

# Hydrograph estimation at upstream ungauged sections on the Secchia River (Italy) by means of a parallel Bayesian inverse methodology

Alessia Ferrari<sup>1\*</sup>, Marco D’Oria<sup>1</sup>, Renato Vacondio<sup>1</sup>, Paolo Mignosa<sup>1</sup> and Maria Giovanna Tanda<sup>1</sup>

<sup>1</sup>Department of Engineering and Architecture, University of Parma, Parco Area delle Scienze 181/A, 43124 Parma, Italy

**Abstract.** In this work, we present a reverse flow routing procedure, which allows estimating discharge hydrographs at upstream ungauged stations by means of information available at downstream monitored sites. The reverse routing problem is solved adopting a Bayesian Geostatistical Approach (BGA). In order to capture the complex hydrodynamic field typical of many real cases of rivers including large floodable areas, meanwhile overcoming the computational time limitations, we adopted as forward model a self-developed 2D-SWE parallel numerical model (PARFLOOD) that allows achieving ratio of physical to computational time of about 500-1000. To exploit the computational capabilities of modern GPU cluster, a parallel procedure to estimate the Jacobian matrix required by the BGA approach has been implemented. The inflow hydrograph in a river reach with several meanders and floodplains has been estimated in “only” 13 hours using a HPC cluster with 10 P100 Nvidia GPUs.

## 1. Introduction

The knowledge of discharge hydrographs in natural river sections is fundamental for water resource and risk management, design of new structures, etc. Furthermore, due to the different scopes, discharge information may potentially be required in any section along the river reach. Particularly challenging is the discharge reconstruction at the entrance of a river reach, where no observations (about discharge or water levels) are available neither in the section of interest nor in any upstream ones.

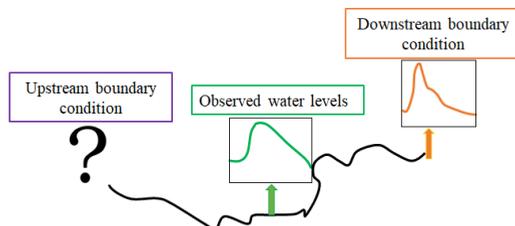
One of the approaches to solve this problem consists in the reverse flow routing process, which is a numerical procedure that allows evaluating the upstream flood wave starting from the knowledge of a downstream recorded hydrograph, together with the hydraulic characteristics of the river reach. [1] estimated discharge hydrographs in ungauged sections of natural channels by means of a Genetic algorithm coupled with a simplified hydrological routing model. However, the assumption of a Pearson type III distributed hydrograph limits the application of the procedure in reconstructing real flood waves which often have irregular shapes. [2] coupled a Genetic Algorithm with a 1D forward hydraulic model under the

---

\* Corresponding author: [alessia.ferrari@unipr.it](mailto:alessia.ferrari@unipr.it)

kinematic approximation, in order to identify the upstream hydrograph, given the downstream one, of a river reach. However, the estimations are limited to a rectangular prismatic channel and no errors were added to the downstream observations. Conversely, [3] solved the reverse flow routing problem for a real river and an irregular discharge hydrograph. They adopted a Bayesian Geostatistical Approach (BGA), introduced by [4] for groundwater problems, which considers a flow hydrograph as a statistical continuous random function that presents autocorrelation and accounts for uncertainties. In the framework of 1D numerical schemes, the authors showed the capability of the BGA methodology to estimate the inflow in an upstream ungauged section having discharge information (also corrupted) only in a downstream one. [5] further extended the method in order to adopt, as downstream observations, stage hydrographs instead of discharge ones. However, in many real cases of rivers including large floodable areas, where the 1D schematization does not hold, it is necessary to adopt a 2D Shallow Water model to capture the complex hydrodynamic field, even if this poses the drawback of the high computation cost, with respect to a 1D scheme.

This work extends the BGA methodology for reverse flow routing to 2D test cases, in order to model natural rivers with complex geometry, including flood plains and meanders. The transition of the application of the inverse procedure from 1D to 2D forward models is not straightforward, due to both the choice of the 2D Shallow Water scheme and the optimization techniques necessary for reducing the otherwise unfeasible computational time. Therefore, the stable, accurate and fast PARFLOOD 2D GPU finite volume model [6,7] is adopted as forward model and it is efficiently coupled with the Bayesian methodology. In this way, the inverse procedure allows estimating the discharge hydrograph in an upstream ungauged river section, having water level information only in a downstream observation site (Figure 1) also for 2D applications.



