Towards a reliable decomposition of predictive uncertainty in hydrological modeling: Characterizing rainfall errors using conditional simulation

Benjamin Renard, D. Kavetski, E. Leblois, M. Thyer, G. Kuczera, S.W. Franks

Author-produced version of the article published in
Water Resources Research (2011) vol. 47, doi : 10.1029/2011WR010643
The original publication is available at http://www.agu.org/journals/wr/

Towards a reliable decomposition of predictive uncertainty in hydrological modeling: Characterizing rainfall errors using conditional simulation

Benjamin Renard[1], Dmitri Kavetski[2], Etienne Leblois[1],
Mark Thyer[3], George Kuczera[2] and Stewart W. Franks[2]

1. Cemagref, UR HHLY
3 bis quai Chauveau
CP 220, F-69336 Lyon
FRANCE

E-mail: benjamin.renard@cemagref.fr

2. School of Engineering
University of Newcastle
Callaghan NSW 2308
AUSTRALIA

3. School of Civil, Environmental and Mining Engineering
The University of Adelaide
Adelaide SA 5005
AUSTRALIA

Author-produced version of the article published in
Water Resources Research (2011) vol. 47, doi : 10.1029/2011WR010643
The original publication is available at http://www.agu.org/journals/wr/

Abstract

This study explores the decomposition of predictive uncertainty in hydrological modeling into its contributing sources. This is pursued by developing data-based probability models describing uncertainties in rainfall and runoff data, and incorporating them into the Bayesian Total Error Analysis methodology (BATEA). A case study based on the Yzeron catchment (France) and the conceptual rainfall-runoff model GR4J is presented. It exploits a calibration period where dense raingauge data is available to characterize the uncertainty in the catchment-average rainfall using geostatistical conditional simulation. The inclusion of information about rainfall and runoff data uncertainties overcomes ill-posedness problems and enables simultaneous estimation of forcing and structural errors as part of the Bayesian inference. This yields more reliable predictions than approaches that ignore or lump different sources of uncertainty in a simplistic way (e.g., standard least squares). It is shown that independently-derived data quality estimates are needed to decompose the total uncertainty in the runoff predictions into the individual contributions of rainfall, runoff and structural errors. In this case study, the total predictive uncertainty appears dominated by structural errors. Although further research is needed to interpret and verify this decomposition, it can provide strategic guidance for investments in environmental data collection and/or modeling improvement. More generally, this study demonstrates the power of the Bayesian paradigm to improve the reliability of environmental modeling using independent estimates of sampling and instrumental data uncertainties.

Author-produced version of the article published in
Water Resources Research (2011) vol. 47, doi : 10.1029/2011WR010643
The original publication is available at http://www.agu.org/journals/wr/

1 Introduction

1.1 Hydrological modeling in the presence of rainfall and runoff errors

Data and model errors conspire to make reliable and robust calibration of hydrological models a difficult task. Consequently, a multitude of paradigms for model estimation and prediction have been proposed and used over the last few decades, ranging from optimization approaches to probabilistic inference schemes [e.g., see the review by *Moradkhani and Sorooshian*, 2008].

The use of raingauges to estimate catchment-average precipitation is currently prevalent in hydrological modeling [*Moulin et al.*, 2009]. A major source of uncertainty is then the poor representativeness of an often small set of gauges of the entire areal rainfield, which is highly variable in both space and time [e.g., *Severino and Alpuim*, 2005; *Villarini et al.*, 2008]. The raingauges themselves are subject to both systematic and random measurement errors, including mechanical limitations, wind effects and evaporation losses, all of which are design-specific and can vary substantially with rainfall intensity [*Molini et al.*, 2005]. Methods for quantifying rainfall uncertainty include geostatistical approaches such as kriging [e.g., *Goovaerts*, 2000; *Kuczera and Williams*, 1992] and conditional simulation [e.g., *Clark and Slater*, 2006; *Gotzinger and Bardossy*, 2008; *Onibon et al.*, 2004; *Vischel et al.*, 2009], or approaches based on dense raingauge networks [e.g., *Villarini et al.*, 2008; *Willems*, 2001].

