Advances in Water Resources 63 (2014) 143–151

Contents lists available at ScienceDirect

Advances in Water Resources

journal homepage: www.elsevier.com/locate/advwatres

Bayesian estimation of inflow hydrographs in ungauged sites of multiple reach systems

Marco D'Oria *, Paolo Mignosa, Maria Giovanna Tanda

Department of Civil, Environmental, Land Engineering and Architecture, University of Parma, Parco Area delle Scienze 181/A, 43124 Parma, Italy

ARTICLE INFO

Article history:
Received 4 June 2013
Received in revised form 18 November 2013
Accepted 20 November 2013
Available online 28 November 2013

Keywords:
Reverse routing
Discharge estimation
Ungauged sites
Bayesian Geostatistical Approach
HEC-RAS

ABSTRACT

A Bayesian Geostatistical Approach to evaluate unknown upstream flow hydrographs in multiple reach systems is implemented. The methodology was, firstly, tested through three synthetic examples of river confluences, that differ in the available data, boundary conditions and number of the estimated inflow time series. Input discharge hydrographs were routed downstream by means of the widely known HEC-RAS river analysis system to obtain the downstream stage hydrographs used as known observations for the reverse procedure. In almost all cases, the observed water levels were corrupted with random errors to highlight the reliability of the methodology in preventing instabilities and overfitting. Then the procedure was applied to the real case study of the Parma–Baganza river confluence located at the city of Parma (Italy) to assess the tributary Baganza River inflow hydrograph (supposed completely ungauged) using water level data collected downstream on the main reach. The results show that the methodology properly reproduces the unknown inflows even in presence of errors affecting the downstream water levels. The practical applicability of the proposed approach is also demonstrated in complex river systems.

1. Introduction

Knowledge of discharge hydrographs in natural rivers is fundamental for water resource management, flood frequency analysis, design of new structures, calibration of rainfall-runoff models and calibration and testing of existing gauged stations, among other purposes. Despite their importance, usually few or sometime none of the river sites are equipped for measuring discharge over time. In fact, continuous direct measurement of discharge in open channels is quite difficult or even impossible and, in any way, time and money consuming. For these reasons, most frequently, river stations are monitored with level gauges that, at a disadvantage, require a reliable rating curve (not easy to assess or sometimes even indeterminable) to convert the water levels into discharge values [1].

In this context, the development of methodologies able to evaluate discharge time series in ungauged river sections based on the knowledge of flow or, even better, stage hydrographs at gauged stations is essential. To address this issue, methods based on flow routing models (e.g. [2]) or lumped approaches (e.g. [1,3–5]) can be applied. These methodologies typically allow for: (1) the estimation of the downstream hydrograph assuming the discharge known at an upstream site; (2) the evaluation of the discharge hydrographs in two river sections where the water levels are simultaneously known; (3) the estimation of upstream discharge by assuming the flow known at a downstream site and stages measured at both the ends. Nevertheless, the above mentioned methods do not contemplate the estimation of flow hydrographs in ungauged sections upstream the gauged stations. The assessment of an upstream flow hydrograph, starting from the knowledge of downstream discharges or water levels, is known as reverse routing process that, belonging to the inverse problem category, deals with existence, non-uniqueness and instability of the solution [6]. As a consequence, the reverse routing problem is particularly sensitive to errors present in the available data and/or in the model that are amplified during the inverse procedure and can cause instabilities and spurious oscillation of the solution [7].

In literature two main approaches are available to solve the reverse routing problem in open channels: the inverse solution of the de Saint Venant equations and the application of the Muskingum model in a reverse form [7–15]. A detailed review of the literature can be found in [16]. In almost all the previous studies, the reverse routing procedures have been applied in single prismatic channels with simplified flow conditions; the presence of multiple reaches, irregularities such as compound cross sections, contractions, expansions and structures was never considered.

This study extends the Bayesian inverse methodology applied in [16] to more complex synthetic and real cases where, for the first time, the presence of multiple reaches has been considered. The

* Corresponding author. Tel.: +39 (0) 521 905931; fax: +39 (0) 521 905912.
E-mail addresses: marco.doria@unipr.it (M. D'Oria), paolo.mignosa@unipr.it (P. Mignosa), mariagiovanna.tanda@unipr.it (M.G. Tanda).

