VARIATIONAL DATA ASSIMILATION FOR 2D FLUVIAL HYDRAULICS SIMULATIONS

Marc Honnorat, Xijun Lai, Jerome Monnier, Francois-Xavier Le Dimet

VARIATIONAL DATA ASSIMILATION FOR 2D FLUVIAL HYDRAULICS SIMULATIONS

MARC HONNORAT[1], XIJUN LAI, JÉRÔME MONNIER AND FRANÇOIS-XAVIER LE DIMET

[1] Laboratoire de Modélisation et Calcul
B.P. 53, 38041 Grenoble Cedex 9
Projet IDOPT, INRIA Rhône-Alpes
marc.honnorat@imag.fr

Abstract

A numerical method for model parameters identification is presented for a river model based on a finite volume discretization of the bidimensional shallow water equations. We use variational data assimilation to combine optimally physical information from the model and observation data of the physical system in order to identify the value of model inputs that correspond to a numerical simulation which is consistent with reality. Two numerical examples demonstrate the efficiency of the method for the identification of the inlet discharge and the bed elevation. An application to real data on the Pearl River for the identification of boundary conditions is presented.

1. INTRODUCTION

The numerical simulation of river flows requires a precise modelling of the underlying physics. The bidimensional shallow water equations can describe accurately many free surface hydraulic configurations. However, in order to carry out a realistic simulation of a particular system, the numerical model requires information related to the physical domain. Model inputs such as bed elevation, roughness coefficients, initial and boundary conditions determine the state of the flow and should therefore be determined accurately. Unfortunately, they are usually not well known. Variational data assimilation makes it possible to optimally combine information from the model and observations of the flow in order to identify model parameters that minimize the discrepancy between simulation results and physical measurements. Here, observations of water elevation and velocities are considered.

The method is applied to the identification of boundary conditions and bed elevation in academic test cases and a case with real data. The bidimensional conservative shallow water equations and their discretization using a finite volume scheme are presented in Section 2, the variational data assimilation procedure is described in Section 3. Numerical results on the identification of inlet discharge and bed topography with two academic test cases are presented in Section 3 as well as an application to real data on the Pearl River for the identification of boundary water level. The conclusion is drawn in Section 5.

2. SHALLOW WATER MODEL

2.1. **Shallow Water equations.** We use the bidimensional shallow water equations in a conservation formulation. The state variables are the water depth h and the local discharge $\mathbf{q} = h\mathbf{u}$, where $\mathbf{u}$ is the depth-averaged velocity vector. On a computational domain Ω and for a computational time interval $[0, T]$, the shallow water equations can be written as

$$\left\{ \begin{array}{lll} \partial_t h \ + \ \mathrm{div}(\,\mathbf{q}\,) \ = \ 0 & & \text{in } \Omega\times]\,0,T] \\ \partial_t \mathbf{q} \ + \ \mathrm{div}\left(\frac{1}{h}\,\mathbf{q}\otimes\mathbf{q}\right) \ + \ \frac{1}{2}g\nabla h^2 + \ gh\nabla z_b \ + \ g\frac{n^2\|\mathbf{q}\|_2}{h^{7/3}}\,\mathbf{q} \ = \ 0 & & \text{in } \Omega\times]\,0,T] \\ h(0) \ = \ h_0, \qquad \mathbf{q}(0) \ = \ \mathbf{q}_0 & & \end{array} \right. \tag{1}$$

where g is the magnitude of the gravity, z_b the bed elevation, n the Manning coefficient, h_0 and $\mathbf{q}_0$ are the initial conditions for the state variables. In the following, the variable $c = \sqrt{gh}$ will denote the local wave celerity.

To complete this set of equations, one must define boundary conditions. Let Γ be the boundary of the domain Ω, it can be split up in the following way: $\Gamma = \Gamma_q \cup \Gamma_z \cup \Gamma_t \cup \Gamma_w$. On boundary Γ_q, we prescribe a discharge $\bar{q}$ and a homogeneous Neumann condition on the water depth. Boundary Γ_w corresponds to a wall, we prescribe a slip condition on the velocity and a homogeneous Neumann condition on the water depth. A water elevation $\bar{z_s}$ is prescribed on boundary Γ_z and homogeneous Neumann condition for all state variables are prescribed on boundary Γ_t:

