

Prof. Roberto Greco,

Editor of Hydrology and Earth System Sciences

02/08/2018

**Re: Manuscript hess-2018-118**

Thank you for your comments regarding our manuscript entitled “Discharge hydrograph estimation at upstream-ungauged sections by coupling a Bayesian methodology and a 2D GPU Shallow Water model” by A. Ferrari, M. D’Oria, R. Vacondio, A. Dal Palù, P. Mignosa, and M.G. Tanda. We have now updated the manuscript following the suggestions of the Reviewers after their analysis of the paper.

Please find attached to this letter a description of changes and our replies to each comment along with an updated manuscript.

We hope that all the points raised by the Reviewers have been satisfactorily addressed. We wish to kindly thank the referees for their careful reviews and invaluable comments and hope to hear from you again on the status of the manuscript.

Yours sincerely,

Alessia Ferrari

PhD, Research Assistant

Department of Engineering and Architecture

University of Parma

## RESPONSE TO REFEREE #1: Dr. A. D. Koussis

**The authors gratefully acknowledge the positive and constructive review of Dr. A.D. Koussis. In this document the comments provided by Dr. A.D. Koussis are reported in italic, whereas the authors' response and indications about the original paper modification are marked in bold fonts.**

*Estimating an unknown discharge hydrograph at an upstream cross-section is useful in flood hydrology both as a forensic activity (to find the inflow that caused a flood event observed at a certain downstream section) as well as operationally (to determine the operational mode of a reservoir in order to protect a downstream area). Such (rather special, but not rare) problems are tackled either by reverse routing the observed hydrograph to the upstream cross-section (an inverse problem, the solution of which exists, but is not unique and must be regularised; the authors should note, in their relevant section, that the solution does exist), or via optimisation. Both inversion approaches are subject to instabilities that must be controlled (e.g., smoothing). Past research has been referenced properly.*

**We thank the Referee for this comment and we agree that the solution of this kind of inverse problems exists. As a consequence, we have reformulated the sentence about the ill-posedness of the inverse problem in the Introduction, clarifying that the challenges we are dealing with are the non-uniqueness and the instabilities of the solution and not its existence, as follows:**

**“In the literature, this approach is known as reverse flow routing (D’Oria and Tanda (2012)), an ill-posed inverse problem that presents two main challenges: the solution may be non-unique, and instabilities may arise during the inversion.”**

*The submitted research opts for an optimisation approach: the procedure applies a Bayesian geostatistical methodology coupled with forward routing that solves the full 2-D shallow water equations. Using a 2-D flow model in the context of inverse flood routing is an advance beyond the state of the art. But the computational load caused by the necessary multiple 2-D flow runs is heavy. Therefore, the authors have carried out their inversion procedure by parallelising the evaluation of the Jacobian matrix (it assesses the solution sensitivity to each unknown flow value), taking advantage of the floating point calculation capabilities of an array of Graphical Processing Units grouped in a remote High Performance Computing cluster.*

*The testing and validation of the method is sound and thorough; it includes simulations of generic floods with perfect (error-free) and with corrupted data, as well as of real flood events. The achieved accuracy is very good, including the peak region. Large oscillations of the inverted flow (recovered inflow) hydrograph occurring near its end are explained (Figs. 10 and 16); oscillations occurring at the start of the flood (e.g., Fig. 9a) seem to be due to the somewhat abrupt initiation of the transient from the steady state, while oscillations in the peak region are likely due to the change from a rising to a falling flood flow (Figs. 13a and 14a). The largest oscillations of the stage hydrograph occur at the start of the flood (Figs. 9b and 17) and should be also attributable to the somewhat abrupt initiation of the transient from the steady state (please*

*comment). These oscillations are, of course, stronger in the simulations with corrupted data. Relevant comments by the authors would be appreciated; they would help the reader, too.*

**We thank the Referee for his comment since this is an excellent point to discuss. In the Bayesian Geostatistical Approach, the main mechanism by which soft knowledge about the unknown parameter function is imparted is through the prior information. In our approach, this soft knowledge is intentionally limited to the choice of a parameterized covariance model (the structural parameters, which control the balance between smoothness and misfit, are also estimated during the process) such that significant flexibility is available to the algorithm. Nevertheless, the behaviour raised by the Referee in correspondence of abrupt changes of the inflow hydrograph is due to the regularization imposed by the prior-information. In order to comment and justify this behaviour, in Sect. 4.1 we have added the following sentences:**

**“In addition to this behaviour at the end of the discharge hydrograph (that can be postponed extending the hydrograph total duration), very small differences between the observed and modelled variables appear when abrupt changes in the inflow function are present (e.g. the initial transition from the steady state to the flood wave). This behaviour is due to the regularization introduced into the solution by the prior information that imposes some degree of continuity and/or smoothness to the estimated hydrograph. However, the residuals are practically negligible and abrupt discontinuities in the inflow hydrographs are not common in natural floods.”**

*It is noted, as an aside, that evidence is not conclusive as to which approach, reverse routing or optimisation, is more prone to spurious oscillations; a specific comparative investigation, under identical conditions, is required.*

**We really thank the Referee for this comment. Due to the fact that the 2D Shallow Water Equations in their complete and conservative formulation, which govern the motion of the fluid, cannot be inverted, for the 2D applications presented in the paper no comparison between the reverse flow routing technique and the Bayesian one can be performed. However, D’Oria et al. (2012a, 2012b) compared the two strategies for the level pool routing algorithm used to compute the inflow hydrograph in a reservoir. They showed that in presence of corrupted observations, the reverse routing procedure amplifies the errors, whereas the regularization provided by the optimization Bayesian procedure avoids spurious oscillations in the solution.**

*The paper is structured well. The theory is presented succinctly, with adequate mathematics, and contains all relevant information; the same holds for the (important) computational aspects of the modelling approach. Figures and tables add significantly to the understanding of the textual account, and figures are of good quality. The language is generally quite good, yet the paper would benefit from careful editing (e.g., most ‘which’ should be ‘that’, ‘resulted’ should be ‘resulting’ etc.); some indentations must be corrected. (I will mail my marked up manuscript to the corresponding author for the consideration of the team of authors).*

**The authors wish to thank the Referee for having provided his marked copy of the manuscript: the suggested corrections have been included in the revised paper.**

*The Conclusions section could be enhanced. Particularly, given that the computing facilities and arrangements required for the inverse modelling approach reported in the manuscript are currently tailored to research rather than to the work of professional hydrologists, the authors should comment on how they envision their model finding its way to the hydrological practice.*

**We really appreciate this useful advice since it allows us to better remark the practical aspects of our work. The definition of a discharge hydrograph in an ungauged river section is a relevant issue for professional hydrologists involved for example in the design of hydraulic infrastructures as well as for engineers working on water resource management (i.e. irrigation system, hydroelectric power stations) or forensic activities. With the aim of solving this problem, we propose an application that requires supercomputer and High Performing Computer clusters. These tools are mostly used for University research activities, but they are not only reserved to these environments. In fact, clouds of GPUs or on-line mini cluster are now common and thus everyone can manage to access these facilities. Moreover, the adopted Bayesian software (bgaPEST) is open access and 2D Shallow Water Equations models are nowadays a quite common tools for practitioners. Therefore, in the concluding section the following sentences have been added:**

**“The test cases were simulated taking advantage of the HPC cluster of the University of Parma. However, since the implemented procedure is general, it is possible to adopt clouds of GPUs or on-line mini clusters, which are now common and accessible to everyone. The adopted Bayesian software (bgaPEST) is open access and 2D-SWE models are a quite common tools for practitioners, even if till now few of them are fast enough to perform the necessary simulations with a reasonable computing time. Therefore, the 2D coupled methodology here proposed can be adopted in the near future also by professional hydrologists involved for example in the design of hydraulic infrastructures as well as for engineers working on water resource management (i.e. irrigation systems, hydroelectric power stations, etc.) or forensic activities.”**

*Assessment: The paper addresses in a novel way an interesting topic (for specialists) that is within the scope of HESS, is scientifically sound and methodologically solid. It is very good and should be published after minor revision.*

**The authors wish to thank Dr. A.D. Koussis for his suggestions and considerations.**

## RESPONSE TO REFEREE #2:

**The authors gratefully acknowledge the positive and constructive review of the anonymous Referee. In this document the comments provided by the Referee are reported in italic, whereas the authors' response and indications about the original paper modification are marked in bold fonts.**

### *General comments*

*The manuscript applies a Bayesian geostatistical methodology to the solution of the inverse problem aiming to estimate the upstream flood hydrograph at an un-gauged river section. The downstream routing of the hydrograph is pursued by means of a 2D shallow water model. This leads to a computationally intensive problem, for which a parallel implementation is designed. The most computationally intensive operation (i.e.: the evaluation of the Jacobian matrix) is demanded to a multi-GPU HPC, and also the forward model exploits the opportunities of GPU-parallelization.*

*The adoption of two-dimensional hydraulic model represents a step forward compared with both the previous research developed by the Authors and with the state-of-the-art. The resulting complication arising from the increased computational effort is handled properly. Therefore, the research described in the paper appears to be sufficiently innovative, well-designed and of interest to the readers of HESS.*

*I am rather supportive of the publication of the manuscript, provided that the Authors put some additional effort in improving the quality of the presentation (especially of the English) and in addressing some issues in order to make their outcomes more conclusive. I provide in the following few specific comments to be considered in the revision, as well as some minor issues that could contribute to improve the quality of the manuscript.*

**The authors wish to thank the anonymous Referee for his positive overview about the manuscript.**

### *Specific comments*

*• I appreciate that the presentation of the Bayesian Geostatistical Approach (BGA) is concise but complete of every detail: however I found it not very clear at some points, detailed below:*

- 1. The “prior mean” defined in eq. (9) should be better commented, explaining why the vector reduces to “a single value” (do the Authors mean the same value for each parameter?), and why the matrix  $X$  reduces to “a single vector of ones”.*

**We appreciate this comment and we agree with the Referee that more information about the prior mean is needed to facilitate the readers in figuring out the Bayesian Geostatistical Approach philosophy. As a result, at Sect. 2.1.2 of the revised paper, the involved paragraph commenting the terms that form the prior mean has been reword as follows:**

“The prior mean is defined as:  $E[s] = X\beta$ , where  $E$  is the expected value,  $\beta$  is the vector of drift coefficients, and  $X$  is a known matrix of basis functions. In our case  $\beta$  is a single unknown scalar, but different drift coefficients can be used to introduce discontinuities in the stochastic function to be estimated (e.g. when the unknown parameters are likely to form distinct populations). For example, in the context of reverse flow routing problems, multiple values of  $\beta$  are adopted if more than one inflow hydrograph must be estimated at the same time (e.g. the inflow on both the upstream branches of a river confluence). The matrix of basis function,  $X$ , links each unknown parameter with the corresponding element of  $\beta$  and, at the same time, specifies the model of the mean (e.g. constant mean, mean with a trend, etc.); in our case the mean is constant and therefore  $X$  is a single vector of ones (Fienen et al., 2008).”

2. *The separation distance  $d$  should be defined explicitly.*

We really appreciate this comment and we acknowledge the potential confusion that arises from the use of the term separation distance. This is a legacy from the fact that geostatistics is mainly used in estimating spatial parameter fields rather than time functions. In Sect. 2.1.2 of the revised paper we have explicitly defined the variable as follows:

“...  $d$  represents the vector of the separation times between all the parameter pairs ( $d_{i,j} = t_i - t_j$  with  $i, j = 1, \dots, N_p$ ,  $t$  denoting the time associated with each parameter and  $N_p$  the total number of unknowns).”

3. *I wonder about the opportunity of defining  $Q_{ss}$  as  $Q_{ss}(\theta)$  since the r.h.s. of eq. (6) does not contain  $\theta$ .*

We acknowledge the mistake in the original version of our manuscript. The prior covariance matrix in Eq. (6) is not influenced by the slope parameter  $\theta$  but by the variance  $\sigma_s^2$  and the integral scale  $l$ ; we have corrected  $Q_{ss}(\theta)$  as  $Q_{ss}(\sigma_s^2, l)$ .

4. *I could not find the definition of  $\xi$  appearing in eq. (9) and eq. (13).*

We agree with the Referee and the definition of  $\xi$ , which was missing in the original paper, has been included at Sect. 2.1.3 as follows:

“In case a linear relationship between parameters and observations (linear forward model) holds, a computationally efficient method to find the best estimate  $\hat{s}$  of vector  $s$  (and  $\hat{\beta}$  of  $\beta$ ) is obtained introducing the vector  $\xi = (HQ_{ss}H^T + R)^{-1}(y - HX\hat{\beta})$  and solving the following linear system of equations (Fienen et al. (2009)):

5. *The Authors should better explain what they mean with “a flat solution”.*

We thank the Referee for this comment and we agree that the term “flat” should be better explained. For this reason, we completely reworded the sentence making clear what we mean with “a flat solution” as follows:

“... the starting values for the structural parameters are assigned so that the variability between contiguous parameters is small (flat solution, with a high degree of correlation); complexity is then introduced during the optimization process if supported by the data. The variance of the epistemic errors is assumed close to the expected one.”

• *In the scheme depicting the BGA in figure 3, I could not find the condition corresponding to the parameters convergence, which is claimed in the text. According to the scheme, the inner cycle terminates only when the maximum number of iterations  $N_i$  is reached. The Authors should clarify this point and modify accordingly the manuscript and/or the figure. Assuming that also convergence causes termination, the Authors should explain how did they check the convergence.*

The Referee is right. We confirm that both the inner loop to estimate the model parameter and the outer one to estimate the structural parameters iterate until convergence or the assumed maximum number of iterations is reached. Therefore, the 2<sup>nd</sup> (inner  $> N_i$ ) and 3<sup>rd</sup> (outer  $> N_o$ ) decision blocks in Fig. 3 of the manuscript do not only check if the maximum number of iterations is reached, but also verify if convergence is achieved. The flow chart in Fig. 3 has been corrected. Additionally, we have included in Sect. 2.1.3 the definition about convergence, as follows:

“Recalling that the aim of the inverse procedure is to obtain the vector of the unknown parameters  $s$ , as well as to quantify the uncertainty in the estimation, the solution is found by maximizing the posterior pdf or, more conveniently, minimizing its negative logarithm (objective function) (Fienen et al., 2013).”

“The linearization process ends if the improvement (absolute difference between two successive iterations) in the objective function is below a user defined value or if the maximum number of iterations  $N_i$  is reached. The structural parameter iteration loop (outer loop) progresses until the  $L_2$ -norm of the differences between structural parameter values at consecutive iterations is below a user defined value or if the maximum number of iterations  $N_o$  is reached (Fienen et al. 2013).”

