NFS \sim dt_{obs}/2$. Again, as a practical guide, a simple rule would be to set NFS such that $\frac{T}{NFS} = \frac{dt_{obs}}{2}$ i.e. considering one intermediate point (and only one) between two observations, Fig. 3.3.

The advantages of identifying $Q_{in}(t)$ as a Fourier series are the following: the control vector is smaller, the frequency imposed a-priori can be quite easily estimated, and the identified inflow discharge remains smooth (this circumvents the potential oscillations obtained in the case IDbasic with $dt_a << dt_{obs}$ for example).

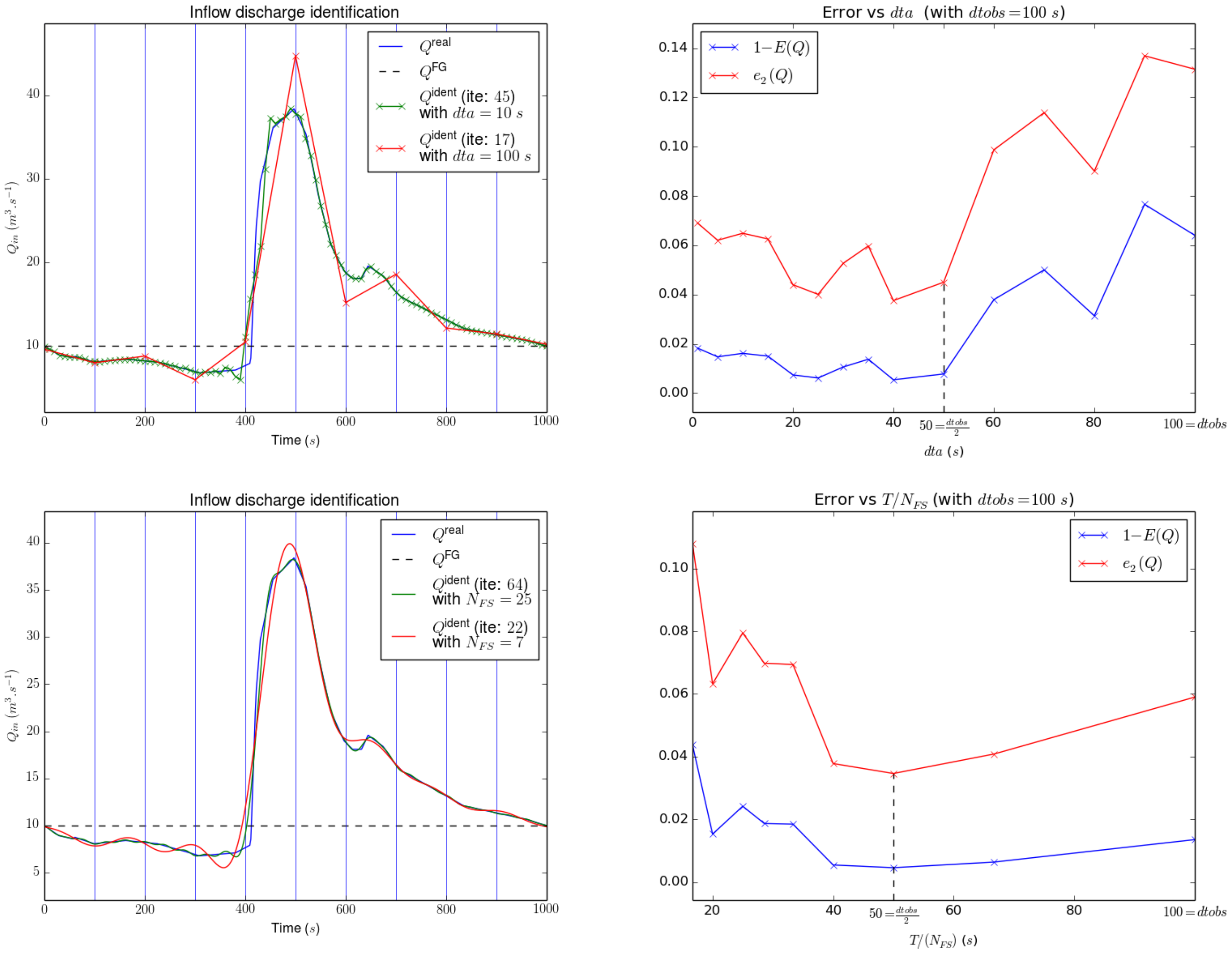

FIGURE 3.3. (Top) Discharge identification: IDbasic approach with $dt_{obs} = 100\ s$. (Left, top) Discharge identification with $dt_a = dt_{obs} = 100\ s$ and $dt_a = dt_{obs}/10 = 10\ s$. (Right, top) Normalized gradient $\nabla_Q J$ with $dt_a = dt_{obs}$ and $dt_a = dt_{obs}/10 = 10\ s$. (Middle) Errors vs dt_a; $dt_a = dt_{obs}/2$ is almost the optimal value. (Bottom) Discharge identification: IDFourier case (Fourier series reconstruction) with $dt_{obs} = 100\ s$: (Bottom Left) Discharge identification with $N_{FS} = 7$ and $N_{FS} = 25$. (Bottom Right) Errors vs T/N_{FS}

3.3. **Identification robustness vs observation sparsity.** A VDA process is global in time. The previous numerical experiments demonstrate that refining too much the identification time grid dt_a compared to dt_{obs} (typically $dt_a = dt_{obs}/10$) deteriorates the identification accuracy. In other words, given an observation time grid, the identification of the time dependent inflow discharge cannot be obtained at much finer time scale. All these previous experiments have been performed with observations available on the whole domain (case OD1, see Section 3). In a real case (e.g. SWOT data of Garonne river test, see Section 2.4) the observations are not available for the whole domain, nor all at the same time. Thus in the present experiments, the robustness and accuracy of the discharge identification is investigated if considering real-like SWOT data hence much sparse observations.

The inflow identification are performed with a pseudo-optimal assimilation time step $dt_a = 25\,s$ (still with $dt_{obs} = 100\,s$) for the three cases OD1, OD2, OD3, see Fig. 3.1.

As discussed in Section 3.1, the identifiability maps (see Fig. 3.2) indicate that in the three cases the identification should be accurate. Indeed, the numerical results obtained by VDA confirm this a-priori analysis since the error is extremely low, typically $E > 0.99$, see Tab. 2. The identified inflows are not plotted since the results are similar to the previous case.

	OD 1	OD 2	OD 3
Nash-Sutcliffe coefficient (E)	0.993	0.994	0.991

Table 2. Academic test case, Nash-Sutcliffe coefficient (E) for $dt_a = 25\,s$ in function of the observations availability: cases OD1, OD2, OD3.

4. Discharge and roughness identification in the academic test case

In the previous section, $Q_{in}(t)$ only was infered. In the present section both the time-dependent inflow discharge and the Strickler coefficient K (time-independent) are infered by the VDA process. Let us recall that K is defined by: $K = \alpha h^{\beta}$. Then the control vector reads: $\mathbf{k} = (Q_{in,1}, Q_{in,2}, ..., Q_{in,p}, \alpha, \beta)^T$ in the IDbasic case and $\mathbf{k} = (a_0, a_1, b_1..., a_{N_{FS}}, b_{N_{FS}}, \alpha, \beta)^T$ in the IDFourier case. In the present experiments the bathymetry is given. The synthetic observations are generated from the same hydrograph (inflow discharge) as previously and an uniform coefficient $K = 25$ i.e. $\alpha = 25$ and $\beta = 0$ in Eqn (2.2). First guesses are respectively chosen equal to $Q_{in}(t) = 100\,m^3.s^{-1}$ for all t, and to $(\alpha, \beta) = (23.5, 0.1)$ (hence considering K depending on h). The observations are available in the whole domain: OD1 scenario.

4.1. **Identifications in the IDbasic and IDFourier cases.** The identified inflow discharge with a basic linear reconstruction (IDbasic case) is as accurate as in the previous case i.e. while identifying Q_{in} only. The identified discharge are plotted in Fig. 4.1 Left top in the case $dt_a = 10\,s$ and $dt_a = 100\,s$.

For $dt_a = 10\,s$, the identification of the roughness parameters α and β is accurate, see Fig. 4.1 Bottom; the minimization algorithm has converged in 64 iterations. For $dt_a = 100 = dt_{obs}$, the minimization algorithm has more difficulties to converge, see Fig. 4.1 Top right. In the IDFourier case, the results are similar.

In both cases (IDbasic and IDFourier), the identified quantities are accurate if the identification time step dt_a is small enough compared to dt_{obs}, or if the Fourier mode number N_{FS} is large enough. In such cases, the identification of $Q_{in}(t)$ is as accurate and robust as in the previous case (when $Q_{in}(t)$ only was identified).

But if $dt_a = dt_{obs}$ or equivalently if N_{FS} is small, then the minimization algorithm has more difficulties to converge, hence the VDA process provides less accurate quantities.

The errors on the roughness coefficients are plotted in Fig. 4.3. Since the error made on the identified discharge are very similar than in the previous case they are not plotted. The value of dt_a (resp. N_{FS}) such that $dt_a = dt_{obs}/2$ (resp. $T/NFS = dt_{obs}/2$) are almost the optimal values. Thus the basic practical rule consisting to set the assimilation frequency equal to the double of the observation frequency is relevant.

