Hydrological Sciences Journal

ISSN: 0262-6667 (Print) 2150-3435 (Online) Journal homepage: https://www.tandfonline.com/loi/thsj20

A Bayesian method for missing rainfall estimation using a conceptual rainfall–runoff model

Siao Sun, Günther Leonhardt, Santiago Sandoval, Jean-Luc Bertrand-Krajewski & Wolfgang Rauch

To cite this article: Siao Sun, Günther Leonhardt, Santiago Sandoval, Jean-Luc Bertrand-Krajewski & Wolfgang Rauch (2017) A Bayesian method for missing rainfall estimation using a conceptual rainfall–runoff model, Hydrological Sciences Journal, 62:15, 2456-2468, DOI: 10.1080/02626667.2017.1390317

To link to this article: https://doi.org/10.1080/02626667.2017.1390317

Accepted author version posted online: 09 Oct 2017.
Published online: 31 Oct 2017.

HYDROLOGICAL SCIENCES JOURNAL – JOURNAL DES SCIENCES HYDROLOGIQUES, 2017
VOL. 62, NO. 15, 2456–2468
https://doi.org/10.1080/02626667.2017.1390317

A Bayesian method for missing rainfall estimation using a conceptual rainfall–runoff model

Siao Sun[a], Günther Leonhardt[b,c], Santiago Sandoval[d], Jean-Luc Bertrand-Krajewski[d] and Wolfgang Rauch[b]

[a]Key Laboratory of Regional Sustainable Development Modelling, Institute of Geographical Sciences and Natural Resource Research, Chinese Academy of Sciences, Beijing, People's Republic of China; [b]Unit of Environmental Engineering, University of Innsbruck, Innsbruck, Austria; [c]Department of Civil, Environmental and Natural Resources Engineering, Luleå University of Technology, Luleå, Sweden; [d]INSA Lyon, DEEP, Université de Lyon, Villeurbanne, France

ABSTRACT
The estimation of missing rainfall data is an important problem for data analysis and modelling studies in hydrology. This paper develops a Bayesian method to address missing rainfall estimation from runoff measurements based on a pre-calibrated conceptual rainfall–runoff model. The Bayesian method assigns posterior probability of rainfall estimates proportional to the likelihood function of measured runoff flows and prior rainfall information, which is presented by uniform distributions in the absence of rainfall data. The likelihood function of measured runoff can be determined via the test of different residual error models in the calibration phase. The application of this method to a French urban catchment indicates that the proposed Bayesian method is able to assess missing rainfall and its uncertainty based only on runoff measurements, which provides an alternative to the reverse model for missing rainfall estimates.

ARTICLE HISTORY
Received 8 March 2017
Accepted 23 August 2017

EDITOR
A. Castellarin

ASSOCIATE EDITOR
A. Langousis

KEYWORDS
Bayesian method; conceptual rainfall–runoff model; missing rainfall data; rainfall estimates; uncertainty

1 Introduction

Rainfall data are of great importance in hydrology, since rainfall is a major driver of the hydrological process. However, data gaps occur frequently in measured rainfall time series due to various reasons such as measurement device failure, equipment maintenance and data storage problems. Data gaps in rainfall time series or other formats (e.g. spatial fields) can present a serious obstacle to data analysis and modelling studies in hydrology. Filling gaps in rainfall data is therefore a crucial issue in both research and practice.

Numerous schemes for filling missing rainfall data have been proposed. The most commonly used techniques are based on spatial information from surrounding stations, mainly including weighted average methods and function fitting methods. The weighted average methods such as the Thiessen polygons method, inverse distance weighting method and kriging method use a weighted average of surrounding values to replace missing data (e.g. Lebel *et al.* 1987, Mair and Fares 2011). Alternatively, the function fitting methods such as the spline-surface fitting method (e.g. Tait *et al.* 2006) and the neural network algorithms (e.g. Kuligowski and Barros 1998, Teegavarapu and Chandramouli 2005) employ a functional relationship between data from target and surrounding stations for estimating missing rainfall data. Furthermore, the probability density function preservation approach, aiming to preserve statistical properties of time series, was developed to fill data gaps in precipitation series (Simolo *et al.* 2010). Overall, the above methods involve data analysis techniques based on other existing rainfall data series.