0309-1708/$ - see front matter © 2013 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.advwatres.2013.11.007

inverse procedure uses as known observations water level data recorded downstream. In a Bayesian framework, the unknown discharge time series is expressed as a random function defined by means of its statistical properties. Uncertainty into the unknowns and errors in the available data can be considered. Prior information, in form of geostatistical models, is interjected into the solution, imposing some degree of continuity and/or smoothness on the shape of the estimated hydrograph regularizing the solution.

The applied methodology requires a calibrated numerical model of the considered river system. The hydraulic model is strictly connected with the inverse approach since it relates the unknown parameters (upstream discharge values) with the downstream water levels used as observations. As a consequence, the flow model must be able to account for all the relevant hydraulic processes in the river reach: resistance, bed slope, compound cross sections, confluences, structures, etc. It is beyond the scope of this work to discuss the hydraulic model setup and its calibration; hereafter these processes are considered already accomplished. In this paper, the used forward numerical model is the widely known HEC-RAS river analysis system [17], but any model able to solve analogous problems could be adopted in the same way.

A summary of the Bayesian Geostatistical Approach is presented in Section 2. Then, after a brief description of the forward routing model (Section 3), the results of the inverse procedure, applied to estimate the inflow hydrographs of multiple reach systems using synthetic downstream level data, are reported in Section 4. Three examples of river confluences, that differ in the downstream boundary conditions, the water level data used during the inversion and the number of the estimated discharge hydrographs, were considered. In Section 5, the BGA is applied to the real case study of the Parma-Baganza river confluence at the city of Parma (Italy). Conclusions are, finally, drawn in the last section.

2. Bayesian Geostatistical Approach

The Bayesian Geostatistical Approach (BGA) is a flexible highly parameterized inversion method suitable to estimate unknown parameters that exhibit auto-correlation, either in space or time. The term Bayesian refers to the inverse theory aspect, whereas geostatistical deals with the way of enforcing prior information into the solution.

A flow hydrograph, that involves rainfall-runoff processes, is a continuous function of the time that shows a serial correlation and its shape is adequately described by means of a covariance function. These characteristics make BGA suitable to estimate an unknown inflow hydrograph expressing it as a random function defined through its statistical properties and, at the same time, incorporating uncertainty in the unknowns and errors in the observations.

In this section a brief description of the inverse methodology is reported; more details can be found in [18–24].

The basis of BGA is the Bayes' theorem:

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

where $\mathbf{s}$ is the unknown parameter vector (the time values of the estimated inflow hydrographs in this work), $\mathbf{y}$ is the measured data vector (downstream water levels), $p(\mathbf{s}|\mathbf{y})$ is the posterior probability density function (*pdf*), $p(\mathbf{s})$ represents the prior *pdf* of $\mathbf{s}$ and $L(\mathbf{y}|\mathbf{s})$ is the likelihood function. The prior *pdf* of $\mathbf{s}$ is assumed with Gaussian distribution with unknown mean $E[\mathbf{s}] = \mathbf{X}\boldsymbol{\beta}$ and covariance $E[(\mathbf{s}-\mathbf{X}\boldsymbol{\beta})(\mathbf{s}-\mathbf{X}\boldsymbol{\beta})^{\mathrm{T}}] = \mathbf{Q}_{\mathbf{ss}}$, where E designates the expected value, $\mathbf{X}$ is a known matrix of basis functions, $\boldsymbol{\beta}$ is a vector of drift coefficients and $\mathbf{Q}_{\mathbf{ss}}$ indicates the parameter covariance matrix. The prior *pdf* represents soft or expert knowledge about the structure of the unknowns $\mathbf{s}$, serves the role of regularization and can eventually be used to enforce non-negativity to the parameters [25,26]. In this work the inflow hydrographs were estimated in a power transformed space [27,28]; even if it was not strictly necessary to impose non-negativity to the unknowns, this transformation avoids negative values in the calculated credibility intervals.

The likelihood function, assumed also Gaussian, characterizes the errors and indicates how likely a candidate set of parameters $\mathbf{s}$ is to reproduce the observations $\mathbf{y}$ through application of the forward model.