Similarly, runoff data also contain significant observational errors, due to discharge gauging errors, extrapolation of rating curves, unsteady flow conditions, flow regime hysteresis and temporal changes in the channel properties. Several approaches have been proposed to quantify this uncertainty [e.g. *Di Baldassarre and Montanari*, 2009; *Herschy*, 1994; *Lang et al.*, 2010; *McMillan et al.*, 2010; *Reitan and Petersen-Overleir*, 2009].

Finally, the characterization of structural uncertainty is a particularly challenging task, and the hydrological community is yet to agree on suitable definitions and approaches for handling structural model errors in the context of model calibration [e.g. see the conceptualizations proposed by *Beven*, 2005; *Doherty and Welter*, 2010; *Kuczera et al.*, 2006].

1.2 Decomposing predictive uncertainty

The focus of this paper is on the decomposition of the total uncertainty in hydrological predictions into its contributing sources. This is important in several scientific and operational contexts:

(i) Operational prediction, especially when data of differing quality are used in calibration and prediction;

Author-produced version of the article published in
Water Resources Research (2011) vol. 47, doi : 10.1029/2011WR010643
The original publication is available at http://www.agu.org/journals/wr/

(ii) Model comparison: separating data and structural uncertainties enables a more meaningful model comparison because structural errors are not obscured by data uncertainty;

(iii) Prediction in ungauged basins: insights into the relative contributions of data and model structural errors may be useful when a calibrated model is transferred to a different catchment. In addition, potential relationships between catchment characteristics and hydrological model parameters may be hidden or biased by data errors.

(iv) Strategic guidance for reducing total predictive uncertainty: insights into the relative contributions of individual sources of error help in more informed research and experimental resource allocation, and, importantly, allow a meaningful a posteriori evaluation of these efforts.

Uncertainty decomposition has a considerable history in the hydrologic forecasting community. For example, the Bayesian Forecasting System (BFS) [*Krzysztofowicz*, 1999; 2002] distinguishes between two sources of uncertainties in hydrologic forecasts:

- "Input uncertainty" refers to the uncertainty in forecasting an unknown future rainfall.
- "Hydrologic uncertainty" collectively refers to all other uncertainties, in particular structural errors of the hydrologic model, parameter estimation errors, input/output measurement and sampling errors [*Krzysztofowicz*, 1999].

This description highlights a major difference between the uncertainty decomposition in *forecasting* mode versus the decomposition in *prediction* mode. In the former, input uncertainty is due to *forecast* errors, while in the latter, input uncertainty is due to errors in the estimation of areal rainfall using *observations*. Note that the word *prediction* is used here to denote an application where the hydrologic model is forced with *observed* inputs (as opposed to forecasted inputs).

This paper focuses on decomposing uncertainty in the prediction context. This can be viewed as an attempt to further decompose what is termed "Hydrologic uncertainty" in *Krzysztofowicz's* BFS framework. Although *Seo et al.* [2006] discussed the potential benefits of such an additional decomposition, it is usually not viewed as a major objective because, at least for forecast lead times exceeding the routing time of the catchment, rainfall forecast uncertainty will usually dominate other sources of error [*Krzysztofowicz*, 1999]. However, the situation is different in a prediction context, where no rainfall forecast is involved. In this case, the relative contributions of rainfall, runoff and structural errors to the total predictive uncertainty are unclear and likely case-specific.

Author-produced version of the article published in
Water Resources Research (2011) vol. 47, doi : 10.1029/2011WR010643
The original publication is available at http://www.agu.org/journals/wr/

In a prediction context, attempts to decompose the total uncertainty into its three main sources have been made using several related methods. Multiple studies have employed recursive data assimilation methods such as extended and ensemble Kalman filters [*Evensen*, 1994; *Moradkhani et al.*, 2005b; *Rajaram and Georgakakos*, 1989; *Reichle et al.*, 2002; *Vrugt et al.*, 2005] or Bayesian filtering [*Moradkhani et al.*, 2005a; *Moradkhani et al.*, 2006; *Salamon and Feyen*, 2009; *Smith et al.*, 2008; *Weerts and El Serafy*, 2006]. In this paper, we consider Bayesian hierarchical approaches [e.g. *Huard and Mailhot*, 2008; *Kuczera et al.*, 2006], which to-date have been implemented in batch estimation form (but can also be formulated recursively). While the distinction between recursive versus batch processing strategies is important from the computational perspective, our focus here is on the fundamental issues of the derivation of informative error models and their incorporation into the inference framework.