4.2. **Sensitivity of identifications to first guesses and observation errors.** The sensitivity of the identified quantities ($Q_{in}(t)$ and (α, β)) with respect to the first guess values $Q_{in,FG}$ and $(\alpha, \beta)_{FG}$ is investigated: OD1 case (complete spatial observations), IDFourier case with $N_{FS} = 20$. For each sensitivity map representing identification errors in the space of first guess values of α and Q_{in} (Fig. 4.4) the parameter β is fixed ($\beta = 0$). They show that the identification of inflow discharge $Q_{in}(t)$ and the roughness coefficients are accurate for a large value range of $Q_{in,FG}$. However the accuracy is important for low values of $Q_{in,FG}$. Thus it is preferable to over estimate the first guess (hence starting from high water levels) than under estimate it. The results are very similar if the fixed parameter is the discharge Q_{in} or the roughness law parameter α, then the corresponding figures are not presented.

Finally the impact of observation errors on the three identified quantities are presented in Fig. 4.5. A Gaussian noise $\mathcal{N}(0, \sigma)$ is added to the water elevation data H^{obs}. In the case $\sigma = 0.1\,m$ (this corresponds to the expected error of the forthcoming SWOT instrument, cf. [46]), the error on the roughness law parameters (α, β) equals approximatively 5% and the Nash–Sutcliffe criteria $E \approx 0.5$. In a bad observational context with $\sigma = 0.5\,m$, the error on the roughness law parameters (α, β) equals approximatively $10 - 25\%$ and the Nash–Sutcliffe criteria $E(Q)$ becomes negative. Therefore,

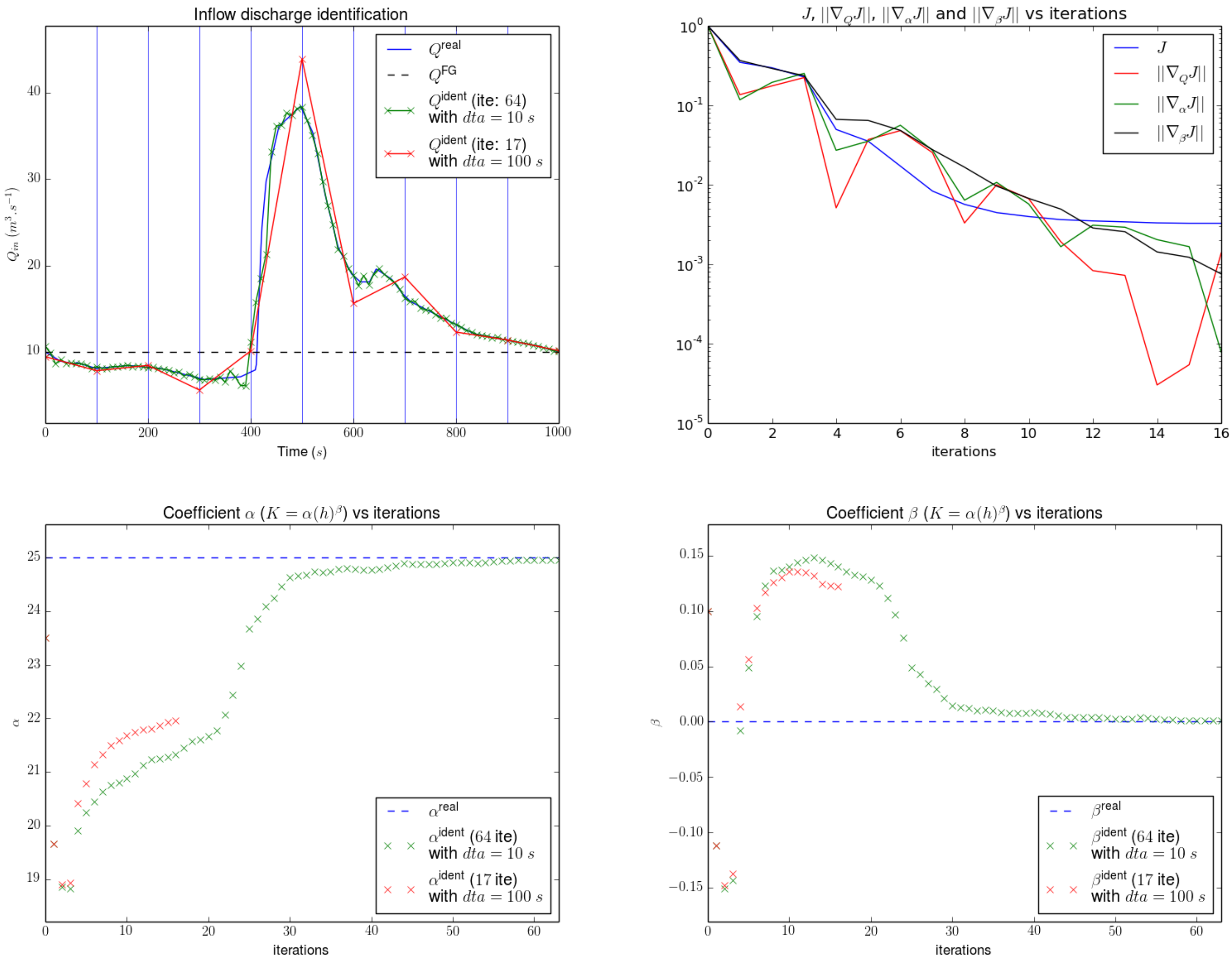

FIGURE 4.1. Discharge and roughness identification in the academic test case (IDbasic case). (Left, top) Discharge identification with $dt_a = dt_{obs} = 100\ s$ and $dt_a = dt_{obs}/10 = 10\ s$. (Right, top) Function cost J,$||\nabla_Q J||$,$||\nabla_\alpha J||$ and $||\nabla_\beta J||$ vs minimization iterations. (Left, bottom) Roughness law coefficient α vs minimization iterations. (Right, bottom) Roughness law coefficient β vs minimization iterations.

the identification of the composite control parameter $(Q_{in}(t); K(h))$ turns out to be quite sensitive to the observation errors but its inference remains accurate in the case of a SWOT-like accuracy $(\sigma = 0.1\,m)$.

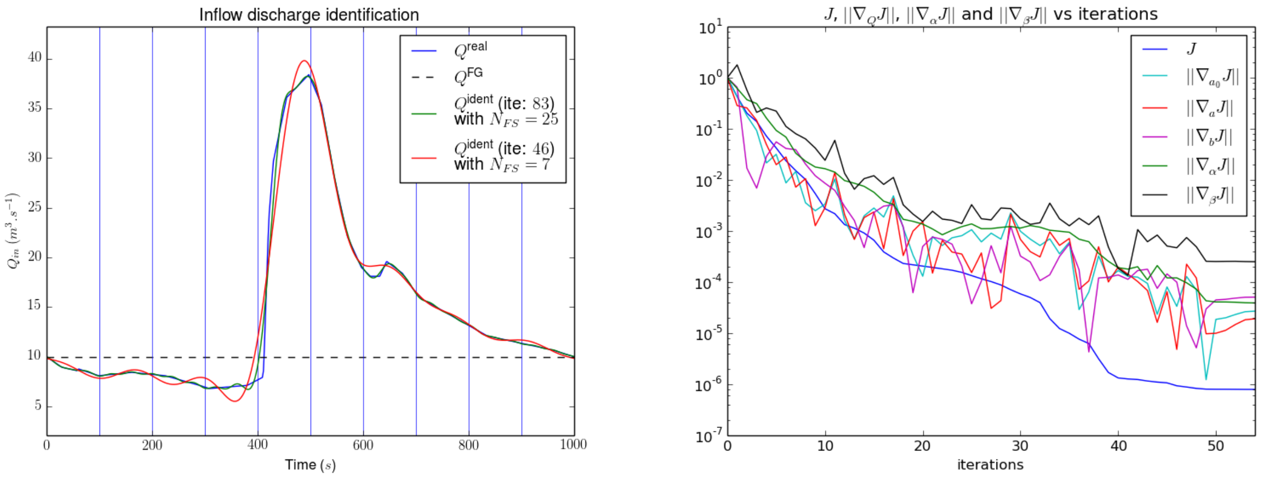

FIGURE 4.2. Discharge and roughness identification in the academic test case (IDFourier case). (Left) Discharge identification with $N_{FS}=7$ and $N_{FS}=25$. (Right) Function cost J,$||\nabla_{a_0}J||$,$||\nabla_{a_n}J||$, $||\nabla_{b_n}J||$,$||\nabla_\alpha J||$ and $||\nabla_\beta J||$ vs minimization iterations.