• *The Authors should explain how the credibility intervals may be evaluated based on the results of the BGA algorithm, or at least provide a reference to previous literature.*

We really appreciate this suggestion and accordingly we have modified the involved paragraph at the end of Sect. 2.1.3 as follows:

“The diagonal elements of this matrix represent the posterior variance ( $\sigma^2$ ) of the estimated parameters and thus, the 95% credibility interval of the solution is evaluated as  $\pm 2\sigma$ .”

• *About the core of the research described in the manuscript, I am mostly concerned about three issues. They should hopefully be addressed in the revised version of the manuscript.*

1. *Since the principal innovation comes from the adoption of a 2D forward hydraulic model, the improvement in terms of the quality of the estimated hydrograph deriving from the use of a more detailed (but also demanding) schematization of the hydraulic process should be explicitly assessed. For instance, how wrong is the estimated hydrograph if one uses a 1D model as the forward routing model in one of the presented examples?*

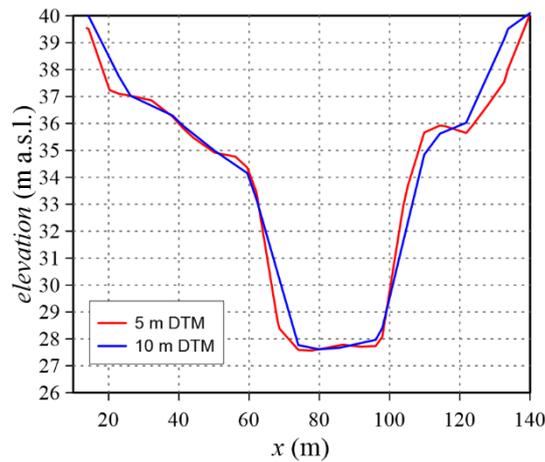
**We really thank the Referee for this comment since it allows us to discuss the motivations that led to enhance the serial Bayesian procedure introduced by D’Oria and Tanda (2012) for 1D cases, to 2D forward models. The choice between 1D and 2D models concerns the classical forward propagation rather than the Bayesian application. In fact, in literature the advantages of 2D-SWEs in comparison with 1D schematizations have been thoroughly discussed (e.g. Costabile et al., 2015), assessing that if river reaches present several floodable areas, meanders and floodplains, as it is typical for lowland streams, only 2D models can properly describe the flood propagation. As shown for example in Fig. 7 of the manuscript, in such rivers the low flow at the beginning of the event follows the meanders and water is contained in the main channel, whereas for high discharge the flow involves the river banks and a continuous mass and momentum exchange occurs between the main channel and the river banks and thus the assumptions of 1D models do not hold.**

**Therefore in our opinion, since the physical phenomena can be only accurately simulated by a 2D numerical scheme, no accurate upstream discharge hydrograph can be obtained by adopting 1D models. Finally, coupling the Bayesian approach with a fast, stable and accurate 2D forward model is the first step for reconstructing the discharge hydrograph during a levee failure and/or overtopping that causes the flooding of the nearest lowlands; the authors are also working in this direction that clearly requires the adoption of a 2D model.**

2. *Could the Authors discuss (hopefully with the aid of some additional results) the effects of the resolution of the DEM and/or of the values of the roughness parameters on the estimated hydrograph?*

**We thank the Referee for this useful comment that allows us to clarify some further aspects of the forward numerical modelling.**

**The mesh design is an issue related to create an accurate forward model. As for every numerical method that aims at describing a physical phenomenon in a spatial domain, the mesh must be chosen considering both the needed accuracy and the required computational effort. Firstly, the mesh must be defined in such a way that the bathymetry of the rivers is adequately resolved. Figure 1 below shows that the adopted mesh (with  $\Delta x=10$  m inside the river) is able to accurately reproduce the river geometry.**



**Figure 1. Ponte Alto section on the Secchia River: comparison between the sections extracted from a 5 m and 10 m resolution DTM.**

Secondly, the grid size must guarantee that the numerical solution is close to the “exact solution” of the SWEs. Convergence analysis can be proficiently performed for simple test cases, in which the mesh can be progressively halved many times with a reasonable computational effort. A similar analysis was done in a previous study conducted by some of the present authors (Aureli et al., 2008) and it is beyond the scope of this work. Anyhow, grid size, roughness estimation and numerical discretization of SWEs, all play an interlaced role on the solution results. First order accurate models, for example, intrinsically introduce more dissipation into the solution and this behaviour must be counterbalanced during the calibration phase, for example reducing the dissipation term due to friction, since a part of the dissipation is already embedded in the intrinsic numerical viscosity of the model. Despite the calibration of the considered river (grid size, roughness and numerical discretization) was already assessed in previous studies (Vacondio et al., 2016), according to the Referee’ suggestion we performed an additional inverse Bayesian estimation with a different roughness coefficient for the real field test case in Sect. 5 (please refer also to the next comment answer). Particularly, the Manning coefficient originally set equal to  $0.05 \text{ s/m}^{1/3}$  was decreased by 15% and assumed equal to  $0.0425 \text{ s/m}^{1/3}$ , as for example can happen due to seasonal changes in vegetation. As shown in Fig. 18 of the revised paper, the estimated flood waves are similar and the highest difference, which is in correspondence with the second peak, is less than 6%. Therefore, the influence of assuming a “wrong” roughness coefficient is less than linear in the discharge estimation. However, we want to stress that the same issue holds for any model setup.

3. *I understand the role of the simulations based on synthetic data-sets, with or without accounting for measure corruption in the validation of the procedure. On the other hand, as far as the “real field application” is concerned, I think that a different test case should have been considered, namely one for which the measured hydrograph was available, in order to compare the estimated with the actual one. This not being the case, the evaluation of the procedure performance cannot go further than the “credibility” (in a statistical sense), and the claims by the Authors in the comments (“This real field*

*application further confirms the capability of the proposed inverse procedure of estimating irregular inflow hydrographs in real rivers”) may sound excessive and not fully supported. Could the Authors take into consideration the addition of such an example?*

**We really appreciate this comment and the suggestion pointed out by the Referee. Since the upstream section A is located immediately downstream a flood control reservoir equipped with water level sensors, we have adopted the classic hydraulic theory of sluice gates and spillways to calculate the “reference” solution. As a result, Sect. 5 has been improved by validating the inflow hydrograph resulted from the inverse procedure as follows:**

**“With the aim of validating the methodology for this real application, it is noteworthy that the upstream section of the river is located immediately downstream a flood control reservoir equipped with water level sensors. Therefore, the "reference" discharge hydrograph has been obtained from the dam geometrical data (i.e. number and dimension of the bottom openings, crest length of the spillway, etc.) and the recorded water levels adopting the classic hydraulic theory of sluice gates and spillways. Due to the uncertainty in evaluating the discharge coefficients and to the fact that during flood events a large amount of wood debris reduces the outflow discharge from the bottom openings (especially during the depletion phase) and interferes with the overflow spillway, the discharge hydrograph has been calculated adopting equally likely coefficients (Fig. 18). The flood wave estimated by the inverse procedure is in good agreement with the one calculated using the flood reservoir data; the main differences are after the highest peak, which is well reproduced, although the inverse methodology provides a smoother solution. For this real application, even if the river roughness coefficient was already calibrated in previous studies (Vacondio et al. (2016)), an additional inverse Bayesian estimation was performed with a different value, in order to assess the effect of this coefficient on the solution. Particularly, the Manning coefficient originally set to  $0.05 \text{ s/m}^{1/3}$  was decreased by 15% ( $0.0425 \text{ s/m}^{1/3}$ ), as for example can happen due to seasonal changes in vegetation. As shown in Fig. 18, the estimated flood waves are similar and the highest difference, which is in correspondence with the main peak, is less than 6%. Therefore, the influence of assuming a “wrong” roughness coefficient is less than linear in the discharge estimation.”**

- *English should be carefully revised throughout the entire manuscript to match the standards of scientific communication.*

**We thank the Reviewer for his suggestion. The entire manuscript has been carefully revised and the language corrections kindly provided by Dr. A. D. Koussis (first Referee) have been integrated in the revised manuscript.**

*technical corrections*

- *Please refer to eq. (5) and (6) as to linear or Gaussian variogram, just the way you did in section 4.2*

**We thank the Reviewer for his technical corrections. The formula reported in eq. (5) and (6) express the linear and Gaussian covariance function, respectively, and not the variogram. Accordingly, at Sect. 4.2 of the revised paper we have referred to covariance functions and not variogram.**

*• Probably in r.h.s. of eq. (14) a “+” sign is missing. Please check.*

**The Referee is right: this has been corrected in the revised paper.**

*• Throughout the manuscript, “non linear” should better read “non-linear”*

**We thank the Reviewer for this suggestion: this has been corrected in the revised manuscript.**

*• Please note that actually the r.h.s. of eq. (12) is not a fraction, therefore referring to “denominator of Eq. (12)” makes sense if you are considering the discrete approximation of the Jacobian.*

**We totally agree with the Referee and, as consequence, we have reformulated the involved paragraph in the revised version of the manuscript. In Sect. 2, where the theory of the Bayesian approach is described, Eq. (12) defines the Jacobian matrix computation, which is not a fraction but a partial derivative. Therefore, in Sect. 3 we have reworded the paragraph as follows:**

**“The simulation of a base run, once a particular set of parameters has been assumed (deriving from the initialization or from previous estimation steps), represents a mandatory step for the Jacobian matrix evaluation, which is performed at this point of the procedure in order to quantify how each observation is influenced by the variation of each estimable parameter. Particularly, Eq. (12) is approximated using a finite difference method, and hence each element of the matrix is evaluated as the ratio between the variation of each observation (numerator) for given variation of each parameter (denominator) with respect to the base run.”**

*• The description of fig. 6 and the figure itself refer to four cross-sections along the river: an upstream ungauged one (A), two intermediate (B and C) where water levels are measured, and a fourth one (D) for downstream boundary condition assignment. However, in the presented examples, only a single intermediate measuring cross section is used, so maybe the description and the figure should be consistently simplified.*

**We really thank the Referee for this comment and we acknowledge that the role of the section D was not clear in the manuscript. However, the presence of section D plays a specific role in setting up the synthetic case to use as benchmark for the inverse procedure. In the revised manuscript the following sentences have been added:**

**“The information in sub-reach C-D is only preparatory for setting up the synthetic cases and it is not used in the inverse procedure. Imposing a rating curve in D allows to obtain water levels with a non-unique stage-discharge relationship in section C, which is more close to the real circumstances when applying the inverse procedure.”**

## References

- Aureli, F., Maranzoni, A., Mignosa, P., and Ziveri, C. A weighted surfacdepth gradient method for the numerical integration of the 2D shallow water equations with topography. *Advances in Water Resources*, 31(7), 962-974, 2008.
- Costabile, P., Macchione, F., Natale, L., and Petaccia, G. Flood mapping using LIDAR DEM. Limitations of the 1-D modeling highlighted by the 2-D approach. *Natural Hazards*, 77(1), 181-204, 2015.
- D’Oria, M. and Tanda, M. G.: Reverse flow routing in open channels: A Bayesian Geostatistical Approach, *Journal of hydrology*, 460, 130–135, 2012.
- D’Oria, M., Mignosa, P., & Tanda, M. G. Reverse level pool routing: Comparison between a deterministic and a stochastic approach. *Journal of hydrology*, 470, 28-35, 2012a.
- D’Oria, M., Mignosa, P., & Tanda, M. G. Response to Comment on “Reverse level pool routing: Comparison between a deterministic and a stochastic approach” by H. Md. Azamathulla, *J. Hydrol.* (2012), <http://dx.doi.org/10.1016/j.jhydrol.2012.09.005>. *Journal of Hydrology*, 475, 499-501, 2012b.
- Fienen, M. N., Clemo, T., and Kitanidis, P. K.: An interactive Bayesian geostatistical inverse protocol for hydraulic tomography, *Water Resources Research*, 44, 2008.
- Fienen, M., Hunt, R., Krabbenhoft, D., and Clemo, T.: Obtaining parsimonious hydraulic conductivity fields using head and transport observations: A Bayesian geostatistical parameter estimation approach, *Water resources research*, 45, 2009.
- Fienen, M. N., D’Oria, M., Doherty, J. E., and Hunt, R. J.: Approaches in highly parameterized inversion: bgaPEST, a Bayesian geostatistical approach implementation with PEST: documentation and instructions, Tech. rep., US Geological Survey, 2013.
- Vacondio, R., Aureli, F., Ferrari, A., Mignosa, P., Dal Palù, A. Simulation of the January 2014 flood on the Secchia River using a fast and high-resolution 2Dparallel shallow-water numerical scheme. *Natural Hazards*, 80(1), 103-125, 2016.

# Discharge hydrograph estimation at upstream-ungauged sections by coupling a Bayesian methodology and a 2D GPU Shallow Water model

Alessia Ferrari<sup>1</sup>, Marco D'Oria<sup>1</sup>, Renato Vacondio<sup>1</sup>, Alessandro Dal Palù<sup>2</sup>, Paolo Mignosa<sup>1</sup>, and Maria Giovanna Tanda<sup>1</sup>

<sup>1</sup>Department of Engineering and Architecture, University of Parma, Parma, Italy

<sup>2</sup>Department of Mathematical, Physical and Computer Sciences, University of Parma, Parma, Italy

*Correspondence to:* Alessia Ferrari (alessia.ferrari@unipr.it)

**Abstract.** ~~In this~~ This paper ~~presents~~ a novel methodology for estimating the unknown discharge hydrograph at the entrance of a river reach, ~~where when~~ no information is available, ~~is presented~~. The methodology ~~is obtained by coupling~~ couples an optimization procedure, based on the Bayesian Geostatistical Approach (BGA), with a forward self-developed 2D hydraulic model ~~of the stream~~. In order to accurately describe the flow propagation in real rivers characterized by large floodable areas, the forward model solves the 2D Shallow Water Equations (SWEs) by means of a Finite Volume explicit shock-capturing algorithm. The 2D-SWE forward code exploits the computational power of Graphics Processing Units (GPUs), achieving ratio of physical to computational time up to 1000. With the aim of enhancing the computational efficiency of the inverse estimation, the Bayesian technique is parallelized developing a procedure based on the Secure Shell (SSH) protocol that allows to take advantage of remote High Performance Computing clusters (including those available on the Cloud) equipped with GPUs. The capability of the coupled models methodology is assessed by estimating irregular and synthetic inflow hydrographs in real river reaches, taking into account also the presence of downstream corrupted observations. Finally, the ~~capability to adopt this methodology for real cases is demonstrated by reconstructing~~ procedure is applied to reconstruct a real flood wave in a river reach located in Northern Italy.

