RANDOMLY SUPPORTED VARIATION OF DETERMINISTIC MODELS AND ITS APPLICATION TO ONE-DIMENSIONAL SHALLOW WATER FLOWS

E. G. Birgin, Ph.D.[1], M. R. Correa, Ph.D.[2], V. A. González-López, Ph.D.[3], J. M. Martínez, Ph.D.[4], and D. S. Rodrigues, Ph.D.[5]

[1]Professor, Dept. of Computer Science, Institute of Mathematics and Statistics, University of São Paulo, Rua do Matão, 1010, Cidade Universitária, 05508-090, São Paulo, SP, Brazil. Email: egbirgin@ime.usp.br. Corresponding author.

[2]Associate Professor, Dept. of Applied Mathematics, Institute of Mathematics, Statistics, and Scientific Computing, University of Campinas, 13083-859, Campinas, SP, Brazil. Email: maicon@ime.unicamp.br

[3]Associate Professor, Dept. of Statistics, Institute of Mathematics, Statistics, and Scientific Computing, University of Campinas, 13083-859, Campinas, SP, Brazil. Email: veronica@ime.unicamp.br

[4]Emeritus Professor, Dept. of Applied Mathematics, Institute of Mathematics, Statistics, and Scientific Computing, University of Campinas, 13083-859, Campinas, SP, Brazil. Email: martinez@ime.unicamp.br

[5]Assistant Professor, School of Technology, University of Campinas, 13484-332, Limeira, SP, Brazil. Email: diego.rodrigues@ft.unicamp.br

ABSTRACT

This paper deals with the prediction of flows in open channels. For this purpose, models based on partial differential equations are used. Such models require the estimation of constitutive parameters based on available data. After this estimation, the solution of the equations produces

predictions of flux evolution. In this work, we consider that most natural channels may not be well represented by deterministic models for many reasons. Therefore, we propose to estimate parameters using stochastic variations of the original models. There are two types of parameters to be estimated: constitutive parameters (such as roughness coefficients) and the parameters that define the stochastic variations. Both types of estimates will be computed using the maximum likelihood principle, which determines the objective function to be used. After obtaining the parameter estimates, due to the random nature of the stochastic models, we are able to make probabilistic predictions of the flow at times or places where no observations are available.

INTRODUCTION

The Saint-Venant equations are often used to predict river flows (Ding and Wang 2005; Ayvaz 2013; Ying et al. 2004). To solve these partial differential equations, one must know the initial state of the channel at time $t = t_0$, including the depth, fluid velocity, and/or flow rate at as many points as possible in the one-dimensional channel (Emmett et al. 1979). Information about boundary conditions, that is, values of the main description variables at the beginning and/or the end of the channel, may also be required. Additionally, topographic information such as bed slopes and shapes of transversal areas, as well as a roughness parameter, known as Manning's coefficient (Ding et al. 2004; Ding and Wang 2005; Pappenberger et al. 2005; Agresta et al. 2021), are needed. With this information at hand, we can solve the Saint-Venant equations using numerical methods and we can predict the physical characteristics of the flow in unknown positions or in the future.

If the channel's geometry is well determined, the Manning coefficients can be accurately estimated using available data. In general, standard least-squares procedures are successful for the minimization of differences between real observations and model predictions. Many papers have been devoted to the problem of estimating Manning's parameters using this approach. In Pappenberger et al. (2005) the performance of the HEC-RAS software (Brunner 1994) for predicting inundation data was analyzed as a function of Manning's roughness coefficient and weighting discretization parameters to produce dynamic probability maps of flooding during the event. HEC-RAS was also used in Agresta et al. (2021), where different heuristic methods were employed for optimizing the Manning coefficient. Ding and Wang (2005) solved the Saint-Venant equations to simulate flows in channel networks and used the resulting deterministic model to compute the optimal Manning's coefficient using standard quasi-Newton methods. Askar and Al-Jumaily (2008) estimated the Manning coefficient using Saint-Venant equations as predictors and sequential quadratic programming for optimization purposes. Ebissa and Prasad (2017) used the GVF (Gradually Varied Flow) equations for simulations and genetic algorithms for deterministic optimization of the roughness parameter. In Birgin and Martínez (2022), a secant derivative-free optimization method was developed for determining the Manning coefficient in synthetic experiments. In the Data Assimilation approach with joint state parameter estimation (Ziliani et al. 2019), at each time level one has estimations of the state variables, the constitutive parameters, and the process noise. Using simulation, forecasts of the state variables for the next time level are computed and the distribution of noise is updated.

In this paper we propose a modification of the original one-dimensional shallow water (Saint-Venant) deterministic model by introducing stochastic variations in order to add variability to, and in some cases also improve, the already proven successful estimations based on least-squares minimization of errors. Within this approach, there are two types of parameters to be estimated: Manning's coefficients and parameters that define perturbations of the (Saint-Venant) deterministic models (essentially, standard deviations). These parameters are coupled and are computed using the maximum likelihood principle.

Although stochastic modeling is, of course, not new, the idea presented in this paper for estimating distribution parameters using simulations and maximal likelihood has not been attempted in the past. In many regression problems, it is necessary not only to predict a value but also to give confidence or uncertainty intervals. For example, in Gaussian process-based models one looks for the type of dependence between two successive states with which a good fit to the available data and, sometimes, an adequate satisfaction of a physical law is produced. See Rasmussen and Williams (2005). Our approach is different in the sense that we start from the physical law and postulate that the observations are the result of a random perturbation of it. The magnitude of such perturbation, in our case, is estimated by maximizing the likelihood function.

This methodology can be useful for irregular rivers for which a one-dimensional simplification roughly corresponds to reality, and the available data are sparse both in time and space. These cases are very frequent in Brazil and other Latin American countries. We illustrate with numerical experiments that the proposed method works well when the observed data come from laboratory tests and in tests that involve a real river reach.

Related approaches to the one presented in this paper can be found in the Biostatistics literature in connection to growth processes (Chao and Huisheng 2016; Delgado-Vences et al. 2023; Jiang and Shi 2005; Lillacci and Khammash 2010; Román-Román et al. 2010). For a comprehensive treatment of Stochastic Differential Equations, see Panik (2017). Kalman filter and its nonlinear variations (Kalman 1960) should also be evoked in this context as they produce stable estimations of a system's present state as a combination of observation and prediction. In Gaussian Processes, one models the evolutionary physical phenomenon as a stochastic process whose covariance needs to be estimated and where PDE relations are incorporated to feed the estimation process. In some sense, our approach is the inverse of the one adopted in the Gaussian Process. In fact, in our case, we start from the discretized PDE equation incorporating random variation as an essential part of the evolution model.

The rest of this paper is organized as follows. The Saint-Venant equations, selected as the basic model to describe the flux of water in one-dimensional channels, are described in Section 2. The proposed model considering inadequacies and the parameters estimation strategy based on a likelihood function is introduced in Section 3. In Section 4 we describe the optimization procedure. An extensive set of numerical experiments describing different open channel flow scenarios and comparing the results obtained from the deterministic and the stochastic models is reported in Section 5. This includes the application of the proposed method to a real irregular river. The last section presents the conclusions and lines for future research.

SAINT-VENANT EQUATIONS

The Saint-Venant equations (Saint-Venant 1871) are usually employed for river-flow simulations.

These equations are given by

$$\frac{\partial A}{\partial t} + \frac{\partial Q}{\partial x} = 0 \tag{1}$$

and

$$\frac{\partial Q}{\partial t} + \frac{\partial}{\partial x}\left(\frac{Q^2}{A}\right) + gA\frac{\partial Z}{\partial x} + \frac{n_g^2 Q|Q|}{AR^{4/3}} = 0 \tag{2}$$

for $t \in [0, T]$ and $x \in [x_I, x_F]$, with x_I and x_F representing the initial and final representative points of the analyzed part of the channel. $Z(x,t)$ is the water surface elevation, $z_b(x)$ is the channel bed elevation, $h(x,t) = Z(x,t) - z_b(x)$ is the depth of the river, $A(x,t)$ is the transversal wetted area, $P(x,t)$ is the wetted perimeter, $R(x,t) = A(x,t)/P(x,t)$ is the hydraulics radius, $V(x,t) = Q(x,t)/A(x,t)$ is the average speed of the fluid, and g is the acceleration of gravity taken as $9.81\,\mathrm{m/s^2}$. Equation (1) describes mass conservation and equation (2) represents balance of the linear momentum. The coefficient n_g is known as Yen-Manning roughness coefficient, introduced in Yen (1992) and Yen (1993), which has units $\mathrm{m}^{1/6}$ in SI. This parameter relates to the classical Manning's coefficient n through the relation $n_g = \sqrt{g}\,n$. Typically, this roughness coefficient depends on x due to the morphological aspects of the river along its course. Sediment deposition can also affect the roughness coefficients over time.

Other forms of the Saint-Venant equations can be considered as well. For example, in Ding and Wang (2005) and Chaudhry (2022) it is considered a more general form of equation (2) in order to take into account the non-uniformity of velocity in cross-sections. In their approach, the momentum equation takes the form

$$\frac{\partial}{\partial t}\left(\frac{Q}{A}\right) + \frac{\partial}{\partial x}\left(\frac{\beta Q^2}{2A^2}\right) + g\frac{\partial Z}{\partial x} + \frac{n_g^2 Q|Q|}{A^2 R^{4/3}} = 0, \tag{3}$$

where β is a momentum correction factor. If we consider that equation (3) represents a more accurate representation of the balance of momentum, the employment of (2) may represent an "error" in the modeling that persists throughout the time horizon. Our approach is intended to deal with all kinds of errors that arise in the description of evolutionary systems. Of course, in general, it is better to use the physical model of the phenomena that best corresponds to reality. However, the approach supported in this paper is intended to be applied to "inaccurate" models.

The development of accurate, efficient, and robust numerical schemes for calculating approximate solutions of the hyperbolic systems (1,2) and (1,3) is still a challenging issue that has already been extensively investigated (Cockburn 1999; Correa 2017; Khan and Lai 2014; Kurganov 2018; Ying et al. 2004). In the present work, we assume that the Saint-Venant system is numerically solved using a stable and accurate numerical scheme, for which the discrete in space and time formulation can be written as

$$\vec{\mathcal{U}}^{n+1} = \vec{\mathcal{F}}(t_n, t_{n+1}, \vec{\mathcal{U}}^n, \vec{n}_g). \tag{4}$$

In (4), $\vec{\mathcal{U}}^{n+1}$ is a vector containing the numerical solution at $t = t_{n+1}$, $\vec{\mathcal{F}}$ is a vector function depending on the previous solution $\vec{\mathcal{U}}^n$ and on a vector of parameters $(n_g)_i$, $i = 1, \ldots, n_{n_g}$, representing different values of the Yen-Manning coefficient n_g in space and time. The discrete form given by equation (4) is typical of explicit numerical schemes.