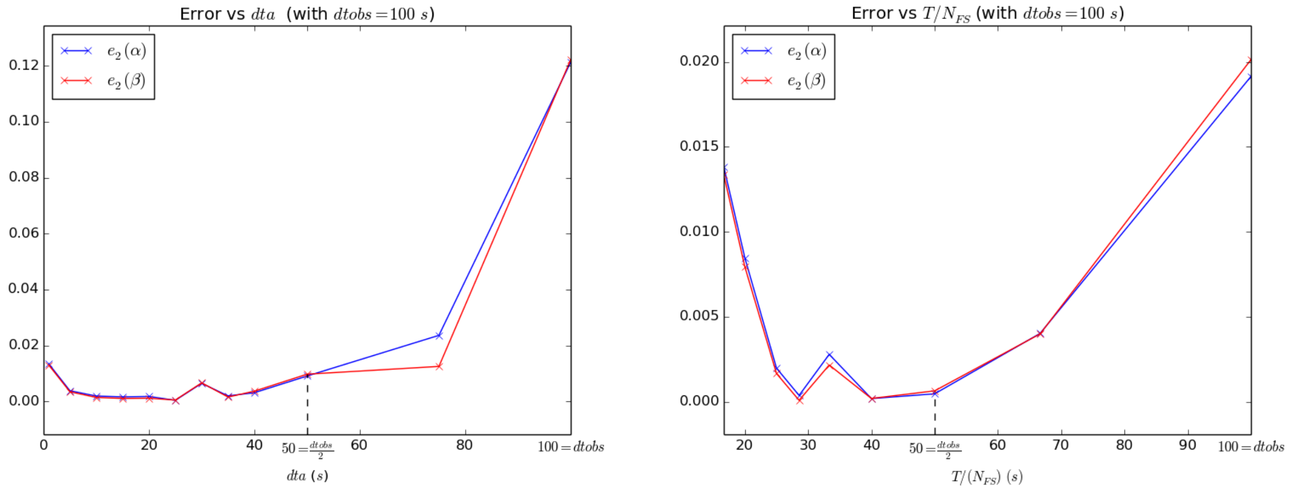

FIGURE 4.3. Roughness identification in the academic test case: errors e_2 on the coefficients (α,β). (Left) IDbasic case: errors vs dt_a. (Right) IDFourier case: errors vs T/N_{FS}.

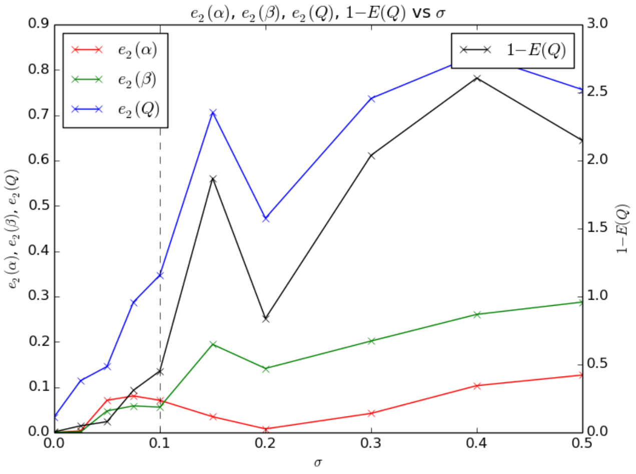

FIGURE 4.5. Error on the identified quantities with $\mathbf{k}=(a_0,\ a_1,\ b_1...,\ a_{N_{FS}},\ b_{N_{FS}},\ \alpha,\ \beta)^T$ vs the observation error σ (standard deviation of the Gaussian noise). The vertical dashed line represents the expected error of the SWOT mission, both in norm 2 and Nash-Sutcliffe criteria.

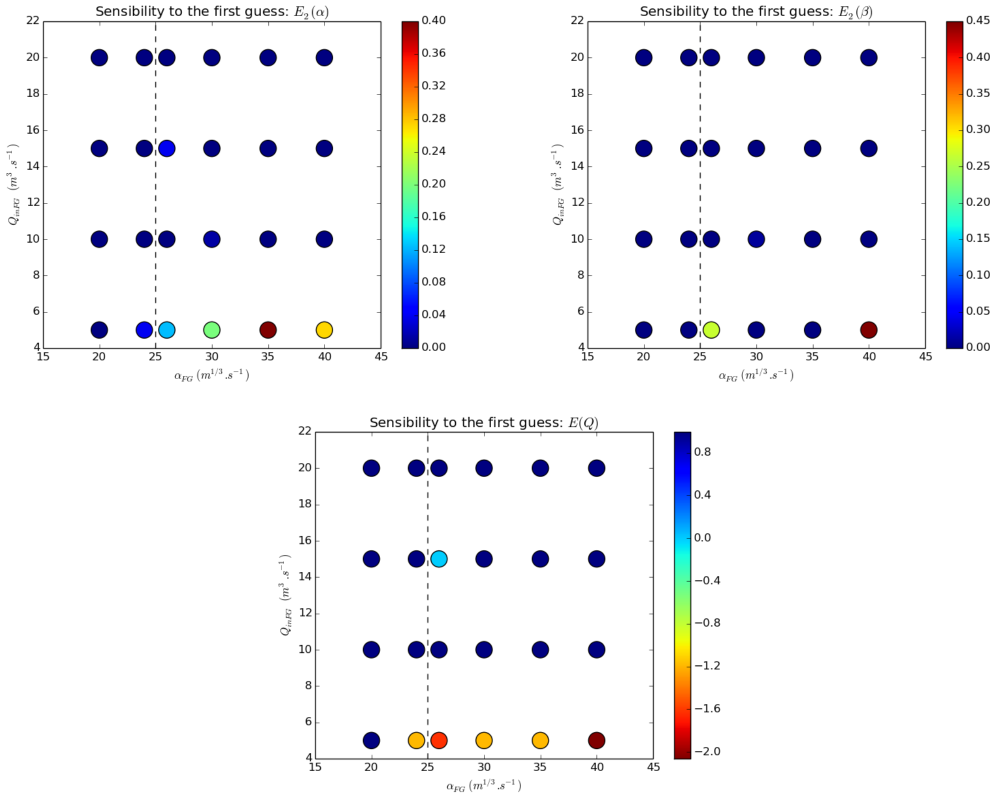

FIGURE 4.4. Sensitivity to the first guess: errors on the identified quantities vs α_{FG} and $Q_{in,FG}$ (β is fixed). (Left, top) Error $e_2(\alpha)$. (Right, top) Error $e_2(\beta)$. (bottom) Error on Q_{in} : Nash–Sutcliffe criteria E.

5. GARONNE RIVER TEST CASE

The accuracy and the robustness of the VDA process, see sections 2.2 and 2.5, is investigated in a realistic data context. The test case is the Garonne river (portion downstream of Toulouse) described in Section 2.3.2. The considered hydrograph is presented on Fig. 5.2. The SWOT-like observations are generated by the model following the method presented in Section 2.4. For the VDA computations the first guess $Q_{in,FG}$ is chosen constant and equal to 268 m^3/s (the mean value of the true hydrograph), see the horizontal dotted lines in the inflow discharge graphs, Fig. 5.2.

As a first step and following Section 3.1, the identifiability maps are computed. Scenario 1 (Section 5.2) consists to consider a SWOT temporal sampling as defined in Section 2.4. The repeat period is 21 days and the simulation time is $T = 80$ days. Scenario 2 is based on a densified SWOT temporal sampling by a factor 100: the repeat period is 0.21 day and the simulation time $T = 0.8$ day. This theoretical scenario would correspond to a combination of observations provided by different satellites. Also during the SWOT CalVal period (the first weeks after the launch), the satellite will be on a lower orbit and will offer a ~ 1 day repeat period on some rivers.

It has been shown in the previous section (academic test case) that the error made on the identified inflow discharge $Q_{in}(t)$ is similar if identifying $Q_{in(t)}$ only or the composite control vector ($Q_{in}(t)$, $K(h)$). Moreover still in terms of error on the identified inflow discharge $Q_{in}(t)$ only, the accuracy obtained from the true effective bathymetry or from the low Froude effective bathymetry are very similar. Obviously the corresponding identified value of $K(h)$ differ between the two cases. This illustrates again the equifinality issue related to the bed properties, that is the pair (bathymetry, friction).

Observe that the VDA process could be performed for the complete unknown parameter ($Q_{in}(t)$, $K(h)$) and $Z_b(x)$ (this has been done and its fine analysis is out of the scope of the present article). However, it may be not the best strategy to calibrate a river dynamic flow model since the equifinality issue on the bed properties (K, Z_b). That is the reason why in the present study we do focus on the inversion with respect to $Q_{in}(t)$ (or equivalently with respect to ($Q_{in}(t)$, $K(h)$)), and we investigate into details the reliability and accuracy of the obtained results. The eqbathymetryuifinality issue is complex; it is the main purpose of an on-going study and likely next article.

5.1. **Identifiability maps.** The identifiability maps are computed from the observations following the method described in Section 3.1 for both scenarios, see Fig. 5.1. On the contrary to the academic test case, no observation is available at $t = 0$ hence the wave velocity $(\bar{u} + \bar{c})$ propagating from $t = 0$ cannot be estimated. Fig. 5.1 Left shows that in the SWOT sampling case, the identifiability of $Q_{in}(t)$ is approximatively limited to the observation “day time”, hence preventing to infer in-between inflow variations (since no constraining information). The lack of constraining observation is accentuated here since a single quite short river portion is considered with its hydraulic propagation time $T_{wave} \sim 3.4$ h only, see Tab. 1, hence an extremely low identifiability index $I_{ident} \sim 6.7\,10^{-3}$.

The next scenario (Scenario 2) is a 100 times greater revisit frequency: $dt_{obs} = 0.21$ day. Keeping the same hydrograph but rescaled in time, the hydraulic propagation time T_{wave} is the same (~ 3.4 h) but the observation frequency equals 0.21 day, hence the identifiability index is 100 times greater: $I_{ident} \sim 0.67$. This rough analysis informs that almost the complete wave set traveling within the river portion should be captured by the sensor.

In the identifiability maps Fig. Fig. 5.1 the inflow discharge identifiability is represented by the vertical dashed lines at $x = 0$: in red the characteristics feet provided from the “far” green observed reaches (hence an identifiability likely less accurate); in black the characteristics feet provided by the “close” blue observed reaches (hence an identifiability likely very accurate).