## 1 Introduction

The definition of discharge hydrographs in specific river sections is still a relevant hydraulic problem not only for flood modelling purposes, but also for more practical issues related to flood protection measures, hydropower plants, water resource management, design of new structures, etc. Flood routing techniques, either hydrological or hydraulic, are extensively studied and widely used to estimate discharge hydrographs in downstream ungauged sites based on data available at upstream gauged stations (forward propagation). However, often, the flow hydrograph is required in a river section ~~which that~~ is completely ungauged and does not have upstream useful information for its definition. In these cases, discharge hydrographs at specific sites can be estimated by coupling rainfall-runoff and forward flood propagation models. However, rainfall-runoff models (Beven (2011)) present several uncertainties associated, for example, with the choice of the model for the basin schematization, with

the evaluation of the effective rainfall, and with the calibration procedure. An alternative approach is to assess the upstream unknown flow hydrograph using only the information, in terms of discharge values or water levels, available downstream the selected site and, possibly, the characteristics of the river reach. In the literature, this approach is known as reverse flow routing (D’Oria and Tanda (2012)), ~~an ill-posed inverse problem that presents three main challenges: the solution may not exist, or it may be non-unique, and instabilities may arise during the inversion.~~ **an ill-posed inverse problem that presents two main challenges: the solution may be non-unique, and instabilities may arise during the inversion.** The traditional attempts of solving the reverse flow routing problem are based on two main approaches: the solution of a reverse form of the Saint Venant equations (e.g. Eli et al. (1974), Szymkiewicz (1993), Dooge and Bruen (2005), Bruen and Dooge (2007)) and the back oriented application of hydrological routing schemes (e.g. Das (2009), Koussis et al. (2012), Koussis and Mazi (2016)).

Beyond the approximations introduced by the hydrological routing schemes, the above procedures were applied to simplified reach geometries and flow conditions. In almost all cases, especially considering downstream information affected by errors, instabilities and spurious oscillations appeared; low-pass filters, with subjective parameters, were sometime used to damp the estimated inflow fluctuations. D’Oria and Tanda (2012) and Zucco et al. (2015) provide additional references and details on the reverse flow routing problem.

In addition to the above procedures, the estimation of an unknown upstream flow hydrograph, based only on downstream information (observations), can be performed via optimization methods. These techniques aim at finding the upstream flow hydrograph that, routed downstream, best matches the available observations. D’Oria and Tanda (2012) solved the reverse flow routing problem adopting, **as optimization procedure,** a novel Bayesian Geostatistical Approach (BGA) ~~as optimization procedure, which that~~ considers the flow hydrograph as a ~~statistical~~ continuous random function that presents autocorrelation ~~and accounts for uncertainties.~~ The authors showed the capability of the BGA methodology, in combination with a forward hydraulic model, to estimate the discharges in an upstream ungauged section based only on an available downstream flow hydrograph: the **solution was stable** ~~procedure evidenced no instabilities,~~ also in **the** presence of corrupted downstream flow values. The forward model, which solves the 1D Saint Venant equations, was considered already implemented and calibrated and able to describe, with sufficient accuracy, the hydraulic routing process. The BGA method was further extended in order to adopt, as downstream observations, stage hydrographs instead of discharge ones (D’Oria et al. (2014)). Saghafian et al. (2015) identified the upstream hydrograph of a river reach, given the downstream one, by using a Genetic Algorithm coupled with a forward hydraulic model ~~,which that~~ solves the 1D Saint-Venant equations under the kinematic wave approximations. Only some minor oscillations and instabilities occurred during the inversion, but the Authors applied the procedure to a rectangular prismatic channel and no errors were added to the downstream observations. Zucco et al. (2015) investigated the reverse flow routing process in natural channels, and estimated the discharge hydrograph in ungauged sections, by means of a Genetic Algorithm coupled with a simplified routing model. The parametric forward model was based on the continuity equation written in a characteristic form, lumped over the entire river reach, and on simplified rating curves at the channel ends. In addition, the unknown inflow hydrograph was assumed distributed in time as a Pearson type III function with three parameters, **thus** preventing the possibility of estimating ~~of~~ real flood waves with irregular shapes (e.g. multi-peak hydrographs).

All the previously cited works adopted 1D hydraulic models or simplified hydrological routing ~~models schemes~~, in combination with different optimization procedures. Nevertheless, in many real cases, the complex hydrodynamic field generated by the flood propagation cannot be accurately described under 1D assumptions and it is necessary to adopt schemes based on the 2D Shallow Water Equations, even if this poses the drawback of the computational ~~burden efficiency~~ and requires a detailed terrain survey. However, nowadays, bathymetric data can be easily obtained from high-resolution Digital Terrain Models (DTM) and fast 2D numerical models have been developed. With the purpose of estimating the discharge hydrograph in an upstream ungauged river section, having water level information only in a downstream observation site, this paper extends the BGA methodology for reverse flow routing of D’Oria and Tanda (2012) and D’Oria et al. (2014) to a 2D forward ~~modelization algorithm~~ in order to model natural rivers with complex geometry, including flood plains and floodable areas. With this aim, the stable, accurate and fast PARFLOOD GPU code (Vacondio et al. (2014), Vacondio et al. (2017), Vacondio et al. (2016)), which solves the conservative form of the 2D Shallow Water Equations on a finite volume scheme, is adopted as forward model and coupled to the inverse estimation procedure. In order to reduce the computational times, the Jacobian matrix estimation procedure, which is the key point of the BGA method, has been parallelized. Additionally, a host-server data management procedure has been implemented, so as to exploit the computational power of remote large modern supercomputer and/or cloud HPC resources. The capability of the optimization procedure has been tested by estimating real or pseudo-real inflow hydrographs in natural river reaches, where 1D models cannot accurately describe the flood propagation. Moreover, during the discharge estimation, the presence of downstream corrupted observations has also been taken into account, since registered data at gauging stations are quite often affected by instrumental errors. ~~Dealing with real recorded data and real field application, the discharge parameter values have been estimated in a logarithmic space, in order to prevent the rise of negative values.~~

The paper is organized as follows: in Sect. 2 the theory of the Bayesian Geostatistical Approach is illustrated. A step-by-step description of the inverse procedure is provided in Sect. 3: the parallel implemented scheme, the forward model optimization for reducing the run times and the iteration management between the local host and the remote server are ~~detailed~~ described in detail. Section 4 is dedicated to the ~~procedure validation~~ application of the procedure to synthetic test cases concerning, ~~which concerns~~ the estimation of inflow hydrographs with different shapes in two rivers in Northern Italy. The ~~application~~ ~~practicability~~ of the inverse procedure for reconstructing a historical flooding event is presented in Sect. 5. Some concluding remarks are finally outlined in Sect. 6.

## 2 Theory of the Bayesian Geostatistical Approach

The optimization software adopted to solve the reverse flow routing problem is the bgaPEST (Fienen et al. (2013)), which implements the Bayesian Geostatistical Approach of Kitanidis (1995) and it is developed according to the PEST (Model Independent Parameter Estimation) parameter estimation software (Doherty (2016)). The bgaPEST is appropriate ~~to solve for~~ ~~solving~~ inverse problems (in a context of a highly parametrized inversion), which are characterized by unknown parameters that are correlated one another in space or time, as for example the discharge values of a flow hydrograph. The first applications of the inverse methodology ~~are were~~ related to the estimation of spatial parameter fields in a groundwater context (Kitanidis

and Vomvoris (1983), Hoeksema and Kitanidis (1984), among others) but later the method **has been was** adopted to evaluate unknown time functions in different areas (e.g. Snodgrass and Kitanidis (1997), Michalak et al. (2004), Butera et al. (2013), D’Oria and Tanda (2012), D’Oria et al. (2015), Leonhardt et al. (2014)).

## 2.1 The Bayes’ theorem

5 The crux of the adopted bgaPEST, as well as other methods based on the Bayesian Approach, is **the** Bayes’ theorem, which reads:

$$p(\mathbf{s}|\mathbf{y}) \propto L(\mathbf{y}|\mathbf{s})p(\mathbf{s}), \quad (1)$$

where  $\mathbf{s}$  is the vector of the unknown parameters,  $\mathbf{y}$  is the vector of the measured data,  $p(\mathbf{s}|\mathbf{y})$  is the posterior probability density function (pdf) of  $\mathbf{s}$  given  $\mathbf{y}$ ,  $L(\mathbf{y}|\mathbf{s})$  is the likelihood function and  $p(\mathbf{s})$  is the prior probability density function of  $\mathbf{s}$ . Since the present work aims at estimating an upstream hydrograph in an ungauged section, assuming the knowledge of downstream water levels,  $\mathbf{s}$  represents the discharge values over time of the unknown inflow hydrograph, whereas  $\mathbf{y}$  denotes the downstream water level observations. Following Eq.(1), the posterior pdf **, which represents the parameter knowledge after the observations,** can be seen as a combination between a priori knowledge on the parameters (prior pdf), where a priori means that the observed data are still not considered, and information about parameters contained in the measured data (likelihood function) (Glickman and Van Dyk (2007)). In the BGA method proposed by Kitanidis (1995), the prior pdf and the likelihood function are described by means of Gaussian distributions and the best set of parameter  $\mathbf{s}$  is obtained by maximizing the posterior pdf.

### 2.1.1 The likelihood function

**Focusing on the terms of the Bayes theorem**The likelihood function  $L(\mathbf{y}|\mathbf{s})$  in Eq. (1) characterizes the **misfit deviation** between observed data and model results (Fioren et al. (2013)). Starting from the results of the forward model,  $L(\mathbf{y}|\mathbf{s})$  delineates how a particular set of parameters  $\mathbf{s}$  is able to reproduce the observations  $\mathbf{y}$  in space and/or time, **therefore thus** accounting for the epistemic errors. The investigated inverse problem presents different sources of errors **that , which** are related to the conceptual schematization of the inverse procedure, to the numerical forward model and to the data measurement. In the likelihood function, the errors are assumed to be independent and identically distributed, with **null zero** mean and covariance matrix expressed as follows:

$$25 \quad \mathbf{R} = \sigma_R^2 \mathbf{I}, \quad (2)$$

where  $\sigma_R^2$  denotes the variance that **regulates expresses** the misfit between observed and modeled data, and  $\mathbf{I}$  is the identity matrix.

### 2.1.2 The prior probability density function

The prior knowledge about  $\mathbf{s}$  ( $p(\mathbf{s})$  in Eq.(1)) is limited to the **definition assignment** of a mean value (unknown and estimated during the procedure) and a characteristic about **the continuity and/or smoothness of the parameter field described through a**

covariance function implemented as covariance matrix. It furnishes a soft knowledge about the structure/shape of the unknowns and provides a regularization of the solution; the prior pdf can also be used to enforce non-negativity to the parameters (D’Oria and Tanda (2012)). The prior mean is defined as:

$$E[\mathbf{s}] = \mathbf{X}\boldsymbol{\beta}, \quad (3)$$

- 5 where  $E$  is the expected value,  $\boldsymbol{\beta}$  is the vector of drift coefficients, and  $\mathbf{X}$  is a known matrix of basis functions. In our case  $\boldsymbol{\beta}$  is a single unknown scalar, but different drift coefficients can be used to introduce discontinuities in the stochastic function to be estimated (e.g. when the unknown parameters are likely to form distinct populations). For example, in the context of reverse flow routing problems, multiple values of  $\boldsymbol{\beta}$  are adopted if more than one inflow hydrograph must be estimated at the same time (e.g. the inflow on both the upstream branches of a river confluence). The matrix of basis function,  $\mathbf{X}$ , links each unknown parameter with the corresponding element of  $\boldsymbol{\beta}$  and, at the same time, specifies the model of the mean (e.g. constant mean, mean with a trend, etc.); in our case the mean is constant and therefore  $\mathbf{X}$  is a single vector of ones (Fielen et al. (2008)). ~~(a single vector of ones, in this case), which link each value of  $\mathbf{s}$  with the appropriate element of  $\boldsymbol{\beta}$ .~~

The prior covariance matrix of the unknown parameters  $\mathbf{Q}_{ss}$  is then defined as:

$$\mathbf{Q}_{ss} = E \left[ (\mathbf{s} - \mathbf{X}\boldsymbol{\beta})(\mathbf{s} - \mathbf{X}\boldsymbol{\beta})^T \right]. \quad (4)$$

- 15 In the context of geostatistics, the covariance matrix  $\mathbf{Q}_{ss}$  is a function of the separation distance (in time in this case) between the parameters and describes their deviations from the mean behavior. Different functions models can be adopted to describe the covariance; for example, it can be assumed as a linear function, represented through a limiting case of the exponential covariance function model (Fielen et al. (2008)), according to the following relation:

$$\mathbf{Q}_{ss}(\theta) = \theta l \exp \left( -\frac{|\mathbf{d}|}{l} \right), \quad (5)$$

- 20 where  $\mathbf{d}$  represents the vector of the separation ~~distances in time between the parameters,~~ times between all the parameter pairs ( $d_{i,j} = t_i - t_j$  with  $i,j=1,\dots,N_p$ ,  $t$  denoting the time associated with each parameter and  $N_p$  the total number of unknowns),  $l$  a fixed integral scale ( $l = 10 \max(d)$ ) and  $\theta$  the slope (structural parameter) ~~,which influences that governs~~ the correlation between the discharge values of the unknown hydrograph. A different formulation (D’Oria et al. (2014)) defines the prior covariance matrix  $\mathbf{Q}_{ss}$  by means of a Gaussian function model characterized by two structural parameters ( $\sigma_s^2$  and  $l$ ):

$$25 \mathbf{Q}_{ss}(\sigma_s^2, l) = \sigma_s^2 \exp \left( -\frac{|\mathbf{d}|^2}{l^2} \right), \quad (6)$$

where  $\sigma_s^2$  denotes the variance. The linear function (Eq.(5)) enforces only continuity to the solution whereas the Gaussian function model (Eq.(6)) adds also some degree of smoothness, but the final solution is still driven by the observations.

### 2.1.3 The posterior probability density function

With the assumptions made, the likelihood and prior terms that compose the posterior pdf of Eq.(1) can be rewritten as follows (Fiene et al. (2009); D’Oria and Tanda (2012); D’Oria et al. (2014)):

$$L(\mathbf{y}|\mathbf{s}) = \exp\left(-\frac{1}{2}(\mathbf{y} - \mathbf{h}(\mathbf{s}))^T \mathbf{R}^{-1}(\mathbf{y} - \mathbf{h}(\mathbf{s}))\right) \quad (7)$$

5

$$p(\mathbf{s}) = \exp\left(-\frac{1}{2}(\mathbf{s} - \mathbf{X}\boldsymbol{\beta})^T \mathbf{Q}_{ss}^{-1}(\mathbf{s} - \mathbf{X}\boldsymbol{\beta})\right). \quad (8)$$

The term  $\mathbf{h}(\mathbf{s})$ , in the likelihood function, represents the modeled values in the same place and time ~~of~~ as the available observations  $\mathbf{y}$ . Therefore, to evaluate  $\mathbf{h}(\mathbf{s})$ , a forward model of the considered river reach that is able to describe the hydraulic routing process is required in order to provide, for a given set of parameter  $\mathbf{s}$ , the corresponding downstream water levels.