$$\text{on } \Gamma_q : \qquad (\mathbf{q}\cdot\mathbf{n})_{|\Gamma_q}(t) = -\bar{q}(t)\,,\ \tfrac{\partial h}{\partial \mathbf{n}}_{|\Gamma_q}(t) = 0 \qquad \forall\, t \in]0,T]\,, \tag{2}$$

$$\text{on } \Gamma_w : \qquad \mathbf{u}\cdot\mathbf{n}_{|\Gamma_w}(t) = 0\,,\ \tfrac{\partial h}{\partial \mathbf{n}}_{|\Gamma_w}(t) = 0 \qquad \forall\, t \in]0,T]\,, \tag{3}$$

$$\text{on } \Gamma_z : \quad h_{|\Gamma_z}(t) = \bar{z_s}(t) - z_b{}_{|\Gamma_z},\ \tfrac{\partial(\mathbf{u}\cdot\mathbf{n}+2c)}{\partial \mathbf{n}}_{|\Gamma_z}(t) = 0 \quad \forall\, t \in]0,T]\,, \tag{4}$$

$$\text{on } \Gamma_t : \qquad \tfrac{\partial h}{\partial \mathbf{n}}_{|\Gamma_t}(t) = 0\,,\ \tfrac{\partial \mathbf{q}}{\partial \mathbf{n}}_{|\Gamma_t}(t) = 0 \qquad \forall\, t \in]0,T]\,. \tag{5}$$

The condition (4) on Γ_z is valid only for sub-critical flows, *i.e.* if the local Froude number $Fr = \frac{\|\mathbf{u}\|}{c}$ is strictly less than 1. Otherwise, one should prescribe condition (5).

2.2. **Finite Volume solver.** The bidimensional shallow water equations described above are solved numerically on an unstructured mesh using the finite volume method. The system (1) can be written in a general form as

$$\partial_t U \ + \ \mathrm{div}\, F(U) \ = \ S(U)\,, \tag{6}$$

where $U = (h, \mathbf{q})^T$ is the vector of conservative variables, $F(U) = \big(G(U), H(U)\big)^T$ the flux vector and $S(U)$ the source term

$$G(U) = \begin{pmatrix} q_x \\ \frac{1}{h}q_x^2 + \frac{1}{2}gh^2 \\ \frac{1}{h}q_x q_y \end{pmatrix},\ H(U) = \begin{pmatrix} q_y \\ \frac{1}{h}q_x q_y \\ \frac{1}{h}q_y^2 + \frac{1}{2}gh^2 \end{pmatrix},\ S(U) = \begin{pmatrix} 0 \\ -gh\nabla z_b - g\frac{n^2\|\mathbf{q}\|_2}{h^{7/3}}\,\mathbf{q} \end{pmatrix}.$$

The computational domain Ω is discretized using triangular or quadrangular cells. We define the mean value of the state variable U on an arbitrary cell K_i by

$$U_i = \frac{1}{|K_i|}\int_{K_i} U\, d\Omega\,, \tag{7}$$

where $|K_i|$ denotes the surface of the cell. By integrating Eq. (6) over K_i, using the divergence theorem, we obtain

$$\int_{K_i} \partial_t U \, d\Omega \; + \; \sum_{j=1}^{N_i} \int_{E_{ij}} T_{ij}^{-1} G(T_{ij} U) \, ds \; = \; \int_{K_i} S(U) \, d\Omega \; , \tag{8}$$

where N_i denotes the number of faces of the cell K_i (3 or 4), E_{ij} is the cell interface (see Fig. 1) and T_{ij} is the 3×3 rotation matrix of angle θ_{ij}. The usual flux term derived from the divergence theorem has been replaced by integrals over the cell edges thanks to the rotational invariance property of the shallow water equations (see [Toro, 2001, p. 65]). Hence, the bidimensional problem actually consists in a sum of one-dimensional Riemann problems that can be solved numerically using a Riemann solver. One can write the following semi-discrete scheme

$$\frac{d}{dt} U_i \; + \; \frac{1}{|K_i|} \sum_{j=1}^{N_i} T_{ij}^{-1} \, \widetilde{G}(U_L, U_R) = S_i \; , \tag{9}$$

where $\widetilde{G}(U_L, U_R)$ is an approximation of the flux through the cell interface E_{ij}. Subscripts L and R denote cells respectively to the left and to the right of the interface. To compute numerically the discrete flux $\widetilde{G}(U_L, U_R)$, we use the HLLC approximate Riemann solver [Toro, 2001]. This first order scheme handles correctly the transition between sub-critical and super-critical flows, unlike most other first order scheme (see [Zoppou and Roberts, 2003]). The discretization of the bed slope is actually included in the flux term. A forward Euler scheme is used for time discretization. As a result, the following stability condition on the time step Δt must be satisfied: $\Delta t \leq \frac{\min(d_{L,R})}{\max(\|\mathbf{u}\|+c)}$, where $d_{L,R}$ is the distance between the cell center and the center of interface.