Recall that this identifiability analysis is based on the wave velocities estimations only, while the dissipation due to the friction source term is not taken into account. However these maps indicate that in Scenario 2 a large proportion of inflow values should be accurately identifiable (see the vertical points at $x = 0$).

The forthcoming VDA experiments confirm this a-priori analysis; the dashed vertical lines (red and black) on Fig. 5.1are taken back on the identified discharge graphs on Fig. 5.3.

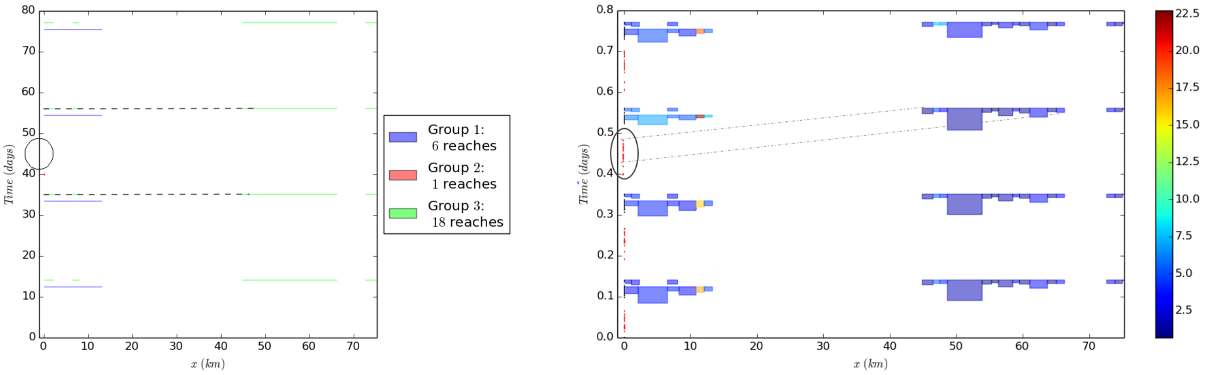

FIGURE 5.1. Identifiability maps in the Garonne river case: (Left) Scenario 1 (SWOT like, 21 days repeat) (Right) Scenario 2 (100 times more frequent, 0.21 day repeat). The circles centered at $t \approx 0.45$ days correspond to the inflow peak. .
In Scenario 1, the identifiability index I_{ident} is so tiny that all the characteristics are almost horizontal and the identifiable times at $x = 0$ corresponds roughly at the “observation day”.
In Scenario 2, the velocity waves $(\bar{u} + \bar{c})$ (dotted lines) are estimated at each reach from the available observations (see sections 3.1 and 5.1). The rectangle heights are proportional to the local value $(\bar{u}+\bar{c})$. The dashed vertical lines at upstream represent the characteristic feet i.e. the sets of points which can be identified in the model without the dissipative source term: in red the information coming from the “far” green observed reaches (hence an identifiability likely less accurate); in black the information coming from the close blue observed reaches (hence an identifiability likely very accurate). These dashed vertical lines (red and black) are taken back from the identified discharge graphs Fig. 5.3.

5.2. **Scenario 1: real SWOT temporal sampling.** The Strickler coefficient K and the bed level z_b are given. The latter is either the effective true bathymetry or the bathymetry estimated by the low-Froude equation presented in [18] and one (1) in-situ measurement. The numerical results presented below are those obtained with the effective bathymetry estimated from the low-Froude equation and the exact lowest wetted area at $x = 40$ km (the so-called reference point in Fig. 2.1). Next the inflow discharge is identified by VDA from the real SWOT space time sampling. Following the preliminary study based on the identifiability map, $Q_{in}(t)$ is decomposed as a Fourier series(IDFourier case) with $N_{FS} = 5$ (Fig. 5.2 Left) and $N_{FS} = 10$ (Fig. 5.2 Right). Then as expected, the identification is accurate in the vicinity of each observation (the vertical colored lines in Fig. 5.2) but inaccurate elsewhere. Indeed, norm 2 of the identified discharge at observation times is $e_2^{Tobs} \sim 16.5\%$ and $e_2 \sim 42\%$ if considering the whole hydrograph (more precisely 41.4% for $NFS = 5$ and 54.5% for $NFS = 10$).As expected increasing the identification frequency (case $N_{FS} = 10$) does not improve the coarser approximation ($N_{FS} = 5$) since the latter already corresponds to an adequate frequency compared to the observation mean frequency, see Fig. 5.2.

As already discussed, the identifiability index is extremely small ($I_{ident} \sim 6.7\,10^{-3}$). This very small index value is due to the important spatiotemporal sparsity of the data and a short river portion (74 km). However the VDA process makes possible to infer quite accurately the inflow discharge roughly at observation day times, but the too small identifiability index prevents to constraint the inflow discharge between the observation times.

All these results corroborate the a-priori analysis made from the identifiability map.

Let us point out that in a complete river network, each observation (at a given location and a given time) is spread into the whole network model (at the various wave velocities) if the hydraulic propagation time is larger than the observation frequency (i.e. with a identifiability greater than 1). Then each satellite overpass can constraint the lowest frequency of the inflow hydrograph in the network.

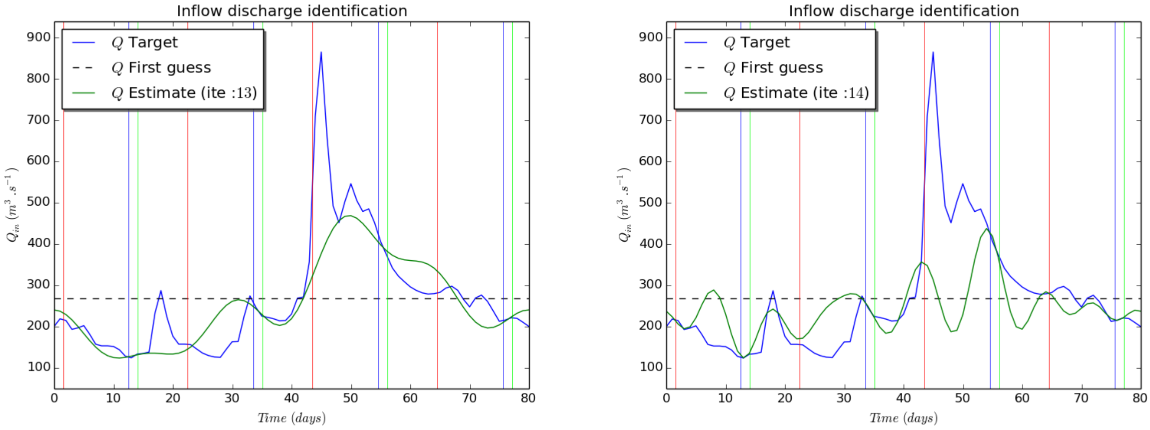

Figure 5.2. Garonne river, Scenario 1. Discharge identification with Fourier series with: (Left) $N_{FS} = 5$, $e_2^{Tobs}(Q_{in}^{estimate}) = 17.1\%$. (Right) $N_{FS} = 10$, $e_2^{Tobs}(Q_{in}^{estimate}) = 16.2\%$. Vertical lines corresponds to the time observations (blue for Group 1, red for Group 2 and green for Group 3). The horizontal dotted line corresponds to the first guess $Q_{in} = 268\, m^3/s$.

5.3. **Scenario 2: densified SWOT temporal sampling by a factor 100.** In the present case, the data sampling and the hydrograph are re-scaled / densified in time by a factor 100. The numerical inversions are strictly the same as the previous ones but the time scale and the values of NFS.

The re-scaled hydrograph remains consistent with the domain length since the peak duration is higher than the response time of the whole river portion; recall $T_{wave} \sim 3.4$ hours. The identifiability index $I_{ident} \sim 0.67$.

As indicated on the identifiability map Fig. 5.1, a majority of the inflow information is observed since it has time enough to travel throughout the domain. This suggests that the inflow values are in majority accurately identifiable but are not during some (a-priori short) time intervals. These more or less accurate time intervals are indicated as the black and red dots in Fig. 5.1 and Fig. 5.3.

The VDA results are presented on Fig. 5.3, read e.g. the case $NFS = 10$, Left-Bottom. The values at the times corresponding to the black identifiability intervals are accurate (as expected). The norm 2 error at observation times equals $\sim 4.5\%$. On the contrary, the peak is partially captured only since it occurs during a red identifiability interval (see the red dots on Fig. 5.1 and Fig. 5.3). However, the identification is globally correct considering the quite low identifiability index I_{ident} value of the scenario. Indeed the index is strictly lower than 1, hence suggesting some "blind" time intervals in terms of identifiability.

As indicated in Fig. 5.3), the VDA process is performed for four values of NFS= $NFS = 5$, 10, 15 and 40. In the numerical method, the value of NFS has to be a-priori set. This can be easily done from the identifiability map analysis and the dt_{obs} value. Indeed it has already been suggested that setting NFS such that: $T/NFS \sim dt_{obs}/2$ (which corresponds here to $NFS \sim 8$) should be quite optimal.

In view to fully analysis the sensitivity with respect to the NFS value, the results obtained from for the four values above are compared. Moreover, to better understand the origin of the identification errors, the approximation of the *exact* inflow discharge $Q_{in}^{target}(t)$ by the same Fourier series is plotted in each case, see the four curves "Exact FS with NFS=..." in Fig. 5.3. This makes possible to analyze the error origin from the Fourier series approximation and from the VDA process (with respect to the present index value).