10 Recalling that the aim of the inverse procedure is to obtain the vector of the unknown parameters  $\mathbf{s}$ , as well as to quantify the uncertainty in the estimation, the solution is found ~~by~~ maximizing the posterior pdf ~~or, more conveniently, minimizing its negative logarithm (objective function)~~ (Fiene et al. (2013)).

In case a linear relationship between parameters and observations (linear forward model) holds, a computationally efficient method to find the best estimate  $\hat{\mathbf{s}}$  of vector  $\mathbf{s}$  (and  $\hat{\boldsymbol{\beta}}$  of  $\boldsymbol{\beta}$ ) is obtained introducing the vector  $\boldsymbol{\xi} = (\mathbf{H}\mathbf{Q}_{ss}\mathbf{H}^T + \mathbf{R})^{-1}(\mathbf{y} - \mathbf{H}\mathbf{X}\hat{\boldsymbol{\beta}})$  and solving the following linear system of equations (Fiene et al. (2009)): ~~In case a linear relationship between parameters and observations (linear forward model) holds, the best estimate  $\hat{\mathbf{s}}$  of vector  $\mathbf{s}$  (and  $\hat{\boldsymbol{\beta}}$  of  $\boldsymbol{\beta}$ ) is obtained by solving the following linear system of equations~~

$$\begin{cases} \hat{\mathbf{s}} = \mathbf{X}\hat{\boldsymbol{\beta}} + \mathbf{Q}_{ss}\mathbf{H}^T\boldsymbol{\xi} \\ \begin{bmatrix} \mathbf{H}\mathbf{Q}_{ss}\mathbf{H}^T + \mathbf{R} & \mathbf{H}\mathbf{X} \\ \mathbf{X}^T\mathbf{H}^T & 0 \end{bmatrix} \begin{bmatrix} \boldsymbol{\xi} \\ \hat{\boldsymbol{\beta}} \end{bmatrix} = \begin{bmatrix} \mathbf{y} \\ 0 \end{bmatrix}, \end{cases} \quad (9)$$

where  $\mathbf{H}$  is the sensitivity (Jacobian) matrix, representing how the observations  $\mathbf{y}$  are influenced by ~~a single each~~ unknown parameter  $s_i$  (D’Oria et al. (2015)). However, for ~~this particular problem~~ the particular problem under investigation,  $\mathbf{h}(\mathbf{s})$  is non-linear and therefore matrix  $\mathbf{H}$  depends on  $\mathbf{s}$ . Following the quasi-linear geostatistical approach (Kitanidis (1995)), ~~the relationship between observations and parameters can be successively linearized about a candidate solution  $\mathbf{s}_k$ , where  $k$  represents each iteration for each iteration  $k$ , the relationship between observations and parameters is successively linearized about a candidate solution:~~

$$25 \quad \mathbf{h}(\mathbf{s}) \approx \mathbf{h}(\mathbf{s}_k) + \tilde{\mathbf{H}}_k(\mathbf{s} - \mathbf{s}_k), \quad (10)$$

and then a correction to the measurements is applied according to the following relation:

$$\mathbf{y}_k = \mathbf{y} - \mathbf{h}(\mathbf{s}_k) + \tilde{\mathbf{H}}_k\mathbf{s}_k. \quad (11)$$

Therefore, the sensitivity matrix is evaluated at each iteration as follows (D’Oria et al. (2014)):

$$\tilde{\mathbf{H}}_k = \left. \frac{\partial \mathbf{h}(\mathbf{s})}{\partial \mathbf{s}} \right|_{\mathbf{s}_k}. \quad (12)$$

Analogously to the linear system in Eq. (9), the linearized system is solved according to:

$$\begin{bmatrix} \tilde{\mathbf{H}}_k \mathbf{Q}_{ss} \tilde{\mathbf{H}}_k^T + \mathbf{R} & \tilde{\mathbf{H}}_k \mathbf{X} \\ \mathbf{X}^T \tilde{\mathbf{H}}_k^T & 0 \end{bmatrix} \begin{bmatrix} \hat{\boldsymbol{\xi}}_{k+1} \\ \hat{\boldsymbol{\beta}}_{k+1} \end{bmatrix} = \begin{bmatrix} \mathbf{y}_k \\ 0 \end{bmatrix}, \quad (13)$$

5 and the next estimate of the parameters is evaluated by means of:

$$\tilde{\mathbf{s}}_{k+1} = \mathbf{X} \hat{\boldsymbol{\beta}}_{k+1} + \mathbf{Q}_{ss} \tilde{\mathbf{H}}_k^T \hat{\boldsymbol{\xi}}_{k+1}. \quad (14)$$

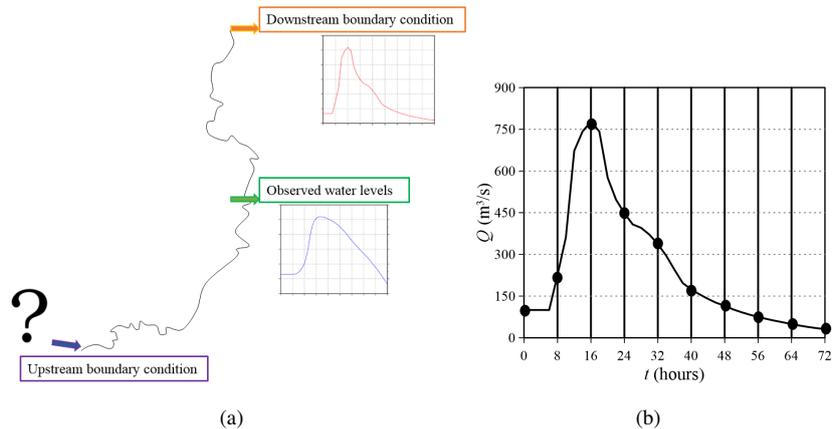
A proper selection of the covariance **function model** structural parameters ( $\theta$ ,  $\sigma_s^2$  and  $l$ ) and optionally of the epistemic error variance  $\sigma_R^2$  is important **in order** to reach a good solution. **However,** ~~the~~ structural parameters are estimated from the data using a Bayesian adaptation of the Restricted Maximum Likelihood (RML) method of Kitanidis (1995) **that ,which**  
 10 **adopts probability functions and** allows reaching the best compromise between the fitting of the modeled data **and with** the observations and the prior information (Fiene et al. (2013)). Dealing with **non-linear** problems, unknowns ( $\mathbf{s}$ ) and structural parameters must be iteratively estimated in successive steps. **The linearization process ends if the improvement (absolute difference between two successive iterations) in the objective function is below a user defined value or if the maximum number of iterations  $N_i$  is reached. The structural parameter iteration loop (outer loop) progresses until the  $L_2$ -norm of the differences**  
 15 **between structural parameter values at consecutive iterations is below a user defined value or if the maximum number of iterations  $N_o$  is reached** (Fiene et al. (2013)). Finally, at the end of the estimation, the linearized uncertainties of the unknowns can be evaluated in terms of the posterior covariance matrix of the estimated parameters (Fiene et al. (2013)). **The diagonal elements of this matrix represent the posterior variance ( $\sigma^2$ ) of the estimated parameters and thus, the 95% credibility interval of the solution is evaluated as  $\pm 2\sigma^2$ .**

### 20 3 Description of the Bayesian estimation procedure

After having described the theory of the Bayesian **Geostatistical** Approach in Sect. 2, some operational information about the BGA inverse procedure is now illustrated. As mentioned in the Introduction and sketched in Fig. 1-a, the goal of the adopted BGA methodology is the estimation of the discharge hydrograph in an upstream-ungauged river section (identified by a question mark in Fig. 1-a), having information about water levels observed in a downstream section (intermediate site in Fig.  
 25 1-a). A boundary condition, downstream of the observation site, must also be specified; this can be based on observed data or can be approximated extending the computational domain faraway from the intermediate section. The inverse method estimates  $N_p$  parameters (the vector of the unknown parameters  $\mathbf{s}$  in Eq. (1)) **that originates ,which derive** from the discretization of the **upstream** discharge hydrograph by means of time intervals, regular in this case (Fig.1-b).

The BGA algorithm solves the inverse problem by means of the following steps.

30 First **ly**, the unknown parameters and the structural ones are initialized. The first ones may be all assumed equal to a constant



**Figure 1.** Definition of the reverse flood routing problem (a) and of the unknown parameters (b).

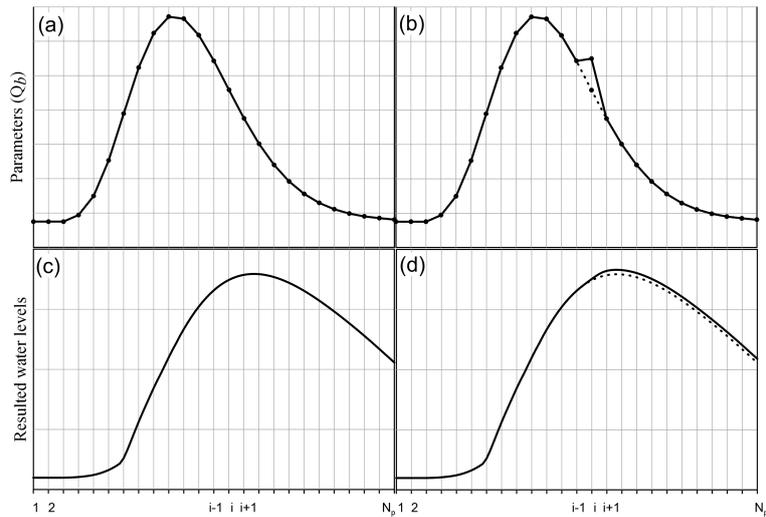
discharge value coherent with the considered river, whereas the starting values for the structural parameters are assigned so that the variability between contiguous parameters is small (flat solution, with a high degree of correlation); complexity is then introduced during the optimization process if supported by the data. The variance of the epistemic errors is assumed close to the expected one. the structural parameters are usually set so as to guarantee a flat solution (complexity is introduced during the optimization process only if supported by the data) and the variance of the epistemic errors is assumed close to the expected one.

Assuming the first guess of the unknown parameters as upstream boundary condition, the hydraulic forward model is run and the resulting water levels are extracted at the observation site. The simulation of a base run, once assumed a particular set of parameters has been assumed (deriving from the initialization or from previous estimation steps), represents a mandatory step for the Jacobian matrix evaluation, which is performed at this point of the procedure in order to quantify how each observation is influenced by the variation of each estimable parameter. The Jacobian matrix quantifies how each observation is influenced by the variation of each estimable parameter, and it is calculated using a finite differences method. According to Eq. (12), each element is evaluated as the ratio between the variation of each observation for given variation of each parameter (numerator) and the variation of the parameter value with reference to the base run (denominator). Particularly, Eq. (12) is approximated using a finite difference method, and hence each element of the matrix is evaluated as the ratio between the variation of each observation (numerator) for given variation of each parameter (denominator) with respect to the base run. Therefore, additionally to the base run, the hydraulic forward model is further run as many times as the number of parameters to estimate  $N_p$ . At each run, a single value of the upstream boundary condition is modified by a known quantity with respect to the previous value, and the hydraulic forward model is run again. Therefore, as a consequence, each simulation tests the sensitivity of the resulted water levels (all the observations at once) to the variation of a single parameter  $i$ .

In order to exemplify this step, Fig. 2-a shows the discharge imposed as upstream boundary condition for a base run of an intermediate set of parameters: after the propagation, the resulting water levels extracted at the observation site are shown in Fig. 2-c. To test the sensitivity to parameter  $i$  Assuming that the Jacobian matrix is testing the sensitivity, in Fig. 2-b

the considered parameter is varied with changed by a known quantity and a new upstream boundary condition is defined (solid line); it is worth noting that the solid and the dotted lines differ only for the parameter  $i$ . The water levels resulting resulted from this single parameter variation are shown in Fig. 2-d (solid line): they are identical to the base run ones until time  $i-1$ , whereas after that time they differ from those of the base run (dotted line). The computation of the differences between the

5 resulting resulted water levels of the simulation  $i$  and of the base run (solid and dotted lines) and the variation of parameter  $i$  allows computing the column  $i$  of the Jacobian matrix, which is a  $N_{obs} \times N_p$  matrix, where  $N_{obs}$  represents the number of the observations. After having collected all the perturbed observations  $N_p$  runs, the Jacobian/sensitivity matrix is evaluated and a new set of parameters  $\mathbf{s}$  is estimated (Eq. (14)).



**Figure 2.** Example of the base run (a) and of the run  $i$  for the Jacobian matrix evaluation (b).

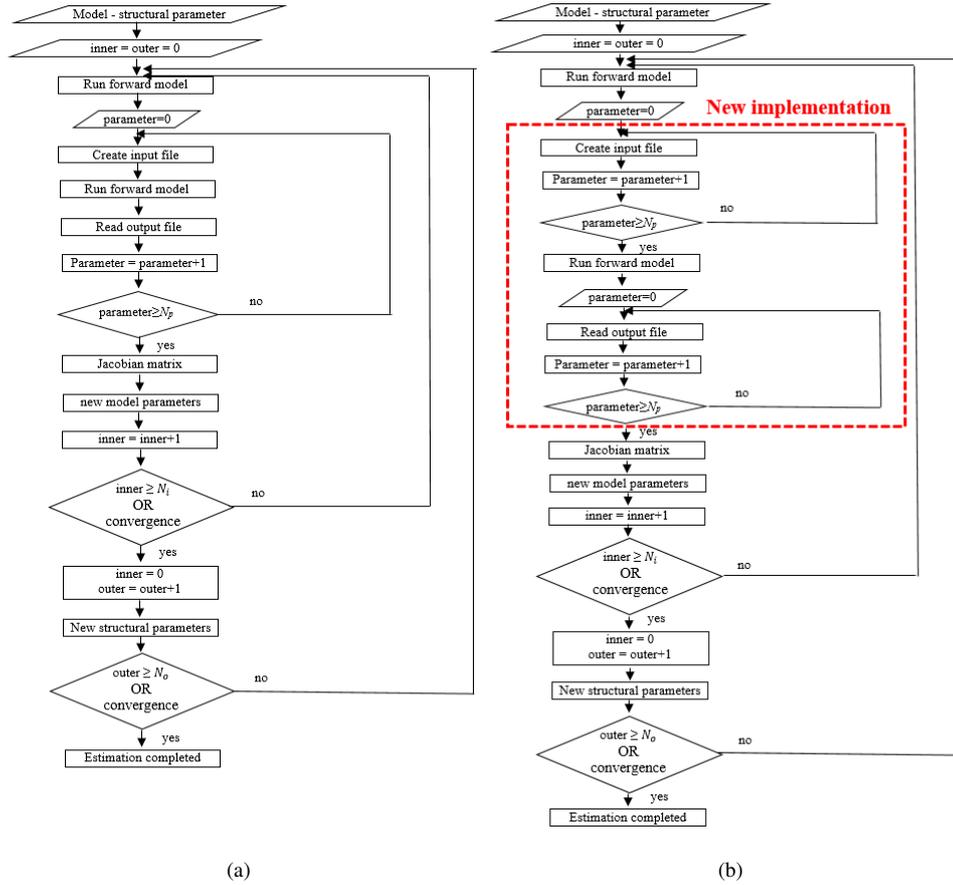
Then, the first set of resulted parameters is used for evaluating a new Jacobian matrix and as a result, a second set of parameters is estimated. This procedure is repeated until convergence or the maximum number of iteration  $N_i$  is reached. Then, the structural parameters are estimated using the last set of parameters  $\mathbf{s}$ . holding the last set of parameters  $\mathbf{s}$  constant, the structural parameters are estimated. Due to the non-linearity of the forward problem, the model and the structural parameter estimation is repeated until convergence of the lasts (or the maximum number of iterations  $N_o$  is reached ). Therefore, the BGA implementation requires running the forward model  $N_t$  times, according to the following relation (Fienen et al. (2013)):

$$15 \quad N_t = (N_p + 1) N_o N_i + 1. \quad (15)$$

The whole BGA procedure previously described is sketched in Fig. 3-a.