3. VARIATIONAL DATA ASSIMILATION

The state of the flow is solution of the equation system (1) combined with the set of boundary conditions (2–5). Hence, it is determined by the initial conditions (Cauchy problem), by the model parameters (z_b and n) and the boundary conditions. The values of these model inputs form the control vector $\mathbf{c} = (h_0, \mathbf{q}_0, z_b, n, \bar{q}, \bar{z}_s)^T$.

In order to carry out a simulation of a real flow, it is necessary to have a good knowledge of these model inputs. Unfortunately, they are incompletely known in practice, and

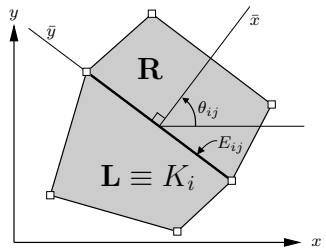


Figure 1: Two adjacent finite volumes

when an approximation is available, it is often subject to large uncertainties. However, some observations of the flow state may be available, such as water elevations or velocity measurements. These should be in accordance with the simulation results. Therefore, the problem to be addressed consists in identifying the control vector $\mathbf{c}$ consistent with both the simulation results and the hydraulic reality represented by observation data.

Data assimilation methods make it possible to combine optimally observation data with numerical model in order to improve the simulation results. Variational data assimilation [Le Dimet and Talagrand, 1986] is based on optimal control theory [Lions, 1971] and consists in identifying the control vector that minimizes a cost function.

3.1. **Cost function.** In order to compare the simulation results and the observations of the flow, we introduce a functional

$$j(\mathbf{c}) \;=\; \int_0^T \Big(\big\| C_h h(t) - h^{obs}(t) \big\|^2 + \big\| C_{\mathbf{q}} \mathbf{q}(t) - \mathbf{q}^{obs}(t) \big\|^2 \Big)\, dt \;, \tag{10}$$

where $\mathbf{c}$ is the control vector, h^{obs} and q^{obs} are observations of the flow, C_h and $C_{\mathbf{q}}$ are operators that map the state variables to the observations. This cost function measures the discrepancy between the computed state variables and available observations. Its minimization is carried out using a descent algorithm that requires the computation of the gradient of the cost function, *i.e.* the vector of its partial derivatives with respect to each control variable. We use the quasi–Newton algorithm M1QN3 written by [Gilbert and Lemaréchal, 1989].

3.2. **Adjoint model.** An efficient computation of the gradient is necessary to carry out an efficient minimization. However, the cost function is not explicitly dependant of the control vector. Indeed, it is an explicit function of the state variable $(h, \mathbf{q})$ which depends implicitly on $\mathbf{c}$ through the shallow water equations. Therefore, the gradient of the cost function cannot be computed directly. A direct computation of the gradient using a tangent linear model would require extensive computations. Conversely, the use of an adjoint model makes it possible to compute exactly all components of the gradient with a single backward integration in time of an adjoint model