The best results is obtained with NFS equals to 10 (the case 15 is good too), providing an error at observation times $\sim 4.2\%$, and $\sim 20\%$ if considering the whole hydrograph, see Fig. 5.3. All errors are detailed in the title of Fig. 5.3.

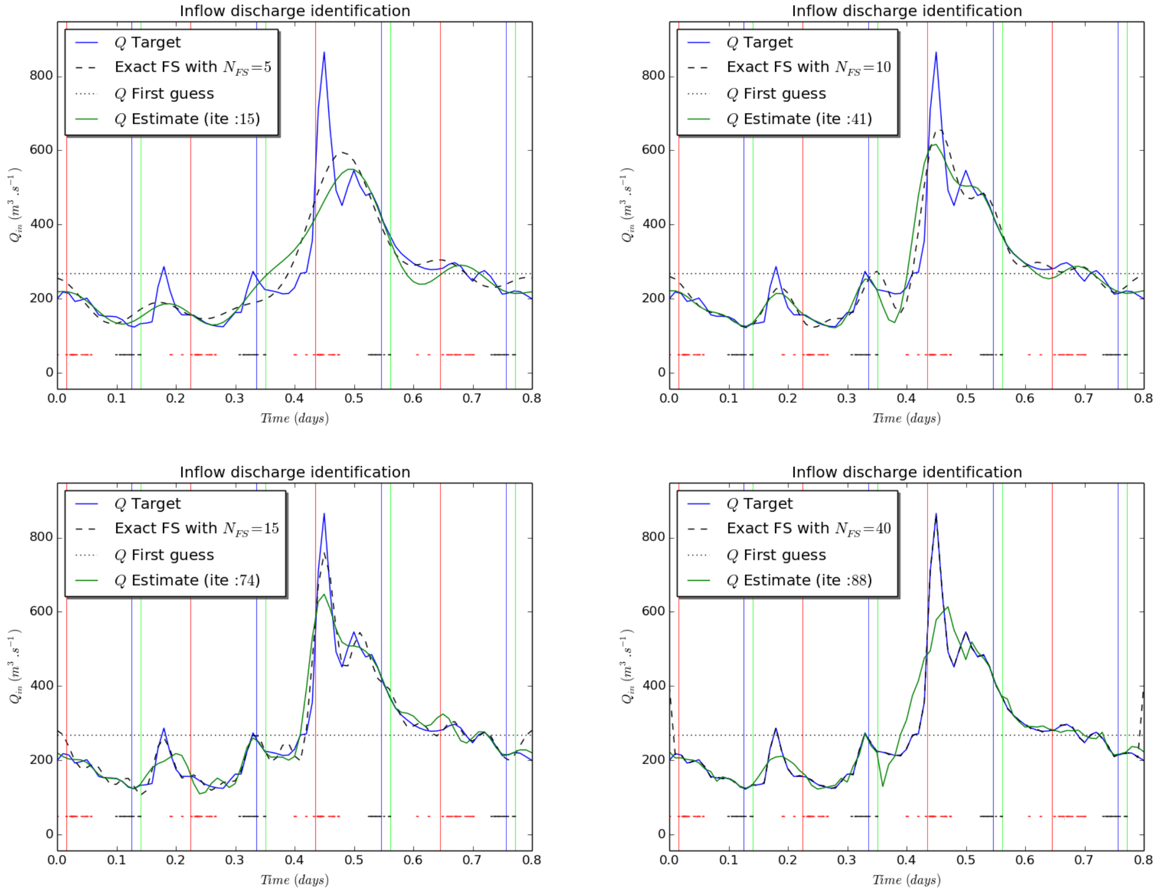

FIGURE 5.3. Garonne river, Scenario 2. Discharge identification with Fourier series with: (Left, top) $N_{FS} = 5$, $e_2(Q_{in}^{estimate}) = 27.6\%$, $e_2^{Tobs}(Q_{in}^{estimate}) = 11.6\%$ (Right, top) $N_{FS} = 10$ $e_2(Q_{in}^{estimate}) = 22.5\%$, $e_2^{Tobs}(Q_{in}^{estimate}) = 4.2\%$ (Left, bottom) $N_{FS} = 15$, $e_2(Q_{in}^{estimate}) = 18.4\%$, $e_2^{Tobs}(Q_{in}^{estimate}) = 4.6\%$ (Right, bottom) $N_{FS} = 40$, $e_2(Q_{in}^{estimate}) = 23.7\%$, $e_2^{Tobs}(Q_{in}^{estimate}) = 8.5\%$. The vertical colored lines correspond to the observation times (blue for Group 1, red for Group 2 and green for Group 3), see Fig. 2.4.

The horizontal colored dashed lines (red and black) at $Q = 50$ correspond to the characteristic feet, lines taken back from the identifiability maps.

The horizontal dotted black line corresponds to the first guess $Q_{in} = 268\ m^3/s$.

6. Conclusion

The identifiability of inflow discharge and roughness coefficients have been investigated into details in the context of SWOT like data (sparse altimetric data of the river surface) and highly frequent revisiting. The bed level was either given or otherwise infered by a lower complexity model. The investigations have been led in for a single (relatively short) river reach, hence in a "high-resolution" context, and at the lowest spatiotemporal limit of the data-model inversion capability. The difficulty of the inverse problem (or equivalently the data-model inversion capability) has been analyzed in terms of the hydraulic propagation time $T_{wave} = \frac{L}{mean(u+c)}$ and the identifiability index $I_{ident} = \frac{Twave}{dt_{obs}}$. Identifiability maps representing the complete information in the (x,t)-plane (the model wave propagation, the observations and the misfit with the Manning-Strikler's law) have been proposed. Their analysis provides a comprehensive overview of the considered inverse problem. Typically in the SWOT data context, the identifiability map of the tested cases suggests that the observations sampling in relation with the characteristic time of the river makes possible to accurately infer the inflow discharge at the "observation day time" but prevents to infer accurately a "continuous" hydrograph, that is inflow discharge values between the observation times.

The numerous numerical VDA experiments (performed both on academic test cases and on a $74\,km$ portion of the Garonne river) have confirmed the preliminary analysis based on the identifiability maps. Moreover it has been shown that in the present case (a single river reach without any additional prior information on the river flow dynamics), the optimal assimilation time step should be set approximatively to the half of the observation time step (one point of identification between two satellite time revisits). From this basic guideline, reducing the control parameter $Q_{in}(t)$ in a Fourier series can be easily done by selecting the lowest identifiable frequency plus a few others. All these numerical results have been analyses for various observation sampling densities hence different identifiability indices. In other respect, sensitivity analyses with respect to the observation errors and with respect to the first guesses values demonstrate the good robustness of the VDA inferences.

It has been demonstrated that infering the roughness values (defined as a power law) simultaneously with the inflow discharge $Q_{in}(t)$ does not affect the accuracy of the identified discharge values. This robustness feature can be partially explained directly from the identifiability map too. Indeed $K(h)$ is a spatially distributed coefficient (the $x-axis$ on the map) while $Q_{in}(t)$ is a point-wise time-dependent coefficient (the $y-axis$ on the map).

Finally the present study completes the previous analyses led on this topic. It investigates the lowest spatiotemporal limit for a given single river reach. It demonstrates the limits of these forthcoming data inversion capability but also their great potential to constraint 1D river flow dynamic models and infer the discharge, including if considering a single relatively short river reach. This study constitutes an important stage before addressing the identifiability and inferences by VDA of multi-satellites, multi-sensors data. Let us point out that if considering a complete river network then the hydraulic propagation time is a-priori larger than if considering a single river reach of the network, then the identifiability index is more important. Indeed in this case each observation (given at one location and one instant) can be spread into the whole network following the wave characteristics. Then if the total hydraulic propagation time is larger than the observation frequency (that is the identifiability index larger than 1) then each satellite overpass can constraint "continuously" the inflow hydrograph of the network.

The VDA process could have been performed for the complete unknown parameter set $(Q_{in}(t), K(h))$ and $Z_b(x)$ (by employing the present computational software DassFlow). However without prior information, the computed optimal solution is not necessarily the correct one since many pairs (K, Z_b) can provide the correct discharge values. In other words the equifinality issue on the bed properties may prevent a correct descriptive model to be predictive. This equifinality issue is a topic of further research, in the present context of sparse altimetric data too. In the future, similar numerical experiments should be performed for a complete river network and for longer time simulations, hence making increase the identifiability index of the considered data-model inversion capability.

Author contributions and acknowledgements. The first author has written the computational software (DassFlow1D) and has performed the numerical results. The second author has elaborated the identifiability map, the frequency analysis, the computational VDA method; he has led the research. The third author has supervised the sensitivity with respect to the first guesses and observation errors; he has greatly contributed to the real data model analysis. The last author has prepared Garonne River data and has helped to set up the numerical hydraulic model of the river portion.

The first author has been funded as a software engineer (2 years contract) by a CNES TOSCA grant related to the forthcoming NASA-CNES SWOT mission. The other authors have been co-funded by CNES TOSCA research project too. The authors would like to thank Dr Sylvain Biancamaria from CNRS - LEGOS Toulouse since he has helped the first author to generate the SWOT tracks, see Fig. 2.4. The corresponding author would like to thank Prof. J-P. Vila from IMT-INSA Toulouse for numerous and rich discussions on hydraulic modeling and numerical schemes during few years. Finally the authors would like to thank four anonymous reviewers for their detailed readings (16 months review process) which made improved the readability of the article.