**Fig. 1.** Definition of the considered inverse problem.

## 2. The Bayesian Geostatistical Approach

The bgaPEST software ([8]) solves inverse problems, in which the unknown parameters are correlated one another in space or time, by maximizing the posterior probability density function, which according to the Bayes' theorem 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 (the time values of the inflow hydrograph, in this case),  $\mathbf{y}$  is the vector of the measured data (the downstream water level observations),  $p(\mathbf{s}|\mathbf{y})$  the posterior probability density function (pdf) of  $\mathbf{s}$  given  $\mathbf{y}$ ,  $L(\mathbf{y}|\mathbf{s})$  the likelihood function and  $p(\mathbf{s})$  the prior probability density function of  $\mathbf{s}$ .

The prior pdf provides a soft knowledge about the structure of the unknown parameters, making use of temporal autocorrelation functions. Meanwhile the likelihood function quantifies how a given set of parameters succeeds in reproducing the observations  $\mathbf{y}$ , by calculating the Jacobian matrix ([5]). Therefore, the evaluation of the likelihood function requires a forward model of the considered river reach that is able to describe the hydraulic routing process, and that provides for a given inflow hydrograph (set of parameters), the modelled water levels extracted in the same place and time of the available observations  $\mathbf{y}$ .

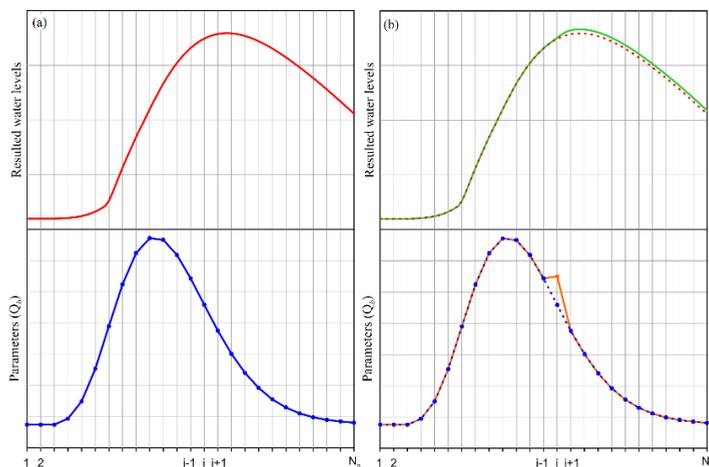
The unknown parameters ( $s$  vector in Equation (1)) consist in the  $N_p$  values that discretize the upstream discharge hydrograph, and represent the main final result of the estimation process. Additionally, so-called structural parameters (also known as hyperparameters), associated with the prior pdf and the likelihood function, are part of the inverse procedure. These parameters controls the shape or structure of the unknowns characterizing the covariance of both the prior and the uncertainty related to the observations ([8]). The structural parameters are also estimated by the inverse algorithm.

The bgaPEST software solves the inverse problem by means of the following steps:

1. The vector of the unknown parameters  $s$  is initialized by assuming starting values (even constant) coherent with the considered river. Additionally, the structural parameters are set as to guarantee a flat solution (complexity is introduced during the optimization process only if supported by the data);
2. The hydraulic forward model propagates the discharge hydrograph defined at point 1 as upstream boundary condition. At the end of the simulation, the resulted water levels in the observation site are extracted;
3. A new set of parameters  $s$  is estimated after having computed the Jacobian matrix, which requires, in a finite differential approximation, as many runs of the forward model as the number of parameters  $N_p$ . The water levels extracted from simulation  $i$  are the result of the routing of the discharge hydrograph, where just the parameter  $i$  has been varied of a known quantity, with respect to the value at point 2 (Figure 2);
4. The parameters resulted at point 3 are adopted as discharge hydrograph and the operations at points 2-3 are repeated until convergence or the maximum number of iteration  $N_i$  (usually  $N_i = 4-5$ ) is reached. Thus, a new set of  $N_p$  parameters is estimated;
5. The structural parameters are estimated until convergence or the maximum number of iteration  $N_o$  (usually  $N_o = 4-5$ ) is reached.