$$\left\{ \begin{array}{l} \partial_t \widetilde{h}(t) - \mathbf{u}\cdot(\mathbf{u}\cdot\nabla)\widetilde{\mathbf{q}} + gh\,\mathrm{div}(\widetilde{\mathbf{q}}) - g\widetilde{\mathbf{q}}\cdot\nabla z_b \\ \qquad\qquad + \frac{7}{3} g \frac{n^2\|\mathbf{u}\|}{h^{4/3}} \mathbf{u}\cdot\widetilde{\mathbf{q}} \;=\; C_h^T (C_h h(t) - h^{obs}(t)) \qquad \forall\, t \in]0,T[\\ \partial_t \widetilde{\mathbf{q}}(t) + (\mathbf{u}\cdot\nabla)\widetilde{\mathbf{q}} + (\nabla\widetilde{\mathbf{q}})\mathbf{u} + \nabla\widetilde{h} - g\frac{n^2\|\mathbf{u}\|}{h^{4/3}}\widetilde{\mathbf{q}} \\ \qquad\qquad - g\frac{n^2}{h^{4/3}\|\mathbf{u}\|}(\mathbf{u}\otimes\mathbf{u})\widetilde{\mathbf{q}} \;=\; C_{\mathbf{q}}^T (C_{\mathbf{q}}\mathbf{q}(t) - \mathbf{q}^{obs}(t)) \quad \forall\, t \in]0,T[\\ \widetilde{h}(T) = 0\,, \quad \widetilde{\mathbf{q}}(T) = 0\,, \quad \widetilde{h}_{|\Gamma_t} = 0\,, \quad \frac{\partial h}{\partial \mathbf{n}}_{|\Gamma_q\cup\Gamma_w} = 0\,, \\ \widetilde{\mathbf{q}}_{|\partial\Omega\setminus\Gamma_z} = 0\,, \quad \widetilde{\mathbf{q}}\cdot\tau_{|\Gamma_z} = 0\,, \quad \big(\widetilde{h} + 2(\mathbf{u}\cdot\mathbf{n})(\widetilde{\mathbf{q}}\cdot\mathbf{n})\big)_{|\Gamma_z} = 0\;. \end{array} \right. \tag{11}$$

The partial derivatives of the cost function are simple functions of the adjoint state variables $\widetilde{h}$ and $\widetilde{\mathbf{q}}$. For example

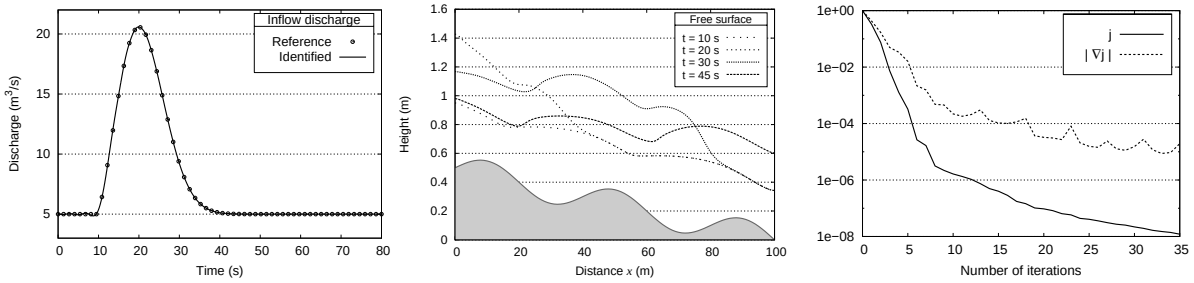
$$\frac{\partial j}{\partial h_0}(\mathbf{c}) = -\widetilde{h}(0)\,, \qquad \frac{\partial j}{\partial \mathbf{q}_0}(\mathbf{c}) = -\widetilde{\mathbf{q}}(0)\,, \qquad \frac{\partial j}{\partial z_b}(\mathbf{c}) = -\int_0^T \mathrm{div}\big(gh(t)\widetilde{\mathbf{q}}(t)\big)\, dt\,,$$
$$\frac{\partial j}{\partial \bar{q}}(\mathbf{c}) = -\widetilde{h}_{|\Gamma_q}, \qquad \frac{\partial j}{\partial \bar{z}_s}(\mathbf{c}) = \Big((\widetilde{\mathbf{q}}\cdot\mathbf{n})\big(c^2 - (\mathbf{u}\cdot\mathbf{n})^2\big) \Big)_{|\Gamma_z}\,.$$

3.3. **Automatic Differentiation.** In practice, there are three main methods to obtain an implementation of the adjoint model. The continuous adjoint model (11) can be discretized using an appropriate numerical scheme which is then implemented. One can also write the adjoint of the direct numerical scheme and implement it. A better way consists in starting directly from the implementation of the direct model, and writing the adjoint of the code. A large part of this extensive task can be automated using algorithmic differentiation [Griewank, 2000]. Here, we use the automatic differentiation tool Tapenade [Hacoët and Pascual, 2004].