### 3.1 Parallelization of the Jacobian matrix evaluation

The most relevant contribution to Focusing on the total computational time required by the inverse estimation procedure, it emerges that the most relevant contribution is ascribed to the run-of-the forward model runs (i.e. the computation of each



**Figure 3.** Scheme of BGA algorithm in the serial (a) and parallel (b) version.

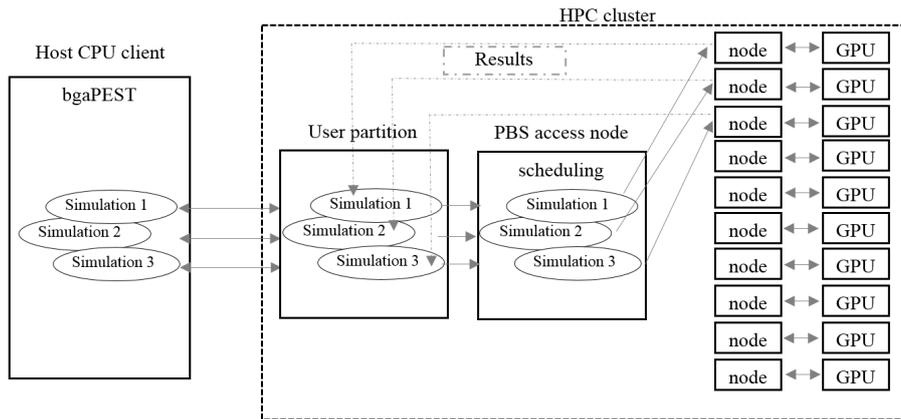
element of the Jacobian matrix), rather than to the bgaPEST operations. However, since each of the  $N_p$  runs in Eq. (15) checks the sensitivity of the observations to the variation of a single parameter, the solution of a run test does not affect the solution of the other ones. Therefore, in order to reduce the computational burden, the independent  $N_p$  runs can be potentially performed in parallel.

- 5 In this work, the PARFLOOD 2D-GPU numerical model presented in Vacondio et al. (2014) and Vacondio et al. (2017) has been adopted for routing the inflow hydrograph. Therefore, the bgaPEST routine to evaluate the Jacobian matrix has been parallelized in order to run simulations taking take advantage of the computational capability of modern High Performance Computing (HPC) clusters, which are usually equipped with many GPUs. The implemented parallel procedure, which is sketched in the flow chart of Fig. 3-b, handles the parallelism among host and GPUs by means of the Secure Shell network protocol (SSH) and manages the most operative parts of the parallelism (login, run, etc.) outside the bgaPEST code. In the serial version (Fig. 3-a), the crucial part of the Jacobian matrix evaluation implementation consists in a do-loop over the parameters. Considering the parameter  $i$ , first by the input file that will be read by the forward model is written, then the model is run and
- 10

finally the **resulting** **resulted** values are read. In the modified version (Fig. 3-b), this main loop is split in three parts: first **ly**, all the input files (equal to  $N_p$ ), **inside in each** of which a particular parameter is modified, are written, then the forward model is run ( $N_p$  times), and finally a second loop is performed to read all the resulted values.

### 5 3.2 The forward model

In the parallel bgaPEST (Fig. 3-b), the “Run forward model” instruction actually runs a shell script **,which that** controls the file transfer between the host (a **standard classical** PC or a single node of a cluster) **,** and the HPC platform, the creation of the  $N_p$  simulations for the Jacobian matrix evaluation, and the run of the 2D-SWE GPU code on the device (GPU). In the present work, a cluster with 10 NVIDIA ® Tesla ® P100 GPUs hosted by the University of Parma was adopted. As shown in Fig. 4, the bgaPEST algorithm runs on the CPU of a computer, where the  $N_p$  simulations (in Fig. 4 assumed equal to three for the sake of simplicity) are first **ly** created and then sent to the server user partition, by means of the SSH protocol. Here, the cluster access node schedules all the jobs submitted by the users, using the HPC scheduler Portable Batch System (PBS). Then, each simulation is assigned to a specific GPU node. At the end of the computation, the observations are extracted and the output files remain on the cluster partition, until the CPU verifies via SSH the end of the simulation and copies the results back. The procedure sketched in Fig. 4 and following described represents **the parallelization of the Jacobian matrix computation one of the  $N_T$  iterations.**



**Figure 4.** Schematization of the data transfer assuming three parameters and thus three parallel simulations.

**Listing 1** provides a detailed description of the “Run forward model” shell file. ~~With the aim of describing in detail the Run forward model, after having clarified the data transfer procedure (Fig.4), in Listing 1 the structure of the shell file is presented.~~ In order to use the Algorithm for different test cases and potentially on different HPC clusters, all the paths are first **ly** declared together with the involved variables (number of parameters to estimate, time interval among parameters, start/end of the simulation) (line 2). Then, the algorithm (line 3) checks if the considered run is one useful for the Jacobian matrix

evaluation, where a given parameter varies, or if it is the base run. Considering the first *if* condition as true (line 3), the script generates and copies all the input files for all the  $N_p$  simulations to the server (lines 5-7). These files tests contain the same bathymetrical, initial conditions (water level and velocity) and roughness configuration, but a different upstream boundary condition; each simulation tests the sensitivity of the observations to the variation of a given model parameter. Moreover, all the simulations adopt the same grid (Cartesian or multiresolution), which is generated only once at the beginning of the procedure. It is relevant to note notice that all the  $N_p$  simulations have not to be run from time  $t_{start}$  to time  $t_{end}$ ; in fact, ~~In effect~~ the variation of parameter  $i$  ~~causes effects~~ affects the observations only after time  $t_{i-1}$  and thus the results until  $t_i$  are still identical to the base run (see Fig. 2). The PARFLOOD model guarantees the possibility of using the results of the base run and starting simulations from time  $t_{i-1}$ . The theoretical physical time  $T$  required to evaluate the Jacobian matrix simulating each of the  $N_p$  runs of simulations run from  $t_{start}$  to  $t_{end}$  is equal to evaluated as follows:

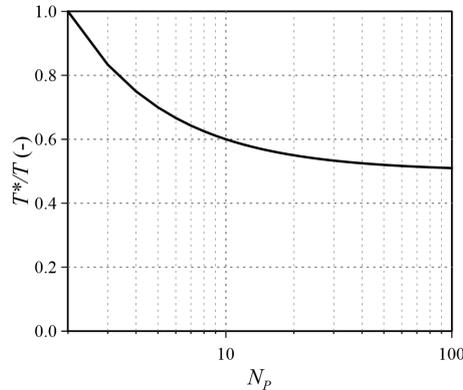
$$T = N_p(N_p - 1) \Delta t, \quad (16)$$

where  $N_p$  denotes the parameter number and  $\Delta t$  denotes the constant time interval between two consecutive parameters  $i$  and  $i+1$ .

Conversely, the physical time  $T^*$  required to simulate all the  $N_p$  runs restarting the  $i$ -th simulation from time  $t_{i-1}$  instead of  $t_{start}$  of simulations run from  $t_{restart}$  to  $t_{end}$  is equal to reads:

$$T^* = (N_p - 1) \Delta t + \sum_{i=2}^{N_p} [N_p - (i - 1)] \Delta t. \quad (17)$$

As pointed out by Eq. (16)-(17) and exemplified in Fig. 5 for a test case with 20 parameters, this simple operation allows reaching a relevant decrease of the total computational time. Therefore, at line 8, the algorithm computes the time useful to restart the simulation.



**Figure 5.** Time reduction  $T^*/T$  as a function of the number of estimable parameters (the x-axis is in logarithmic scale). Comparison of the cumulative computational times obtained simulating the  $N_p$  runs from time  $t_{start}$  until  $t_{end}$  or with restart option.

In order to perform the simulation, the host **logins logs in** to the HPC cluster **server** by means of the SSH protocol (line 9) and a sleep condition ensures the login procedure (line 10). Then the job is submitted to the queue of the cluster using external parameters for passing the name of the simulation folder and the time for restart (line 11): the submitted job contains the reference to the PBS queue and the link to the executable 2D-SWE GPU code. At the end of the simulation, the water levels at the observation site are automatically extracted. Once the job is submitted, the SSH login is closed (line 12). After having submitted all the simulations, for each parameter (line 15) the code regularly (line 18) tests via SSH the presence of the `end_file`, which states the end of the simulation (line 20), and waits in case it is missing (line 25). Once the simulation is finished, the resulting **ed** observations are copied back to the **CPU Host client** (line 28) and the folder is removed from the server (line 29).

10 **On the other side** Conversely, the *else* condition (line 30) is true for the base run, ~~which is necessary for the Jacobian matrix computation~~. The simulation folder ~~with, which contains all~~ the **necessary** input files is copied to the server (line 31) and the job is submitted (line 34). Then, the algorithm periodically verifies the end of the simulation and copies the results back to the **CPU Host client** (lines 39-49). It is relevant to note that the base run is **firstly** performed **first**, whereas the other  $N_p$  ones can be **performed in parallel performed**.

**Listing 1.** “Run forward model” for the parallel bgaPEST scheme

```

15 #!/bin/bash
2 Variable and path declaration
3 if [parameter run];
4     then
5     for (( i=1; i<=Np; i++ ))
26     do
6         Create and copy the simulation folder to the server (simulation_i)
7         Compute the time from which restarting the simulation
8         ssh server name << EOF
9         sleep 15
25        ssh submission: frame number for restart , name of the test case , job to submit
12        exit
13        EOF
14    done
15    for (( i=1; i<=Np; i++ ))
30    do
16        end_file=0
17        while [ end_file -eq 0 ];
18        do
19            ssh server name find server_path -iname end_file.txt | wc -l > end_file
35            if [end_file];
22            then
23                continue
24            else
25                sleep 10

```

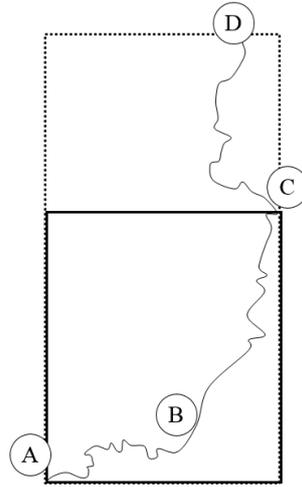
```

26         fi
27     done
28     Copy the file with observation from server to CPU
29     Remove the simulation folder on the cluster
30 else
31     Create and copy the simulation folder to server
32     ssh server name << EOF
33     sleep 15
34     ssh submission: frame number for restart , name of the test case , job to submit
35 exit
36 EOF
37 end_file=0
38 while [ end_file -eq 0 ];
39 do
40     ssh server name find server_path -iname end_file.txt | wc -l > end_file
41     if [end_file ];
42     then
43         continue
44     else
45         sleep 10
46     fi
47 done
48 Copy the file with observation from server to CPU
49 fi

```

#### 25 4 ~~Validation of the inverse methodology~~ Application of the inverse methodology to synthetic test cases

With the purpose In the context of applying validating the BGA method described above before, it is worth noting that reference solutions for inverse problems are by definition unavailable, since the goal of the methodology is the estimation of an upstream inflow hydrograph that is unknown at the beginning of the process. Therefore, in this section the inflow hydrographs in two natural rivers in Northern Italy are estimated and the reference solutions, which are necessary in order to validate the inverse procedure, are obtained as follows (D’Oria et al. (2014)). Considering the domain in Fig. 6, a selected inflow discharge  $Q^{ref}$  is routed from the upstream section A to until the downstream boundary D, where a rating curve is imposed far away from C. The resulting ed water level hydrographs is are extracted in at sites B and C. The inverse procedure is then applied to the sub-domain sketched with solid line in Fig. 6, by assuming the water levels in sites B and C (resulted from derived in step 1) as observations and downstream boundary condition, respectively. The information in sub-reach C-D is only preparatory for setting up the synthetic cases and it is not used in the inverse procedure. Imposing a rating curve in D allows to obtain water levels with a non-unique stage-discharge relationship in section C, which is more close to the real situations when applying the inverse procedure. The methodology estimates the inflow  $Q^{est}$  assuming that no information is available on the discharge (or water stage) at the inflow section A.



**Figure 6.** Exemplification of a test case definition.

Quantitative information about the accuracy of the inverse methodology is [here](#) provided evaluating the differences between the reference  $Q^{ref}$  and the estimated  $Q^{est}$  hydrographs by means of three different indicators. Firstly, the Nash-Sutcliffe efficiency criterion (Nash and Sutcliffe (1970))  $E_h$  was adopted, according to the following relation:

$$E_h = \left[ 1 - \frac{\sum_{i=1}^{N_p} (Q_i^{ref} - Q_i^{est})^2}{\sum_{i=1}^{N_p} (Q_i^{ref} - \bar{Q}^{ref})^2} \right] \cdot 100, \quad (18)$$

- 5 where  $N_p$  is the number of parameters,  $Q_i^{ref}$  and  $Q_i^{est}$  are the  $i$ -th reference and estimated inflow values, respectively, and  $\bar{Q}^{ref}$  is the mean value of the reference hydrograph. Then, the root mean square error, RMSE, was evaluated as follows:

$$RMSE = \sqrt{\frac{\sum_{i=1}^N (Q_i^{ref} - Q_i^{est})^2}{N_p}}. \quad (19)$$

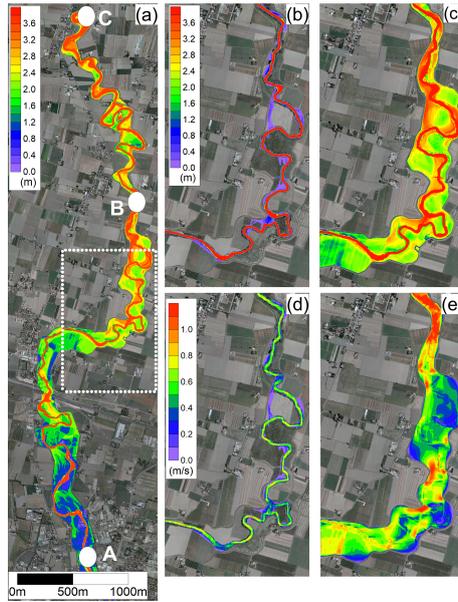
Finally, the **estimation** error in the peak discharge  $E_p$  was assessed as:

$$E_p = \left[ \frac{Q_p^{est}}{Q_p^{ref}} - 1 \right] \cdot 100, \quad (20)$$

- 10 where  $Q_p^{est}$  and  $Q_p^{ref}$  denote the peak discharge value of the estimated and reference hydrographs, respectively.

