

A PDE-informed optimization algorithm for river flow predictions*

E. G. Birgin[†]

J. M. Martínez[§]

February 13, 2023

Abstract

An optimization-based tool for flow predictions in natural rivers is introduced assuming that some physical characteristics of a river within a spatial-time domain $[x_{\min}, x_{\max}] \times [t_{\min}, t_{\text{today}}]$ are known. In particular, it is assumed that the bed elevation and width of the river are known at a finite number of stations in $[x_{\min}, x_{\max}]$ and that the flow-rate at $x = x_{\min}$ is known for a finite number of time instants in $[t_{\min}, t_{\text{today}}]$. Using these data, given $t_{\text{future}} > t_{\text{today}}$ and a forecast of the flow-rate at $x = x_{\min}$ and $t = t_{\text{future}}$, a regression-based algorithm informed by partial differential equations produces predictions for all state variables (water elevation, depth, transversal wetted area, and flow-rate) for all $x \in [x_{\min}, x_{\max}]$ and $t = t_{\text{future}}$. The algorithm proceeds by solving a constrained optimization problem that takes into account the available data and the fulfillment of Saint-Venant equations for one-dimensional channels. The effectiveness of this approach is corroborated with flow predictions of a natural river.

Key words: Flow predictions in natural rivers, Saint-Venant equations, constrained optimization, algorithms.

1 Introduction

Flow prediction in natural rivers is a relevant problem in Hydraulic Engineering because of its impact on agricultural production planning and flood prevention. Moreover, this problem is an interesting challenge for Mathematical Modelling, Numerical Analysis, and Optimization. The main tool regarding the definition of the algorithm introduced in this paper relies on Optimization techniques. We will see that, for each required prediction, a constrained optimization problem must be solved whose complexity is related to the choice of model functions for water elevation.

Most published works for flow prediction in natural rivers are based on the solution of the Saint-Venant equations [17], that require knowledge of the so called Manning roughness coefficients or, simply, Manning coefficients. Since specific Manning coefficients for particular rivers are not known, they need to be estimated using available data. Therefore, the standard procedure consists of applying an optimization method to the minimization, with respect to the Manning coefficients, of the sum of squares of the differences between real observations and observations predicted by the numerical solution of Saint-Venant equations. A lot of research has been devoted to the problem of estimating

*This work was supported by FAPESP (grants 2013/07375-0, 2016/01860-1, and 2018/24293-0) and CNPq (grants 302538/2019-4 and 302682/2019-8).

[†]Department of Computer Science, Institute of Mathematics and Statistics, University of São Paulo, Rua do Matão, 1010, Cidade Universitária, 05508-090, São Paulo, SP, Brazil. e-mail: egbirgin@ime.usp.br

[‡]Corresponding author

[§]Department of Applied Mathematics, Institute of Mathematics, Statistics, and Scientific Computing (IMECC), State University of Campinas, 13083-859, Campinas, SP, Brazil. e-mail: martinez@ime.unicamp.br

Manning parameters using this approach. In [16] the performance of the HEC-RAS software [16] was analyzed and suitable weighting discretization parameters were chosen in order to produce dynamic probability maps of flooding during the event. In [1], HEC-RAS was used in connection with different heuristic procedures for optimizing the Manning coefficient computation. Quasi-Newton methods for the same purpose were used in [10]. Sequential quadratic programming was employed in [4]. Graduate Varied Flow equations and genetic algorithms were used in [13]. In [5], a secant derivative-free optimization tool for determining the Manning coefficient in synthetic experiments was developed.

The approximate solution of Saint-Venant equations requires knowledge of initial conditions and appropriate boundary conditions. Sometimes the approximate solution of these equations is painful due to the need of maintaining very small time steps to avoid numerical instability. In addition, the boundary conditions required for the mathematical solution of the equations are not available in most cases, so artificial boundary conditions are necessary or, equivalently, questionable linear extrapolations on the boundary are employed. Furthermore, as already mentioned, the integration of the Saint-Venant equations requires either prior knowledge of the Manning coefficients, which, in general, is not available, or the estimation of these coefficients using available data. In the latter case, the estimation of the Manning coefficients using an optimization technique requires that the equations be repeatedly solved numerically for different tentative values of the coefficients; see [1, 3, 4, 5, 9, 10, 13, 14, 16].

These inconveniences motivated us to introduce a new method strongly based on available data and informed by the ubiquitous fulfillment of the Saint-Venant equations. The method does not use discretization schemes and it is strongly user-oriented in the sense that provides predictions of flow variables at desired times in the future using forecasts of inlet discharge. The method exploits an empirical relation between inlet discharge and water elevation for each fixed station in the course of the river. This empirical relation allows us to postulate an analytic formula for the transversal area whose coefficients are estimated in order to satisfy Saint-Venant equations as well as possible. This defines a challenging constrained optimization problem, the solution of which is successfully obtained by consolidated constrained optimization software.

The rest of this paper is organized as follows. Section 2 defines the tackled problem. Section 3 introduces the proposed method. Section 4 discusses the optimization problems solved by the main algorithm. Section 5 evaluates the proposed approach applying it to a natural river. Conclusions and lines for future research are given in the last section.

2 Definition of the problem

We consider that a natural river is well represented by state variables defined at spatial positions $x \in [x_{\min}, x_{\max}]$ and time $t \in [t_{\min}, t_{\max}]$. Given data for $t \in [t_{\min}, t_{\text{today}}]$ and $t_{\text{future}} \in (t_{\text{today}}, t_{\max}]$, one wishes to estimate the state variables for all $x \in [x_{\min}, x_{\max}]$ and $t = t_{\text{future}}$. This is the problem that most people need to solve in practice, when t_{today} represents the present instant and they wish to predict what will happen at some point in the future. It is assumed that the river cross-sections are rectangular and that the available data are

- (a) $z_b(x_k)$: bed elevation, in meters, measured from a datum, for a finite number of points $x_k \in [x_{\min}, x_{\max}]$ for $k = 1, \dots, n_{z_b}$.
- (b) $w(x_k)$: width of the river, in meters, for a finite number of points $x_k \in [x_{\min}, x_{\max}]$ for $k = 1, \dots, n_w$.
- (c) $Q(x_{\min}, t_k)$: Inlet discharge (i.e. flow-rate at $x = x_{\min}$), in cubic meters per second, for a finite number of time instants $t_k \in [t_{\min}, t_{\text{today}}]$ for $k = 1, \dots, n_Q$.