7. APPENDICES

7.1. **River model geometry .** The resolution of the Saint-Venant equation (1D shallow water) (2.1) requires the computation of wet surface S and perimeter Pe in function of water depth h and geometrical parameters. Then sequences of wet surface $(S_i)_{0\leqslant i\leqslant I}$ and perimeter $(Pe_i)_{0\leqslant i\leqslant I}$ are introduced with $I \in 0,..,N_p$ where N_p is the maximal number of triplets $(H_i, w_i, Y_i)_{0\leq i\leq N_p}$.

For the notations, the reader should refer to Fig. 7.1.

- The wet surfaces $(S_i)_{0\leqslant i\leqslant I}$ are defined by:

$$\begin{cases} S_0 = (H_0 - z_b)w_0 \\ S_i = \dfrac{1}{2}\left(w_{i-1} + w_i\right)\left(H_i - H_{i-1}\right) \quad \forall i \in [\![1, N_p]\!] \end{cases}$$

- The wet perimeters $(Pe_i)_{0\leqslant i\leqslant I}$ are defined by:

$$\begin{cases} Pe_0 = w_0 + 2(H_0 - z_b) \\ Pe_i = \underbrace{\left(\left(\dfrac{W_i}{2} - \left(\dfrac{W_{i-1}}{2} - y_i\right)\right)^2 + (H_i - H_{i-1})^2\right)^{1/2}}_{=Pe_{1i}} + \underbrace{\left(\left(\dfrac{W_i}{2} - \left(\dfrac{W_{i-1}}{2} + y_i\right)\right)^2 + (H_i - H_{i-1})^2\right)^{1/2}}_{=Pe_{2i}} \quad \forall i \in [\![1, N_p]\!] \end{cases}$$

with $y_i = Y_{i-1} - Y_i$, $i \in [\![1, N_p]\!]$.
Let $m \in \mathbb{N}$ such that: $H_m < h < H_{m+1}$; or equivalently, $\sum_{i=1}^{m} S_i < S < \sum_{i=1}^{m+1} S_i$.
Thanks to the sequences $(S_i)_{0 \leqslant i \leqslant I}$ and $(Pe_i)_{0 \leqslant i \leqslant I}$, it is possible to define the following geometric functions:

- Function $Pe(h)$:

$$Pe(h) = \begin{cases} 0 & \text{if } h = 0 \\ 2h + w_0 & \text{if } 0 < h \leqslant H_0 - z_b \\ (2h + w_0) + \displaystyle\sum_{i=1}^{m} Pe_i + Pe'_m & \text{if } h > H_{0-} z_b \end{cases}$$

with:

$$\begin{aligned} Pe'_m = \quad & \left(\left(\tfrac{W_{m+1}}{2} - \left(\tfrac{W_m}{2} - y_{m+1} \right) \right)^2 + (H_{m+1} - H_m)^2 \right)^{1/2} \left(\tfrac{(h+z_b)-H_m}{H_{m+1}-H_m} \right) \quad + \\ & \left(\left(\tfrac{W_{m+1}}{2} - \left(\tfrac{W_m}{2} + y_{m+1} \right) \right)^2 + (H_{m+1} - H_m)^2 \right)^{1/2} \left(\tfrac{(h+z_b)-H_m}{H_{m+1}-H_m} \right) \end{aligned}$$

- Function $S(h)$:

$$S(h) = \begin{cases} 0 & \text{if } h = 0 \\ hw_0 & \text{if } 0 < h \leqslant H_0 - z_b \\ \displaystyle\sum_{k=0}^{m} s_i + s'_m & \text{if } h > H_{0-} z_b \end{cases}$$

with:

$$s'_m = \frac{1}{2} \left(2w_m + \left(\left(Pe^2_{1(m+1)} - (H_{m+1} - H_m)^2 \right)^{\frac{1}{2}} + \left(Pe^2_{2(m+1)} - (H_{m+1} - H_m)^2 \right)^{\frac{1}{2}} \right) \left(\frac{(h + z_b) - H_m}{H_{m+1} - H_m} \right) \right) ((h+z_b) - H_m)$$

- Function $h(S)$:

$$h(S) = \begin{cases} 0 & \text{if } S = 0 \\ \dfrac{S}{w_0} & \text{if } S \leqslant s_0 \\ H_m - z_b + h'_m & \text{if } S > s_0 \end{cases}$$

with:

$$h'_m = - \left(\frac{w_m - X}{w_{m+1} - w_m} \right) (H_{m+1} - H_m) \text{ , where } X = \sqrt{w_m^2 + 2 \left(\frac{w_{m+1} - w_m}{H_{m+1} - H_m} \right) (S - s_m)}$$

If m is such that $w_{m+1} = w_m$, so the relation is simplify by: $h'_m = \frac{(s - s_m)}{w_m}$

- Function $w(h)$:

$$w(h) = \begin{cases} 0 & \text{if } h = 0 \\ w_0 & \text{if } 0 < h \leqslant H_0 - z_b \\ w_m + \alpha_{1(m+1)}((h + zb) - H_m) + \alpha_{2(m+1)}((h + zb) - H_m) & \text{if } h > H_{0-} z_b \end{cases}$$

With α_{1i} and α_{2i} the slope of trapezium i so:

$$\alpha_{1i}, \alpha_{2i} = \frac{\left(\dfrac{w_i}{2} - \left(\dfrac{w_{i-1}}{2} \pm y_i \right) \right)}{H_i - H_{i-1}}$$

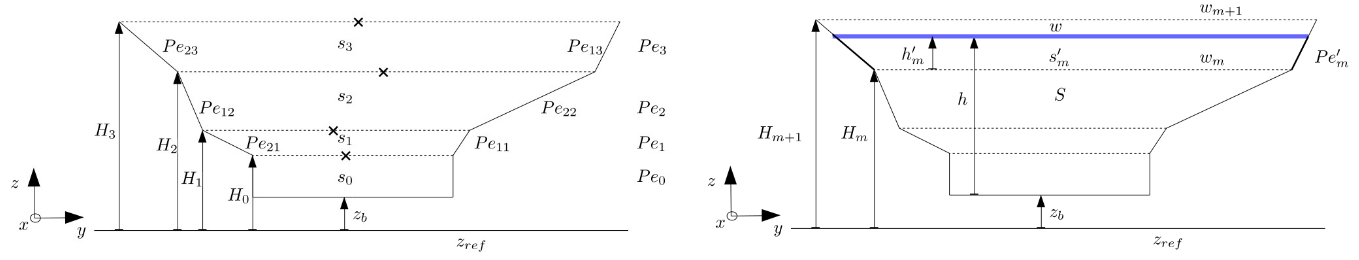

FIGURE 7.1. Notation used for the geometric functions.

7.2. **Finite volume scheme.** The Saint-Venant equation (1D shallow water) (2.1) are computationally solved by the following first order finite volume scheme. The conservative part of the system is written following the form proposed in [54]. The Riemann solver is the classical HLL scheme; the source term is discretized by a classically splitting approach, see e.g. [53]. The resulting numerical scheme is well-balanced in the sense it satisfies the water at rest property (also called C-property in the literature). The computational code has been widely assessed on classical benchmarks (transcritical flow with and without chocks, low Froude flows and of course C-property). The present scheme has been compared with respect to other schemes (the classical Preissmann see e.g. [13] but also an original low-Froude scheme).

7.2.1. *First order scheme.* Eqn (2.1) are rewritten in conservative form as follows:

$$(7.1) \qquad \left\{ \begin{array}{lcll} \dfrac{\partial S}{\partial t} + \dfrac{\partial Q}{\partial x} & = & 0 & (7.1.1) \\ \dfrac{\partial Q}{\partial t} + \dfrac{\partial}{\partial x}\left(\dfrac{Q^2}{S} + P\right) & = & g\displaystyle\int_0^h (h-z)\frac{\partial \tilde{w}}{\partial x} dz - gS\frac{\partial z_b}{\partial x} - gSS_f & (7.1.2) \end{array} \right.$$

where P is a "pressure term" as proposed by [54], next used by [24]. It is defined by:

$$(7.2) \qquad P(x, \tilde{S}, t) = g \int_0^{h(x,t)} (h(x,t) - z)\tilde{w}(x,z,t)dz$$

Then (2.1) is re-written as follows:

$$(7.3) \qquad \frac{\partial \mathbf{U}}{\partial t} + \frac{\partial \mathbf{F}(\mathbf{U})}{\partial x} = \mathbf{S}(\mathbf{U})$$

$$\text{with} \quad \mathbf{U} = \begin{bmatrix} S \\ Q \end{bmatrix}, \ \mathbf{F}(\mathbf{U}) = \begin{bmatrix} Q \\ \dfrac{Q^2}{S} + P \end{bmatrix}, \ \mathbf{S}(\mathbf{U}) = \begin{bmatrix} 0 \\ g\displaystyle\int_0^h (h-z)\frac{\partial \tilde{w}}{\partial x} dz - gS\frac{\partial z_b}{\partial x} - gSS_f \end{bmatrix} \quad \text{and } P = g\int_0^h (h-z)\tilde{w}dz$$

The Jacobian matrix of F reads:

$$\mathbf{J_F} = \begin{bmatrix} 0 & 1 \\ c^2 - u^2 & 2u \end{bmatrix} \text{ since } c = \sqrt{\frac{\partial P}{\partial S}} = gh \text{ and } u = \frac{Q}{S}$$

The eigenvalues of $\mathbf{J_F}$ are: $\lambda_1 = u + c$ and $\lambda_2 = u - c$; their associated eigenvectors are: $r_1 = (1, u + c)^T$ and $r_2 = (1, u - c)^T$.