Since the problem is non-linear, the inverse procedure requires  $N_t$  run of the forward model, according to the following relation ([8]):

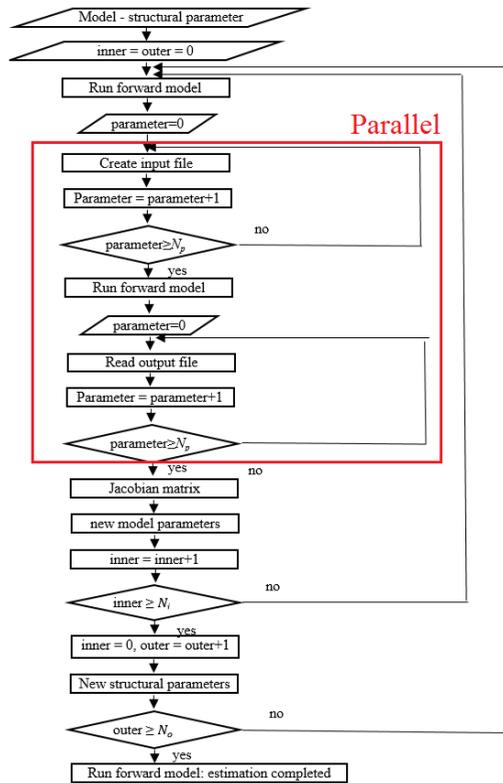
$$N_t = (N_p + 1) \cdot N_o \cdot N_i + 1 \tag{2}$$



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

Most of the  $N_t$  runs in Equation (2) derive from the  $N_p$  ones needed for the Jacobian matrix computation. However, since each run checks the sensitivity of the observations to the variation of a single parameter (run  $i$  tests the sensitivity on parameter  $i$ ), the solution of a test does not have effects on the solution of other tests: as a consequence, the  $N_p$  runs are independent one another and can be potentially performed in parallel. Due to the fact that the

adopted PARFLOOD ([6,7]) forward model solves the 2D-SWE on a finite volume scheme, and it is parallelized on GPU, the computational burden can be reduced by taking advantage of High Performance Computing (HPC) clusters equipped with several GPUs. For this reason, the original bgaPEST software has been modified as to run parallel simulations on any cluster. In the serial version, a main *do* loop over the parameters manages the operations that, for each parameter  $i$ , write the input file of the forward model, execute the numerical forward model and read the results. Conversely, the parallel version here implemented allows running all the  $N_p$  independent simulations at the same time, before computing the Jacobian matrix and then estimating a new set of parameters  $s$ . The parallelism introduced in the bgaPEST procedure is highlighted in the flow chart in Figure 3.



**Fig. 3.** Flow chart of the parallel version of the bgaPEST.

The parallel version of the bgaPEST is executed on the CPU of a classical PC, and it performs, for each of the  $N_i$  simulations, the following operations (Figure 4):

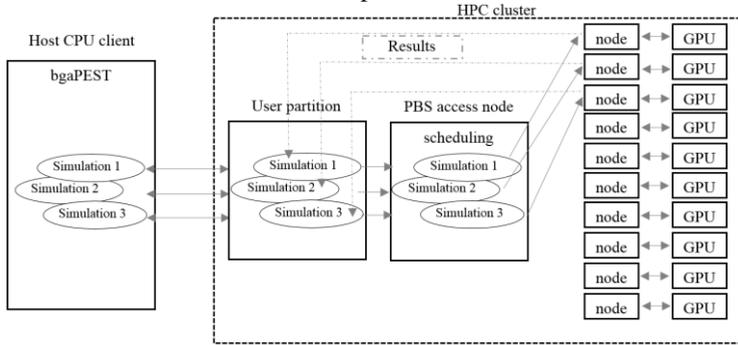
1. The input files for the PARFLOOD model are organized. All the  $N_p$  simulations present the same bathymetrical and roughness maps, initial condition, but a different upstream boundary condition;
2. All the simulations are sent to the server by means of the Secure Shell network protocol (SSH). In the HPC cluster the access node schedules all the submitted jobs and assigns each one to a GPU node where the run will be executed. At the end, the resulted water levels are extracted;
3. All the results are moved from the remote server to the local PC, where a new parameter estimation is done.