#### 4.1 Inflow hydrograph estimation on the Parma River

- The first test concerns the estimation of a **Synthetic hypothetical** discharge hydrograph at the entrance of the Parma River (Northern Italy). Figure 7-a illustrates the studied domain and the locations of the upstream boundary condition A, of the observation site B and of the downstream boundary section C. The domain includes a 20-km long **embanked** reach **that, which**  
 15 is characterized by several meanders and flood plains. As shown in Fig. 7, the flow field significantly varies at low and high



**Figure 7.** Map of the maximum simulated water depths for in the Parma River (a): the upstream (A) and downstream (C) boundary conditions and the intermediate observation site (B) are indicated. With reference to the area marked with dotted white line in (a), (b) and (c) represent the water depths and (d) and (e) the velocity field at low and high discharge values, respectively.

discharge values due to the river morphology. At the beginning of the flood wave, the flow is characterized by both low water depths (7-b) and velocity (7-d). Conversely, at the arrival of the flood peak *piek*, most of the meanders are cut by the flow, as shown in Fig. (7-c) and Fig. (7-e) for water depths and velocity, respectively. This makes the adoption of 1D numerical schemes *not suitable difficult* to accurately describe the flood propagation.

- 5 The bathymetry was derived from a 1-m resolution DTM obtained through a LiDAR survey carried out in drought condition. The domain was discretized by means of a Cartesian grid with cell sizes  $\Delta x = \Delta y = 4$  m and about  $275 \cdot 10^3$  computing cells were adopted. The Manning roughness coefficient was assumed equal to  $0.05 \text{ s/m}^{1/3}$ . The steady-state values of water depth and velocity fields, obtained considering the initial discharge value of the hydrograph, were adopted as initial conditions.

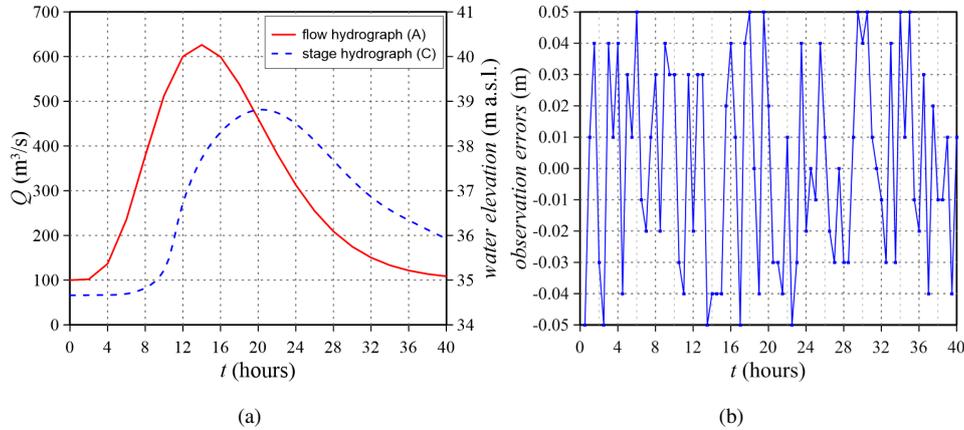
The inflow condition to be estimated *concerns a Synthetic Discharge Hydrograph with gamma distribution that was calculated was assumed* as follows (D’Oria et al. (2015)):

$$Q(t) = A + B \cdot f(t, b, k), \quad (21)$$

where  $t$  denotes the time,  $A$  the base flow (constant value),  $B$  the volume above the base flow (constant value) and  $f$  the gamma distribution, which states:

$$f(t, b, k) = \frac{1}{k^b \Gamma(b)} t^{b-1} e^{-\frac{t}{k}}, \quad (22)$$

where  $\Gamma(b)$  represents the gamma function defined through the parameters  $b$  and  $k$  that denote the shape and the scale parameter, respectively. The parameters of the gamma distribution were set as follows:  $A = 100 \text{ m}^3/\text{s}$ ,  $B = 3 \cdot 10^7 \text{ m}^3$ ,  $b = 6$  and  $k = 10000 \text{ s}$ . The resulted flood wave presented a peak value of about  $630 \text{ m}^3/\text{s}$  at time  $(b-1)k \approx 14$  hours (Fig. 8-a).



**Figure 8.** Parma River inflow: flow and stage hydrographs at referred to sections A and C, respectively (a) and observation error distribution (b).

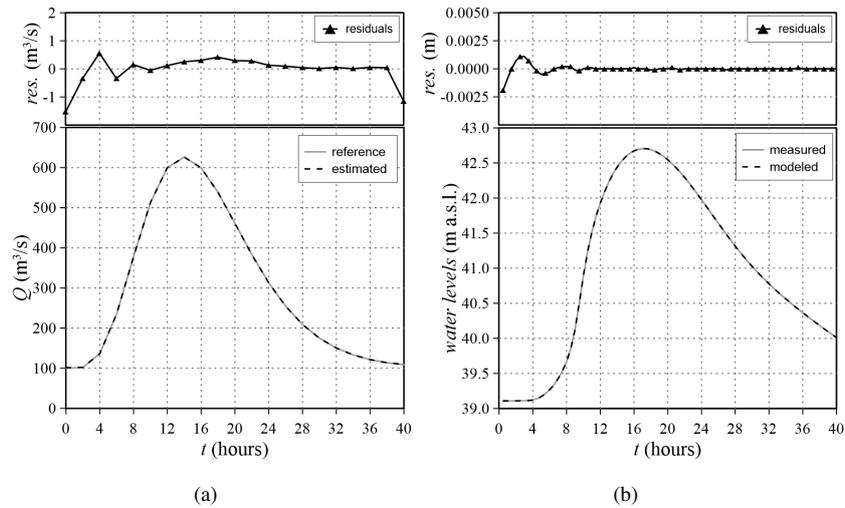
During the estimation, when the sensitivity to the first parameter  $p_1$  is investigated, the steady-state flow for the initial discharge is also recomputed. This means that parameter  $p_1$  determines not only the first value of the estimated flood wave but also it governs the initial condition of the river reach.

The inflow hydrograph duration was limited to 40 hours and it was discretized using 2 hours time interval steps ( $N_p=21$ ), whereas the observation stage hydrographs were discretized every 0.5 hours (80 water levels). The prior pdf was defined by means of a Gaussian covariance function model, and the initial structural parameters were set as reported in Table 1. In order to avoid non-physical discharge values during the computations, non-negativity was enforced to the unknown parameters by performing the estimation in a logarithmic space. The initial model parameter values were defined by applying the linesearch tool of the bgaPEST, which damps the solution between successive iterations (Fienen et al. (2013)), and avoids numerical instabilities that may occur starting from a worse first choice of the parameters too far from the true one.

The inflow hydrograph was estimated first by considering true observations (the variance was set equal to  $10^{-8} \text{ m}^2$  to take into account the truncation error) free of errors and with truncation error resulting in a variance of  $10^{-8} \text{ m}^2$ . Then, the same discharge hydrograph was defined corrupting the observed water levels with random errors uniformly distributed with maximum deviations of  $\pm 0.05 \text{ m}$  and variance  $10^{-3} \text{ m}^2$  (Fig. 8-b).

Qualitative assessment of the inverse methodology is achieved by comparing the reference with the estimated inflow hydrograph, as well as the observed with the modeled water levels in the observation site. Considering the simulation without errors in the observations, Fig. 9 shows that the estimated flood wave overlaps the reference one (a), and the modeled water levels

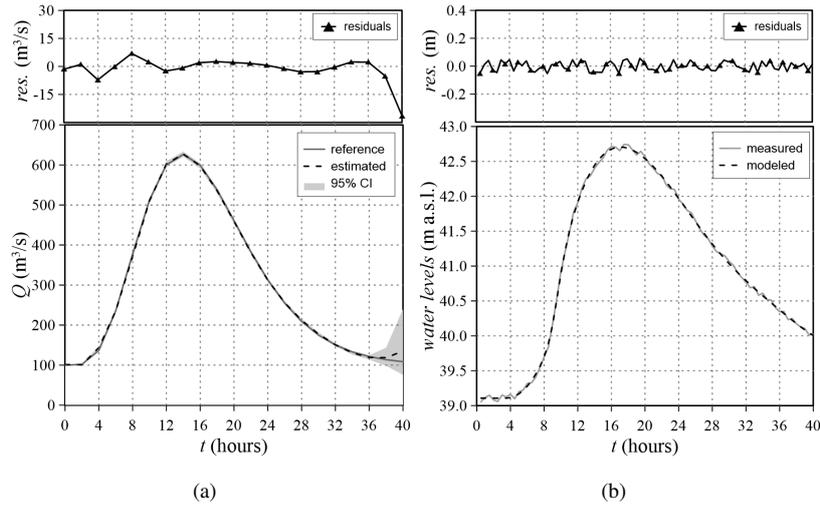
agree almost perfectly with the measured ones (b). Particularly, with reference to the peak value, the estimated flood wave presents the maximum misfit of 0.2%, whereas the modeled water levels differ from the observed ones less than the 0.01%.



**Figure 9.** Parma River inflow and uncorrupted observations: reference and vs estimated inflow hydrograph (a) and observed (uncorrupted) and vs modeled water levels (b). The residuals between reference and estimated values are also reported.

The results of the simulation with random errors corrupting the observations are depicted in Fig. 10. The estimated flood wave well matches again the reference one, presenting a misfit referred relative to the peak value lower than the 5%, and similarly the modeled water levels reproduce the reference ones with residual less than 1%. Only the last value of the reconstructed flood wave it is slightly overestimated, since the more the tested parameter nears the end of the wave, the fewer observations contain information about the related effects, as illustrated by the increasing range of the 95% credibility interval. However, the "true" discharge values are estimated wave is inside the 95% credibility interval, thus confirming the good accuracy results of the solution. In addition to this behaviour at the end of the discharge hydrograph (that can be postponed extending the hydrograph total duration), very small differences between the observed and modelled variables appear when abrupt changes in the inflow function are present (e.g. the initial transition from the steady state to the flood wave). This behaviour is due to the regularization introduced into the solution by the prior information that imposes some degree of continuity and/or smoothness to the estimated hydrograph. However, the residuals are practically negligible and abrupt discontinuities in the inflow hydrographs are not common in natural floods. The structural parameters and the epistemic error variance estimated in the presence and absence of corrupted observations are reported in Table 1.

Quantitative assessment Assessment of the methodology accuracy has been quantified achieved by means of the Nash-Sutcliffe  $E_h$ , root mean square error RMSE and error in the peak discharge  $E_p$  values reported in Table 2. The  $E_h$  values are greater than over the 99%, the  $E_p$  ones values are almost negligible and the RMSE error is less than 0.5 m<sup>3</sup>/s without random errors and reaches the maximum value of 6 m<sup>3</sup>/s with corrupted observations.



**Figure 10.** Parma River inflow and corrupted observations: reference and vs estimated (with 95% credibility interval) inflow hydrograph (a) and observed (corrupted) and vs modeled water levels (b). The residuals between reference and estimated values are also reported.

**Table 1.** Parma River inflow: initial and estimated structural parameters and epistemic error variance.

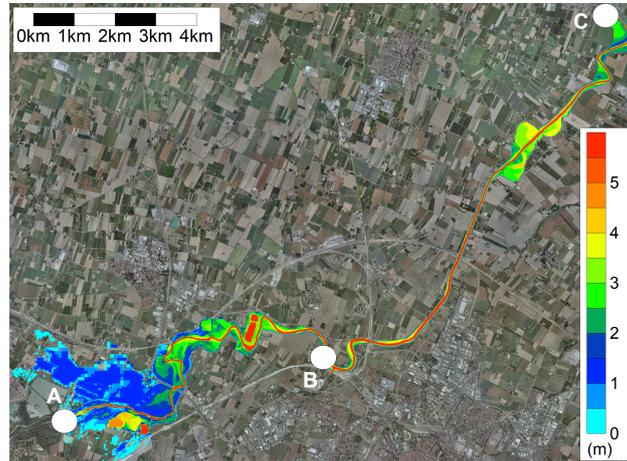
|  |           | No random errors | Random errors |
|--|-----------|------------------|---------------|
| $\sigma_R^2$ (m <sup>2</sup> )                 | Initial   | -                | 1.00E-4       |
|  | Estimated | -                | 1.09E-3       |
| $\sigma_S^2$ (m <sup>6</sup> s <sup>-2</sup> ) | Initial   | 5.00E+2          | 5.00E+2       |
|  | Estimated | 1.07E+3          | 5.36E+1       |
| $l$ (s)  | Initial   | 6.48E+4          | 6.48E+4       |
|  | Estimated | 2.90E+4          | 5.28E+4       |

## 4.2 Inflow hydrograph estimation on the Secchia River

The second test case concerns both a different river reach and shape of the inflow hydrograph. The studied domain includes a 25 km-long reach of the Secchia River (Northern Italy) between the outflow of the flood control reservoir of Rubiera-Campogalliano located at west of Modena town (point A) and the gauging station of Ponte Bacchello (point C) and referring the water level observations to the gauging station of Ponte Alto (point B) (Fig. 11). The modeled river reach is characterized by the presence of many flood plains and floodable areas that influence the flood propagation. The bathymetry was derived from a 1-m resolution DTM obtained through a LiDAR survey carried out in drought condition.

**Table 2.** Parma inflow River: Nash-Sutcliffe  $E_h$ , root mean square error RMSE and error in the peak discharge  $E_p$  values.