RANDOM VARIATION OF A MODEL AND THE ESTIMATION OF ITS PARAMETERS

The numerical solution of Saint-Venant equations provides state variables (transversal areas and flow rates) at a finite number of time instants $\mathcal{T} = \{t_1, \ldots, t_{|\mathcal{T}|}\}$. Our proposal is to perturb (or deviate) the computed states at selected time instants $t \in \mathcal{T}_\gamma \subseteq \mathcal{T}$ with random values in the way described below. More specifically, we postulate that, instead of obeying the evolution dictated by the numerical solution of Saint-Venant equations, the actual evolution of channel flows obeys the stochastic process

$$\vec{\mathcal{U}}^{n+1} = \vec{\mathcal{F}}(t_n, t_{n+1}, \vec{\mathcal{U}}^n, \vec{n}_g) + \delta(t_{n+1})\, \vec{\mathcal{V}}(t_{n+1}, \vec{\mathcal{U}}^n, \sigma). \tag{5}$$

In (5), $\vec{\mathcal{F}}(t_n, t_{n+1}, \vec{\mathcal{U}}^n, \vec{n}_g)$ represents the state variables computed by the chosen numerical Saint-Venant solver at time t_{n+1} and $\vec{\mathcal{V}}(t_{n+1}, \vec{\mathcal{U}}^n, \sigma)$ is a vector whose entries are random variables with zero expectation and standard deviation equal to σ times the modulus of the correspondent component of $\vec{\mathcal{F}}(t_n, t_{n+1}, \vec{\mathcal{U}}^n, \vec{n}_g)$. Moreover, $\delta(t_{n+1})$ is an indicator function which depends on probabilistic parameters to be determined and that takes the values 1 and 0 according to the decision of perturbating the state at time t_{n+1} or not. As a result, the state computed by our randomly perturbed evolution method is defined by $\vec{\mathcal{U}}^{n+1}$.

We consider that perturbations are made at time instants in a set $\mathcal{T}_\gamma = \left\{t_1^\gamma, t_2^\gamma, \dots, t_m^\gamma\right\} \subseteq \mathcal{T}$. To construct $\mathcal{T}_\gamma$, for a given parameter γ, we define $t_1^\gamma = \min\{t \in \mathcal{T} \mid t - t_1 \geq \gamma\}$ and, for $l > 1$, $t_l^\gamma = \min\left\{t \in \mathcal{T} \mid t - t_{l-1}^\gamma \geq \gamma\right\}$. This means that time instants in the set $\mathcal{T}_\gamma$ have intervals of size around γ, or, in other words, that perturbations occur with frequency $1/\gamma$. Accordingly, for $t \in \mathcal{T}$, the indicator function $\delta(t)$ in (5) is defined as

$$\delta(t) = \begin{cases} 1, & \text{if } t \in \mathcal{T}_\gamma \\ 0, & \text{otherwise.} \end{cases}$$

This means that $\delta(t)$ in (5) depends on $\mathcal{T}_\gamma$, which in turn depends on the set of instants $\mathcal{T}$ determined by the method used for the numerical solution of the Saint-Venant equations and on the (unknown probabilistic) parameter γ. Moreover, for further reference, we define the vector $\vec{\tau} \in \mathbb{R}^m$ containing the time instants $t_1^\gamma < t_2^\gamma < \cdots < t_m^\gamma$ at which the second term in the right-hand side of equation (5) is "activated" (i.e., containing all the elements of $\mathcal{T}_\gamma$) as $\tau_i = t_i^\gamma$ for $i = 1, \dots, m$.

In summary, the stochastic model (5) differs from the deterministic discretized model (4) due to the introduction of zero-mean random perturbations in space, on almost equally-spaced-in-time states of the solution. It is important to emphasize that the proposed stochastic model seats upon the determination of three parameters that must be estimated using available data, namely, the Yen-Manning coefficients $\vec{n}_g$, the vector $\vec{\tau}$ (that depends by construction on the parameter γ and the set of instants $\mathcal{T}$ that is built by the chosen Saint-Venant solver) and the deviation σ. (Notice that if $\gamma = +\infty$ or $\sigma = 0$, then no deviation is introduced and the stochastic model coincides with the deterministic model.) The remainder of this section is devoted to the proposal for estimating the parameters of the stochastic model.

For the sake of simplicity, hereinafter, the description corresponds to the case in which the Yen-Manning coefficient is spatially homogeneous and does not change in time. However, there are no complications in extending it to the case in which different roughness coefficients are found for different arguments $x \in [x_I, x_F]$ and $t \in [0, T]$. Moreover, we also assume that the Saint-Venant solver was already chosen and parameter γ is known. Thus, the variables that determine the vector $\vec{\tau}$ are fixed and we focus on the determination of n_g and σ. As well as in the case of a spatially non-homogeneous roughness coefficient, considering $\vec{\tau}$, the vector that contains the time instants at which perturbations occur, an unknown array (of unknown dimension) fits within the scope of the procedure proposed in the present work. In general, an arbitrary number of parameters could be considered both in the deterministic part of the model and in the probabilistic part related to perturbations. The trade-off would be to have a more difficult optimization problem in the parameter adjustment phase.

We assume that N_{obs} observations v_k^{obs} (v may be either A or Q, or any other related quantity such as h or V) at spatial-time coordinates $(x_k^{\text{obs}}, t_k^{\text{obs}}) \in [x_I, x_F] \times [0, T]$ for $k = 1, \dots, N_{\text{obs}}$ are given. In addition, we assume that each observation v_k^{obs} is associated with a quantity $\vartheta_k > 0$, which typically represents the measurement error of the observation. Roughly speaking, ϑ_k is the absolute value of the difference between the kth observation and its simulation below which we consider that the similarity between both is high. Therefore, ϑ_k must take into account the intrinsic measurement error (due to the precision of the instrument) and, possibly, the intrinsic error of the simulation. Of course, in many cases we may consider that the latter is zero, but in other cases it is not. We wish to determine n_g and σ from available data using the maximization of a likelihood function calculated through simulations. To do so, for a given pair (n_g, σ), we consider N_{sim} simulations calculated through runs of the process (5). This means that the simulated values are calculated on a space-time grid. If any $(x_k^{\text{obs}}, t_k^{\text{obs}})$ does not belong to the grid, the corresponding simulated value may be calculated with interpolation.

Let v_{kj}^{sim} be the simulated value at spatial-time coordinate $(x_k^{\text{obs}}, t_k^{\text{obs}})$ for $k = 1, \dots, N_{\text{obs}}$ obtained at simulation j for $j = 1, \dots, N_{\text{sim}}$. The likelihood associated with a pair (n_g, σ) is intended to represent the probability of the given set of observations to be generated by (5). Roughly speaking, the considered likelihood is the ratio of the favorable cases to the total number of simulations N_{sim}. For each simulation j, instead of a binary definition of favorability, we propose the smoothed definition given by

$$\exp\left(-\left(d(v_1^{\text{obs}},\ldots,v_{N_{\text{obs}}}^{\text{obs}},v_{1,j}^{\text{sim}},\ldots,v_{N_{\text{obs}},j}^{\text{sim}})\right)^2\right), \tag{6}$$

where

$$d(v_1^{\text{obs}},\ldots,v_{N_{\text{obs}}}^{\text{obs}},v_{1,j}^{\text{sim}},\ldots,v_{N_{\text{obs}},j}^{\text{sim}}) = \sqrt{\frac{1}{N_{\text{obs}}}\sum_{k=1}^{N_{\text{obs}}}\left(\frac{(v_k^{\text{obs}}-v_{k,j}^{\text{sim}})^2}{\vartheta_k^2}\right)} \tag{7}$$

represents the root mean square deviation of simulation j with respect to observations. Therefore, the favorability of simulation j is 1 if v_k^{obs} coincides with $v_{k,j}^{\text{sim}}$ for $k = 1,\ldots,N_{\text{obs}}$ and is equal to $1/e \approx 0.37$ if the distance between v_k^{obs} and $v_{k,j}^{\text{sim}}$ is equal to ϑ_k for $k = 1,\ldots,N_{\text{obs}}$. Thus, ϑ_k can be chosen in practical cases as a representation (not necessarily a rigorous upper bound) of the measurement error of observation v_k^{obs}. Consequently, the likelihood associated with the pair (n_g,σ) is given by

$$\mathcal{L}_\vartheta(n_g,\sigma) = \frac{1}{N_{\text{sim}}}\sum_{j=1}^{N_{\text{sim}}}\exp\left(-\left(d(v_1^{\text{obs}},\ldots,v_{N_{\text{obs}}}^{\text{obs}},v_{1,j}^{\text{sim}},\ldots,v_{N_{\text{obs}},j}^{\text{sim}})\right)^2\right). \tag{8}$$

The parameters that are considered optimal for model (5) are the ones that maximize the likelihood function (8). A sketch representing at a high level the differences between the usual deterministic process and the procedure proposed in the present work is shown in Figure 1.

OPTIMIZATION PROCEDURE

Given the observed data v_k^{obs} at spatial-time coordinates $(x_k^{\text{obs}}, t_k^{\text{obs}})$ for $k = 1,\ldots,N_{\text{obs}}$, finding optimal n_g^* and σ^* consists of maximizing function (8). Evaluating (8) requires to consider N_{sim} simulations. The value of N_{sim} will be empirically determined for each experiment. Function (8) is a function of two variables, stochastic, and nonlinear. Considering that we know a priori intervals $[n_{g_{\min}}, n_{g_{\max}}]$ and $[\sigma_{\min},\sigma_{\max}]$ within which the optimal values n_g^* and σ^* lie, the simplest way to find these values is to choose steps Δn_g and $\Delta\sigma$ and to perform a global search within the given bounds. According to the desired accuracy of the optimal values, iterative refinements can be performed. A procedure that already includes successive refinements for the computation of n_g^* is detailed in Algorithm 1. In the numerical experiments, the algorithmic parameters

${n_g}_{\min}, {n_g}_{\max}, \sigma_{\min}$ and $\sigma_{\max}$ were established using rough estimates of the parameters sought. The number of simulations N_{sim} was decided empirically, starting from a small value and increasing it until verifying that increasing it does not significantly modify the results.