- (d) $z(x_k, t_k)$: water elevation, in meters, for a finite number of space-time pairs $(x_k, t_k) \in [x_{\min}, x_{\max}] \times [t_{\min}, t_{\text{today}}]$ for $k = 1, \dots, n_{\text{obs}}$.

Let $x_1 < x_2 < \dots < x_{n_{z_b}}$ be the points for which $z_b(x_k)$ is given. We construct a natural cubic spline interpolating the points $(x_k, z_b(x_k))$ for $k = 1, \dots, n_{z_b}$ and, with abuse of notation, we call this spline $z_b(x)$. Then, from here on, we assume that the bed elevation is given by $z_b(x)$ for all $x \in [x_{\min}, x_{\max}]$. It is worth noting that this may correspond to both interpolation and extrapolation of available data. Analogously, the same is done with the width and inlet discharge. Then, from here on, we assume that $w(x)$ gives the river width for any $x \in [x_{\min}, x_{\max}]$ and that $Q(x_{\min}, t)$ gives the inlet discharge for any $t \in [t_{\min}, t_{\text{today}}]$. Using the available data (a)–(d) and a given forecast of $Q(x_{\min}, t_{\text{future}})$, the goal is to predict

- (e) $z(x, t_{\text{future}})$: water elevation, in meters, for all $x \in [x_{\min}, x_{\max}]$, and

- (f) $Q(x, t_{\text{future}})$: flow-rate, in cubic meters per second, for all $x \in [x_{\min}, x_{\max}]$.

It is worth noting that, since the input data $Q(x_{\min}, t_{\text{future}})$ is a forecast, we may be interested in predicting the state of the system at time t_{future} , given by (e) and (f), for several different forecasts of the inlet discharge $Q(x_{\min}, t_{\text{future}})$.

The Saint-Venant equations [17] are usually employed for river-flow simulations. These equations are given by

$$\frac{\partial A}{\partial t} + \frac{\partial Q}{\partial x} = 0 \quad (1)$$

and

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

for $x \in [x_{\min}, x_{\max}]$ and $t \in [t_{\min}, t_{\max}]$, where $h(x, t) = z(x, t) - z_b(x)$ is the depth of the river at (x, t) , $A(x, t) = h(x, t)w(x)$ is the transversal wetted area at (x, t) , $P(x, t) = w(x) + 2h(x, t)$ is the wetted perimeter at (x, t) , $R(x, t) = A(x, t)/P(x, t)$ is the hydraulics radius at (x, t) , $V(x, t) = Q(x, t)/A(x, t)$ is the speed of the fluid at (x, t) , and g is the acceleration of gravity taken as 9.81 m/s^2 . Equation (1) describes mass conservation and equation (2) represents conservation of the linear momentum. The coefficient n_g is known as Manning roughness coefficient. It is unclear in which way this coefficient depends on x or t . On the one hand, the roughness coefficient depends on x due to the morphological differences of the river along its course. In particular, it should be noted that natural rivers are not rectilinear as assumed by the Saint-Venant equations. On the other hand, sediment deposition can also affect the roughness coefficients over time. In (2), n_g has units $\text{m}^{1/6}$. Following [12] we prefer this definition to the one given by $n_g = \eta_{\text{classic}} \sqrt{g}$ because the units of η_{classic} are $\text{s m}^{-1/3}$.

In the present work, the prediction of $z(x, t_{\text{future}})$ in (e) relies on constructing a functional form for $z(x, t)$. On the one hand, this functional form must fit the available data $z(x_k, t_k)$ for $k = 1, \dots, n_{\text{obs}}$ mentioned in (d). On the other hand, from the functional form of $z(x, t)$, we derive functional forms of $A(x, t)$ and $Q(x, t)$ that satisfy the Saint-Venant equations (1,2) for all $x \in [x_{\min}, x_{\max}]$ and $t = t_{\text{future}}$, in such a way that the functional form of $Q(x, t)$ can be used to make a reliable prediction for $Q(x, t_{\text{future}})$ for all $x \in [x_{\min}, x_{\max}]$ as mentioned in (f).

3 PDE-informed regression

Based in observations of the behavior of natural rivers (see [6, 11, 15]), we postulate a particular dependence of $z(x, t)$ with respect to x and $Q(x_{\min}, t)$ of the form

$$z(x, t) = \mathcal{P}(c, x, Q(x_{\min}, t)), \quad (3)$$

where $c \in \mathbb{R}^{n_{\text{coef}}}$ is a vector of coefficients that need to be determined. Specifically, we will postulate that $n_{\text{coef}} = 9$ and \mathcal{P} assumes the functional form given by

$$\begin{aligned} \mathcal{P}(c, x, Q(x_{\min}, t)) &= c_1 + c_2 Q(x_{\min}, t) + c_3 Q(x_{\min}, t)^2 + \\ &10^{-3}x [c_4 + c_5 Q(x_{\min}, t) + c_6 Q(x_{\min}, t)^2] + \\ &(10^{-3}x)^2 [c_7 + c_8 Q(x_{\min}, t) + c_9 Q(x_{\min}, t)^2]. \end{aligned} \quad (4)$$

Given a trial set of coefficients c , the value of $z(x, t)$ may be computed using (3). Consequently, the values of the state variables $h(x, t)$ and $A(x, t)$ follow using the definitions $h(x, t) = z(x, t) - z_b(x)$ and $A(x, t) = h(x, t) w(x)$. Moreover, the values of the flow-rate $Q(x, t)$ follows from (1) using

$$Q(x, t) = Q(x_{\min}, t) - \int_{x_{\min}}^x \frac{\partial}{\partial t} [(\mathcal{P}(c, \xi, Q(x_{\min}, t)) - z_b(\xi)) w(\xi)] d\xi. \quad (5)$$

(Note that the integral in (5) can be analytically computed using the piecewise definition of both splines $z_b(x)$ and $w(x)$.) It remains the question of how to compute adequate coefficients c taking into account available data $z(x_k, t_k)$ for $k = 1, \dots, n_{\text{obs}}$ and fulfillment of (2). Obviously, the fulfillment of (1) is guaranteed by (5).