1.3 Specifying data and structural error models

Although the importance of adequate descriptions of input/output/structural errors is well known, developing quantitative error models is a considerable challenge in hydrological applications. In particular, assigning reasonable values to the variances of rainfall and runoff errors is notoriously difficult [e.g. *Huard and Mailhot*, 2008; *Reichle*, 2008; *Weerts and El Serafy*, 2006]. The characterization of structural errors of hydrological models is also a major research challenge [e.g., see the discussions in *Beven*, 2005; *Doherty and Welter*, 2010; *Renard et al.*, 2010].

As a result, it is currently common to use rule-of-thumb or literature values to fully specify the input, output and structural error models and keep their parameters fixed during the hydrological model calibration. For example, *Huard and Mailhot* [2008] used literature values for rainfall errors and rule-of-thumb values for structural errors (~15% standard error). Similarly, *Salamon and Feyen* [2010] used literature values for runoff errors (~12.5% standard error for large runoff) and rule-of-thumb values for rainfall and structural errors (~15% standard error).

However, recent empirical and theoretical evidence re-emphasizes the need for reliable descriptions of uncertainties in both the forcing and response data if a meaningful decomposition of predictive uncertainty is required [e.g., *Huard and Mailhot*, 2008; *Renard et al.*, 2010]. Since the inference can be sensitive to these specifications [*Renard et al.*, 2010; *Weerts and El Serafy*, 2006], using an unreliable error model will generally yield an unreliable uncertainty decomposition. Hence, using literature values from other studies may not always be adequate. For instance, rating curve errors depend on the hydraulic configuration of the gauging section, the number of gaugings, the degree

Author-produced version of the article published in
Water Resources Research (2011) vol. 47, doi : 10.1029/2011WR010643
The original publication is available at http://www.agu.org/journals/wr/

of extrapolation, etc, all of which are site-specific. Similarly, structural errors of a hydrological model are likely to depend on the catchment, time period, etc, and are difficult to estimate a priori.

An alternative to fixing the error model parameters a priori is to include them in the inference. For instance, the variance of rainfall errors can be estimated during hydrological model calibration, rather than being fixed a priori. Although this distinction may appear a superficial technicality, it is highly pertinent to the inference in the presence of multiple sources of errors [*Huard and Mailhot*, 2008; *Renard et al.*, 2010; *Weerts and El Serafy*, 2006]. In particular, fixing the error model parameters to incorrect values may yield a computationally tractable, yet statistically unreliable inference. On the other hand, the information content of the data may not be sufficient to support the inference of the error model parameters.

The approach of inferring the error model parameters was used in the studies of *Kavetski et al.* [2006c], *Reichert and Mieleitner* [2009] and *Thyer et al.* [2009]. However, these studies did not attempt to fully decompose predictive uncertainty. *Kuczera et al.* [2006] attempted to simultaneously infer rainfall and structural errors, but limited themselves to point-estimates of inferred quantities, thus leaving open questions regarding parameter identifiability and posterior well-posedness. More recently, *Renard et al.* [2010] and *Kuczera et al.* [2010b] quantitatively demonstrated the difficulties of simultaneously identifying rainfall and structural errors from rainfall-runoff data when only vague estimates of data uncertainty are known prior to the hydrological model calibration. This result confirms the earlier discussions by *Beven* [2005; 2006] of potential interactions between multiple sources of error. However, *Renard et al.* [2010] also illustrated that the use of more precise (though still inexact) statistical descriptions of data errors makes the posterior distribution well-posed.