With those assumptions, the posterior *pdf* can be written as:

$$p(\mathbf{s}|\mathbf{y}) \propto \underbrace{\exp\left(-\frac{1}{2}(\mathbf{s}-\mathbf{X}\boldsymbol{\beta})^{\mathrm{T}}\mathbf{Q}_{\mathbf{ss}}^{-1}(\mathbf{s}-\mathbf{X}\boldsymbol{\beta})\right)}_{\text{Prior term}} \times \underbrace{\exp\left(-\frac{1}{2}(\mathbf{y}-\mathbf{h}(\mathbf{s}))^{\mathrm{T}}\mathbf{R}^{-1}(\mathbf{y}-\mathbf{h}(\mathbf{s}))\right)}_{\text{Likelihood term}} \tag{2}$$

where $\mathbf{h}(\mathbf{s})$ represents the modeled values collocated in time with the observed data $\mathbf{y}$ (by means of HEC-RAS in this work) and $\mathbf{R}$ is the covariance matrix of the epistemic errors (a lack of knowledge of the problem as a result of many sources of error such as those in the observed data and/or in the conceptual model).

The best set of parameters $\mathbf{s}$ maximizes the posterior *pdf* (2); for an efficient method to calculate the best estimates see for example [23].

When the relation between parameters and observations (expressed in this work by the forward model that solves the unsteady one dimensional de Saint Venant equations) is non-linear, the function $\mathbf{h}(\mathbf{s})$ can be successively linearized about a candidate solution $\mathbf{s}_k$ following the quasi-linear geostatistical approach [21].

At each iteration k in the linearization process, one approximates $\mathbf{h}(\mathbf{s}) \approx \mathbf{h}(\mathbf{s}_k) + \mathbf{H}(\mathbf{s}-\mathbf{s}_k)$ where $\mathbf{H}$ is the Jacobian matrix (sensitivity of observations $\mathbf{y}$ to unknown parameters $\mathbf{s}$) evaluated at each iteration as $\mathbf{H} = \frac{\partial \mathbf{h}(\mathbf{s})}{\partial \mathbf{s}}\Big|_{\mathbf{s}_k}$ using a finite difference approach. The computation of $\mathbf{H}$ requires as many forward model runs as the number of parameters ($\mathbf{s}$) plus one base run.

The uncertainty of the unknowns can be evaluated in terms of posterior covariance that is expressed through [23]:

$$\mathbf{V} = \mathbf{Q}_{\mathbf{ss}} - \mathbf{Q}_{\mathbf{ss}}\mathbf{H}^{\mathrm{T}}(\mathbf{H}\mathbf{Q}_{\mathbf{ss}}\mathbf{H}^{\mathrm{T}} + \mathbf{R})^{-1}\mathbf{H}\mathbf{Q}_{\mathbf{ss}} \tag{3}$$

The posterior covariance can be used to evaluate the 95% credibility intervals of the estimated parameters; it is worth noting here that, in case of estimation in a power transformed space, the upper and lower 95% intervals are not symmetrical about the best estimate in the not transformed space.

In this work, the prior covariance matrix of the unknown parameters, $\mathbf{Q}_{\mathbf{ss}}$, is assumed to have Gaussian form:

$$\mathbf{Q}_{\mathbf{ss}} = \sigma_s^2 \exp\left(-\frac{d^2}{l^2}\right) \tag{4}$$

where σ_s^2 is the variance, d is the separation time between parameters $\mathbf{s}$ and l is the integral scale. The Gaussian model, in describing the prior information, enforces only continuity and some degree of smoothness to the unknowns but the observations still drive the solution.

The epistemic errors are assumed independent and identically distributed with variance σ_R^2 and covariance matrix $\mathbf{R} = \sigma_R^2\mathbf{I}$ where $\mathbf{I}$ is the identity matrix.

The structure of $\mathbf{s}$, dependent on the prior covariance parameters (σ_s^2 and l), and the epistemic variance (when not *a priori* fixed), that regulate the degree of smoothness of the estimation and the level of fit between simulated and observed quantities, are inferred from the data using a Bayesian adaptation of the Restricted Maximum Likelihood method of [21]. The structural parameters are described through a probability density function and estimated