On the one hand, the requirement to meet the available data translates into the fulfillment of the equations

$$\mathcal{P}(c, x_k, Q(x_{\min}, t_k)) = z(x_k, t_k) \text{ for } k = 1, \dots, n_{\text{obs}}. \quad (6)$$

At this point, it should be noted that, since we are interested in a physically meaningful estimate for $z(x, t_{\text{future}})$ for all $x \in [x_{\min}, x_{\max}]$, we must include the constraint $\mathcal{P}(c, x, Q(x_{\min}, t_{\text{future}})) \geq z_b(x)$ for all $x \in [x_{\min}, x_{\max}]$ that can be given by

$$\mathcal{P}(c, x_\nu, Q(x_{\min}, t_{\text{future}})) \geq z_b(x_\nu) \quad (7)$$

for all x_ν in a grid of n_{grid} points belonging to $[x_{\min}, x_{\max}]$. Of course, this also assures that $h(x_\nu, t_{\text{future}}) \geq 0$ and $A(x_\nu, t_{\text{future}}) \geq 0$ for all x_ν in the grid. On the other hand, the requirement to satisfy equation (2) could be represented in discrete form as

$$\left. \frac{\partial Q}{\partial t} \right|_{(x_\nu, t_{\text{future}})} + \left[\frac{\partial}{\partial x} \left(\frac{Q^2}{A} \right) + gA \frac{\partial z}{\partial x} \right] \Big|_{(x_\nu, t_{\text{future}})} + \left[\frac{n_g^2 Q |Q|}{AR^{4/3}} \right] \Big|_{(x_\nu, t_{\text{future}})} = 0 \quad (8)$$

for the same x_ν in the grid. Due to the analytic form of Q , given by (5), (8) is a set of n_{grid} algebraic equations. Moreover, the well definiteness of (8) requires that $z_b(x)$ and $w(x)$ should be differentiable and $Q(x_{\min}, t)$ should be twice differentiable. These differentiability requirements are satisfied in the present case since the three mentioned functions are cubic splines. However, considering, as mentioned in the introduction, that $n_g \geq 0$ may depend both on x and t , the equation (8) may be written as

$$\begin{cases} \left. \frac{\partial Q}{\partial t} \right|_{(x_\nu, t_{\text{future}})} + \left[\frac{\partial}{\partial x} \left(\frac{Q^2}{A} \right) + gA \frac{\partial z}{\partial x} \right] \Big|_{(x_\nu, t_{\text{future}})} \leq 0, & \text{if } Q(x_\nu, t_{\text{future}}) \geq 0, \\ \left. \frac{\partial Q}{\partial t} \right|_{(x_\nu, t_{\text{future}})} + \left[\frac{\partial}{\partial x} \left(\frac{Q^2}{A} \right) + gA \frac{\partial z}{\partial x} \right] \Big|_{(x_\nu, t_{\text{future}})} \geq 0, & \text{if } Q(x_\nu, t_{\text{future}}) \leq 0. \end{cases} \quad (9)$$

Assuming $Q(x_\nu, t_{\text{future}}) \neq 0$ for all $\nu \in \{1, \dots, n_{\text{grid}}\}$, (9) is equivalent to

$$\left(\left. \frac{\partial Q}{\partial t} \right|_{(x_\nu, t_{\text{future}})} + \left[\frac{\partial}{\partial x} \left(\frac{Q^2}{A} \right) + gA \frac{\partial z}{\partial x} \right] \Big|_{(x_\nu, t_{\text{future}})} \right) Q(x_\nu, t_{\text{future}}) \leq 0. \quad (10)$$

Moreover if, for some empirical reason, it is known that $Q(x, t_{\text{future}}) > 0$ for all $x \in [x_{\min}, x_{\max}]$, we can replace (10) with

$$Q(x_\nu, t_{\text{future}}) \geq 0 \quad (11)$$

plus

$$\left. \frac{\partial Q}{\partial t} \right|_{(x_\nu, t_{\text{future}})} + \left[\frac{\partial}{\partial x} \left(\frac{Q^2}{A} \right) + gA \frac{\partial z}{\partial x} \right] \Big|_{(x_\nu, t_{\text{future}})} \leq 0 \quad (12)$$

for all $\nu \in \{1, \dots, n_{\text{grid}}\}$.

Equations and inequations (6,7,10) or, alternatively, (6,7,11,12) define a system of nonlinear algebraic equalities and inequalities. Due the arbitrary nature of the functional form imposed to $z(x, t)$ in (3), they may be incompatible. Therefore, we consider the constrained optimization problem given by

$$\text{Minimize}_{c \in \mathbb{R}^{n_{\text{coeff}}}} \sum_{k=1}^{n_{\text{obs}}} (\mathcal{P}(c, x_k, Q(x_{\min}, t_k)) - z(x_k, t_k))^2 \quad (13)$$

subject to (7,10) or subject to (7,11,12). It is worth noticing that the objective function as well as the constraints of these optimization problems are all differentiable with respect to the unknowns c .

4 Constrained Optimization problems

For solving the problems (13,7,10) or (13,7,11,12) we used Algencan. This is a well-established software based on Safeguarded Augmented Lagrangians (SAL) for solving possibly large-scale constrained optimization problems; see [2, 7]. In [8], complexity results for Algencan were proved which consider even the case of unbounded penalty parameters. Moreover, in that paper extensive numerical results were reported. We use the name Algencan for both the SAL algorithm and the Fortran subroutine that implements it, which is freely available at <http://www.ime.usp.br/~egbirgin/tango/>. Of course, the mathematical algorithm described in [2, 7, 8] may be implemented by means of different subroutines. We employed a version of Algencan that needs only first derivatives. First derivatives of the objective function (13) and of the constraints (7,10) or (7,11,12) are available because, due to (5), differentiable analytic expressions of the objective function and of the constraints are available. It should be noted, however, that coding analytic derivatives in an efficient and economic way is not free of cumbersome difficulties. Anyway, we developed efficient codes for this purpose, that may be easily adapted in the case of alternative definitions of \mathcal{P} .

For simplicity, let us concentrate in problem (13,7,11,12), which has objective function (13) and constraints (7), (11), and (12). The variables of this problem are the coefficients $c_1, \dots, c_{n_{\text{coeff}}}$. Therefore, the number of variables n depends on the choice of the approximating function \mathcal{P} . In the present work \mathcal{P} is given by (4), so that $n = n_{\text{coeff}} = 9$. The number of constraints is $m = 3n_{\text{grid}}$. Although the number of variables is small, we should be ready to employ more complex approximating functions \mathcal{P} , in which n_{coeff} could be big.