It is therefore vital that priors on individual sources of error reflect actual knowledge, rather than be used as mere numerical tricks to achieve well-posedness. Given the difficulty of obtaining prior estimates of structural errors (especially for highly conceptualized rainfall-runoff models), it may be more practical to first focus on the observational uncertainty in the rainfall-runoff data. Provided the data error models are reliable, they can achieve closure on the total errors, and can allow reliably estimating structural errors as "what remains" once data errors are accounted for.

1.4 Study aims

The aims of this paper are: (i) demonstrate the development and incorporation of uncertainty models for forcing and response data into a Bayesian methodology for hydrological calibration and prediction; (ii) examine the resulting improvements in the predictive performance, (iii) evaluate

Author-produced version of the article published in
Water Resources Research (2011) vol. 47, doi : 10.1029/2011WR010643
The original publication is available at http://www.agu.org/journals/wr/

whether using informative models for data errors enables inference of structural errors as part of the model calibration process; and (iv) evaluate the ability of the inference to provide quantitative insights into the relative contributions of individual sources of uncertainty. Point (iii) above is of primary importance because of the intrinsic difficulty in defining structural error models a priori. This constitutes a major contribution of this paper, since previous attempts at isolating the contribution of structural errors to predictive uncertainty [*Huard and Mailhot*, 2008; *Salamon and Feyen*, 2010] were based on assuming known parameters of the structural error model.

This paper uses the Bayesian Total Error Analysis (BATEA) [*Kavetski et al.*, 2002; *Kavetski et al.*, 2006b; *Kuczera et al.*, 2006]. The Bayesian foundation of BATEA, in particular, its ability to exploit quantitative (though potentially vague) probabilistic insights into individual sources of error, makes it well suited for using independent knowledge to improve parameter inference and predictions, and to quantify individual contributions to predictive uncertainties. However, the development of realistic error models for rainfall and runoff errors is of general interest for any method aiming at decomposing the predictive uncertainty into its three main contributive sources.

Here, the rainfall error model is developed using a geostatistical analysis of the raingauge network coupled with conditional simulation [e.g., *Vischel et al.*, 2009]. For the runoff data, the rating curve data and gaugings are used to derive a heteroscedastic error model [*Thyer et al.*, 2009]. The BATEA framework is then used to explore different calibration schemes for integrating observational uncertainty into the inference, and to evaluate their influence on calibration and validation, focusing on objectives (ii)-(iv) described above.

This work is innovative in several aspects. First, while the characterization of rainfall errors has received considerable attention [e.g., *Krajewski et al.*, 2003; *Villarini et al.*, 2008 and others], a comprehensive integration of this knowledge within a Bayesian statistical inference for hydrological models is yet to be demonstrated in a real catchment case study. More generally, the integration of independently derived data error models into a Bayesian framework for probabilistic predictions, and a stringent verification and refinement of all error models, is of increasing interest not just in hydrology, but elsewhere in environmental sciences [e.g., *Cressie et al.*, 2009]. Finally, a systematic disaggregation of predictive uncertainty into its contributing components in realistic case studies is only in its nascence. Previous studies in this area [e.g., *Huard and Mailhot*, 2008; *Salamon and Feyen*, 2010] were based on assuming known fixed values for the structural error parameters, which is hardly tenable as discussed in section 1.3.

Author-produced version of the article published in
Water Resources Research (2011) vol. 47, doi : 10.1029/2011WR010643
The original publication is available at http://www.agu.org/journals/wr/

Second, this study further develops the BATEA approach. Previous applications of BATEA focused primarily on rainfall errors and lacked a separate characterization of structural errors [*Kavetski et al.*, 2006a; *Thyer et al.*, 2009]. *Kuczera et al.* [2006] explored separate specifications of rainfall, runoff and structural errors, but did not use informative priors on the parameters of their error models, nor carried out a full Bayesian treatment of the posterior distribution (they limited themselves to finding the posterior mode only). *Renard et al.* [2010] illustrated, based on synthetic experiments, the necessity of deriving reliable and precise prior descriptions of data errors to achieve well-posed inferences. The present paper builds on the latter work and proposes a practical strategy towards these objectives. Moreover, it explicitly demonstrates the utility of independent rainfall error analysis for improving the predictive reliability, and for gaining quantitative and qualitative insights into the contribution of different sources of errors in hydrological prediction.