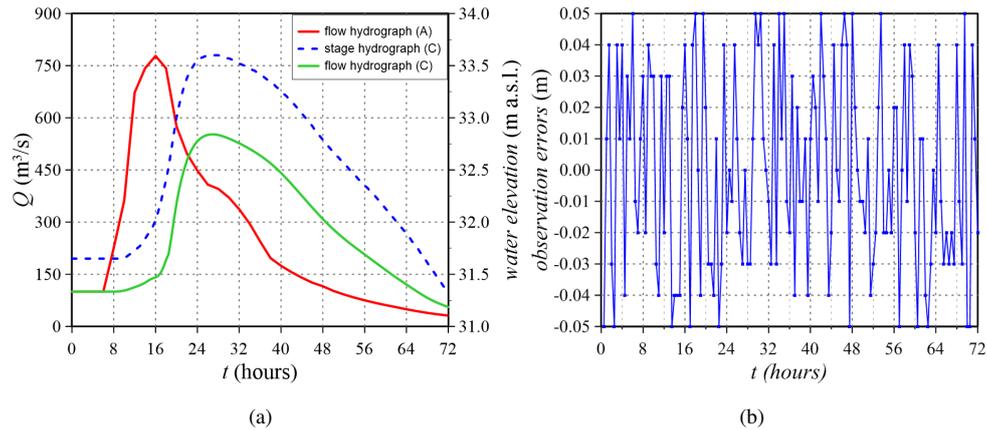
|                  | $E_h$ (-) | RMSE (m <sup>3</sup> /s) | $E_p$ (%) |
|------------------|-----------|--------------------------|-----------|
| No random errors | 99.99     | 0.49                     | -0.04     |
| Random errors    | 99.88     | 6.65                     | 0.15      |



**Figure 11.** Map of the water depths at the flood peak occurrence on the Secchia River, with indication of the upstream (A) and downstream (C) boundary conditions and the intermediate observation site (B).

The domain was discretized by means of a non-uniform BUQ grid (Vacondio et al. (2017)), resulting in  $77 \cdot 10^3$  computing cells. The Manning roughness coefficient in the riverbed was assumed equal to  $0.05 \text{ s/m}^{1/3}$  (Vacondio et al. (2016)).

The discharge hydrograph to be estimated is the **synthetic** flood wave of 20-years-return period of the Secchia River with a peak value of about  $780 \text{ m}^3/\text{s}$  ~~after 18 hours~~. In order to increase the non-smoothness of the wave, a quite abrupt increment that separates the initial steady-state condition ( $100 \text{ m}^3/\text{s}$ ) from the rising limb was introduced (Fig. 12-a). It is noteworthy that this flow hydrograph is characterized by a pseudo-real irregular shape, that cannot be **properly** approximated by an **analytical** parametric function (**i.e.e.g.**, Gamma distribution, Pearson function). The inflow hydrograph ended in 72 hours and ~~it~~ was discretized using 2-hours **time interval steps** ( $N_p=37$ ), whereas the **observed ation** stage hydrograph was discretized every 0.5 hours (~~144 water level values~~). The inflow hydrograph was first **ly** estimated assuming the **true** water levels extracted at section B **free of errors** ( with **only** a truncation error resulting in a variance of  $10^{-8} \text{ m}^2$  ), and **thensecondly** considering corrupted observations with random errors uniformly distributed with maximum deviations of  $\pm 0.05 \text{ m}$  and variance  $10^{-3} \text{ m}^2$  (Fig. 12-b). **Besides the discharge and stage hydrographs** Figure 12-a depicts **also** the discharge hydrograph **resulted in** at the downstream boundary condition section C, in order to highlight the attenuation effect exerted by the flood plains and floodable areas.



**Figure 12.** Secchia River inflow: flow and stage hydrographs referred to sections A and C, respectively flow hydrograph at section A and flow and stage hydrographs at section C (a) and observation error distribution (b).

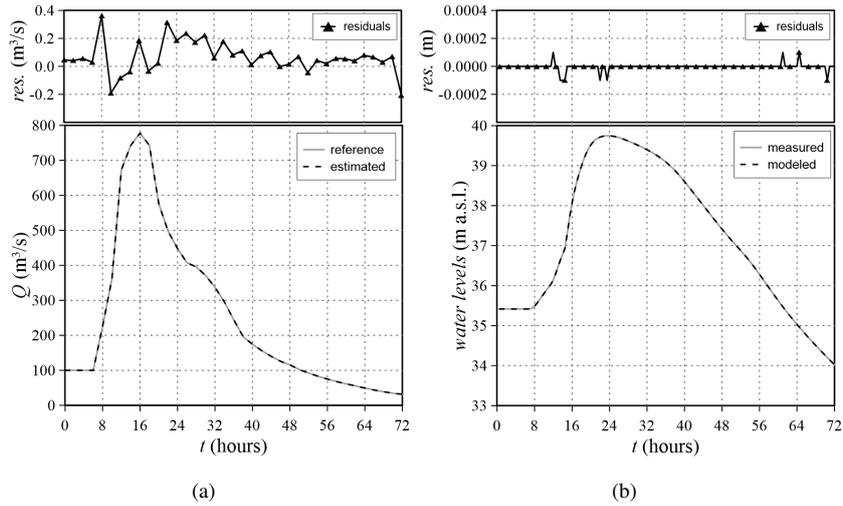
As before, the parameters were estimated in a logarithmic space and their initial values were calculated adopting the line-search tool of the bgaPEST (Fienen et al. (2013)). The prior pdf was described by means of a linear and Gaussian covariance function variogram, in the configuration with and without corrupted observations, respectively (Table 3). The initial model parameter values were calculated adopting the line-search tool of the bgaPEST, and the estimation was performed in a logarithmic space.

As shown in Fig. 13 for the simulation without corrupted observations, the estimated flood wave matches almost perfectly the reference one, as well as and the modeled water levels agree with the measured ones. With reference to the peak value, the flood wave is estimated with less than the 0.03% difference against the reference one, and similarly, the residuals between modeled and observed water levels are less than the 0.01%.

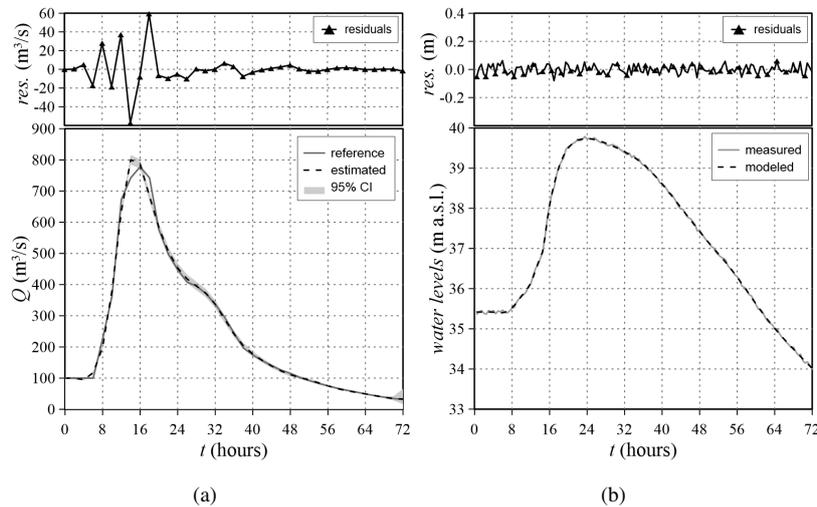
The results of the simulation with corrupted observations depicted in Fig. 14 highlight that both the shape and the peak value are well captured. The residual between reference and estimated discharge, referred to the peak value, is about 8%, whereas the misfit between observed and modeled water levels is less than 0.3%. The small discrepancies of the estimated peak flood wave from the reference one are essentially caused by the fact that the portion with the peak is discretized with only a few parameters and the adopted covariance function variogram smooths the solution.

The structural parameters and the epistemic error variance estimated in the presence and absence of corrupted observations are reported in Table 3.

The resulted indicators used for evaluating the accuracy of the methodology are reported in Table 4. The Nash-Sutcliff efficiency  $E_h$  values exceed the 99%, the errors in the peak flow  $E_p$  are almost negligible and the RMSE error is less than 1  $\text{m}^3/\text{s}$  without random errors and reaches the maximum value of 16  $\text{m}^3/\text{s}$  with corrupted observations: these values highlight the accuracy of the procedure in estimating the overall shape and peak of the inflow hydrograph.



**Figure 13.** Secchia River inflow and uncorrupted observations: reference and vs estimated inflow hydrograph (a) and observed (uncorrupted) and vs modeled water levels (b). The residuals between reference and estimated values are also reported.



**Figure 14.** Secchia River inflow and corrupted observations: reference and vs estimated (with 95% credibility interval) inflow hydrograph (a) and observed (corrupted) and vs modeled water levels (b). The residuals between reference and estimated values are also reported.

With the aim of exemplifying the efficiency of the proposed parallel inverse procedure, some For this case, some details about the computational characteristics times are furnished for this test case, whose main features are reported in Table 5.

The computational time of the whole inflow hydrograph simulation (72 hours) is 9.62 minutes, whereas the simulations for evaluating the Jacobian matrix and testing parameters from 2 till 37 required a computational time progressively lower than 9.62 minutes, thanks to the restart option illustrated in the Sect. 3. In order to evaluate the total time required by the

**Table 3.** Secchia River inflow: initial and estimated structural parameters and epistemic error variance.

|  |           | No random errors | Random errors |
|--|-----------|------------------|---------------|
| $\theta$ (m <sup>6</sup> s <sup>-3</sup> )     | Initial   | 1.00E-10         | -             |
|  | Estimated | 3.97E-6          | -             |
| $\sigma_R^2$ (m <sup>2</sup> )                 | Initial   | -                | 1.00E-4       |
|  | Estimated | -                | 1.11E-3       |
| $\sigma_S^2$ (m <sup>6</sup> s <sup>-2</sup> ) | Initial   | -                | 5.00E+2       |
|  | Estimated | -                | 1.38E+1       |
| $l$ (s)  | Initial   | -                | 4.32E+4       |
|  | Estimated | -                | 3.88E+4       |

**Table 4.** Secchia River inflow: Nash-Sutcliffe  $E_h$ , root mean square error RMSE and error in the peak discharge  $E_p$  values.

|                  | $E_h$ (-) | RMSE (m <sup>3</sup> /s) | $E_p$ (%) |
|------------------|-----------|--------------------------|-----------|
| No random errors | 99.99     | 0.13                     | -0.02     |
| Random errors    | 99.44     | 16.57                    | 2.89      |

inverse procedure, it is noteworthy that dealing with an HPC cluster the global run time depends on the number of the available GPUs. However, this test was performed using 10 GPUs and the computational cost of the 609 runs was about 13 hours. Since the implemented procedure that manages the interaction between host and server can be used for different HPC cluster, the availability of a cluster equipped with  $N_p$  GPUs would have allowed the estimation of the flood wave in about 8 hours. On the other side, the adoption of the serial bgaPEST procedure and the PARFLOOD code as routing model would have required about 4 days of computations that means about 8 times higher slower than the parallel procedure here proposed. Particularly interesting is the hypothetical evaluation of the computational time for a serial BGA procedure and the adoption of a serial CPU code as forward hydraulic model. Vacondio et al. (2014) pointed out that the PARFLOOD code led to speedup up to two order of magnitude if compared to a serial CPU code. Therefore, if a serial BGA procedure and the GPU forward model would have required about 4 computational days, the inverse problem solution with a serial forward code would ended in 400 computational days, making the use of the inverse procedure practically infeasible.

**Table 5.** Secchia River inflow: characteristics of the simulation.

|   |               |
|---|---------------|
| Number of parameters $N_p$  | 37            |
| Physical total time of the inflow hydrograph  | 72 hours      |
| Physical total time of the run testing the 1 <sup>st</sup> parameter $p_1$ , assuming 100 hours for reaching the steady state condition | 172 hours     |
| Computational time of the whole inflow hydrograph simulation (72 hours)   | 9.62 minutes  |
| Computational time of the run testing the 1 <sup>st</sup> parameter (172 hours)   | 19.38 minutes |
| Number of the BGA iterations $N_i$ for the model parameter estimation   | 4             |
| Number of the BGA iterations $N_o$ for the structural parameter estimation  | 4             |
| Total number of simulations $N_t$ (Eq. 15)  | 609           |

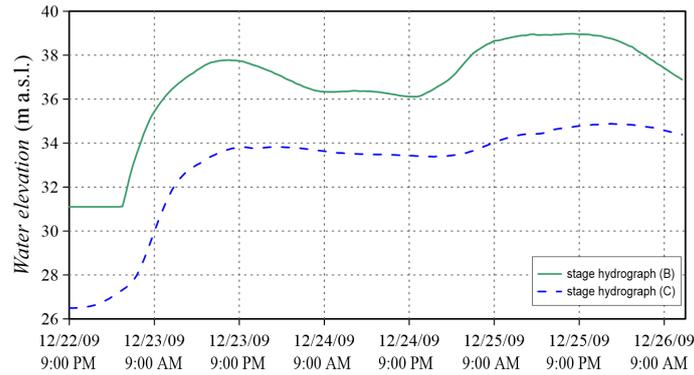
## 5 Reconstruction of a historical event: the December 2009 flood wave on the Secchia River

After the model validation assessed in the previous section, The inverse procedure is now validated adopted in the framework of a real field application, by investigating the December 2009 flooding event on the Secchia River, which is one of the three most significant events occurred in the last ten years in this river. The Interregional Agency for the Po River (AIPo) monitored the river and provided the water stage hydrographs recorded in the two gauging stations indicated in Fig. 11 with letters B and C, respectively. As shown in Fig. 15, the recorded registered water levels present more than a one rising and recession limb, and thus, besides the challenges related to a real field application, this test aims at addressing also the estimation of an inflow with multiple two peaks. In order to estimate the discharge at section A (Fig. 11), the water levels registered recorded at point B and C were assumed as observations and downstream boundary condition, respectively. The event was simulated from 9 p.m. of the 22<sup>nd</sup> December 2009, till 12 a.m. of the 26<sup>th</sup> December, for a total duration of 87 hours. The water levels were recorded every 0.5 hours, and thus the observations consist of 174 values. Conversely, whereas the unknown inflow hydrograph was discretized into 88 parameters (one per hour,  $N_p=88$ ). every hour, in a result of 88 parameters to be estimated ( $N_p=88$ ). The stage hydrographs, adopted as observations and downstream boundary condition, respectively, are shown in Fig. 15.

The studied domain is analogous to the one the same previously adopted for a synthetic inflow, and thus, the reader is kindly referred to Sect. 4.2 for the information about bathymetry, initial condition, and roughness configuration.

As before, the parameters were estimated in a logarithmic space and their initial values were calculated adopting the line-search tool of the bgaPEST (Fienen et al. (2013)). The prior pdf was described by means of a Gaussian variogram covariance function; the initial and estimated structural parameters are reported in Table 6.

Figure 16 shows the estimated flood wave (and the 95% credibility interval), which presents an irregular shape and two main peaks, as it could be expected from the observed stage hydrograph. Moreover, an additional small intermediate peak is captured



**Figure 15.** December 2009 registered recorded stage hydrographs on the Secchia River at, referred to sections B and C, respectively.