Since the variation of parameter  $i$  causes effects only after time  $t_i$ , each of the  $N_p$  simulations can be performed from time  $t_i$ , further reducing the computational times. The

decrease of the theoretical computational time can be evaluated as the difference between the run time for simulations performed from  $t_{start}$  until  $t_{end}$  and those run from  $t_i$  to  $t_{end}$ , as follows:

$$T = N_p(N_p - 1)\Delta t - \left[ (N_p - 1)\Delta t + \sum_{i=2}^{N_p} [N_p - (N_i - 1)]\Delta t \right] \quad (3)$$

where  $\Delta t$  represents the time interval between parameters  $i$  and  $i+1$ .



**Fig. 4.** Schematization of the data transfer from the client to the server assuming three parameters and thus three simulations to be parallel performed.

### 3. Test cases

The method was tested by estimating upstream discharge hydrographs in the reach of the Secchia River (Northern Italy) connecting the flood control reservoir of Rubiera (point A in Figure 5) to the gauging station of Ponte Bacchello (point C in Figure 5). In this reach the river is characterized by wide flood plains and floodable areas. The water levels at the intermediate gauging station of Ponte Alto (point B in Figure 5) were adopted as observations. Historical events recorded at those stations allowed the model calibration ([7]).

The overall studied domain was discretized by means of a non-uniform grid ([6]), in a result of  $77 \cdot 10^3$  computational cells.

The procedure proposed by [9] was adopted in order to obtain a reference solution for the method validation. Firstly, the selected inflow discharge  $Q^{act}$  was routed from the upstream section A until a section located far downstream C, and the resulted water level hydrographs were extracted in sites B and C. Then, in the inverse procedure, the water levels in sites B and C (resulted from the previous step) were assumed as observations and downstream boundary condition, respectively; the inverse methodology estimated the inflow discharge  $Q^{est}$  assuming that no information was available on the discharge/stage time series in site A.

The availability of a reference solution allows providing quantitative information about the accuracy of the inverse methodology, by calculating two indicators. Firstly, the root mean square error,  $RMSE$  was evaluated as follows:

$$RMSE = \sqrt{\frac{\sum_{i=1}^N (Q_i^{act} - Q_i^{est})^2}{N}} \quad (4)$$

where  $N$  is the total number of parameters (data of the discharge hydrograph),  $Q_i^{act}$  and  $Q_i^{est}$  are the  $i$ -th actual and estimated inflow values, respectively.

Secondly, the error in the peak discharge  $E_p$  was assessed as:

$$E_p = \left[ \frac{Q_p^{est}}{Q_p^{act}} - 1 \right] \cdot 100 \quad (5)$$

where  $Q_p^{est}$  and  $Q_p^{act}$  denote the peak discharge value of the estimated and actual hydrographs, respectively.

### 3.1. Synthetic Gamma-distributed hydrograph

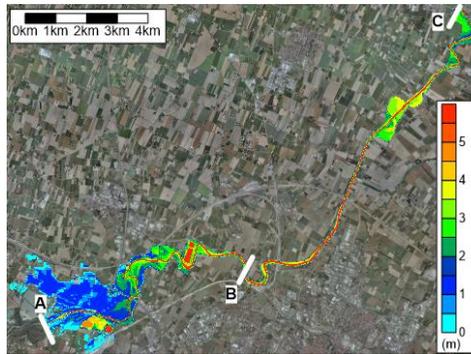
The first test concerns the estimation of an inflow hydrograph assumed with Gamma distribution ([5]):

$$Q(t) = A + B \cdot \frac{1}{k^b \Gamma(b)} t^{b-1} e^{-\frac{t}{k}} \tag{6}$$

where  $t$  denotes the time,  $A$  the base flow (150 m<sup>3</sup>/s),  $B$  the volume above the base flow (75·10<sup>6</sup> m<sup>3</sup>),  $\Gamma(b)$  the gamma function defined through the parameters  $b$  (8.5) and  $k$  (9000 s) that denote the shape and the scale parameter, respectively.