Lumped conceptual rainfall–runoff models are often applied in urban hydrology. These models usually require average areal rainfall as the model input to produce catchment runoff as the output. Nevertheless, point rainfall measurements (gauge data) are often used as model input. The calculation of runoff from measured (preferably areal but usually point) rainfall input is a "forward" application of a rainfall–runoff model, while the estimation of rainfall from runoff data can be viewed as a reverse application of the typical rainfall–runoff problem. Given that the runoff is an aggregated measurement for the entire catchment, a reverse application of a rainfall–runoff model provides the spatial average rainfall over the catchment. The literature provides a variety of approaches for reverse modelling of different hydrological systems (e.g. Zoppou 1999, D'Oria *et al.* 2012, D'Oria and Tanda 2012, Koussis *et al.* 2012). In catchment hydrology, Hino and Hasebe (1981) addressed effective precipitation estimation from daily runoff data based on

CONTACT Siao Sun suns@igsnrr.ac.cn

autoregressive forward models. The reverse modelling approach was later adapted and applied to hourly data (Hino and Hasebe 1984, Hino 1986). In addition, Kirchner (2009) derived storage–flow relationships from hourly streamflow data and applied them to estimate effective areal precipitation, given that the hydrological process can be described by a first-order nonlinear differential equation. Leonhardt *et al.* (2014b) developed a reverse rainfall–runoff model to estimate rainfall from runoff data, based on an inverse formulation of an original rainfall–runoff model, which in the end can be formulated as a set of linear equations. As a reverse model amplifies high-frequency variations such as measurement noise (e.g. Leonhardt *et al.* 2014b), it is always important to discuss reverse modelling results in the context of uncertainty in data and models.

The Bayesian method has long been applied in hydrology, most frequently for model calibration, which assesses parameters of a hydrological model by adjusting them to make model outputs close to measurements. In addition to an optimal set of model parameters, the Bayesian method also provides an estimate of uncertainty in model parameters by giving their posterior probability distributions. Rainfall data usually contain significant uncertainty and methods have been developed in recent research to incorporate rainfall uncertainty in model calibration. One of the most significant contributions is the Bayesian total error analysis (BATEA) framework developed by Kavetski *et al.* (2006a, 2006b). The BATEA is a hierarchical Bayesian model for estimating uncertainty in variables (e.g. rainfall) by using latent variables that need to be identified in model calibration as other model parameters. The BATEA framework has been applied mainly in the context of model calibration (e.g. Thyer *et al.* 2009, Renard *et al.* 2010), with the advantage of having consistent model parameter estimations by taking input uncertainty into account in calibration, in contrast to the more common application of the Bayesian method for model parameter estimation. In addition, the Bayesian geostatistical approach (BGA), originally developed in subsurface hydraulics (e.g. Kitanidis 1995), is able to solve highly parameterized inversion problems; however, it requires a parameter covariance model as prior information in the Bayesian formula. This method was recently applied to address reverse flow routing (e.g. D'Oria and Tanda 2012, Leonhardt *et al.* 2014a).

In this paper, we develop a Bayesian approach to estimate missing data in rainfall time series based on a lumped conceptual rainfall–runoff model. Using the original forward conceptual rainfall–runoff model, the Bayesian method is employed in the search for the optimal rainfall values by achieving the best fit between the forward model output and measured runoff data. Different from most Bayesian applications, the method presented in this paper is applied independently of calibration: the model input, i.e. rainfall, is considered as an unknown variable and is estimated based on a pre-calibrated hydrological model. This approach provides areal rainfall estimates (rather than point rainfall) over the catchment, because the runoff flows are an aggregate output from the areal rainfall. In comparison to the methods in previous literature that estimated rainfall in a Bayesian framework, the novelty of this study lies in the missing rainfall estimates, while Leonhardt *et al.* (2014b) requires measured rainfall as a prerequisite and only provides updates of uncertainty in rainfall. This method provides an alternative to the approach of missing rainfall estimates based on rainfall measurements of surrounding sites, when such data are not available.

With the advancement of monitoring equipment and techniques, more and more measurement data in short time steps are available in urban drainage systems nowadays. The application in this study focuses on rainfall and runoff data in a small urban catchment with a high temporal resolution, where the measurement time steps are in the order of a few minutes, in accordance with actual data availability.

2 Methodology

2.1 Conceptual hydrological model

A conceptual hydrological model is used in this study to simulate the relation between rainfall and runoff. It consists of a rainfall loss model, a runoff routing model and a baseflow component. All these model concepts are commonly applied in urban hydrology (e.g. Sarma *et al.* 1973). The conceptual model used in this study has been successfully applied in previous studies (e.g. Sun and Bertrand-Krajewski 2013, Leonhardt *et al.* 2014b).