Algorithm 1: Estimation of the stochastic model parameters with successive refinements

---

Input: The observed data v_k^{obs} at spatial-time coordinates $(x_k^{\text{obs}}, t_k^{\text{obs}})$ and the measurement errors ϑ_k for $i = 1, \dots, N_{\text{obs}}$, the frequency of perturbations $1/\gamma$, the number of simulations N_{sim}, an initial interval $[{n_g}_{\min}, {n_g}_{\max}]$ for the Manning coefficient and its number of subdivisions n_{div,n_g}, the precision ε_{n_g} required for the Manning coefficient, a fixed interval $[\sigma_{\min}, \sigma_{\max}]$ for the dispersion parameter and the number $n_{\text{div},\sigma}$ of equidistant trials. In addition, an algorithm to solve equations (1,2) is given.

Output: Optimal values n_g^* and σ^* for the Manning parameter and the dispersion parameter, respectively.

- $\mathcal{L}_{\max} \leftarrow 0$
- $\Delta\sigma = (\sigma_{\max} - \sigma_{\min})/n_{\text{div},\sigma}$
- **while** $\left({n_g}_{\max} - {n_g}_{\min}\right) > \varepsilon_{n_g}$ **do**
  - $\Delta n_g \leftarrow \left({n_g}_{\max} - {n_g}_{\min}\right)/n_{\text{div},n_g}$
  - **for** $p = 0, 1, \dots, n_{\text{div},n_g}$ **do**
    - $n_g^{\text{trial}} \leftarrow {n_g}_{\min} + p\,\Delta n_g$
    - **for** $q = 0, 1, \dots, n_{\text{div},\sigma}$ **do**
      - $\sigma^{\text{trial}} \leftarrow \sigma_{\min} + q\,\Delta\sigma$
      - **for** $j = 1, 2, \dots, N_{\text{sim}}$ **do**
        - By solving (5), with $n_g \equiv n_g^{\text{trial}}$, using the algorithm chosen to solve (1,2), and perturbing with dispersion parameter $\sigma \equiv \sigma^{\text{trial}}$ and frequency $1/\gamma$, compute $v_{k,j}^{\text{sim}}$ for $k = 1, 2, \dots, N_{\text{obs}}$.
      - Evaluate the likelihood $\mathcal{L}_\vartheta(n_g^{\text{trial}}, \sigma^{\text{trial}})$ defined in (8).
      - **if** $\mathcal{L}_\vartheta(n_g^{\text{trial}}, \sigma^{\text{trial}}) > \mathcal{L}_{\max}$ **then**
        - $\mathcal{L}_{\max} \leftarrow \mathcal{L}_\vartheta(n_g^{\text{trial}}, \sigma^{\text{trial}})$
        - $n_g^* \leftarrow n_g^{\text{trial}}$
        - $\sigma^* \leftarrow \sigma^{\text{trial}}$
  - ${n_g}_{\min} \leftarrow n_g^* - \Delta n_g$
  - ${n_g}_{\max} \leftarrow n_g^* + \Delta n_g$

NUMERICAL EXPERIMENTS

In this section, we present numerical experiments to illustrate the performance of the proposed approach. Different open channel scenarios are considered, namely, the formation of a hydraulic jump in a horizontal flume, the simulation of a partial dam break, and simulations of a real river. In all experiments, the stochastic process (5), based on equations (1,2), is solved with the upwind conservative finite volume scheme proposed inYing et al. (2004). Moreover, the frequency $1/\gamma$ of the perturbations is assumed to be known and, thus, parameters to be determined from observed data are the Yen-Manning coefficient $n_g \in \mathbb{R}$ and the deviation parameter $\sigma \in \mathbb{R}$. It should be noted that the ultimate goal of the numerical experiments is not to determine these parameters but to deliver predictions and approximations of unavailable data. These predictions will include information related to the lack of fidelity with which the model represents reality. For the sake of completeness, the estimated values for the parameter n_g are compared with the values obtained using the deterministic model (which corresponds to considering $\gamma = +\infty$ and/or $\sigma = 0$ in the stochastic model).

It is important to highlight that Manning's coefficient may adequately approximate randomness in the geometry of channels. Our contribution is not in contradiction with this statement and does not seek a better approximation of the Manning coefficients than the ones obtained by other methods. What we aim to do is, given a model, a method of resolution and, perhaps, the result of a deterministic estimation of constitutive parameters (as the Manning roughness coefficient), to determine the probabilistic variation intrinsic either to the model or to the solution method that makes it possible probabilistic predictions (for example, confidence intervals) in situations not contemplated in the observation data. Although we consider the joint estimation of constitutive parameters and probabilistic parameters (dispersion), this is not essential to our approach. In other words, we may consider that constitutive parameters are well established by other methods. The fact that the calibrated n_g is, in many cases, the same as the one obtained by other, previously stated, methods is not a surprise. Our method may be interpreted in terms of the analysis of the error originated in the optimization of parameters of arbitrary models using arbitrary methods.

Algorithm 1 was implemented in Fortran. The code was compiled by the GFortran compiler of GCC (version 9.3.0) with the -O3 optimization directive enabled. Tests were conducted on a computer with a 4.5 GHz Intel Core i7-9750H processor and 16GB 1600 DDR4 2666 MHz RAM memory, running Ubuntu 20.04.

Hydraulic jump in a rectangular channel

In this first experiment, we study the formation of a hydraulic jump by taking as reference the experiment reported in Gharangik and Chaudhry (1991). There, the authors performed experimental investigations of the steady-state location of the hydraulic jump in a horizontal 14.0 m long and 0.46 m wide flume, for different Froude numbers $\mathrm{Fr} = V/\sqrt{gh}$, by starting from a supercritical flow in the entire channel and then controlling the tailwater depth by an adjustable downstream gate. The bottom of the flume is made up of metal and the walls are made up of glass for $0\,\mathrm{m} \leq x \leq 3.05\,\mathrm{m}$ and of metal for $3.05\,\mathrm{m} < x \leq 14\,\mathrm{m}$. The Saint-Venant equations were then solved in uniform meshes of $n_x = 50$ cells with Courant number $\mathrm{Cr} = 0.1$ in the CFL-type condition that defines the time scale parameter τ (Ying et al. 2004) and $\gamma = 0.5$ s. Empirically, we considered $N_{\mathrm{sim}} = 100$ in the evaluation of (8) and $\mathrm{Fr} = 4.23$.

The initial condition is a steady-state flow with water height $h(x,0) = 0.043$ m and velocity $V(x,0) = 2.737\,\mathrm{m/s}$ for all x. The upstream boundary condition is given by these same values, while the downstream boundary condition for the water depth changes according to $h(14,t) = \min\{0.222, 0.043 + 0.00358\,t\}$. The observed values h_k^{obs} at the points x_k^{obs}, $k = 1, \ldots, N_{\mathrm{obs}}$ were selected from the experimental measurements of Gharangik and Chaudhry (1991). According to Gharangik and Chaudhry (1991) these values correspond to the steady state of the system. However, it is not clear for which value of t_k^{obs} they are obtained. Therefore, in our experiments we consider two possibilities for t_k^{obs}: (i) $t_i^{\mathrm{obs}} = 60$ s for $k = 1, \ldots, N_{\mathrm{obs}}$ and (ii) $t_k^{\mathrm{obs}} = 180$ s for $k = 1, \ldots, N_{\mathrm{obs}}$. The variation of the water depth at the downstream boundary ceases at 50 s. Thus, we may expect that in case (i) the solution is still transient, while the simulation for case (ii) is more likely to match the observed values.

In this experiment, as in all the others that follow, we considered that $\vartheta_k = \vartheta$ for all k, i.e., that all observations were measured with the same instrument. Table 1 shows the results for different values of $\vartheta \in \{0.05, 0.01, 0.005, 0.001\}$. These values were chosen because they were considered to represent plausible values for the error of the observation-measuring instrument. (The units of ϑ correspond to the observations' units of measurement.) In the table, column n_g^* shows the optimal value found for the Yen-Manning coefficient n_g, column σ^* shows the optimal value found for the deviation σ of the random effect, and $\mathcal{L}_\vartheta(\sigma^*, n_g^*)$ corresponds to the optimal likelihood. As a reference, the table also includes (in the column named $n_g|_{\sigma=0}$) the optimal value of n_g that is obtained when the condition $\sigma = 0$ is imposed, as well as the corresponding likelihood $\mathcal{L}_\vartheta(0, n_g|_{\sigma=0})$. These values correspond to the least-squares approximation of n_g. The smaller likelihood values obtained for $t = 60$ s can be explained by the fact that the solution is still transient at this instant, as expected. In the four scenarios on ϑ, the probability that the observed data was generated by the distribution defined by σ^* and n_g^* was higher when $t^{\text{obs}} = 180$ s, compared to the case where $t^{\text{obs}} = 60$ s. So, it is sensible to conclude that the published data were obtained at $t^{\text{obs}} = 180$ s or later.

Let us concentrate on the case defined by $t^{\text{obs}} = 180$ s. As we mentioned above, we made four assumptions on the precision with which the observations were obtained. Note that the estimated σ^* increases when ϑ decreases. This means that, as expected, if the observations are made with maximal precision ($\vartheta = 0.001$ in this case) their probability in the case of the deterministic model ($\sigma = 0$) is smaller than the probability in the case of the stochastic model with $\sigma^* = 0.022$. On the other hand, the probability $\mathcal{L}_\vartheta(\sigma^*, n_g^*)$ decreases very quickly with ϑ. Again, this is the expected behavior as far as the assumption of extremely good precision in observations decreases the probability that observations come from mathematical (obviously inexact) models. These results are illustrated in Figure 2, for $\vartheta = 0.005$ and $\vartheta = 0.001$, where we compare the results of the deterministic case ($\sigma = 0$) with the superposition of all the ($N_{\text{sim}} = 100$) simulations obtained for the optimal parameter σ^*. As an illustration, in this figure we also plot the ensemble mean of the N_{sim} simulations. The total CPU time required for the determination of the roughness coefficient $n_g^* = 0.03959$ and the dispersion parameter $\sigma^* = 0.022$ for $\vartheta = 0.001$ with $n_{\text{sim}} = 100$, $T = 180$ s, $n_{g_{\min}} = 0.039$, $n_{g_{\max}} = 0.040$, $n_{\text{div},n_g} = 100$, $\sigma_{\min} = 0.00$, $\sigma_{\max} = 0.025$, and $n_{\text{div},\sigma} = 25$ was 13,099 s.

Partial dam break

In this numerical experiment, we evaluate the performance of the proposed methodology in a scenario of a partial dam break. In this case, we take as reference the experimental investigation performed by the U.S. Army Corps of Engineers in 1960 (USACE 1960), where it was studied the extent and magnitude of floods induced by the breaching of a 0.305 m (1 ft) high dam, located in the middle of a 121.92 m (400 ft) long and 1.219 m (4 ft) wide model flume with a bed slope of 0.005 and rectangular cross-section. From this investigation, we took the stage-time measurements of Test Condition 11.1, which is characterized by an initial state with the upstream side of the channel full of water and the downstream dry, and by the sudden opening of a 0.732 m (2.4 ft) wide and 0.183 m (0.6 ft) breach, from the top of the dam, at $t = 0$. This test was also used in Ying et al. (2004), in order to verify the robustness of their Upwind scheme, which is employed in this paper.