The requirements (7), (11), and (12) define algebraic constraints for each point $x_1, \dots, x_{n_{\text{grid}}}$ in a grid contained in $[x_{\min}, x_{\max}]$. The presence of multiple local minimizers forced us to define an opportunistic two-phases scheme for using Algencan. In the first phase, starting from the initial approximation $c_1 = \dots = c_{n_{\text{coeff}}} = 0$, we solved the optimization problem disregarding the constraints (12). In the second phase, starting from the final point obtained in the first case, we incorporated the constraints (12) and we run again Algencan in order to obtain a suitable solution. This process is repeated from scratch each time a prediction for a different instant t_{future} is sought. As the starting point of the first phase, instead of $c_1 = \dots = c_{n_{\text{coeff}}} = 0$, we could have used the unconstrained minimizer of (13),

which corresponds to a linear least squares problem. However, that supposedly better starting point brought no additional advantage.

In both phases, we used Algencan with all its default parameters. These parameters include tolerances $\varepsilon_{\text{feas}} > 0$ and $\varepsilon_{\text{opt}} > 0$ for feasibility and optimality, respectively. If we define

$$\begin{aligned} \mathbf{f}(c) &= \sum_{k=1}^{n_{\text{obs}}} (\mathcal{P}(c, x_k, Q(x_{\min}, t_k)) - z(x_k, t_k))^2 \\ \mathbf{g}_\nu(c) &= z_b(x_\nu) - \mathcal{P}(c, x_\nu, Q(x_{\min}, t_{\text{future}})) \text{ for } \nu \in \{1, \dots, n_{\text{grid}}\} \\ \mathbf{g}_{n_{\text{grid}}+\nu}(c) &= -Q(x_\nu, t_{\text{future}}) \text{ for } \nu \in \{1, \dots, n_{\text{grid}}\} \\ \mathbf{g}_{2n_{\text{grid}}+\nu}(c) &= \left. \frac{\partial Q}{\partial t} \right|_{(x_\nu, t_{\text{future}})} + \left[\frac{\partial}{\partial x} \left(\frac{Q^2}{A} \right) + gA \frac{\partial z}{\partial x} \right] \Big|_{(x_\nu, t_{\text{future}})} \text{ for } \nu \in \{1, \dots, n_{\text{grid}}\} \end{aligned}$$

then the problem of the second phase is given by

$$\text{Minimize } \mathbf{f}(c) \text{ subject to } \mathbf{g}_j(c) \leq 0 \text{ for } j = 1, \dots, 3n_{\text{grid}}.$$

The Algencan stopping criterion for this particular problem corresponds to finding $c \in \mathbb{R}^n$ and $\mu \in \mathbb{R}^m$ (Lagrange multipliers) with $n = n_{\text{coeff}}$ and $m = 3n_{\text{grid}}$ such that

$$\begin{aligned} \left\| \nabla \mathbf{f}(c) + \sum_{j=1}^m \mu_j \nabla \mathbf{g}_j(c) \right\|_\infty &\leq \varepsilon_{\text{opt}}, \\ \mathbf{g}_j(c) &\leq \varepsilon_{\text{feas}} \text{ for } j = 1, \dots, m, \\ |\mathbf{g}_j(c) \mu_j| &\leq \varepsilon_{\text{feas}} \text{ for } j = 1, \dots, m. \end{aligned}$$

In the numerical experiments, we considered $\varepsilon_{\text{feas}} = 10^{-8}$ and $\varepsilon_{\text{opt}} = 10^{-4}$, which are “typical values” in solving problems that are “well scaled” and for which only first order derivatives are available.

The entire process is coded in the freely available code ALGENFLOW code, improved versions of which will be posted on the website <http://www.ime.usp.br/~egbirgin/> as soon as they become available. Communication with potential users will be vital for making this software increasingly useful and friendly.

5 Numerical corroboration with real data

The development of ALGENFLOW is motivated by the objectives of CRIAB (acronym for “Dams Conflicts, Risks and Impacts” in Portuguese), a research group of the State of São Paulo, in Brazil, aimed at investigating, understanding, and mitigating the consequences of technological disasters caused by the rupture of dams, which, unfortunately, are occurring both in Brazil and in the rest of the world with increasing frequency.

In order to assess the reliability of ALGENFLOW, we addressed the prediction of the state variables of the East Fork River, a tributary of approximately 45 miles (72 km) of the New Fork River in the U.S. state of Wyoming, for which it is possible to find the necessary data in the literature; see [11, 15]. Specifically, the data extracted from [11, 15] correspond, for a time period in 1979, to 39 values of $z_b(x)$ and $w(x)$ at the same non-equidistant points ranging from $x = 0$ to $x = 3213$ m and to 94 values of $Q(x, t)$ corresponding to $x = -39$ m and values of t starting at $t = 0$ and increasing at intervals of 12 hours. We extracted still 122 observations of $z(x, t)$. Of these 122 observations, 60 correspond to $x = 751$ m and t starting at hour zero on day 3 and increasing by 12 hours. The other 62 observations correspond to $x = 3256$ m, starting at the same time instants and with the same interval. With these

data, we defined $x_{\min} = -39$ m and $x_{\max} = 3256$ m. We also defined $t_{\min} = 3$ days and, assuming it would not make much sense to make predictions without using at least 5 days of observations, we stipulated that t_{future} could vary from day 8 to day 32. Figures 1, 2, and 3 show $z_b(x)$ and $w(x)$ for $x \in [x_{\min}, x_{\max}]$ and $Q(x, t)$ for $x = -39$ m and t varying from day 3 to day 33, respectively. The figures display the data as well as the computed splines. Figure 4 displays the available water level observations.

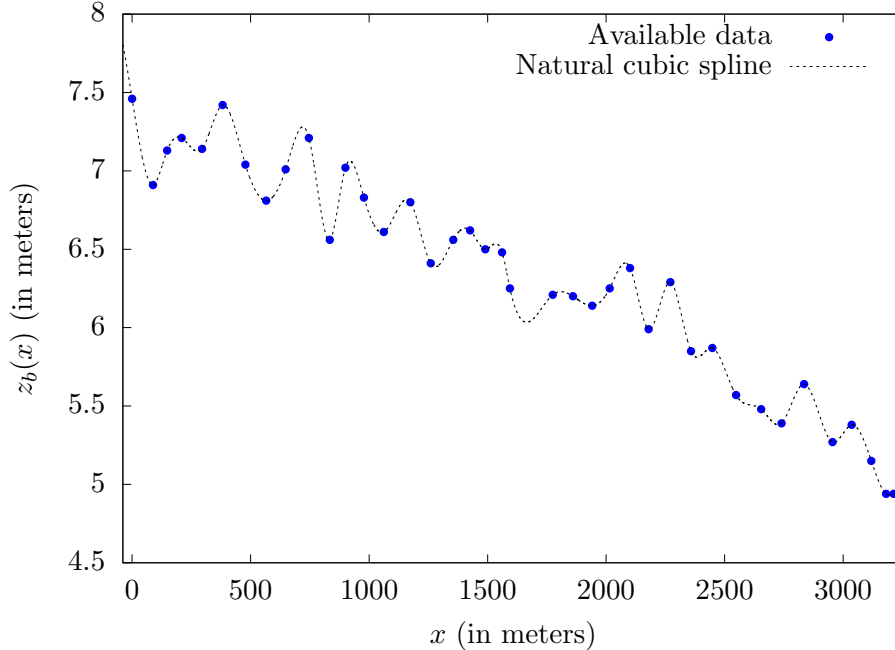


Figure 1: Available data of the East Fork River bed elevation $z_b(x)$ and computed natural cubic spline for $x \in [x_{\min}, x_{\max}]$.