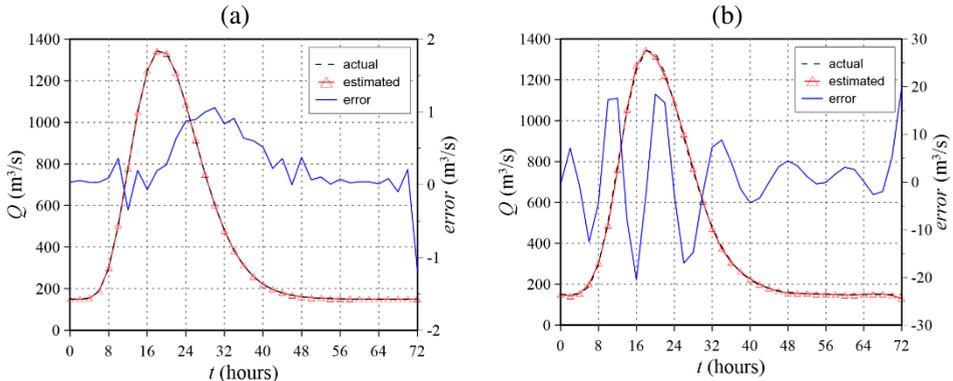
The flow hydrograph is characterized by a peak value of about 1350 m<sup>3</sup>/s, its duration was limited to 72 hours and it was discretized using 2 hours steps, resulting in 37 parameters to be estimated ( $N_p=37$ ). The observations, which were discretized every 0.5 hours, were assumed in a first estimation as free of errors, whereas in a second estimation they were corrupted with random errors uniformly distributed with variance 10<sup>-3</sup> m<sup>2</sup>.

Figure 5 shows the map of the water depth at the arrival of the flood wave pick: most of the areas near the upstream section are flooded, as well as the flood plains along the river.



**Fig. 5.** Map of the water depth at the arrival of the flood pick. The upstream (A) and downstream (C) boundary conditions and the intermediate observation site (B) are indicated.

Figure 6 depicts the comparison between the actual and the estimated inflow hydrographs using both the free of error observations (Figure 6-a) and the corrupted ones (Figure 6-b). In both cases, the estimated flood wave matches very well the actual one: only the last values in (b) slightly differ because fewer downstream observations contain information about the unknown inflow hydrograph.



**Fig. 6.** Gamma-distributed inflow discharge: comparison between actual and estimated hydrographs with free of error (a) and corrupted (b) observations. The differences between actual and estimated values (error) are also reported.

Beside this qualitative assessment of the inverse methodology, quantitative information is provided by means of the  $RMSE$  and  $E_p$  indicators (Table 1). The errors in the peak values  $E_p$  are almost negligible in both cases and the  $RMSE$  is less than  $1 \text{ m}^3/\text{s}$  with free of errors observations and about  $9 \text{ m}^3/\text{s}$  with corrupted observations: these values confirm the accuracy of the procedure in estimating the overall shape and peak of the inflow hydrographs.

**Table 1.** Gamma-distributed inflow:  $RMSE$  and  $E_p$  values.

	$RMSE \text{ (m}^3/\text{s)}$	$E_p \text{ (%)}$
<i>Uncorrupted observations</i>	0.49	-0.014
<i>Corrupted observations</i>	9.20	0.199

Each inflow hydrograph estimation required 609 runs of the PARFLOOD forward model ( $N_p = 37, N_i = 4, N_o = 4$ ) and 13 hours of computational time. The simulations were performed taking advantage of the HPC cluster of the University of Parma, equipped with 10 NVIDIA® Tesla® P100 GPU. Conversely, the adoption of the serial bgaPEST would have required about 4 computing days, by retaining the same accelerated forward model. Finally, the adoption of both a serial bgaPEST procedure and a 2D-CPU forward code, and assuming a speed-up of two order of magnitude between PARFLOOD and serial CPU codes, would have required about 400 computing days, meaning the practical impossibility to resolve the inverse problem.