From the experimental measurements given in Test Condition 11.1 of USACE (1960), we selected a set of observed values of the water height that represents a stage-time hydrograph placed at $x = 68.58$ m, consisting of $N_{\text{obs}} = 10$ observations for $t \leq 20$ s. As a reference, the center of the dam is located at $x = 60.96$ m. In all the simulations, we adopted a uniform mesh of $n_x = 400$ cells, Courant number Cr $= 0.1$ in the CFL-type condition and $\gamma = 0.5$ s. In this experiment, we considered $N_{\text{sim}} = 400$ in the evaluation of (8). The treatment of the dry bed was done as described in Ying et al. (2004), with $h_{\text{dry}} = 10^{-5}$ m. Also, due to the sensitivity of the numerical model to the dry bed treatment, in this numerical experiment we only consider perturbations due to the parameter σ on the flow rate Q (obviously, the water height h is indirectly affected by these perturbations).

The results are shown in Table 2. They indicate that the uncertainties associated with the data are small. The optimal Yen-Manning's coefficient was $n_g^* = 0.02931\,\text{m}^{1/6}$ with zero deviation parameter σ^* for $\vartheta \geq 0.005$, indicating a behavior identical to the deterministic one. For $\vartheta = 0.001$, however, the methodology returned $n_g^* = 0.02927\,\text{m}^{1/6}$ with $\sigma^* = 0.050$. The plots of the solutions obtained for $\vartheta = 0.001$, compared with the observed data, are shown in Figure 3. This figure is a nice illustration of a situation in which the superposition of the simulations provides a better representation of the known (and potentially also the unknown) data when compared to the prediction provided by the deterministic model or by the average of the simulations. Finally, in Figure 4, we show the simulations for $t = 30\,\mathrm{s}$ and $t = 60\,\mathrm{s}$, compared with the measured data also taken from the Test Condition 11.1 of USACE (1960). These results show that the simulations performed with the parameters calculated with the time-stage information obtained at a single spatial point, for $t \leq 20\,\mathrm{s}$, can provide a good prediction of the flood induced by the dam break, at future time instants. The total CPU time required for the determination of the roughness coefficient $n_g^* = 0.02927$ and the dispersion parameter $\sigma^* = 0.050$ for $\vartheta = 0.001$ with $n_{\mathrm{sim}} = 400$, $T = 20\,\mathrm{s}$, $n_{g_{\mathrm{min}}} = 0.0290$, $n_{g_{\mathrm{max}}} = 0.0295$, $n_{\mathrm{div},n_g} = 50$, $\sigma_{\mathrm{min}} = 0.040$, $\sigma_{\mathrm{max}} = 0.050$, and $n_{\mathrm{div},\sigma} = 10$ was 11,536 s.

It must be emphasized that in this experiment the estimates of Yen-Manning's coefficient and variance were obtained using data from a single hydrograph at $x = 68.58\,\mathrm{m}$. Despite this, the forecasts (including uncertainty) shown in Figure 4 were very good at predicting actual data that, in the estimation process, were considered unknown.

East Fork River

In this experiment, we simulate the flood in the 3.3 km flow reach of East Fork River, Wyoming, USA during a part of the high-flow season in late May of 1979, by using the randomly supported one-dimensional shallow water model and several data from the technical reports Emmett et al. (1979) and Meade et al. (1979). A total of 41 sections, ranging from section 0000 to section 3295, were considered. These numbers indicate their center-line distance upstream from $x = 0.0\,\mathrm{m}$. For more details, see Emmett et al. (1979, Fig.1 therein). Thus, in our numerical model, we considered a non-uniform mesh in which each section represents the center of a cell. The simulation period was from 1AM on May 20 to 1PM on May 31, and we considered the mean bed elevations and the mean cross-sections (assumed to be rectangular) measured at 1AM on May 20, taken from Meade et al. (1979, Tables 41 and 42 therein). The discharge values at section 3295, measured at 1AM and 1PM of each day, were used as inflow boundary condition Emmett et al. (1979, Table 7 therein). As outflow boundary condition, we considered the water surface values at section 0000 taken from Emmett et al. (1979, Table 1 therein), also measured at 1AM and 1PM of each day. The water surface measured at 1AM on May 20 was used to set the initial wetted area and the initial flow was assumed to be $8.76\,\mathrm{m}^3/\mathrm{s}$ throughout the entire river. In all simulations ($N_{\mathrm{sim}} = 100$), the Courant number was set to $\mathrm{Cr} = 0.1$ and $\gamma = 256\,\mathrm{s}$. In what follows, we consider that the Yen-Manning coefficient n_g is homogeneous in space, even though the geometric characteristics of the river (width and bed elevation, for example) are heterogeneous.

From Emmett et al. (1979, Table 2 therein) we collected $n_{\mathrm{obs}} = 22$ values of the water surface at section 2505 measured at 1AM and 1PM, from May 21 to May 31. These data, illustrated in Figure 5, represent a stage-time hydrograph at section 2505 and are used as the set of observations in the experiments. We considered two main cases of study: Case A, with one value of n_g for the whole simulation period and Case B, with time-varying n_g. These two cases are described below:

Case A: In this simulation, we proceeded as in the previous experiments and used the set of 22 observations to find the optimal Yen-Manning's coefficient $n_g^* = 0.07749\,\mathrm{m}^{1/6}$ with $\sigma^* = 0.05$ for $\vartheta = 0.01$.

Case B: As reported in Emmett et al. (1979), the width and the bed elevation of the river change in time. Taking this observation into account, in this simulation we subdivide the set of observations in four subsets and we found the optimal Yen-Manning's coefficient for each one of these subsets. The optimal values obtained are: $n_{g1} = 0.09258\,\mathrm{m}^{1/6}$ for observations from 1AM on May 21 to 1PM on May 23; $n_{g2} = 0.07895\,\mathrm{m}^{1/6}$ for 1AM/May 24 to 1PM/May 26; $n_{g3} = 0.06971\,\mathrm{m}^{1/6}$ for 1AM/May 27 to 1PM/May 29; $n_{g4} = 0.07916\,\mathrm{m}^{1/6}$ for 1AM/May 30 to 1PM/May 31. With these values, we compose a time-varying Yen-Manning' coefficient, as illustrated in Figure 6. In this case, the optimal deviation parameter so far obtained was $\sigma^* = 0.033$ for $\vartheta = 0.01$. Notice that, in this case, we assume an abrupt change in the values of n_g on May 23, 26, and 29. The consequences of such an assumption can be appreciated in terms of the random variations estimated by our method in this case. An adequate interpolation in time of the data may be used to describe the sedimentation process's time-varying physics accurately.

Figure 7 displays the results of these experiments. Recall that the elevation data at $x = 2505$ m were used to estimate the parameters n_g^* and σ^*. Therefore these data represent the "training set" of our study. On the other hand, the surface elevation data at $x = 3295$ m were not employed at all in the process of parameter estimation. So, using the machine learning terminology, these data play the role of a "test set" or "validation set".

It is remarkable that in Cases A and B the data from the validation set are recovered with an accuracy similar to that of the training sets. This indicates that training our model using a single hydrograph is enough for obtaining good predictions over all the domain of interest. Moreover, the simulation clouds around deterministic solutions of the Saint-Venant equations seem to provide adequate uncertainty regions for the purpose of taking decisions. As expected, data were better recovered in Case B than in Case A, although the difference between both cases is not very impressive. Finally, the symmetric distribution of random variations involving observations probably indicates that systematic errors in the model are not meaningful in these two cases. The results for the water surface are shown in Figure 8. Once determined the Yen-Manning's coefficients, the total CPU time required for the evaluation of the dispersion parameter σ, with $T = 11.5$ days, $n_{\text{sim}} = 100$, $\sigma_{\min} = 0.00$, $\sigma_{\max} = 0.06$, and $n_{\text{div},\sigma} = 60$ was 23,426 s for Case A and 13,887 s for Case B.

The proposed stochastic model has the potential to deliver better results than the underlying deterministic model in situations where there are no constitutive parameters with which the observed data can be accurately produced by the deterministic model. This may occur because the deterministic model does not adequately describe the real problem or because the observed data contain measurement errors. The non-occurrence of either of these situations would correspond to obtaining an optimal dispersion σ equal to zero. When the optimal σ is strictly positive, the stochastic model is saying that an overlap of simulations better represents the observed data than the deterministic model solution. And the higher the value of the optimal dispersion, the less reliability should be attributed to the deterministic model solution. Graphically, in situations like that, for each x, the deterministic model predicts a value while the stochastic model predicts an interval within which the predicted value may lie. The fact that, as shown in Figure 7, the known observations (both in the training and the test set) are within the predicted interval gives credibility to the prediction. To summarize, the advantage of the stochastic model is to deliver an interval within which the unknown value lies, instead of returning a single prediction without any information about its plausibility.

Spatially heterogeneous roughness

In this last experiment, we proceed as in Ding et al. (2004) and identify the distribution of the Manning coefficient according to a partition of the computation domain of the Fork River into five stretches. The Manning coefficients are assumed to be homogeneous inside each stretch. Thus the roughness parameter structure of the study reach is known, and we aim to identify the roughness values within each partition. A similar study was done in Ayvaz (2013), using equations (1,2).

Unlike the previous example, here we assume steady-state flow and use the cross sections, stream-bed elevations, and water surface elevations on June 28 (discharge at inlet equals to 2.37 m^3/s and water level at the downstream section equals to 5.41 m) obtained from the reports Emmett et al. (1979) and Meade et al. (1979). The set of observations was subdivided into sections, and the respective optimal Yen-Manning coefficients were calculated using each one of these subsets. The sections and the optimal coefficients are: Sections 3168 to 3256, $n_{g1} = 0.07560\,\mathrm{m}^{1/6}$; Sections 2961 to 3108, $n_{g2} = 0.08755\,\mathrm{m}^{1/6}$; Sections 0898 to 2874, $n_{g3} = 0.10698\,\mathrm{m}^{1/6}$; Sections 0220 to 0808, $n_{g4} = 0.08965\,\mathrm{m}^{1/6}$; Sections 0075 to 0137, $n_{g5} = 0.16700\,\mathrm{m}^{1/6}$. The total CPU time required for the evaluation of the dispersion parameter σ, with $T = 1$ day, $n_{\mathrm{sim}} = 100$, $\sigma_{\mathrm{min}} = 0.00$, $\sigma_{\mathrm{max}} = 0.05$, and $n_{\mathrm{div},\sigma} = 50$ was 727 s. The results for the water surface, shown in Figure 9, are in good agreement with the observed data. The comparison with the equivalent Yen-Manning coefficients $n_g = \sqrt{g}n$ from Ding et al. (2004) and Ayvaz (2013) (n are their original values of the Manning coefficients) is presented in Table 3. The results are plausible, and two possible reasons for the differences are the collected data and the simplified cross-sectional geometry (rectangular) used in our modeling.