In these experiments we considered the objective function (13), the constraints (7,11,12), and the first derivatives of both in Fortran 90. We used the version 3.1.1 of Algencan.

All tests were conducted on a computer with a 5.2 GHz 12th Gen Intel(R) Core(TM) i9-12900K processor and 128GB 3200 MHz DDR4 RAM memory, running Ubuntu 22.04.1 LTS. Code was compiled by the GFortran compiler of GCC (version 11.3.0) with the `-O3` optimization directive enabled. The program reads the aforementioned required data (numbered from (a) to (d) in Section 2) from files and calculates the corresponding splines. Then it asks for values of x_{\min} , x_{\max} , t_{\min} , t_{today} , t_{future} , and a forecast for $Q(x_{\min}, t_{\text{future}})$. In all the experiments, considering the available data, we used $x_{\min} = -39$ m, $x_{\max} = 3256$ m, $t_{\min} = 3$ days. We arbitrarily set $n_{\text{grid}} = 100$ and used, as forecast for $Q(x_{\min}, t_{\text{future}})$, the available data. In a first experiment, we consider $t_{\text{today}} = 10$ days and t_{future} varying from day 11 to day 33. This determines $33 - 11 + 1 = 23$ different optimization problems. In each of these problems, we used the water level observations available between day t_{\min} and day t_{today} as training data to find the best value of the coefficients c such that the prediction for day t_{future} has physical sense by satisfying the imposed constraints and agrees as well as possible with the training data. The observations of the water level for day t_{future} are not used in the optimization process and are only used as test data to verify the accuracy of the predictions. Table 1 shows the results. Mean absolute and relative error of the predictions are, respectively, $1.69\text{E-}01$ meters and $2.12\text{E-}02$ for $x = 751$ m and $5.00\text{E-}02$ meters and $8.47\text{E-}03$ for $x = 3256$ m. Figure 5 shows in a graphical

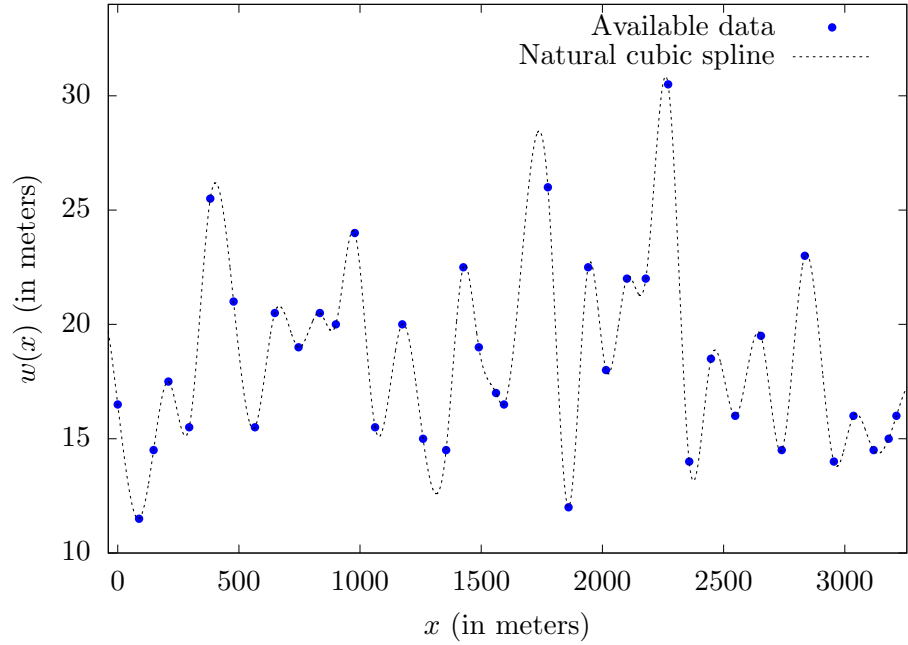


Figure 2: Available data of the East Fork River width $w(x)$ and computed natural cubic spline for $x \in [x_{\min}, x_{\max}]$.

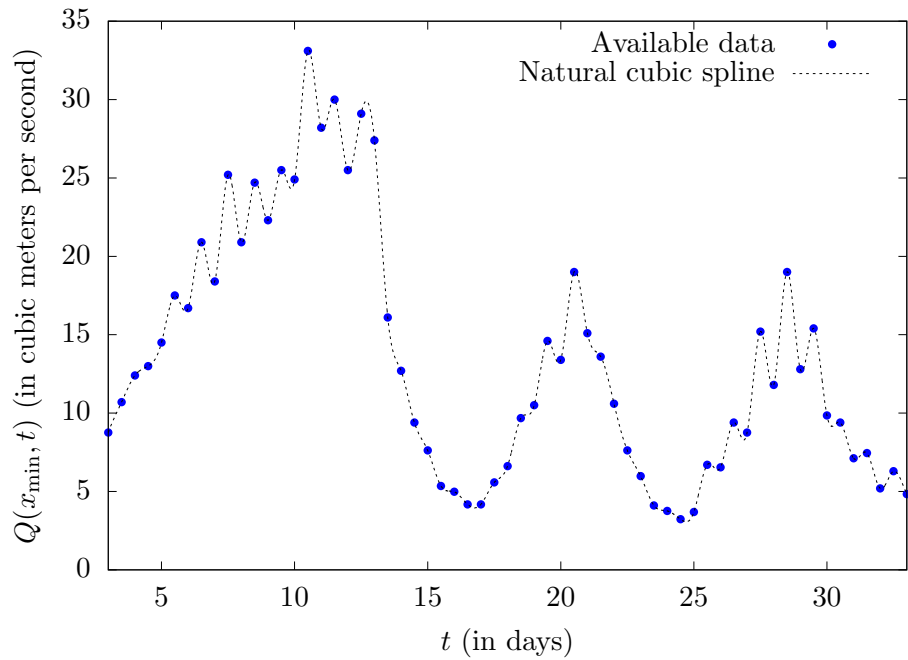


Figure 3: Available data of the East Fork River inlet discharge $Q(x_{\min}, t)$ for t ranging from day 3 to day 33.