4. NUMERICAL RESULTS

In order to validate the data assimilation method, we carry out twin experiments. Using a defined set of parameters $\mathbf{c}^{ref}$, some observations are created from the state variable generated by the direct model. They are called $(h^{obs}, \mathbf{u}^{obs})$. The goal is to identify the value of $\mathbf{c}^{ref}$, from an *a priori* hypothesis $\mathbf{c}^0$ with data assimilation.

4.1. **Validation case: identification of input discharge.** Here, we consider the identification of the inlet discharge $\bar{q}$, which is a boundary condition in the shallow water equations (1). The domain is a 100×8 meters rectangle with an irregular topography defined by: $z_b(x,y) = \frac{1}{2} - \frac{x}{200} + \frac{1}{10}\sin\left(\frac{x\pi}{20}\right)\cos\left(\frac{(y-2)\pi}{6}\right)$. The domain is discretized by a regular mesh made up of 100×10 cells. The upstream boundary corresponds to the edge $x = 0$. The Manning coefficient is set to $n = 0.025$ and a constant discharge $\bar{q} = 5\ m^3/s$ drives the flow to a steady state, which is used as an initial condition (h_0^{ref}, q_0^{ref}) for the reference flow. A flood is simulated by increasing substantially the discharge at the inlet over a short period of time. From the aforementioned steady state, we carry out a simulation of $80\,s$ with a time step $\Delta t = \frac{5}{100}\,s$ and a reference discharge defined by: $\bar{q}^{ref}(t) = 5 + 2(t-10)\exp\big(-\frac{1}{100}(t-15)^2\big)\mathbb{1}_{t\geq 10\,s}$. The corresponding hydrograph is represented with circles in Fig. 2 (a). We can see the propagation of the wave on Fig. 2 (b), where a cut of the free surface in the plane defined by $y = 1.6\,m$ is plotted for several simulation steps. The water is flowing in the direction of increasing x. Observations h_m^{obs} of the water depth are measured at each time step in a unique observation point located at (x_m, y_m). As an initial guess for the discharge, we choose the constant value $\bar{q}^0 = 5\ m^3/s$.



(a) Boundary condition (b) Free surface evolution (c) Cost function

Figure 2: Boundary condition identification: configuration

All other inputs are left unchanged. We introduce the following cost function

$$j_1(\bar{q}) \;=\; \frac{1}{2}\int_0^T \left| h(x_m, y_m; t) - h_m^{obs}(t)\right|^2 \, dt \;, \tag{12}$$

where $h(x_m, y_m; t)$ denotes the water depth at the point (x_m, y_m) at time t. The value of the cost function as well as the norm of its gradient are plotted against the number of iterations of the minimization procedure in Fig. 2 (c). The identified discharge is represented with a continuous line in Fig. 2 (a). We can see that in 35 iterations, the value of the cost function has been divided by 10^8 and that the identified boundary condition perfectely match the reference value.

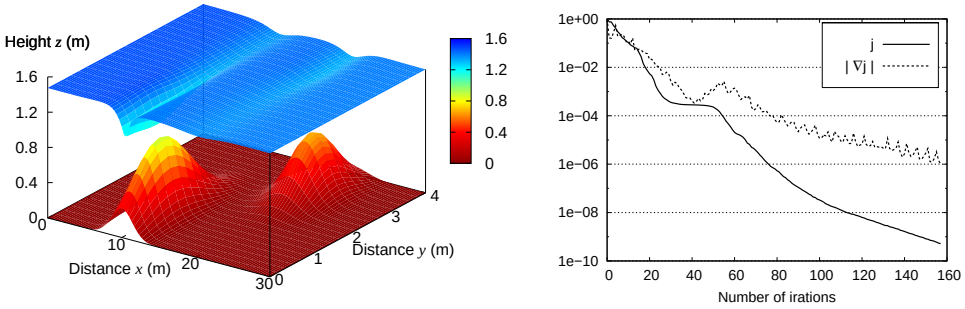
4.2. **Validation case: identification of topography.** Variational data assimilation is now applied to the identification of topography, *ie.* the value of the parameter z_b in equations (1). We consider a rectangular domain with dimensions 30×4 meters and a reference topography defined by $z_b^{ref}(x,y) = 0.9\exp\big(-\frac{1}{4}(x-10)^2\big)\exp\big(-(y-1)^2\big) + 0.7\exp\big(-\frac{1}{8}(x-20)^2\big)\exp\big(-2(y-3)^2\big)$. The domain is discretized by a regular rectangular mesh made of 90×20 cells. The inflow boundary is located at $x = 0$, the outflow at $x = 30$. The boundaries defined by $y = 0$ and $y = 4$ are walls. Thus, the water flows in the direction of increasing x. The Manning coefficient is set to $n = 0.025$, the water elevation at the outlet is $\bar{z}_s = 1.4\ m$ and a constant discharge $\bar{q} = 8\ m^3/s$ drives the flow to a steady state after about $80s$ of simulation. Figure 3 (a) shows the reference topography z_b and the elevation of the free surface in this configuration. For a simulation period of $T = 20\,s$, observations of water depth observations h^{obs} and velocity $\mathbf{u}^{obs}$ are recorded for every cell of the mesh and for every time step.