CONCLUSIONS

One of the most common reasons for using mathematical models is to extract information not directly contained in the data. Except in very rare circumstances, mathematical models, cannot provide such knowledge with absolute certainty. Putting too much faith in model predictions, regardless of their flaws, can lead to fatal judgments. Therefore, models that suggest alternative possibilities for the predicted variables along with the associated probabilities can be useful.

In the fundamental fields of physics, deterministic models are widely known for their accurate predictions. These models often consist of systems of partial differential equations, the numerical solution of which has been the subject of extensive research in the literature. Therefore, it is sensible to rely on these models to generate stochastic counterparts that allow us to make reasonable predictions while accounting for fluctuations and uncertainties. The physical problem under examination in this study was water flux in channels, and the Saint-Venant equations provided the deterministic model on which we built the stochastic counterpart.

Examples from the hydraulic literature were examined to verify the reliability of our approach. These examples demonstrate how effective Saint-Venant equations were in defining a trustworthy underlying deterministic model. Moreover, the stochastic approach's simulations were able to create reasonable bundles of possibilities for unknown variables, including useful confidence intervals and probabilities. In addition, the examples involving Fork River indicated the availability of reasonable bed elevation data is crucial for obtaining reliable predictions. Due to nonlinearity and theoretical intrinsic difficulties, we are not able to determine theoretical properties of the estimators introduced in this paper. Concerning the Probability Density Function, further formalization is necessary which is beyond our present objectives. We plan to address this issue in future research.

In this paper we considered that the model that deserves random perturbation is defined by the discretization of the Saint-Venant differential equations. A different alternative should be to incorporate random perturbations directly on the differential equations employing, perhaps, different methods for their solution. See, for example Man and Tsai (2007). This alternative will be subject of future research. The extension of our approach to 2-D open channel flow and, in fact, to every process governed by evolution equations does not seem to offer specific complications and should be the subject of future research as well.

In the present work, we illustrated the application of the proposed method with stochastic models that had only one constitutive parameter (Manning's coefficient) and one probabilistic parameter (dispersion). The fact that the adjustment of these parameters consisted in solving an optimization problem with only two variables led us to opt for a simple coordinate search algorithm, which turned out to be somewhat costly in terms of computational time. Parallelism (which was not used in this research) could be fully employed with obvious advantage, since simulations could be conducted independently. Its use could decrease the computational time of the presented experiments by at least two orders of magnitude. Besides that, for problems with more than two parameters to be adjusted, the use of more sophisticated optimization algorithms would be recommended. This will be a line of future work.

Finally, it is important to remark that, in the proposed model, the errors due to the numerical solution of the Saint-Venant equations are treated as part of the overall error, with the grid size acting as a constitutive parameter of the model. Thus, the estimated parameters and uncertainty of predictions are related to the discretization and may change with mesh refinement. Numerical results (not presented in the paper) indicate that the likelihood increases with mesh refinement due to the reduction of the approximation error.

Data availability statement: All data, models, and code that support the findings of this study are available from the corresponding author upon reasonable request.

Acknowledgements: This work was supported by FAPESP (grants 2013/07375-0, 2018/24293-0, and 2022/05803-3) and CNPq (grants 304192/2019-8, 302538/2019-4, and 302682/2019-8).

REFERENCES

Agresta, A. and. Baioletti, M., Biscarini, C., Caraffini, F., Milani, A., and Santucci, V. (2021). "Using optimisation meta-heuristics for the roughness estimation problem in river flow analysis." *Applied Sciences*, 11, 10575.

Askar, M. K. and Al-Jumaily, K. K. (2008). “A nonlinear optimization model for estimating Manning’s roughness coefficient.” *Proceedings of the Twelfth International Water Technology Conference, IWTC12*, Alexandria, Egypt, 1299–1306.

Ayvaz, M. T. (2013). “A linked simulation–optimization model for simultaneously estimating the Manning’s surface roughness values and their parameter structures in shallow water flows.” *Journal of Hydrology*, 500, 183–199.

Birgin, E. G. and Martínez, J. M. (2022). “Accelerated derivative-free nonlinear least-squares applied to the estimation of Manning coefficients.” *Computational Optimization and Applications*, 81, 689–715.

Brunner, W. G. (1994). “HEC river analysis system (HEC-RAS).” *Report No. 147*, US Army Corps of Engineers, Hydrologic Engineering Center, Davis, CA, USA.

Chao, W. and Huisheng, S. (2016). “Maximum likelihood estimation for the drift parameter in diffusion processes.” *Stochastics*, 88(5), 699–710.

Chaudhry, M. H. (2022). *Open channel flow*. Springer, 3rd edition.

Cockburn, B. (1999). “Discontinuous galerkin methods for convection-dominated problems.” *High-Order Methods for Computational Physics*, J. T. Barth and H. Deconinck, eds., Vol. 9 of *Lecture Notes in Computational Science and Engineering*, Springer, Berlin, Heidelberg, 69–224.

Correa, M. R. (2017). “A semi-discrete central scheme for incompressible multiphase flow in porous media in several space dimensions.” *Mathematics and Computers in Simulation*, 140, 24–52.

Delgado-Vences, F., Baltazar-Larios, F., Ornelas-Vargas, A., Morales-Bojórquez, E., Cruz-Escalona, V. H., and Salomón Aguilar, C. (2023). “Inference for a discretized stochastic logistic differential equation and its application to biological growth..” *Journal of Applied Statistics*, 50(6), 1231–1254.

Ding, Y., Jia, Y., and Wang, S. S. Y. (2004). “Identification of Manning’s roughness coefficients in shallow water flows.” *Journal of Hydraulic Engineering*, 130(6), 501–510.

Ding, Y. and Wang, S. S. Y. (2005). “Identification of Manning’s roughness coefficients in channel network using adjoint analysis.” *International Journal of Computational Fluid Dynamics*, 19, 3–13.

Ebissa, G. K. and Prasad, K. S. H. (2017). "Estimation of open channel flow parameters by using optimization techniques." *International Journal of Engineering Development and Research*, 5, 1049–1073.

Emmett, W. W., Myrick, R. M., and Meade, R. H. (1979). "Field data describing the movement and storage of sediment in the east fork river, wyoming, part i, *River Hydraulics and Sediment Transport*." *Report No. 1*.

Gharangik, A. M. and Chaudhry, M. H. (1991). "Numerical simulation of hydraulic jump." *Journal of Hydraulic Engineering*, 117, 1195–1211.

Jiang, D. and Shi, N. (2005). "A note on nonautonomous logistic equation with random perturbation." *Journal of Mathematical Analysis and Applications*, 303(1), 164–172.

Kalman, R. E. (1960). "A new approach to linear filtering and prediction problems." *Journal of Basic Engineering*, 82(35-45).

Khan, A. A. and Lai, W. (2014). *Modeling Shallow Water Flows Using the Discontinuous Galerkin Method*. CRC Press, Boca Ratón, London, New York.

Kurganov, A. (2018). "Finite-volume schemes for shallow-water equations." *Acta Numerica*, 27, 289–351.

Lillacci, G. and Khammash, M. (2010). "Parameter estimation and model selection in computational biology." *PLOS Computational Biology*, 6(3), 1–17.

Man, C. and Tsai, C. W. (2007). "Stochastic partial differential equation-based model for suspended sediment transport in surface water flows.." *Journal of Engineering Mechanics*, 133(4), 422–430.

Meade, R. H., Myrick, R. M., and Emmett, W. W. (1979). "Field data describing the movement and storage of sediment in the east fork river, wyoming, part ii, *River Hydraulics and Sediment Transport*." *Report No. 2*.

Panik, M. J. (2017). *Stochastic Differential Equations: An Introduction with Applications in Population Dynamics Modeling*. John Wiley & Sons.

Pappenberger, F., Beven, K., Horrit, M., and Blazkova, S. (2005). "Uncertainty in the calibration of effective roughness parameters in HEC-RAS using inundation and downstream level observations." *Journal of Hydrology*, 302, 46–69.

Rasmussen, C. E. and Williams, C. K. I. (2005). *Gaussian Processes for Machine Learning*. The MIT Press (11).

Román-Román, P., Romero, D., and Torres-Ruiz, F. (2010). "A diffusion process to model generalized von bertalanffy growth patterns: Fitting to real data." *Journal of Theoretical Biology*, 263(1), 59–69.

Saint-Venant, A. J. C. (1871). "Théorie du mouvement non-permanent des eaux, avec application aux crues des rivière at à l'introduction des marées dans leur lit." *Comptes Rendus des Séances de Académie des Sciences*, 73, 147–154.

USACE (1960). "Floods resulting from suddenly breached dams, conditions of minimum resistance, hydraulic model investigation." *Report No. Miscellaneous Paper No. 2-374, Report 1*, U.S. Army Engineer Waterways Experimental Station, Corps of Engineers, Vicksburg, Mississippi, <https://babel.hathitrust.org/cgi/pt?id=mdp.39015095029370&view=1up&seq=3&skin=2021>. Accessed February 11, 2022.

Yen, B. C. (1992). "Dimensionally homogeneous Manning's formula." *Journal of Hydraulic Engineering*, 118(9), 1326–1332.

Yen, B. C. (1993). "Closure to "Dimensionally homogeneous Manning's formula"." *Journal of Hydraulic Engineering*, 119(12), 1443–1445.

Ying, X., Khan, A. A., and Wang, S. Y. (2004). "Upwind conservative scheme for the Saint Venant equations." *Journal of Hydraulic Engineering*, 130, 977–987.

Ziliani, M. G., Ghostine, R., Ait-El-Fquih, B., McCabe, M. F., and Hoteit, I. (2019). "Enhanced flood forecasting through ensemble data assimilation and joint state-parameter estimation." *Journal of Hydrology*, 577, 123924.