To solve the homogeneous form of (7.3), the classical scheme based on the Euler time scheme is used:

$$(7.4) \qquad \mathbf{U}_i^{n+1} = \mathbf{U}_i^n - \Delta t^n \frac{\mathbf{F}_{i+1/2}^n - \mathbf{F}_{i-1/2}^n}{\Delta x_i}$$

The numerical flux $\mathbf{F}_j^n$ are computed by the standard HLL formula, such as derived in [25], see also e.g. [53] and references therein.

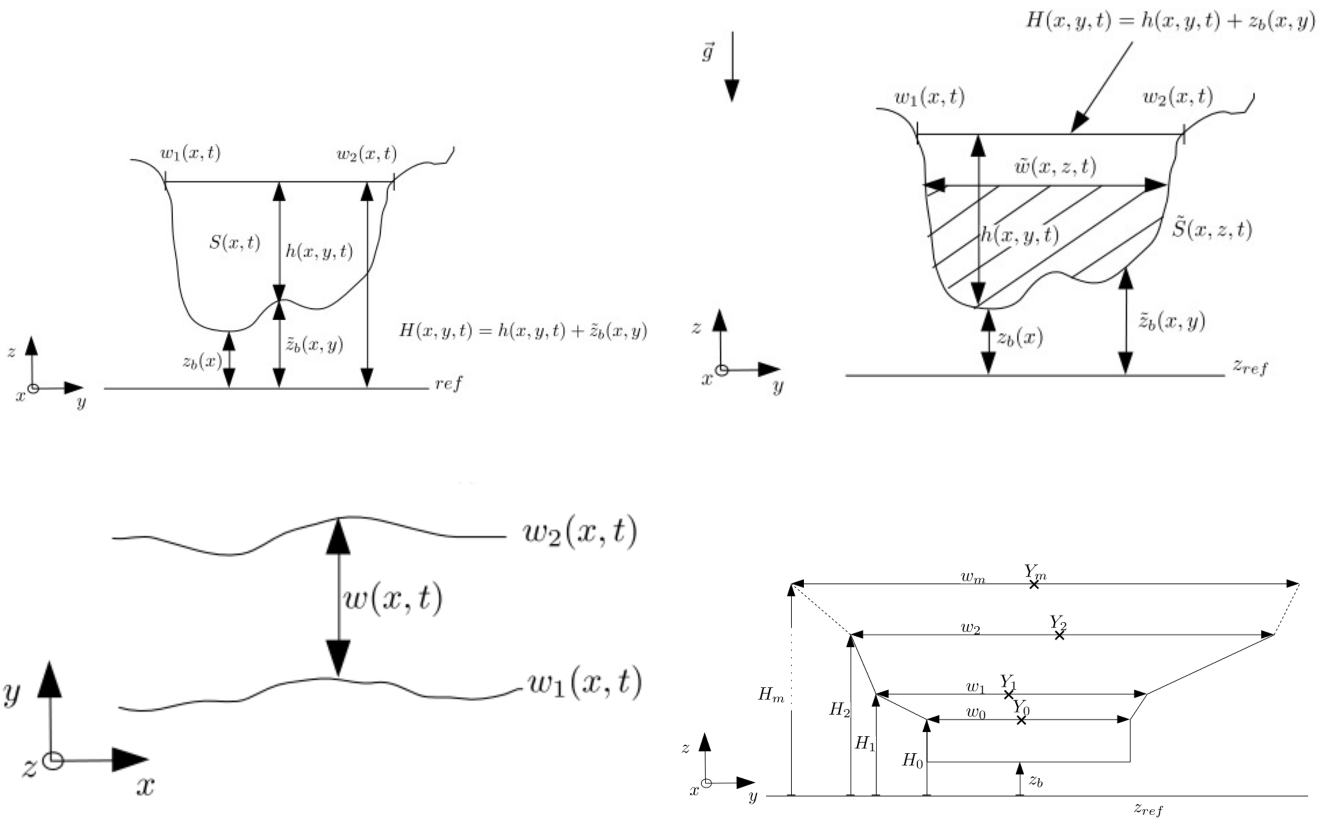

FIGURE 7.2. Notations. (Left, top) Notations for the river cross sections in (yz-view). (Right, top) Variational notations for the river cross sections in (yz-view). (Left, bottom) Notations for the river cross sections in (xy-view). (Right, bottom) Effective geometry considered for each cross section: superimposition of m trapeziums. For the Garonne river case, $m = 150$ (yz-view).

7.2.2. ***Pressure term discretization.*** The pressure term $P = g\int_0^h (h-z)\tilde{w}dz$ has to be correctly discretized to obtain the convergence of the HLL scheme. Thanks to the particular geometry, it is possible to compute the pressure term piecewise. This computational step is CPU time consuming if the number of trapezium is high (recall 150 for the For the Garonne river case).

Let P_i^n be the discrete pressure term with i the cross section number; let j be the trapezium layer number. Let us denotes: $h_i^n \equiv h$, $H_{j,i} \equiv H_j$, $z_{bi} \equiv z_b$, $\alpha_{1j,i} \equiv \alpha_{1i}$, $\alpha_{2j,i} \equiv \alpha_{2i}$, $w_{j,i} \equiv w_j$ and $h_j = (H_j - z_b)$ with $h_{-1} = 0$. Then,

- If $h(x,t) = 0$, $P_i^n = 0$.
- If $0 < h(x,t) \le H_0$,

$$P_i^n = \frac{1}{2} g w_{0,i} (h_i^n)^2$$

- Else,

$$\begin{aligned} P_i^n = \quad & g\sum_{j=0}^{m}\left(w_{j-1}\left(h(h_j-h_{j-1})+\tfrac{h_{j-1}^2-h_j^2}{2}\right)\right) + g w_m\left(h(h-h_m)+\tfrac{h_m^2-h^2}{2}\right)\\ & + g(\alpha_{1j}+\alpha_{2j})\left(\tfrac{h_{j-1}^3-h_j^3}{3}+(h+h_{j-1})\tfrac{h_j^2-h_{j-1}^2}{2}+h h_{j-1}(h_{j-1}-h_j)\right)\\ & + g(\alpha_{1(m+1)}+\alpha_{2(m+1)})\left(\tfrac{h_m^3-h^3}{3}+(h+h_m)\tfrac{h^2-h_m^2}{2}+h h_m(h_m-h)\right) \end{aligned}$$

7.2.3. ***Source term discretization.*** In order to solve the non-homogeneous problem (7.3), a classical splitting method is used, see e.g. [53].

Let us denote $\bar{\mathbf{U}}_i^{n+1} = [\bar{S}_i^{n+1}, \bar{Q}_i^{n+1}]^T$ the solution of the homogeneous problem (7.3) at point x_i and time t^{n+1} ; let us denote $\mathbf{U}_i^{n+1} = [S_i^{n+1}, Q_i^{n+1}]^T$ the solution of the non-homogeneous problem at x_i and t^{n+1}. Then the complete numerical scheme to solve (7.3) reads:

$$\begin{cases} \bar{\mathbf{U}}_i^{n+1} = \mathbf{U}_i^n - \Delta t^n \dfrac{\mathbf{F}_{i+1/2}^n - \mathbf{F}_{i-1/2}^n}{\Delta x_i} \\ \mathbf{U}_i^{n+1} = \bar{\mathbf{U}}_i^{n+1} + \Delta t^n \mathbf{S}(\bar{\mathbf{U}}_i^{n+1}) \end{cases} \tag{7.5}$$

REFERENCES

[1] K. Andreadis and G. Schumann. Estimating the impact of satellite observations on the predictability of large-scale hydraulic models. *Advances in Water Resources*, 73:44–54, 2014.

[2] E. Bélanger and A. Vincent. Data assimilation (4d-var) to forecast flood in shallow-waters with sediment erosion. *Journal of Hydrology*, 300(1-4):114 – 125, 2005.

[3] S. Biancamaria, K. M. Andreadis, M. Durand, E. A. Clark, E. Rodriguez, N. M. Mognard, D. E. Alsdorf, D. P. Lettenmaier, and Y. Oudin. Preliminary characterization of SWOT hydrology error budget and global capabilities. *Selected Topics in Applied Earth Observations and Remote Sensing, IEEE Journal of*, 3:6–19, 2010. 1.

[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, Mar 2016.

[5] D. M. Bjerklie. Estimating the bankfull velocity and discharge for rivers using remotely sensed river morphology information. *Journal of hydrology*, 341(3):144–155, 2007.

[6] F Bouttier and P Courtier. Data assimilation concepts and methods march 1999. *Meteorological training course lecture series. ECMWF*, page 59, 2002.

[7] S. Calmant, J.-F. Crétaux, and F. Rémy. 4 - principles of radar satellite altimetry for application on inland waters. In Nicolas Baghdadi and Mehrez Zribi, editors, *Microwave Remote Sensing of Land Surface*, pages 175 – 218. Elsevier, 2016.

[8] R. Canizares, H. Madsen, H.R. Jensen, and H.J. Vested. Developments in operational shelf sea modelling in danish waters. *Estuarine, Coastal and Shelf Science*, 53(4):595 – 605, 2001.

[9] M. Carlier. *Hydraulique générale et appliquée*. Eyrolles, Paris, France, 1982.