The aim of the experiment is to retrieve the value of the reference topography z_b^{ref} the initial hypothesis of a flat bed $z_b \equiv 0$. To that purpose, the data assimilation scheme is used with the following cost function to be minimized:

$$j_2(z_b) \;=\; \frac{1}{2}\int_0^T \left(\left\| h(t) - h^{obs}(t)\right\|_\Omega^2 + \left\| \mathbf{u}(t) - \mathbf{u}^{obs}(t)\right\|_\Omega^2 \right) dt \;, \tag{13}$$

where $\|\cdot\|_\Omega$ denotes the L^2 norm on the whole domain. The value of the cost function and the norm of its gradient, both normalized by their initial values, are plotted against the number of iterations of the minimization process in Fig. 3 (b). One can notice that convergence has been achieved, and that the original shape of the topography has been well retrieved.

4.3. **Real case: the Pearl River.** In the previous sections, two academic twin experiments demonstrate the good performance of variational data assimilation to identify control parameters using sufficient observations. Now, the method is applied to a real case: the downstream reaches of the Pearl River, Southern China. In the study area (see Fig. 4), the flow is mainly driven by the tidal force. Synchronous measurements of water level and discharge are partly or fully available with a one hour time interval for the six open boundaries (denoted BC_1 to BC_6) as well as for three cross-sections denoted A, B and C. Both discharge and water level observations are available at boundaries BC_4, BC_5, BC_6 and cross sections A, B and C, yet only water level has been measured at boundaries BC_1 and BC_2. The domain is discretized using an hybrid triangular/quadrilateral mesh,



(a) Bed and free surface elevation (b) Convergence of the cost function

Figure 3: Topography identification

consisting of 1684 cells and 1784 nodes. According to the available data, we consider the following data assimilation experiment: given coarse estimates of the initial and boundary conditions, we seek to identify the water level at boundaries BC_1, BC_2 and BC_6 as well as a better initial state using the observed water level hydrograph h^{obs} at the stations A, B and C. To that purpose, we defined the corresponding cost function to be minimized

$$j_3(\bar{z_s}, h_0, \mathbf{q}_0) \;=\; \frac{1}{2}\int_0^T \sum_{i=1}^{3} |h_i(x_m, y_m; t) - h_i^{obs}(t)|^2 dt \tag{14}$$

where $\bar{z_s}$ is the water level at the control boundaries BC_1, BC_2 and BC_6, h_0 and $\mathbf{q}_0$ denote the initial state variables and $h_i(t)$ $(i = 1, 2, 3)$ denotes the water depth at the three observation stations along each cross section at time t. As coarse estimates for water level at the boundaries BC_1, BC_2, and BC_6, we choose a constant value $\bar{z_s} = -0.5\,m$ instead of the observation data. Variational data assimilation is then carried out over a period of 36 hours. The decrease of the cost function is rapid in the first three minimization iterations, then slows down with small oscillating gradients (see Fig. 5 (a)). The water elevation at boundary BC_6 is identified successfully, while the identified water level for BC_1 and BC_2 differs significantly from the observations (see Fig. 5 (b)). From the study of the gradient of the cost function, it has been found that its value is one order of magnitude larger at BC_6 than at BC_1 or BC_2. It means that these two control variables are much less sensitive to the observations at the gauging stations A, B and C than BC_6. This should be put in parallel with the distance of the measurement locations to the boundaries. Figure 5 (c) shows that when using the identified boundary conditions for a direct simulation, the state variable perfectly match the measurements of water elevations at the three gauging stations.