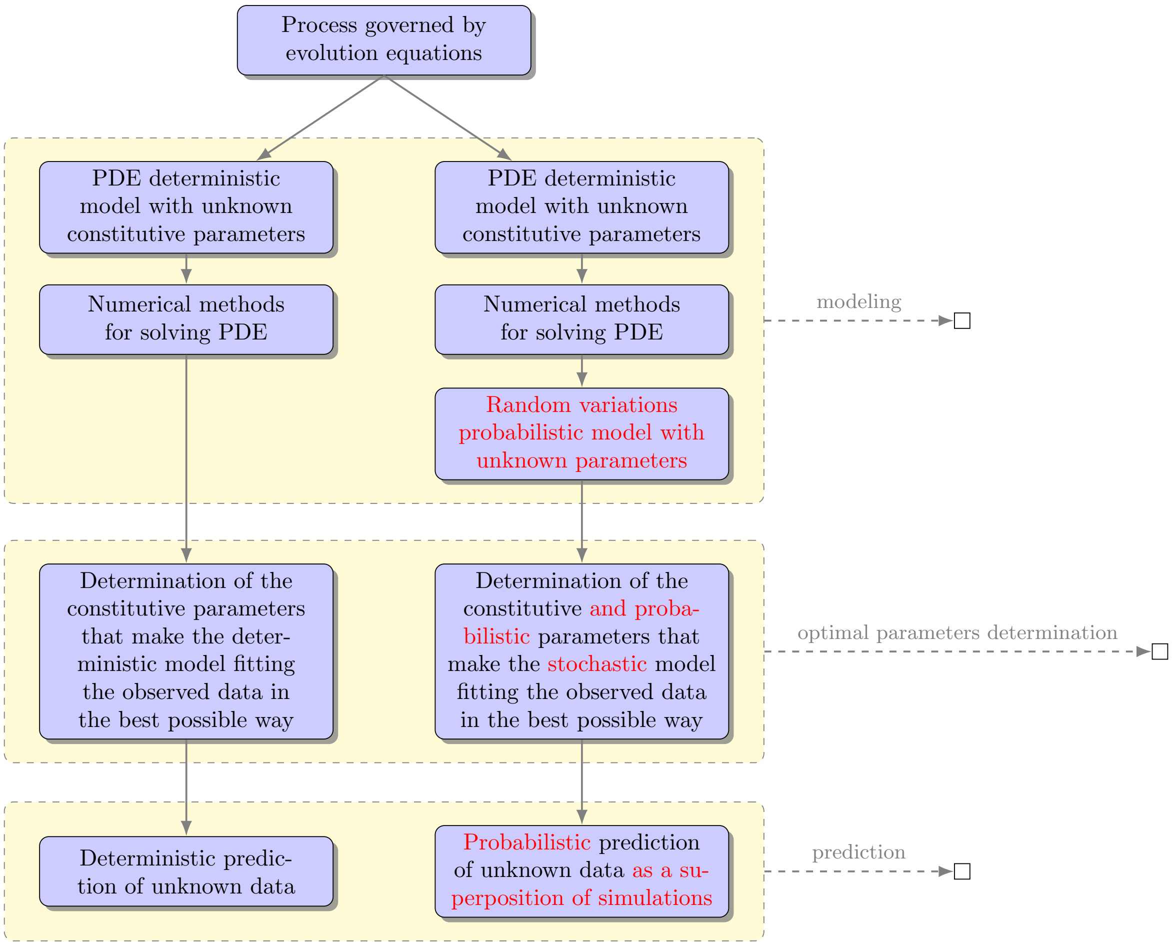


Fig. 1. Sketch representing at a high level the differences between the usual deterministic process and the procedure proposed in the present work. On the left side, the main stages of the usual deterministic procedure are described. On the right side of the figure, the proposed procedure, highlighting the main differences in red, is presented.

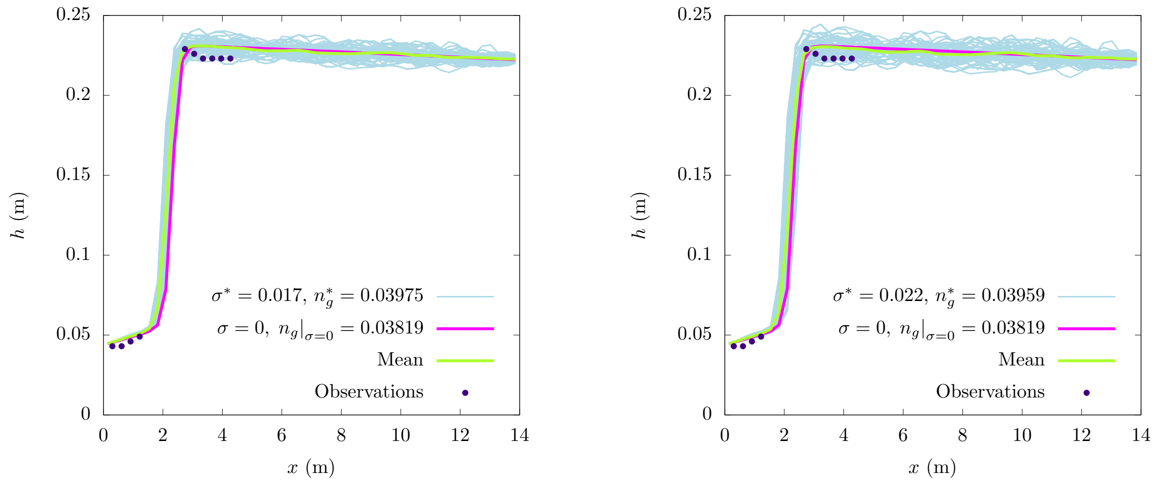


Fig. 2. Simulations of the hydraulic jump problem constructed with the optimal parameters that were obtained with $\vartheta = 0.005$ (left) and $\vartheta = 0.001$ (right), assuming that $t^{\text{obs}} = 180\,\text{s}$. The graphics display the superposition of all the $N_{\text{sim}} = 100$ simulations associated with the optimal Yen-Manning's roughness coefficient n_g^* and deviation parameter σ^*. The pictures also show the least-squares solution, that corresponds to the case $\sigma = 0$, and the ensemble mean of the simulations. The Yen-Manning coefficients have units $\text{m}^{1/6}$.

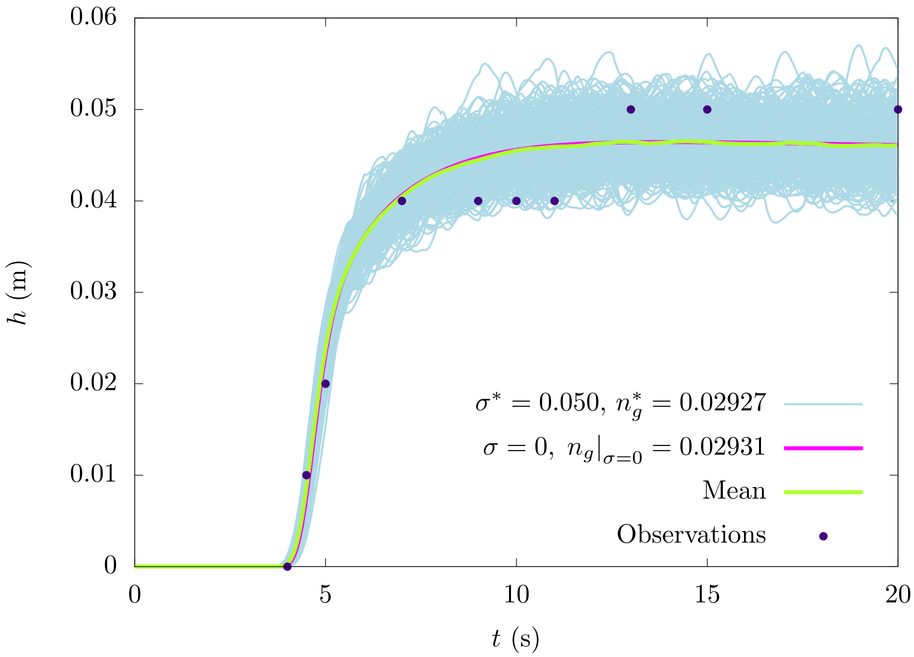


Fig. 3. Simulations of the partial dam break constructed with the optimal parameters that were obtained with $\vartheta = 0.001$. The graphic displays the superposition of all the $N_{\rm sim} = 400$ simulations associated with the optimal Yen-Manning's roughness coefficient n_g^* and deviation parameter σ^*. The pictures also show the deterministic solution, that corresponds to the case $\sigma = 0$ and the ensemble mean of the simulations. In this case the graphic represents an hydrograph at $x = 68.58$ m for $t^{\rm obs} \leq 20$ s. The Yen-Manning coefficients have units $m^{1/6}$.