1.5 Outline of presentation

The Bayesian inference framework is outlined in Section 2. Section 3 describes the specific data and methods used in this case study: the hydrological model and catchment data are described in Section 3.1; Section 3.2 describes the geostatistical raingauge analysis, the development of an error model for the catchment-average rainfall data, and its incorporation into the Bayesian inference; Section 3.3 describes the runoff error model and Section 3.4 discusses the treatment of structural errors. Section 4 presents the results of a case study that evaluates the utility of this information in improving the quantification and decomposition of the runoff predictive uncertainty, with an emphasis on posterior scrutiny of the hypotheses made during calibration. The results are discussed in Section 5, followed by a summary of key conclusions in Section 6.

2 Theory: Bayesian framework

2.1 General setup: data and model

In general, a rainfall-runoff (RR) model $\boldsymbol{H}()$ hypothesizes a mapping between rainfall and runoff, given a set of (usually time-invariant) parameters $\boldsymbol{\theta}$. Let $\boldsymbol{R} = R_{1:N_t} = \left(R_t\right)_{t=1,..,N_t}$ and $\boldsymbol{Q} = Q_{1:N_t} = \left(Q_t\right)_{t=1,..,N_t}$ denote, respectively, the true rainfall and true runoff time series of length N_t. Let $\hat{\boldsymbol{Q}}$ denote the runoff predicted by the RR model, such that

$$\hat{\boldsymbol{Q}} = \boldsymbol{H}\left(\boldsymbol{R}, \boldsymbol{\theta}\right) \tag{1}$$

Hydrological models are usually also forced with potential evapotranspiration (PET). However, sensitivity to PET random errors is minor, and the impact of PET systematic errors remains much

Author-produced version of the article published in
Water Resources Research (2011) vol. 47, doi : 10.1029/2011WR010643
The original publication is available at http://www.agu.org/journals/wr/

smaller than that of rainfall errors [e.g., *Oudin et al.*, 2006]. We therefore exclude PET uncertainty from the analysis and notation. The influence of initial conditions is minimized using a warm-up.

2.2 Data uncertainty

The uncertainty in the rainfall/runoff data can be characterized using statistical error models, which describe what is known about the true values given the observations,

$$\boldsymbol{R} \sim p\left(\boldsymbol{R} \mid \tilde{\boldsymbol{R}}, \Theta_R\right) \tag{2}$$

$$\boldsymbol{Q} \sim p\left(\boldsymbol{Q} \mid \tilde{\boldsymbol{Q}}, \Theta_Q\right) \tag{3}$$

where Θ_R and Θ_Q are error model parameters describing the statistical properties of the rainfall and runoff errors respectively (e.g., means, variances and autocorrelations of observation errors). The specification of these error models is a major focus of this paper. It will be described in details in sections 3.2 (rainfall) and 3.3 (runoff).

2.3 Structural errors of rainfall-runoff models

Unlike data errors, which can be estimated independently from the hydrological model by analyzing the observational network, no widely accepted approaches exist for characterizing structural uncertainty [e.g., see *Beven*, 2005; 2006; *Doherty and Welter*, 2010; *Kennedy and O'Hagan*, 2001; *Kuczera et al.*, 2006]. The most common approach is to use an exogenous structural error term [e.g., *Huard and Mailhot*, 2008; *Kavetski et al.*, 2006b]

$$\boldsymbol{Q} = \hat{\boldsymbol{Q}} + \boldsymbol{\xi} = \boldsymbol{H}\left(\boldsymbol{R}, \boldsymbol{\theta}\right) + \boldsymbol{\xi} \tag{4}$$

$$\boldsymbol{\xi} \sim p\left(\boldsymbol{\xi} \mid \Theta_\xi\right) \tag{5}$$

where $\boldsymbol{\xi}$ is an additive error. For instance, standard least squares regression corresponds to assuming $\boldsymbol{\xi} \sim N\left(\boldsymbol{\xi} \mid 0, \sigma_\xi^2\right)$, and assuming that this term also accounts for input/output errors.