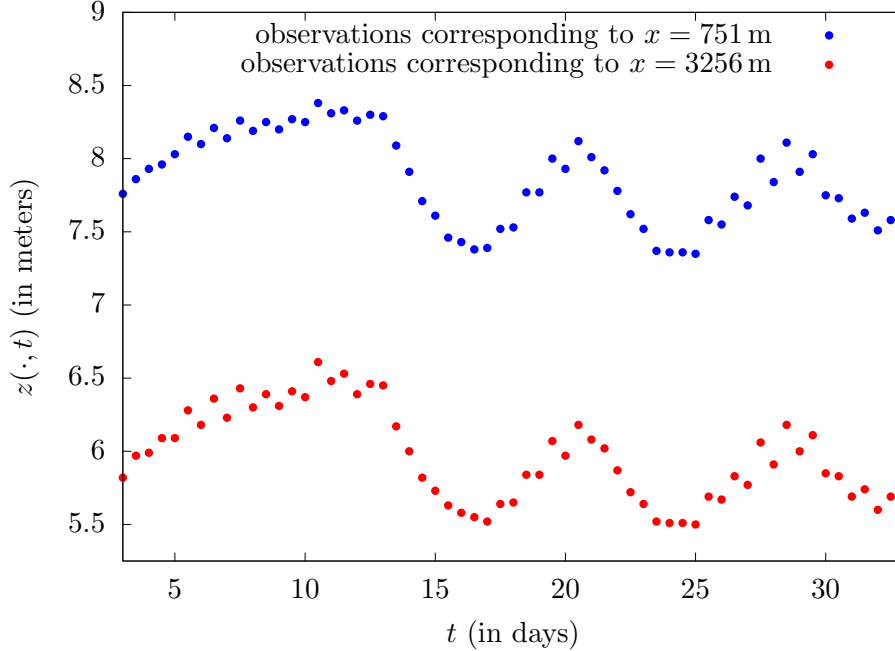


Figure 4: Available observations of the water level $z(x, t)$ for $x = 751$ m (in blue) and $x = 3256$ m (in red). For the first case there are 60 observations starting at hour zero on day 3 at intervals of 12 hours. For the latter case, there are 62 observations starting at the same time instants and with the same interval.

form the same information presented in Table 1. As an illustration, Figures 6 and 7 show, for all $x \in [x_{\min}, x_{\max}]$, the predicted value of the water height $h(x, t_{\text{future}})$ and flow-rate $Q(x, t_{\text{future}})$ for day $t_{\text{future}} = 14$. Massive numerical experiments, varying t_{today} between 7 and 32 and t_{future} between $t_{\text{today}} + 1$ and 33 showed similar results in all cases. As a whole, we solved 351 problems similar to the one described above. The solution of each problem (including phases 1 and 2) took, on average, 15.52 seconds in the described computational environment. The results were in all cases analogous to the one described and the mean absolute and relative error of the predictions were 8.18E-02 meters and 1.16E-02, respectively.

6 Conclusions

In this work, we presented a new algorithm for flow predictions in natural rivers based on an arbitrary number of past observations and ubiquitous fulfillment of Saint-Venant equations. This defined a challenging constrained optimization problem, whose solution was successfully obtained by employing Algencon.

As a consequence, we developed the software ALGENFLOW, in a fully user-oriented form and we tested it using real data, obtaining satisfactory results. We took into account the specific need of the users and the inherent limitations to the feasibility of reliable predictions. For example, when we talk about inflow forecasts, we are aware that, in general, the accuracy of such forecasts is quite uncertain, perhaps with a degree of uncertainty of no less than 10%. Similar levels of uncertainty appear for level and width of the river, apart from the fact that the cross-sectional areas are often far from rectangular. These warnings are necessary to understand that high precisions in the prediction

t_{future}	$x = 751 \text{ m}$				$x = 3256 \text{ m}$			
	observation	prediction	absolute	relative	observation	prediction	absolute	relative
11	8.31	9.04	7.34E-01	8.84E-02	6.48	6.39	-8.66E-02	-1.34E-02
12	8.26	8.85	5.86E-01	7.09E-02	6.39	6.35	-3.95E-02	-6.18E-03
13	8.29	9.04	7.51E-01	9.06E-02	6.45	6.38	-6.99E-02	-1.08E-02
14	7.91	7.95	4.12E-02	5.21E-03	6.00	6.04	3.70E-02	6.17E-03
15	7.61	7.70	8.63E-02	1.13E-02	5.73	5.77	4.21E-02	7.36E-03
16	7.43	7.56	1.29E-01	1.73E-02	5.58	5.61	2.96E-02	5.31E-03
17	7.39	7.52	1.32E-01	1.78E-02	5.52	5.55	3.35E-02	6.07E-03
18	7.53	7.64	1.06E-01	1.40E-02	5.65	5.71	6.13E-02	1.08E-02
19	7.77	7.85	8.01E-02	1.03E-02	5.84	5.93	9.06E-02	1.55E-02
20	7.93	7.98	5.29E-02	6.68E-03	5.97	6.07	9.82E-02	1.64E-02
21	8.01	8.10	9.32E-02	1.16E-02	6.08	6.14	5.54E-02	9.11E-03
22	7.78	7.85	7.41E-02	9.53E-03	5.87	5.93	6.34E-02	1.08E-02
23	7.52	7.61	8.53E-02	1.13E-02	5.64	5.67	3.10E-02	5.50E-03
24	7.36	7.50	1.43E-01	1.94E-02	5.51	5.54	2.65E-02	4.81E-03
25	7.35	7.50	1.46E-01	1.98E-02	5.50	5.52	2.18E-02	3.96E-03
26	7.55	7.63	8.17E-02	1.08E-02	5.67	5.71	3.92E-02	6.91E-03
27	7.68	7.76	8.11E-02	1.06E-02	5.77	5.84	6.72E-02	1.16E-02
28	7.84	7.91	7.11E-02	9.07E-03	5.91	6.00	8.51E-02	1.44E-02
29	7.91	7.96	4.53E-02	5.73E-03	6.00	6.04	4.16E-02	6.93E-03
30	7.75	7.82	6.56E-02	8.47E-03	5.85	5.90	4.69E-02	8.03E-03
31	7.59	7.67	7.63E-02	1.01E-02	5.69	5.74	5.25E-02	9.22E-03
32	7.51	7.57	5.79E-02	7.71E-03	5.60	5.63	2.60E-02	4.65E-03
33					5.60	5.60	4.29E-03	7.66E-04

Table 1: In the table, “observation” correspond to an observation of the water level $z(x, t)$ for $x = 751 \text{ m}$ or $x = 3256 \text{ m}$ and $t = t_{\text{future}}$ days, “prediction” corresponds to the prediction of the observed data, “absolute” corresponds to the difference prediction minus observation, and “relative” corresponds to the same difference divided by the observed value. Note that observed data $z(x, t)$ with $t > 10$ are being used in the table to be contrasted with the predictions but were not used for the prediction calculation, which used only observed data $z(x, t)$ with $t \leq 10$.

of the state variables are meaningless.