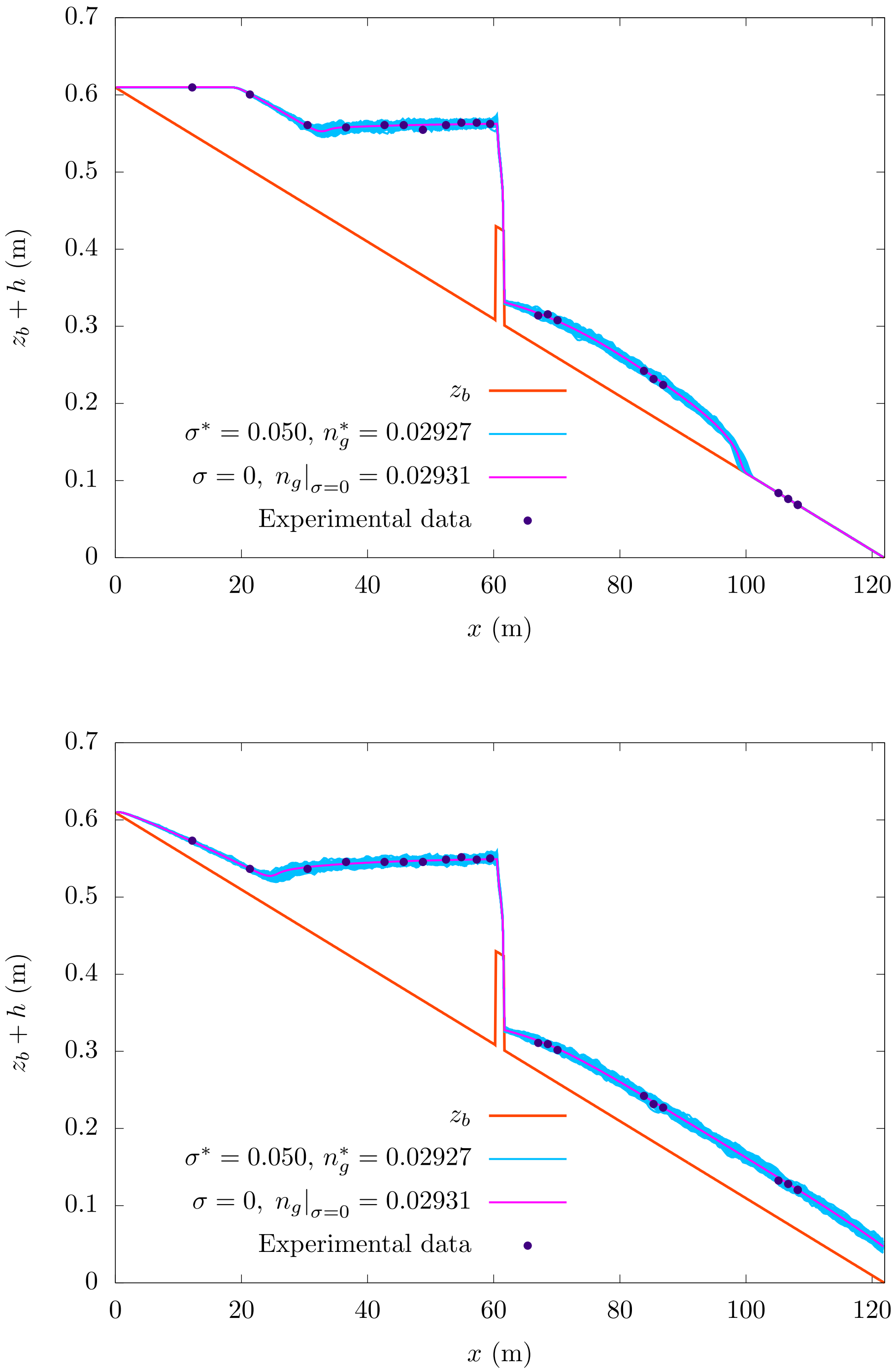


Fig. 4. Simulations of the partial dam break at $t = 30$ s (top) and $t = 60$ s (bottom), constructed with the optimal parameters that were obtained with $\vartheta = 0.001$. The graphics display the superposition of the water level $Z = h + z_b$ of all the $N_{\text{sim}} = 400$ simulations associated with the optimal Yen-Manning's roughness coefficient n_g^* and deviation parameter σ^*. The pictures also show the deterministic solution, that corresponds to the case $\sigma = 0$. In this case, simulations represent predictions, since observed data correspond to $t \leq 20$ s and simulations correspond to $t = 30$ s (top) and $t = 60$ s (bottom). The Yen-Manning coefficients have units $\mathrm{m}^{1/6}$.

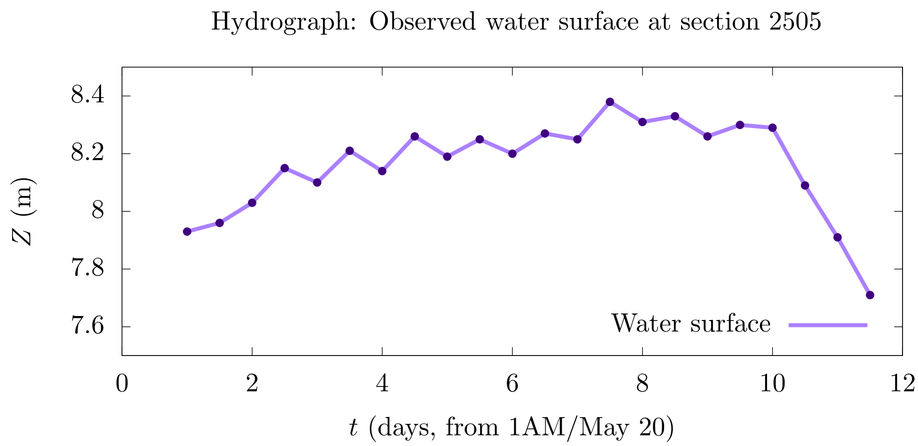


Fig. 5. Water surface at section 2505 measured at 1AM and 1PM, from May 21 to May 31. These $n_{\mathrm{obs}} = 22$ values were used as the set of observations in the experiments.

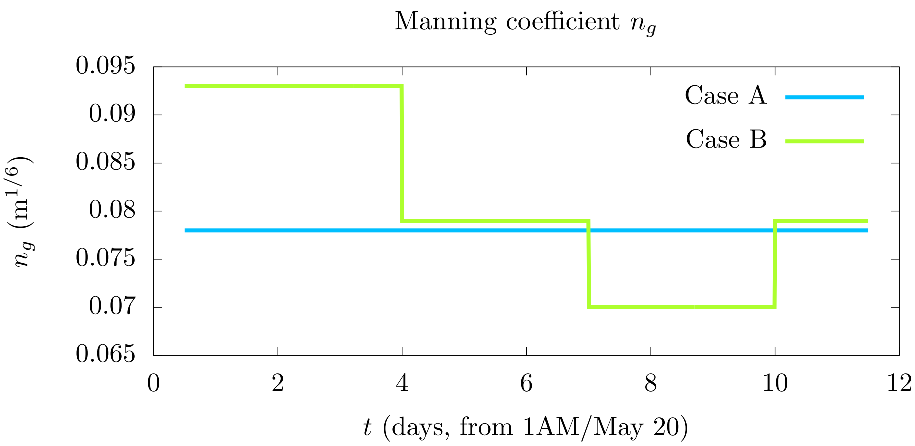


Fig. 6. Yen-Manning's coefficients for the Cases A and B.

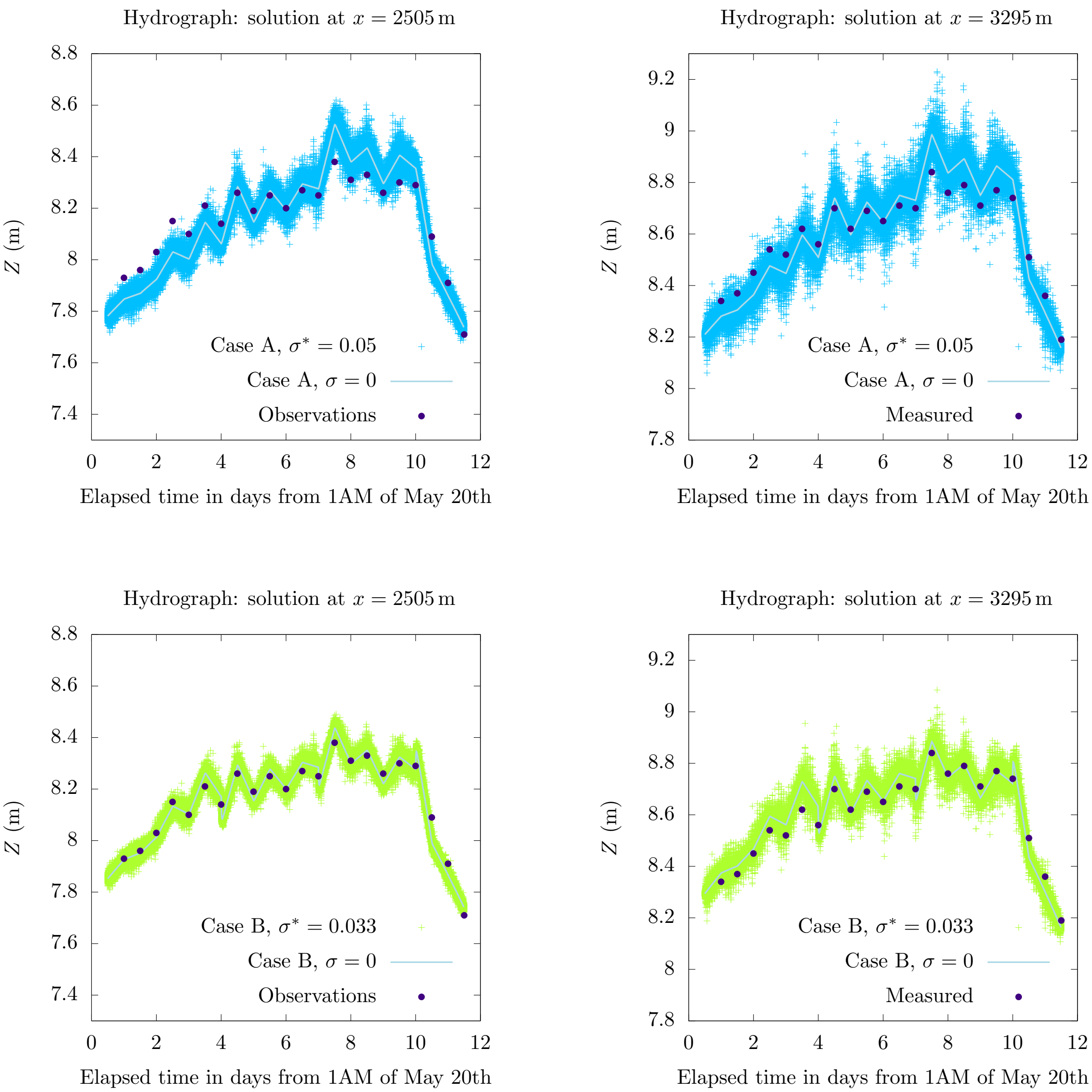
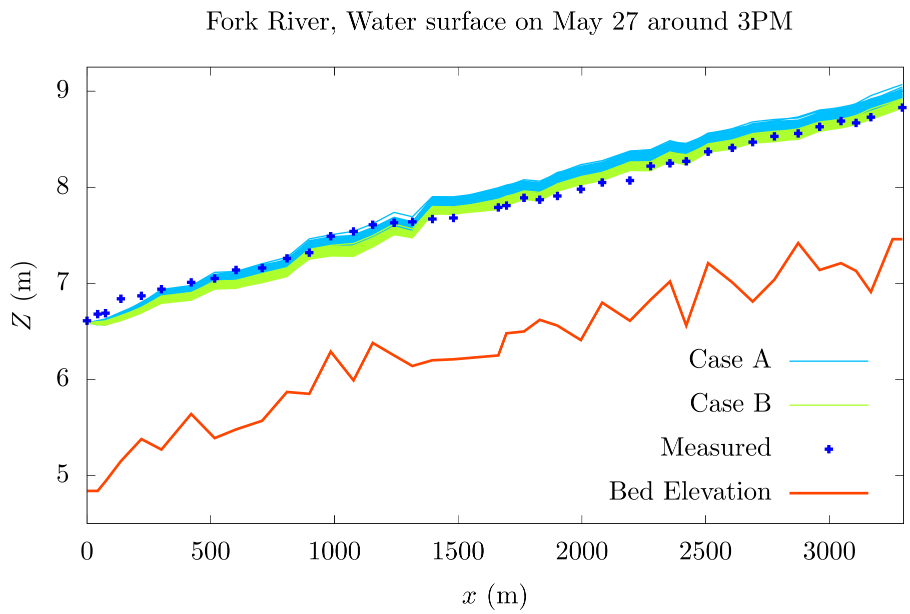


Fig. 7. Simulations of stage-time hydrographs for the Fork river for $\vartheta = 0.01$. The left column contains the computed water surfaces at station 2505 for Cases A (top) and B (bottom) and the observed data. The right column contains predictions at station 3295 for the two cases, compared with measured data (not used for estimating n_g and σ).