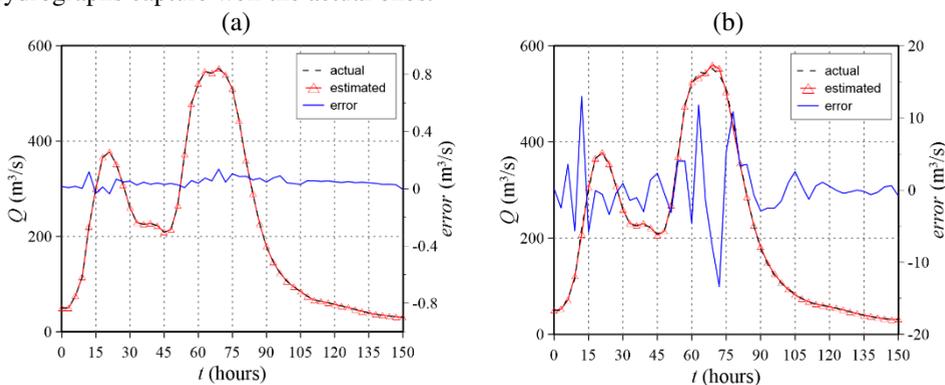
### 3.2. Syntetic hydrograph with two peaks

The second test focuses on the estimation of a real flood wave with a bimodal shape, occurred in December 2009 on the Secchia River ([7]).

The discharge hydrograph (Figure 7) is characterized by a total duration of 150 hours and presents a first peak value of about  $380 \text{ m}^3/\text{s}$  and a second one of about  $550 \text{ m}^3/\text{s}$  45 hours later. As for the previous test, the observations were firstly assumed free of errors and then corrupted by random errors uniformly distributed with variance  $9.7 \cdot 10^{-4} \text{ m}^2$ .

The inflow hydrograph was discretized using 3 hours steps ( $N_p = 51$ ), whereas the observation stage hydrograph was discretized every 0.5 hours.

Figure 7 shows the comparison between the actual and the estimated inflow hydrographs with free of error (Figure 7-a) and corrupted (Figure 7-b) observations. Again, the estimated hydrographs capture well the actual ones.



**Fig. 7.** Two peaks inflow: comparison between actual and estimated hydrographs with free of error (a) and corrupted (b) observations. The differences between actual and estimated values (error) are also reported.

In both estimations, the error in the peak flow  $E_p$  is around 1% and the  $RMSE$  about 4 m<sup>3</sup>/s: these values show the capability of the procedure in reconstructing also flood waves with more than one peak.

**Table 2.** Two peak inflow:  $RMSE$  and  $E_p$  values.

	$RMSE$ (m <sup>3</sup> /s)	$E_p$ (%)
<i>Uncorrupted observations</i>	0.05	-0.025
<i>Corrupted observations</i>	4.29	1.447

## 4. Conclusions

In the present work a Bayesian methodology, that allows the estimation of discharge hydrographs in ungauged sections located at the upstream section of a river reach, has been extended to 2D cases by adopting a 2D-SWE parallel forward routing model. The validation on a natural river has shown that the procedure well estimates unknown inflow hydrographs with different shapes and also in presence of corrupted observations. The parallelization of the Jacobian matrix computation and the possibility of taking advantage of any HPC cluster with GPUs, allowed achieving the highest reduction in the computational costs: the more GPUs are available, the less time is required for computation. For a considered case, this parallel procedure reduced the computational time of a factor 8 against running the 2D-SWE code on a single GPU. Furthermore, the analysis of the runtimes has highlighted that the adoption of a serial code would lead to inadmissible computational times.

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.

## References

1. G. Zucco, G. Tayfur, T. Moramarco, *Water Resour. Manag.* **29(12)**, 4241-4267 (2015)
2. B. Saghafian, M.H. Jannaty, N. Ezami, *Eng. Optim.* **47(8)**, 1031-1042 (2015)
3. M. D'Oria, M.G. Tanda, *J. Hydrol.* **460**, 130-135 (2012)
4. P.K. Kitanidis, *Water Resour. Res.* **31(10)**, 2411-2419 (1995)
5. M. D'Oria, P. Mignosa, M.G. Tanda, *Adv. Water Resour.* **63**, 143-151 (2014)
6. R. Vacondio, A. Dal Palù, A. Ferrari, P. Mignosa, F. Aureli, S. Dazzi, S., *Environ. Modell. Softw* **88**, 119-137 (2017)
7. R. Vacondio, F. Aureli, A. Ferrari, P. Mignosa, A. Dal Palù, *Nat. Hazards* **80(1)**, 103-125 (2016)
8. M.N. Fienen, M. D'Oria, J.E. Doherty, R.J. Hunt, *US Geological Survey*, No. 7-C9 (2013)
9. M. D'Oria, P. Mignosa, M.G. Tanda, *Adv. Water Resour.* **82**, 166-175 (2015)