One of the objectives of this paper is to bring developers of optimization methods in contact with relevant hydraulic prediction problems. The optimization problems at the core of the ALGENFLOW algorithm define an interesting family for testing optimization techniques.

ALGENFLOW has been written in such a way that its application to real-life situations is quite accessible. Therefore, we expect a strong interaction with users who can provide us with useful feedback on the effectiveness of ALGENFLOW in real cases. Nevertheless, according to our own experience, many improvements of ALGENFLOW are expected in the near future. Let us mention a few of them.

1. Different alternatives with respect to the definition of \mathcal{P} need to be defined and tested. The impact of the new definitions on the difficulty of the resulting nonlinear programming problems, and on Algencan’s ability to solve them, needs to be analyzed.
2. The assumption regarding rectangular cross-areas may be excessively strict in some cases. The software should be adapted to consider more arbitrary shape areas.

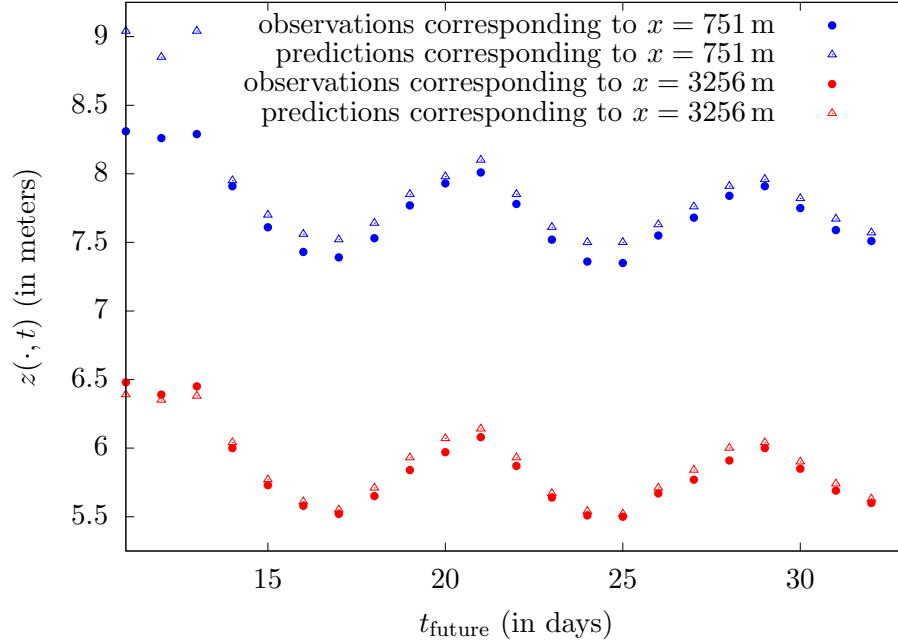


Figure 5: Observations and predictions of the water level $z(x, t)$ for $x = 751$ m (in blue), $x = 3256$ m (in red) and $t_{\text{future}} \in \{11, 12, \dots, 33\}$. The predictions used available data until $t_{\text{today}} = 10$ days. The observations available for $t > 10$ were not used in the prediction process and are just being used in the plot so that they can be compared with the predictions.

3. The variability of Manning’s coefficients requires accounting for causality with respect to the physical parameters.
4. Erosion and deposition of sediments in channels should be reflected in the main analytic assumptions concerning the definition of \mathcal{P} .
5. Tests are needed to evaluate the reliability of the new schemes in the presence of extreme climate events.
6. Bi-dimensional Saint-Venant models should be addressed under the optimization proposal presented in the present paper.

Declaration

Ethical Approval

Not applicable.

Data availability

The datasets generated during and/or analysed during the current study are available from the corresponding author on reasonable request.

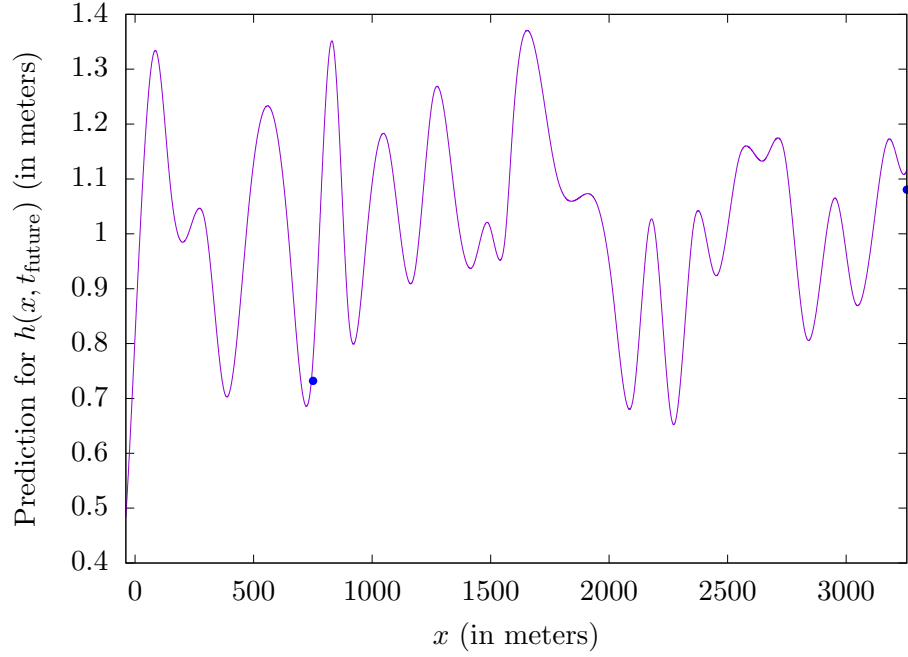


Figure 6: Prediction of $h(x, t_{\text{future}})$ for all $x \in [x_{\text{min}}, x_{\text{max}}]$ for day $t_{\text{future}} = 14$ made on day $t_{\text{today}} = 10$. The two blue dots in the figure represent available observations (for $x = 751$ m and $x = 3256$ m) being used as testing data only; i.e. being considered unknown in the prediction process.