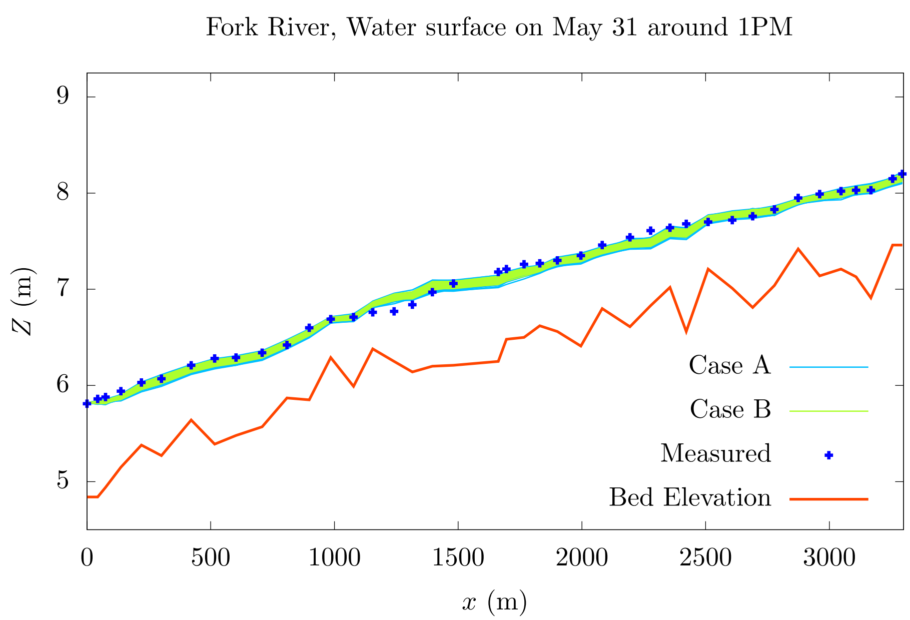


Fig. 8. Cases A and B: Computed and measured water surface around 3PM on May 27 and 1PM on May 31.

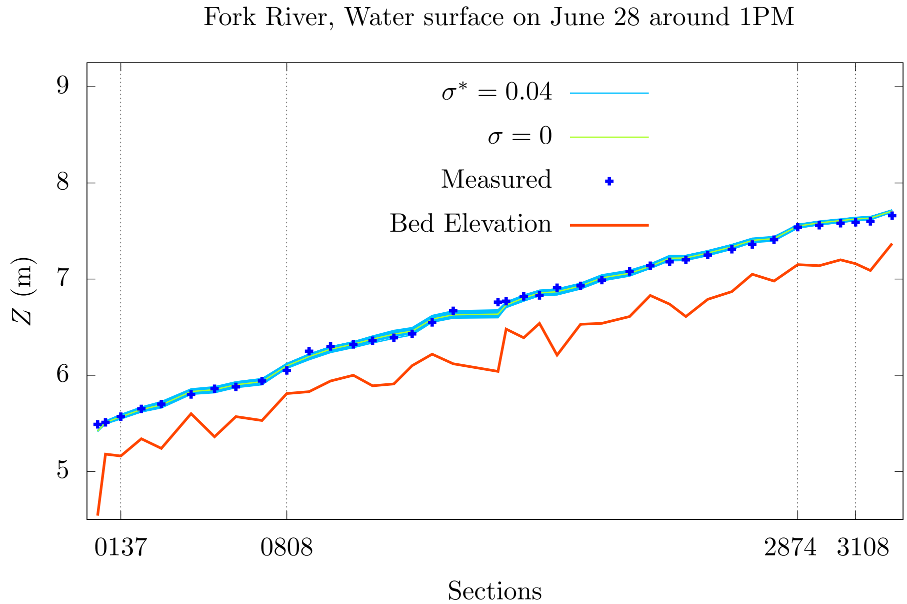


Fig. 9. Heterogeneous case: Computed steady-state solution and measured water surface with the data from June 28. The graphic in blue is the superposition of the water surface of all the $N_{\text{sim}} = 100$ simulations associated with deviation parameter $\sigma^* = 0.04$ and $\gamma = 200\,\text{s}$.

TABLES SECTION (ONE PER PAGE AS REQUIRED)

	ϑ	σ^*	n_g^*	$n_g\|_{\sigma=0}$	$\mathcal{L}_\vartheta(\sigma^*, n_g^*)$	$\mathcal{L}_\vartheta(0, n_g\|_{\sigma=0})$
$t_{\text{obs}} = 60$	0.050	0.000	0.04352	0.04352	9.414×10^{-1}	–
	0.010	0.000	0.04352	0.04352	2.211×10^{-1}	–
	0.005	0.049	0.04330	0.04352	1.570×10^{-2}	2.389×10^{-3}
	0.001	0.048	0.04324	0.04352	1.340×10^{-15}	2.840×10^{-66}
$t_{\text{obs}} = 180$	0.050	0.000	0.03819	0.03819	9.881×10^{-1}	–
	0.010	0.000	0.03819	0.03819	7.417×10^{-1}	–
	0.005	0.017	0.03975	0.03819	3.139×10^{-1}	3.026×10^{-1}
	0.001	0.022	0.03959	0.03819	1.553×10^{-4}	1.053×10^{-13}

TABLE 1. Optimal deviation parameter σ^*, Yen-Manning's coefficient n_g^* (m$^{1/6}$), and likelihood $\mathcal{L}_\vartheta$ obtained for varying values of the precision-related parameter ϑ in the hydraulic jump problem.

ϑ	σ^*	n_g^*	$n_g\big\rvert_{\sigma=0}$	$\mathcal{L}_\vartheta(\sigma^*, n_g^*)$	$\mathcal{L}_\vartheta(0, n_g\big\rvert_{\sigma=0})$
0.050	0.000	0.02931	0.02931	9.941×10^{-1}	–
0.010	0.000	0.02931	0.02931	8.627×10^{-1}	–
0.005	0.000	0.02931	0.02931	5.540×10^{-1}	–
0.001	0.050	0.02927	0.02931	3.575×10^{-5}	3.872×10^{-7}

TABLE 2. Optimal deviation parameter σ^*, Yen-Manning's coefficient n_g^* ($\mathrm{m}^{1/6}$), and likelihood $\mathcal{L}_\vartheta$ obtained for varying values of the precision-related parameter ϑ in the partial dam break problem.

Methodology	n_{g1}	n_{g2}	n_{g3}	n_{g4}	n_{g5}
Ding et al. (2004)	0.18288	0.01532	0.07909	0.12513	0.27923
Ayvaz (2013)	0.12344	0.01854	0.08109	0.13240	0.27882
Present	0.07560	0.08755	0.10698	0.08965	0.16700

TABLE 3. Comparison of optimal Yen-Manning's coefficients $n_g = \sqrt{g}n$ ($\mathrm{m}^{1/6}$) reported in the literature and in the present work.

FIGURE CAPTIONS LIST

Caption of Figure 1: Sketch representing at a high level the differences between the usual deterministic process and the procedure proposed in the present work. On the left side, the main stages of the usual deterministic procedure are described. On the right side of the figure, the proposed procedure, highlighting the main differences in red, is presented.

Caption of Figure 2: Simulations of the hydraulic jump problem constructed with the optimal parameters that were obtained with $\vartheta = 0.005$ (left) and $\vartheta = 0.001$ (right), assuming that $t^{\text{obs}} = 180\,\text{s}$. The graphics display the superposition of all the $N_{\text{sim}} = 100$ simulations associated with the optimal Yen-Manning's roughness coefficient n_g^* and deviation parameter σ^*. The pictures also show the least-squares solution, that corresponds to the case $\sigma = 0$, and the ensemble mean of the simulations. The Yen-Manning coefficients have units $\text{m}^{1/6}$.

Caption of Figure 3: Simulations of the partial dam break constructed with the optimal parameters that were obtained with $\vartheta = 0.001$. The graphic displays the superposition of all the $N_{\text{sim}} = 400$ simulations associated with the optimal Yen-Manning's roughness coefficient n_g^* and deviation parameter σ^*. The pictures also show the deterministic solution, that corresponds to the case $\sigma = 0$ and the ensemble mean of the simulations. In this case the graphic represents an hydrograph at $x = 68.58\,\text{m}$ for $t^{\text{obs}} \leq 20\,\text{s}$. The Yen-Manning coefficients have units $\text{m}^{1/6}$.

Caption of Figure 4: Simulations of the partial dam break at $t = 30\,\text{s}$ (top) and $t = 60\,\text{s}$ (bottom), constructed with the optimal parameters that were obtained with $\vartheta = 0.001$. The graphics display the superposition of the water level $Z = h + z_b$ of all the $N_{\text{sim}} = 400$ simulations associated with the optimal Yen-Manning's roughness coefficient n_g^* and deviation parameter σ^*. The pictures also show the deterministic solution, that corresponds to the case $\sigma = 0$. In this case, simulations represent predictions, since observed data correspond to $t \leq 20\,\text{s}$ and simulations correspond to $t = 30\,\text{s}$ (top) and $t = 60\,\text{s}$ (bottom). The Yen-Manning coefficients have units $\text{m}^{1/6}$.

Caption of Figure 5: Water surface at section 2505 measured at 1AM and 1PM, from May 21 to May 31. These $n_{\text{obs}} = 22$ values were used as the set of observations in the experiments.

Caption of Figure 6: Yen-Manning's coefficients for the Cases A and B.

Caption of Figure 7: Simulations of stage-time hydrographs for the Fork river for $\vartheta = 0.01$. The left column contains the computed water surfaces at station 2505 for Cases A (top) and B (bottom) and the observed data. The right column contains predictions at station 3295 for the two cases, compared with measured data (not used for estimating n_g and σ).

Caption of Figure 8: Cases A and B: Computed and measured water surface around 3PM on May 27 and 1PM on May 31.

Caption of Figure 9: Heterogeneous case: Computed steady-state solution and measured water surface with the data from June 28. The graphic in blue is the superposition of the water surface of all the $N_{\mathrm{sim}} = 100$ simulations associated with deviation parameter $\sigma^* = 0.04$ and $\gamma = 200\,\mathrm{s}$.