[10] W. Castaings, D. Dartus, M. Honnorat, F.-X. Le Dimet, Y. Loukili, and J. Monnier. Automatic differentiation: a tool for variational data assimilation and adjoint sensitivity analysis for flood modeling. In *Automatic Differentiation: Applications, Theory, and Implementations*, pages 249–262. Springer, 2006.

[11] D. Chertok and R. Lardner. Variational data assimilation for a nonlinear hydraulic model. *Applied mathematical modelling*, 20(9):675–682, 1996.

[12] V.T. Chow. Handbook of applied hydrology. *McGraw-Hill Book Co., New-York, 1467 pages*, 1964.

[13] J. A. Cunge, M. Holly, F., and A. Verwey. *Practical Aspects of Computational River Hydraulics*. Pitam Publishing, 1980.

[14] Cacuci DG., Navon IM., and Ionescu-Bugor M. *Computational Methods for Data Evaluation and Assimilation*. Taylor and Francis CRC Press: Boca Raton, 2013.

[15] 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.

[16] M. Durand, J. Neal, E. Rodríguez, K. Andreadis, L. Smith, and Y. Yoon. Estimating reach-averaged discharge for the river severn from measurements of river water surface elevation and slope. *Journal of Hydrology*, 511:92–104, 2014.

[17] R. Frasson, R. Wei, M. Durand, J.-T. Minear, A. Domeneghetti, G. Schumann, B.-A. Williams, E. Rodriguez, C. Picamilh, C. Lion, T. Pavelsky, and P.-A. Garambois. Automated river reach definition strategies: Applications for the surface water and ocean topography mission. *Water Resources Research*, 53(10):8164–8186, 2017.

[18] 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.

[19] I. Gejadze and P.-O. Malaterre. Discharge estimation under uncertainty using variational methods with application to the full saint-venant hydraulic network model. *International Journal for Numerical Methods in Fluids*, 83(5):405–430, 2017. fld.4273.

[20] I. Gejadze and J. Monnier. On a 2d zoom for the 1d shallow water model: Coupling and data assimilation. *Computer methods in applied mechanics and engineering*, 196(45-48):4628–4643, 2007.

[21] A. F. Gessese, M. Sellier, E. Van Houten, and G. Smart. Reconstruction of river bed topography from free surface data using a direct numerical approach in one-dimensional shallow water flow. *Inverse Problems*, 27(2):025001, 2011.

[22] A. F. Gessese, K. Wa, and M. Sellier. Bathymetry reconstruction based on the zero-inertia shallow water approximation. *Theoretical and Computational Fluid Dynamics*, 27(5):721–732, 2013.

[23] J. C. Gilbert and C. Lemaréchal. Some numerical experiments with variable-storage quasi-newton algorithms. *Mathematical programming*, 45(1-3):407–435, 1989.

[24] N. Goutal and F. Maurel. A finite volume solver for 1D shallow-water equations applied to an actual river. *International Journal for Numerical Methods in Fluids*, 38(1):1–19, 2002.

[25] A. Harten, PD. Lax, and B. Van Leer. On upstream differencing and godunov-type scheme for hyperbolic conservation laws. *SIAM Review*, 25(1):35–61, 1983.

[26] L. Hascoët and V. Pascual. The Tapenade Automatic Differentiation tool: Principles, Model, and Specification. *ACM Transactions On Mathematical Software*, 39(3), 2013.

[27] M. Honnorat, X. Lai, F-X. le Dimet, and J. Monnier. Variational data assimilation for 2D fluvial hydraulics simulation. *CMWR XVI-Computational Methods for Water Ressources. Copenhagen, june 2006.*, 2006.

[28] M. Honnorat, J. Monnier, and F.-X. Le Dimet. Lagrangian data assimilation for river hydraulics simulations. *Computing and Visualization in Science*, 12(5):235–246, 2009.

[29] M. Honnorat, J. Monnier, N. Rivière, E. Huot, and F.-X. Le Dimet. Identification of equivalent topography in an open channel flow using lagrangian data assimilation. *Computing and visualization in science*, 13(3):111–119, 2010.

[30] R. Hostache, X. Lai, J Monnier, and C. Puech. Assimilation of spatially distributed water levels into a shallow-water flood model. Part II: Use of a remote sensing image of Mosel River. *Journal of Hydrology*, 390:257–268, 2010. 3-4.

[31] X. Lai and J. Monnier. Assimilation of spatially distributed water levels into a shallow-water flood model. Part I: mathematical method and test case. *Journal of Hydrology*, 377:1–11, 2009. 1-2.

[32] K. Larnier. *Modélisation thermohydraulique d'un tronçon de Garonne en lien avec l'habitat piscicole : Approches statistique et déterministe*. PhD thesis, Toulouse, 2010.

[33] François-Xavier Le Dimet and Olivier Talagrand. Variational algorithms for analysis and assimilation of meteorological observations: theoretical aspects. *Tellus A: Dynamic Meteorology and Oceanography*, 38(2):97–110, 1986.

[34] J. Marin and J. Monnier. Superposition of local zoom models and simultaneous calibration for 1d–2d shallow water flows. *Mathematics and Computers in Simulation*, 80(3):547–560, 2009.

[35] J. Monnier. Variational data assimilation: from optimal control to large scale data assimilation. Open online course, University of Toulouse - INSA. https://www.math.univ-toulouse.fr/ jmonnie.

[36] J. Monnier, P. Brisset, K. Larnier, and P-A. Garambois. Dassflow: Data assimilation for free surface flows. Open source computational software, University of Toulouse - INSA - CNES - CNRS. http://www.math.univ-toulouse.fr/DassFlow.

[37] J. Monnier, F. Couderc, D. Dartus, K. Larnier, R. Madec, and J-P. Vila. Inverse algorithms for 2D shallow water equations in presence of wet dry fronts. application to flood plain dynamics. *Advances in Water Resources*, 97:11–24, 2016.

[38] V.A. Morozov. The error principle in the solution of operational equations by the regularization method. *USSR Computational Mathematics and Mathematical Physics*, 8(2):63–87, 1968.

[39] R. Moussa and C. Bocquillon. Approximation zones of the Saint-Venant equations for flood routing with overbank flow. *Hydrology and Earth System Sciences*, 4:251–261, 2000.

[40] S. Munier, A. Polebistki, C. Brown, G. Belaud, and D. P. Lettenmaier. Swot data assimilation for operational reservoir management on the upper niger river basin. *Water Resources Research*, 51(1):554–575, 2015.

[41] J.E. Nash and J.V. Sutcliffe. River flow forecasting through conceptual models. Part I - A discussion of principles. *J. of Hydrology*, 10:282–290, 1970.

[42] I.M. Navon. Practical and theoretical aspects of adjoint parameter estimation and identifiability in meteorology and oceanography. *Dynamics of Atmospheres and Oceans*, 27(1):55 – 79, 1998.

[43] C. Ottlé and J.-M. Mahfouf. 8 - data assimilation of satellite observations. In Nicolas Baghdadi and Mehrez Zribi, editors, *Microwave Remote Sensing of Land Surface*, pages 357 – 382. Elsevier, 2016.

[44] V.G. Panchang and J.J. O'Brien. On the determination of hydraulic model parameters using the adjoint state formulation. *Modeling marine system*, 1:5–18, 1989.

[45] S. Ricci, A. Piacentini, O. Thual, E. Le Pape, and G. Jonville. Correction of upstream flow and hydraulic state with data assimilation in the context of flood forecasting. *Hydrol. Earth Syst. Sci.*, 15:3555–3575, 2011. 11.

[46] E. Rodríguez. SWOT Science requirements document. JPL document, JPL, 2012.

[47] H. Roux and D. Dartus. Parameter identification using optimization techniques in open-channel inverse problems. *J. of Hyd. Res.*, 43:311–320, 2005. 3.

[48] H. Roux and D. Dartus. Use of parameter optimization to estimate a flood wave: Potential applications to remote sensing of rivers. *J. of Hydrology*, 328:258–266, 2006.

[49] B.F. Sanders and N.D. Katopodes. Control of canal flow by adjoint sensitivity method. *Journal of irrigation and drainage engineering*, 125(5):287–297, 1999.

[50] Y. Sasaki. *Some basic formalisms in numerical variational analysis.* Citeseer, 1970.

[51] S. Simeoni-Sauvage. *Modélisation hydrobiogéochimique de la Garonne à l'étiage estival : Cas de l'azote entre Toulouse et Agen (120 kilomètres).* PhD thesis, Toulouse, INPT, 1999.

[52] O. Thual. *Hydrodynamique de l'environnement.* Ecole polytechnique, http://thual.perso.enseeiht.fr/xsee/ch5/allpdf/00main-frangl.pdf, 2010.

[53] EF. Toro. *Shock-capturing methods for free-surface shallow flows.* Wiley, 2001.

[54] J-P. Vila. *Théorie et approximation numérique de problèmes hyperboliques non linéaires applications aux équations de Saint-Venant et à la modélisation des avalanches de neige dense.* PhD thesis, Université Paris VI, March 1986.

[55] Y. Yoon, M. Durand, C.J. Merry, E.A. Clark, K.M. Andreadis, and Alsdorf D.E. Estimating river bathymetry from data assimilation of synthetic swot measurements. *Journal of Hydrology*, 464 - 465(0):363 – 375, 2012.

[56] Y. Yoon, P-A. Garambois, R. Paiva, M. Durand, H. Roux, and E. Beighley. Improved error estimates of a discharge algorithm for remotely sensed river measurements: Test cases on Sacramento and Garonne Rivers. *Water Resources Research*, 52(1):278–294, 2016.