In the rainfall loss model, the net rainfall, I_{net} (mm/h), is calculated by subtracting an initial loss, L_{ini} (mm), and a proportional loss, P_{prop} (-), from gross rainfall, I (mm/h):

$$I_{\mathrm{net}}(t) = \begin{cases} 0 & \text{if } \int\limits_{t=0}^{t} I\mathrm{d}t \leq L_{\mathrm{ini}} \\ I(t)\left(1 - P_{\mathrm{prop}}\right) & \text{if } \int\limits_{t=0}^{t} I\mathrm{d}t > L_{\mathrm{ini}} \end{cases} \tag{1}$$

In the runoff routing model, which employs the linear reservoir concept (see e.g. Chow *et al.* 1988), the net rainfall is shifted with a lag time, T_{lag}, and is then routed through two linear reservoirs in series. The outflow, Q_{out}, depends linearly on the storage volume, V, with the reservoir constant, K:

$$Q_{out}(t) = V(t)/K \tag{2}$$

The continuity equation is:

$$dV/dt = Q_{in}(t) - Q_{out}(t) \tag{3}$$

The outflow Q_{out} is calculated from inflow, Q_{in}, based on the analytical solution of the linear reservoir model (Equations (2) and (3)) by integrating the differential equation over the time interval $[t - \Delta t, t]$:

$$Q_{out}(t) = \exp\left(-\frac{\Delta t}{k}\right)Q_{out}(t-\Delta t) + \left[1 - \exp\left(-\frac{\Delta t}{k}\right)\right]Q_{in}(t) \tag{4}$$

To avoid strong correlation between parameters during model calibration, the values of the reservoir constants, K, of the two reservoirs are set identical. The baseflow, q, which accounts for dry weather flow in the drainage system from various sources (e.g. dry weather flow), is simply added to the outflow from the reservoir model.

2.2 Bayesian method for missing rainfall estimates

A typical forward hydrological model with measured output (runoff), $\boldsymbol{Q}$, and input (rainfall intensity), $\boldsymbol{r}$, can be written as:

$$\boldsymbol{Q} = f(r, \theta) + \boldsymbol{e} \tag{5}$$

where $f(\cdot)$ represents a hydrological model, θ represents the vector of model parameters that are determined in model calibration (see below for details) and $\boldsymbol{e}$ is the vector of residuals usually consisting of errors in input and output measurements and the model structural error.

Applying Bayesian inference for estimating the posterior probability of rainfall, $p(\hat{\boldsymbol{r}}|\boldsymbol{Q}, \theta)$, gives:

$$p(\hat{\boldsymbol{r}}|\boldsymbol{Q}, \theta) \propto p(\boldsymbol{Q},|\hat{\boldsymbol{r}}, \theta)p(\hat{\boldsymbol{r}}) \tag{6}$$

where $p(\boldsymbol{Q}|\hat{\boldsymbol{r}}, \theta)$ is the likelihood function of measured runoff with simulated rainfall $\hat{\boldsymbol{r}}$ and $p(\hat{\boldsymbol{r}})$ is the prior probability distribution of rainfall. The likelihood function $p(\boldsymbol{Q}|\hat{\boldsymbol{r}}, \theta)$ depends on the residual error models $\boldsymbol{e}$. The form of residual error models is discussed in the next section. In the absence of prior information on rainfall due to missing data, a uniform distribution over a reasonable range of rainfall intensities (e.g. 0–60 mm/h) is assumed. If rainfall measurements are available with their uncertainty, the estimation of rainfall posterior probability taking this information into account then becomes the Bayesian approach used in Leonhardt *et al.* (2014b):

$$p(\hat{\boldsymbol{r}}|\boldsymbol{Q}, \boldsymbol{r}, \theta) \propto p(\boldsymbol{Q}|\hat{\boldsymbol{r}}, \theta)p(\boldsymbol{r}|\hat{\boldsymbol{r}})p(\hat{\mathbf{r}}) \tag{7}$$

where $p(\boldsymbol{r}|\hat{\boldsymbol{r}})$ is the likelihood function of measured rainfall data $\boldsymbol{r}$ with simulated rainfall $\hat{\boldsymbol{r}}$. Leonhardt *et al.* (2014b) demonstrated that additional information included (rainfall measurements) can significantly reduce uncertainty in areal rainfall estimates as compared to the reverse model application. The application of Equations (6) and (7) is essentially different: Equation (6) estimates missing rainfall in the absence of rainfall data, whereas Equation (7) "corrects" rainfall data based on available rainfall data.

In Equation (6), in principle, each rainfall value at one time step (1 min in the case study, see below) can be viewed as an independent variable that needs to be estimated. However, the dimensionality of the problem is massive and it might also be susceptible to over-parameterization (Vrugt *et al.* 2009). Therefore, the dimension of this problem is reduced by grouping a number of missing rainfall values, with one single parameter representing the average rainfall intensity of one group, which will be inferred by the Bayesian method (e.g. Sun and Bertrand-Krajewski 2013).

The Bayesian method usually relies on iterative search algorithms such as the Markov Chain Monte Carlo (MCMC) method to identify the posterior probability distributions of the variables. In this study, the DREAM (differential evolution adaptive metropolis) algorithm based on the MCMC method is applied. The DREAM algorithm was especially designed for efficient estimation of hydrological model parameters (Vrugt *et al.* 2008), and its efficiency has been demonstrated for many complex and high-dimensional problems (e.g. Schoups *et al.* 2010).