A more recent strategy seeks to represent structural uncertainty as a stochastic variation of one or more RR model parameters [e.g., *Kuczera et al.*, 2006; *Rajaram and Georgakakos*, 1989; *Reichert and Mieleitner*, 2009; *Smith et al.*, 2008] or states [e.g. *Moradkhani et al.*, 2005a; *Moradkhani et al.*, 2005b; *Moradkhani et al.*, 2006; *Salamon and Feyen*, 2009; *Vrugt et al.*, 2005; *Weerts and El Serafy*, 2006]. Time- and state- varying parameters have also been explored within the instrumental variable literature [e.g., *Young*, 1998; *Young et al.*, 2001].

Author-produced version of the article published in
Water Resources Research (2011) vol. 47, doi : 10.1029/2011WR010643
The original publication is available at http://www.agu.org/journals/wr/

In this paper, we use a hierarchical structural error model that hypothesizes a single stochastic RR parameter $\boldsymbol{\Lambda}$, which varies on a characteristic time scale represented using epochs ω,

$$\hat{Q}_t = H\left(\boldsymbol{R}_{1:t}, \boldsymbol{\Lambda}_{1:\omega(t)}, \boldsymbol{\theta}\right) \tag{6}$$

$$\boldsymbol{\Lambda}_{\omega(t)} \sim p\left(\boldsymbol{\Lambda} \mid \boldsymbol{\Theta}_{\Lambda}\right) \tag{7}$$

where $\omega(t)$ is the epoch associated with the t^{th} time step and $\boldsymbol{\Theta}_{\Lambda}$ are parameters describing the statistical properties of the stochastic parameters (e.g., $\boldsymbol{\Theta}_{\Lambda}$ could contain the mean and variance of storm-dependent parameters).

A key challenge in using approach (7) is the meaningful specification of $\boldsymbol{\Theta}_{\Lambda}$. Since structural error remains the least understood source of uncertainty, scarce guidance exists for specifying anything other than vague priors, whether on exogenous structural error terms or on stochastic parameters.

2.4 Remnant errors

In addition to error models developed for particular error sources, we also account for "remnant" errors [*Renard et al.*, 2010; *Thyer et al.*, 2009]. These are related to the notions of "model inadequacy" [*Kennedy and O'Hagan*, 2001] and "model discrepancy" [*Goldstein and Rougier*, 2009], but are intended to capture not only unaccounted structural errors of the hydrological model, but also inevitable imperfections and omissions in the descriptions of data uncertainty.

Here, we assume additive Gaussian remnant errors ε_t with unknown variance σ_{ε}^2,

$$Q_t = \hat{Q}_t + \varepsilon_t; \quad \varepsilon_t \sim N\left(0, \sigma_{\varepsilon}^2\right) \tag{8}$$

Note that in traditional regression, remnant errors such as (8) represent the lumped effects of all sources of error and correspond to "residual" errors.

2.5 Posterior distribution

When derived using the approach of *Kavetski et al.* [2002] and *Kuczera et al.* [2010b], the BATEA posterior distribution is given by Bayes' theorem as follows

$$p\left(\boldsymbol{\theta}, \boldsymbol{R}, \boldsymbol{\Lambda}, \boldsymbol{\Theta} \mid \tilde{\boldsymbol{R}}, \tilde{\boldsymbol{Q}}\right) = p\left(\tilde{\boldsymbol{R}}, \tilde{\boldsymbol{Q}} \mid \boldsymbol{\theta}, \boldsymbol{R}, \boldsymbol{\Lambda}, \boldsymbol{\Theta}\right) p\left(\boldsymbol{\theta}, \boldsymbol{R}, \boldsymbol{\Lambda}, \boldsymbol{\Theta}\right) \Big/ p\left(\tilde{\boldsymbol{R}}, \tilde{\boldsymbol{Q}}\right) \tag{9}$$