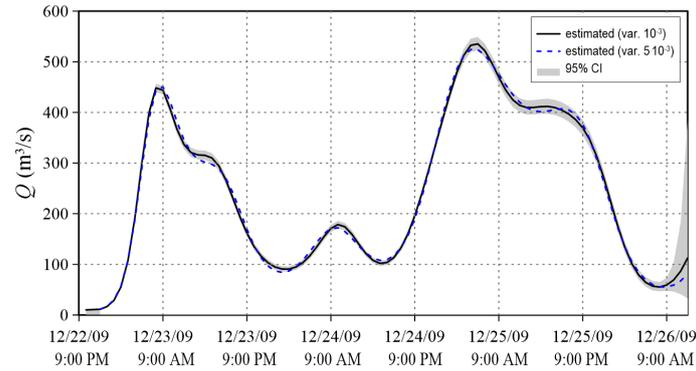
**Table 6.** Secchia 2009 event: initial and estimated structural parameters.

|           | $\sigma_S^2$ ( $\text{m}^6 \text{s}^{-2}$ ) | $l$ (s) |
|-----------|---|---------|
| Initial   | 5.00E+2                                     | 6.48E+4 |
| Estimated | 1.49E+1                                     | 3.36E+4 |

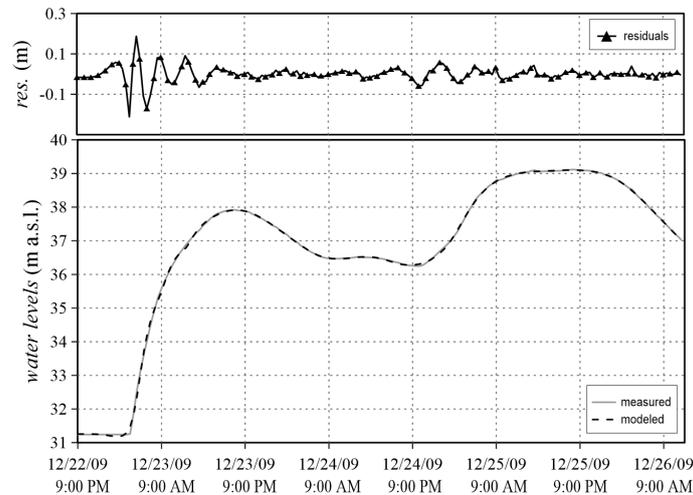
that was not so evident from the registered water levels at section B (Fig. 15), even if a little pronounced local maximum can be seen around 3 p.m. of the 24<sup>th</sup> December 2009. The resulting ed flood wave is included in the 95% credibility interval, and moreover the solution presents neither instabilities nor oscillations. During the computation, the variance of the epistemic error was assumed equal to  $10^{-3} \text{ m}^2$ ; as shown in Sect. 4, this means considering the observed water levels corrupted with random errors with maximum deviations of  $\pm 0.05 \text{ m}$ . In Fig. 16, the flood wave estimated by increasing reducing the variance of half an order of magnitude is also depicted (dotted line): the solution appears slightly smoothed in a few points, but substantially similar to the inflow resulting with the higher smaller variance, which is thus considered as the estimated inflow of the studied event. The comparison between modeled and measured water levels at section B is presented in Fig. 17: it is relevant to note that the residuals between the two hydrographs trends are mostly less than 2 cm and only in a few points of the first rising limb they reach the highest value of 18 cm.

With the aim of validating the methodology for this real application, it is noteworthy that the upstream section of the river is located immediately downstream a flood control reservoir equipped with water level sensors. Therefore, the "reference" discharge hydrograph has been obtained from the dam geometrical data (i.e. number and dimension of the bottom openings, crest length of the spillway, etc.) and the recorded water levels adopting the classic hydraulic theory of sluice gates and spillways.

Due to the uncertainty in evaluating the discharge coefficients and to the fact that during flood events a large amount of wood debris reduces the outflow discharge from the bottom openings (especially during the depletion phase) and interferes with the overflow spillway, the discharge hydrograph has been calculated adopting equally likely coefficients (Fig. 18). The flood wave



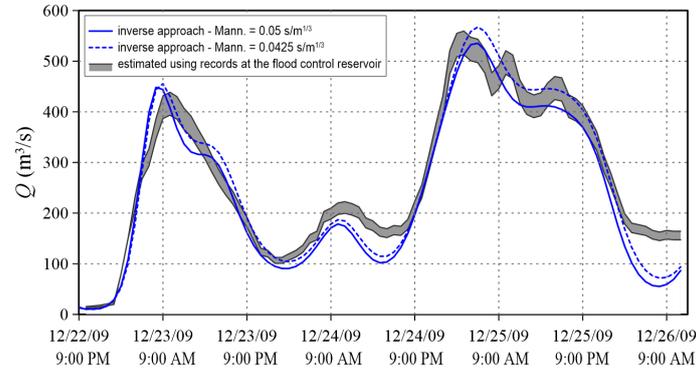
**Figure 16.** Secchia 2009 event: estimated inflow hydrographs assuming the epistemic error variance equal to  $10^{-3} \text{ m}^2$  and  $5 \cdot 10^{-3} \text{ m}^2$ , respectively. The 95% credibility interval is referred to the simulation with the epistemic variance equal to  $10^{-3} \text{ m}^2$ .



**Figure 17.** Secchia 2009 event: observed and modeled water levels at section B. The residuals between recorded reference and estimated values are also reported.

estimated by the inverse procedure is in good agreement with the one calculated using the flood reservoir data; the main differences are after the highest peak, which is well reproduced, although the inverse methodology provides a smoother solution. For this real application, even if the river roughness coefficient was already calibrated in previous studies (Vacondio et al. (2016)), an additional inverse Bayesian estimation was performed with a different value, in order to assess the effect of this coefficient on the solution. Particularly, the Manning coefficient originally set to  $0.05 \text{ s/m}^{1/3}$  was decreased by 15% ( $0.0425 \text{ s/m}^{1/3}$ ), as for example can happen due to seasonal changes in vegetation. As shown in Fig. 18, the estimated flood waves are similar and the highest difference, which is in correspondence with the main peak, is less than 6%. Therefore, the influence of assuming a “wrong” roughness coefficient is less than linear in the discharge estimation. Despite all the involved approximations, this

comparison confirms that the proposed inverse procedure is capable of estimating inflow hydrographs with multiple peaks and irregular shapes in real rivers.



**Figure 18.** Secchia 2009 event: comparison among the inflow hydrographs obtained from the inverse procedure using two different Manning coefficients, and the envelope of different solutions obtained using the records at the flood control reservoir.

## 6 Conclusions

5 In this work the inverse problem of estimating the unknown inflow hydrograph in an upstream-ungauged section, having water level information only in downstream sites, has been solved by means of a Bayesian methodology. The key aspects in the solution of this problem have been the adoption of a parallel 2D-SWE code running on GPUs and the performance of the simulations over a HPC cluster. The parallelization of the runs useful for the Jacobian matrix computation and the implementation of an *ad hoc* procedure, which allows taking advantage of any HPC cluster with GPUs, ~~by means of the~~

10 ~~protocol SSH~~, have provided a remarkable reduction of the computational costs: ~~the more GPUs are available on the cluster, the less time is required for the parameter estimation~~. For a ~~considered~~ test case, this parallel procedure reduced the computational time ~~by of~~ a factor of 8 against running the 2D-SWE code on a single GPU. Furthermore, the analysis of the runtimes has highlighted that the use of a parallel hydraulic forward routing model is the *conditio sine qua non* for solving this

15 type of inverse problem, whereas the adoption of a serial code would lead to inadmissible computational times. The inverse procedure has been validated considering two different natural rivers; in both tests, no instabilities, due to the adopted inverse procedure or to the availability of a stable, fast and accurate forward hydraulic model, arose. Moreover, the obtained results have highlighted that the implemented procedure well estimates the unknown inflow hydrographs with different and irregular shapes and in presence of corrupted observations: quantitative indicators have proved the accuracy of the methodology. In all

20 the presented tests, the ~~resulting ed~~ Nash-Sutcliff efficiency criterion exceeded 99%, the error in the peak discharge was less than 3% and the RMSE error less than 2%. Finally, the proposed inverse procedure allowed the estimation of a historical flood wave characterized by the presence of ~~reaching causing~~ instabilities in the solution. ~~The test cases~~

were simulated taking advantage of the HPC cluster of the University of Parma. However, since the implemented procedure is general, it is possible to adopt clouds of GPUs or on-line mini clusters, which are now common and accessible to everyone. The adopted Bayesian software (bgaPEST) is open access and 2D-SWE models are a quite common tools for practitioners, even if till now few of them are fast enough to perform the necessary simulations with a reasonable computing time. Therefore, the 2D coupled methodology here proposed can be adopted in the near future also by professional hydrologists involved for example in the design of hydraulic infrastructures as well as for engineers working on water resource management (i.e. irrigation systems, hydroelectric power stations, etc.) or forensic activities. Future development of the methodology will focus on the possibility of reconstructing the flood waves also in presence of levee breaches and flooding outside the river region, where the adoption of a 2D-SWE model is mandatory.

10 *Competing interests.* The authors declare that they have no conflict of interest.

*Acknowledgements.* This work was partially supported by Ministry of Education, Universities and Research under the Scientific Independence of young Researchers project, grant number RBSI14R1GP, CUP code D92I15000190001. This research benefits from the HPC (High Performance Computing) facility of the University of Parma, Italy. Interregional Agency for the Po River (AIPo) is also gratefully acknowledged for providing data. The authors are grateful to the editor, the anonymous reviewer and Dr. A.D. Koussis for the valuable suggestions on the early version of this manuscript.

## References

- Beven, K. J.: Rainfall-runoff modelling: the primer, John Wiley & Sons, 2011.
- Bruen, M. and Dooge, J.: Harmonic analysis of the stability of reverse routing in channels, *Hydrology and Earth System Sciences*, 11, 559–568, 2007.
- 5 Butera, I., Tanda, M. G., and Zanini, A.: Simultaneous identification of the pollutant release history and the source location in groundwater by means of a geostatistical approach, *Stochastic Environmental Research and Risk Assessment*, 27, 1269–1280, 2013.
- Das, A.: Reverse stream flow routing by using Muskingum models, *Sadhana*, 34, 483–499, 2009.
- Doherty, J. E.: PEST, Model-Independent Parameter Estimation – User Manual, sixth ed., Tech. rep., Watermark Numerical Computing, Brisbane, Australia, 2016.
- 10 Dooge, J. and Bruen, M.: Problems in reverse routing, *Acta Geophysica Polonica*, 53, 357, 2005.
- D’Oria, M. and Tanda, M. G.: Reverse flow routing in open channels: A Bayesian Geostatistical Approach, *Journal of hydrology*, 460, 130–135, 2012.
- D’Oria, M., Mignosa, P., and Tanda, M. G.: Bayesian estimation of inflow hydrographs in ungauged sites of multiple reach systems, *Advances in Water Resources*, 63, 143–151, 2014.
- 15 D’Oria, M., Mignosa, P., and Tanda, M. G.: An inverse method to estimate the flow through a levee breach, *Advances in Water Resources*, 82, 166–175, 2015.
- Eli, R., Wiggert, J., and Contractor, D.: Reverse flow routing by the implicit method, *Water Resources Research*, 10, 597–600, 1974.
- Fienen, M., Hunt, R., Krabbenhoft, D., and Clemo, T.: Obtaining parsimonious hydraulic conductivity fields using head and transport observations: A Bayesian geostatistical parameter estimation approach, *Water resources research*, 45, 2009.
- 20 Fienen, M. N., Clemo, T., and Kitanidis, P. K.: An interactive Bayesian geostatistical inverse protocol for hydraulic tomography, *Water Resources Research*, 44, 2008.
- Fienen, M. N., D’Oria, M., Doherty, J. E., and Hunt, R. J.: Approaches in highly parameterized inversion: bgaPEST, a Bayesian geostatistical approach implementation with PEST: documentation and instructions, Tech. rep., US Geological Survey, 2013.
- Glickman, M. E. and Van Dyk, D. A.: Basic bayesian methods, *Topics in Biostatistics*, pp. 319–338, 2007.
- 25 Hoeksema, R. J. and Kitanidis, P. K.: An application of the geostatistical approach to the inverse problem in two-dimensional groundwater modeling, *Water Resources Research*, 20, 1003–1020, 1984.
- Kitanidis, P. K.: Quasi-linear geostatistical theory for inversing, *Water Resources Research*, 31, 1995.
- Kitanidis, P. K. and Vomvoris, E. G.: A geostatistical approach to the inverse problem in groundwater modeling (steady state) and one-dimensional simulations, *Water resources research*, 19, 677–690, 1983.
- 30 Koussis, A., Mazi, K., Lykoudis, S., and Argiriou, A.: Reverse flood routing with the inverted Muskingum storage routing scheme, *Natural Hazards and Earth System Sciences*, 12, 217, 2012.
- Koussis, A. D. and Mazi, K.: Reverse flood and pollution routing with the lag-and-route model, *Hydrological Sciences Journal*, 61, 1952–1966, 2016.
- Leonhardt, G., D’Oria, M., Kleidorfer, M., and Rauch, W.: Estimating inflow to a combined sewer overflow structure with storage tank in real time: evaluation of different approaches, *Water Science and Technology*, 70, 1143–1151, 2014.
- Michalak, A. M., Bruhwiler, L., and Tans, P. P.: A geostatistical approach to surface flux estimation of atmospheric trace gases, *Journal of Geophysical Research: Atmospheres*, 109, 2004.

- Nash, J. E. and Sutcliffe, J. V.: River flow forecasting through conceptual models part I—A discussion of principles, *Journal of hydrology*, 10, 282–290, 1970.
- Saghafian, B., Jannaty, M., and Ezami, N.: Inverse hydrograph routing optimization model based on the kinematic wave approach, *Engineering Optimization*, 47, 1031–1042, 2015.
- 5 Snodgrass, M. F. and Kitanidis, P. K.: A geostatistical approach to contaminant source identification, *Water Resources Research*, 33, 537–546, 1997.
- Szymkiewicz, R.: Solution of the inverse problem for the Saint Venant equations, *Journal of Hydrology*, 147, 105–120, 1993.
- Vacondio, R., Dal Palù, A., and Mignosa, P.: GPU-enhanced finite volume shallow water solver for fast flood simulations, *Environmental modelling & software*, 57, 60–75, 2014.
- 10 Vacondio, R., Aureli, F., Ferrari, A., Mignosa, P., and Dal Palù, A.: Simulation of the January 2014 flood on the Secchia River using a fast and high-resolution 2D parallel shallow-water numerical scheme, *Natural Hazards*, 80, 103–125, 2016.
- Vacondio, R., Palù, A. D., Ferrari, A., Mignosa, P., Aureli, F., and Dazzi, S.: A non-uniform efficient grid type for GPU-parallel Shallow Water Equations models, *Environmental Modelling & Software*, 88, 119 – 137, 2017.
- Zucco, G., Tayfur, G., and Moramarco, T.: Reverse flood routing in natural channels using genetic algorithm, *Water resources management*, 15 29, 4241–4267, 2015.