2.3 Residual error models

In Equation (6), the likelihood function of measured runoff, $p(\boldsymbol{Q}|\boldsymbol{r}, \theta)$, can be described by a residual error model. The most common assumption for residual error models is independent and identically distributed (i.i.d.) normal distributions with zero mean and a constant variance. However, this assumption rarely holds in hydrological settings, as residual errors often exhibit temporal correlation and heteroscedasticity (e.g. Schoups and Vrugt 2010, Del Giudice *et al.* 2013). In this study, different residual error models are tested in the model calibration phase, and the most appropriate one will be selected as the likelihood function $p(\boldsymbol{Q}|\boldsymbol{r}, \theta)$ for missing rainfall estimation. Four different residual error models are tested:

(a) The first error model assumes i.i.d. normal distributions for the residuals:

$$e_i = \sigma_i N(0, 1) \tag{8}$$

where $N(0,1)$ is a standard normal distribution, σ_i^2 is the variance of the model residuals. The value of σ_i is assumed identical for all the runoff flows, Q_i, and is determined in calibration together with other hydrological model parameters. This model is hereafter referred to as the i.i.d. error model.

(b) In the second error model, the total error is further divided into two components: the runoff measurement error and model structural error. The runoff measurement errors apply to independent normal distributions with zero mean and variances evaluated based on the measuring technique and method, the value of which is related to the measured runoff rate (see Section 3 for details). The structural error is assumed to apply to i.i.d. normal distributions. Still using Equation (8), instead of identical σ_i in the i.i.d. error model, σ_i^2 has the form:

$$\sigma_i^2 = \sigma_{mi}^2 + \sigma_{si}^2 \tag{9}$$

where σ_{mi}^2 and σ_{si}^2 represent the measurement error variance and the model structural error variance, respectively. The value of σ_{mi} is known for each runoff measurement, while σ_{si} is assumed identical for all runoff flows, and is determined in calibration. This model is hereafter referred to as the two-component error model.

(c) The third residual error model considers the heteroscedasticity of the model structural error based on the two-component error model. The standard deviation σ_{si} is assumed to increase linearly with runoff rates:

$$\sigma_{si} = \sigma_0 Q_i + \sigma_1 \tag{10}$$

where σ_0 and σ_1 are parameters that are inferred in calibration. Previous studies have considered model residual heteroscedasticity in a similar way (e.g. Thyer *et al.* 2009, Schoups and Vrugt 2010). This model is hereafter referred to as the heteroscedastic error model.

(d) The fourth residual error model accounts for dependence of model residuals, assuming that the lumped model residual error applies to a first-order autoregressive model:

$$e_i - \varphi e_{i-1} = \sigma_\varphi N(0,1) \tag{11}$$

where φ is the autocorrelation coefficient and σ_φ is the standard deviation of residuals after excluding the autocorrelation effect. Both φ and σ_φ are inferred in model calibration. This model is hereafter referred to as the correlated error model.

3 Case study

In this study, missing rainfall data are estimated on an event scale. The event-based approach is computationally efficient when runoff is restricted to a short period after a storm event (Maneta *et al.* 2007), which is often the case in urban catchments equipped with storm sewer systems. A computationally efficient rainfall–runoff model is important for missing rainfall estimation since the Bayesian method requires solving the conceptual model many times by iteration. However, the method can easily be applied to continuous missing rainfall estimates using continuous runoff measurements.

The Chassieu catchment, located to the east of the city of Lyon in France (see Fig. 1 for an aerial view of the catchment), covers an industrial area of 185 ha with an imperviousness coefficient of 0.72. The runoff coefficient of this catchment is around 0.4. The impervious area is drained by a separate storm sewer system and the pervious area is not connected to the sewer system. In this study, discharges from impervious areas are used in rainfall–runoff modelling.

The rainfall was measured by a 0.2 mm tipping-bucket rain gauge located in the catchment and the measured data were converted to a one-minute time series. The water depth and mean flow velocity in the sewer pipe at the catchment outlet were measured and used to compute the runoff flow rate. Different sources of uncertainty in measurements are quantified either theoretically or empirically. Uncertainty in runoff flow is then identified by propagating all uncertainty sources through the formula for computing runoff flow from measured values using the law of propagation of uncertainties (refer to Bertrand-Krajewski and Bardin 2002, and Muste *et al.* 2012, for details). The errors in runoff flow measurements vary depending on the magnitude of discharges, with the

Figure 1. Map of the Chassieu catchment from Google Earth (the catchment area is enclosed with the yellow lines).