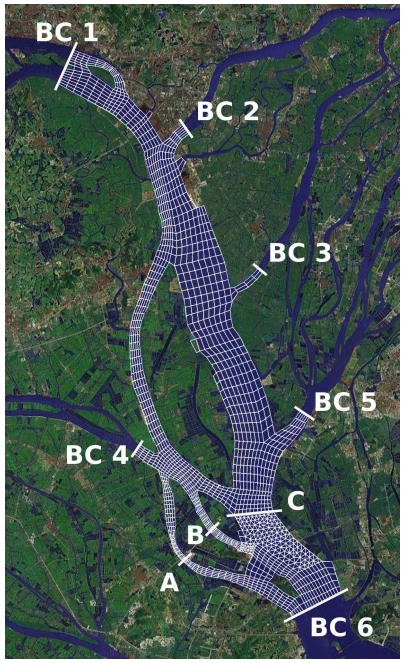
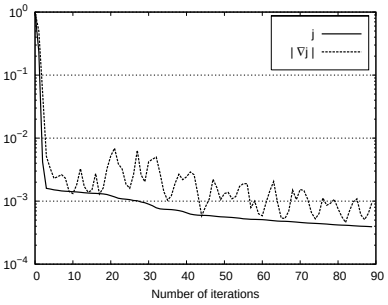
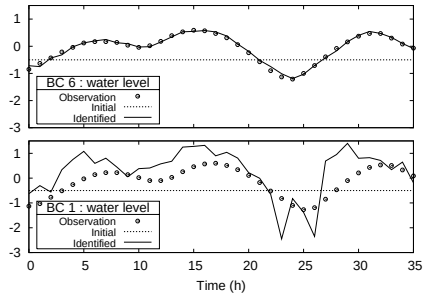


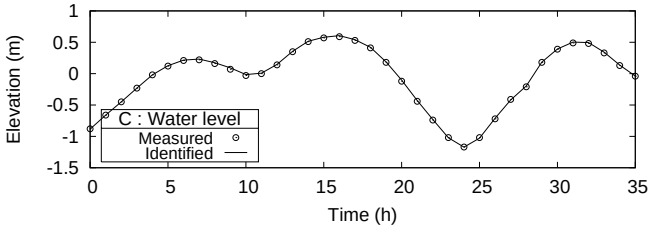
Figure 4: The Pearl river



(a) Convergence of the cost function



(b) Identified water level at BC_1 and BC_6



(c) Comparison of identified water level and observations at C

Figure 5: Pearl river numerical results

5. CONCLUSION

We have presented variational data assimilation for a river model based on a finite volume discretization of the bidimensional shallow water equations. Academic test cases based on twin experiments demonstrate the ability of the method to successfully identify model parameters in various configurations. The inlet discharge in a channel can be identified when an observation of water elevation is available in a single point of the domain. The identification of bed topography is also possible when a dense set of observations of water elevation and velocities is considered. In a configuration with real data on the Pearl River, the method has proved to be able to identify water elevation boundary conditions from sparse measurements of the water level on three cross-sections.

References

Courtier, P., and O. Talagrand (1990), Variational assimilation of meteorological observations with the direct and adjoint shallow water equations, *Tellus, 42A*, 531–549.

Gilbert, J.-C., and C. Lemaréchal (1989), Some numerical experiments with variable storage quasi-Newton algorithms, *Mathematical programming, 45*, 407–435.

Griewank, A. (2000), *Evaluating Derivatives: Principles and Techniques of Algorithmic Differentiation*, *Frontiers in Appl. Math., SIAM*, vol. 19, Philadelphia, PA.

Hascoët, A., V. Pascual (2004), Tapenade 2.1 user's guide, Technical Report RT–300, INRIA.

Le Dimet, F.-X., and O. Talagrand (1986), Variational algorithms for analysis and assimilation of meteorological observations: theoretical aspects, *Tellus, 38A*, 97–110.

Lions, J.-L. (1971), *Optimal control of systems governed by partial differential equations*, Springer-Verlag.

Toro, E.F. (2001), *Shock-capturing methods for free-surface shallow flows*, J. Wiley and Sons.

Zoppou, C., and S. Roberts (2003), Explicit schemes for dam-break simulations, *Journal of Hydraulic Engineering, 129(1)*, 11–34.