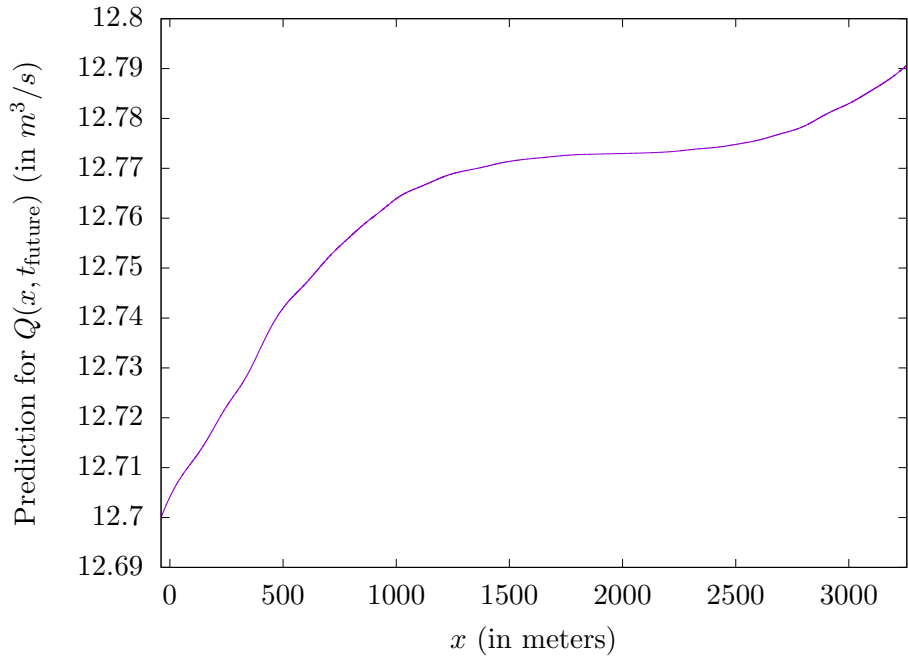


Figure 7: Prediction of $Q(x, t_{\text{future}})$ for all $x \in [x_{\text{min}}, x_{\text{max}}]$ for day $t_{\text{future}} = 14$ made on day $t_{\text{today}} = 10$.

Competing interests

Author J. M. Martínez is a member of the journal’s editorial board.

Funding

This work was supported by FAPESP (grants 2013/07375-0, 2016/01860-1, and 2018/24293-0) and CNPq (grants 302538/2019-4 and 302682/2019-8).

Authors’ contributions

The two authors of this work worked together and equally in all its stages.

Acknowledgments

Not applicable at this time.

References

- [1] A. Agresta, M. Baiocchi, C. Biscarini, F. Caraffini, A. Milani, and V. Santucci, Using optimisation meta-heuristics for the roughness estimation problem in river flow analysis, *Applied Sciences* 11, 10575, 2021.
- [2] R. Andreani, E. G. Birgin, J. M. Martínez and M. L. Schuverdt, On Augmented Lagrangian methods with general lower-level constraints, *SIAM Journal on Optimization* 18, pp. 1286–1309, 2008.
- [3] M. T. Ayvaz, 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, pp. 183–199, 2013.
- [4] M. Kh. Askar and K. K. Al-jumaily, A nonlinear optimization model for estimating Manning’s roughness coefficient, in *Proceedings of the Twelfth International Water Technology Conference, IWTC12*, Alexandria, Egypt, 2008, pp. 1299–1306.
- [5] E. G. Birgin and J. M. Martínez, Accelerated derivative-free nonlinear least-squares applied to the estimation of Manning coefficients, *Computational Optimization and Applications* pp. 689–715, 2022.
- [6] E. G. Birgin, M. R. Correa, V. González-López, J. M. Martínez, and D. S. Rodrigues, Randomly supported models for the prediction of flows in channels, *submitted*.
- [7] E. G. Birgin and J. M. Martínez, *Practical Augmented Lagrangian Methods for Constrained Optimization*, Society for Industrial and Applied Mathematics, Philadelphia, 2014.
- [8] E. G. Birgin and J. M. Martínez, Complexity and performance of an Augmented Lagrangian algorithm, *Optimization Methods and Software* 35, pp. 885–920, 2020.
- [9] Y. Ding, Y. Jia, and S. S. Y. Wang, Identification of Manning’s roughness coefficients in shallow water flows, *Journal of Hydraulic Engineering*, pp. 501–510, 2004.

- [10] Y. Ding and S. S. Y. Wang, Identification of Manning’s roughness coefficients in channel network using adjoint analysis, *International Journal of Computational Fluid Dynamics* 19, pp. 3–13, 2005.
- [11] W. W. Emmett, W. W. Myrick, and R. H. Meade, Field data describing the movement and storage of sediment in the East Fork River, Wyoming, Part 1, *River Hydraulics and Sediment Transport*, Report No. 1, 1979.
- [12] G. Gioia and F. A. Bombardelli, Scaling and similarity in rough channel flows, *Physical Review Letters* 88, 014501, 2001.
- [13] K. Guta and K. S. H. Prasad, Estimation of open channel flow parameters by using optimization techniques, *International Journal of Science and Research* 6, pp. 1295–1304, 2018.
- [14] W. A. Marcus, K. Roberts, L. Harvey, and G. Tackman, An evaluation of methods for estimating Manning’s n in small mountain streams, *Mountain Research and Development* 12, pp. 227–239, 1992.
- [15] R. H. Meade, W. W. Myrick, and W. W. Emmett, Field data describing the movement and storage of sediment in the East Fork River, Wyoming, Part 2, *River Hydraulics and Sediment Transport*, Report No. 2, 1979.
- [16] F. Pappenberger, K. Beven, M. Horrit, and S. Blazkova, Uncertainty in the calibration of effective roughness parameters in HEC-RAS using inundation and downstream level observations, *Journal of Hydrology* 302, pp. 46–69, 2005.
- [17] A. J. C. Saint-Venant, Théorie du mouvement non-permanent des eaux, avec application aux crues des rivières et à l’introduction des marées dans leur lit, *Comptes Rendus des Séances de Académie des Sciences* 73, pp. 147–154, 1871.