$$\begin{aligned} p\left(\boldsymbol{\theta}, \boldsymbol{R}, \boldsymbol{\Lambda}, \boldsymbol{\Theta} \mid \tilde{\boldsymbol{R}}, \tilde{\boldsymbol{Q}}\right) \propto\ & p\left(\tilde{\boldsymbol{Q}} \mid \boldsymbol{\theta}, \boldsymbol{\Lambda}, \boldsymbol{R}, \boldsymbol{\Theta}_Q, \boldsymbol{\Theta}_{\varepsilon}\right) p\left(\tilde{\boldsymbol{R}} \mid \boldsymbol{R}, \boldsymbol{\Theta}_R\right) p\left(\boldsymbol{\Lambda} \mid \boldsymbol{\Theta}_{\Lambda}\right) \times \\ & p\left(\boldsymbol{R}\right) p\left(\boldsymbol{\Theta}_R\right) p\left(\boldsymbol{\Theta}_Q\right) p\left(\boldsymbol{\Theta}_{\Lambda}\right) p(\boldsymbol{\Theta}_{\varepsilon}) p\left(\boldsymbol{\theta}\right) \end{aligned} \tag{10}$$

Author-produced version of the article published in
Water Resources Research (2011) vol. 47, doi : 10.1029/2011WR010643
The original publication is available at http://www.agu.org/journals/wr/

The BATEA posterior in eqn (10) explicitly represents individual sources of uncertainty in the hydrological model-data system as follows:

1) The "runoff likelihood" $p\left(\tilde{\boldsymbol{Q}} \mid \boldsymbol{\theta}, \boldsymbol{\Lambda}, \boldsymbol{R}, \Theta_Q, \Theta_\varepsilon\right)$ describes runoff and remnant errors. We refer to *Kuczera et al.* [2010b] for a fully general derivation of this likelihood, and to section 3.3 for its derivation with the specific error models used in the case study.

2) The "rainfall likelihood" $p\left(\tilde{\boldsymbol{R}} \mid \boldsymbol{R}, \Theta_R\right)$ describes rainfalls errors;

3) The "stochastic-parameter term" $p\left(\boldsymbol{\Lambda} \mid \Theta_\Lambda\right)$ characterizes structural errors.

In addition, independent information on any quantity of inference can be supplied via the priors:

1) $p\left(\Theta_R\right)$ and $p\left(\Theta_Q\right)$: priors on the parameters describing, respectively, rainfall and runoff data uncertainties;

2) $p\left(\boldsymbol{\theta}\right)$: prior on the time-invariant RR parameters;

3) $p\left(\Theta_\Lambda\right)$: prior on the parameters Θ_Λ of the probability model of the stochastic parameters $\boldsymbol{\Lambda}$;

4) $p\left(\boldsymbol{R}\right)$: prior distribution of the true rainfall time series; note that the product of this prior with the "rainfall likelihood" $p\left(\tilde{\boldsymbol{R}} \mid \boldsymbol{R}, \Theta_R\right)$ is proportional to the rainfall error model (2).

5) $p(\Theta_\varepsilon)$: prior on the parameters of the remnant error model. Here, $\Theta_\varepsilon = \sigma_\varepsilon$ in eqn (8).

The posterior in eqn (10) can be explored using Markov Chain Monte Carlo (MCMC) sampling. In this study, we use a multi-stage limited-memory MCMC strategy detailed by Kuczera et al. [2010a]. Also note that eqn (10) can be modified to use joint priors on any quantity of inference. This would be needed, for example, if BATEA was applied recursively as new data arrives.

The key scientific (as opposed to computational) challenge in using BATEA, or any other Bayesian approach, for the decomposition of individual sources of error, is to develop accurate and precise probabilistic models for the individual terms in the posterior (10). This will generally require independent information to augment and constrain the inference. Illustrating these developments in a practical study is a major objective of this paper.

Author-produced version of the article published in
Water Resources Research (2011) vol. 47, doi : 10.1029/2011WR010643
The original publication is available at http://www.agu.org/journals/wr/

2.6 Calibration schemes

The BATEA framework can be used to derive several calibration schemes, differing in the type of error models and the amount of prior knowledge utilized in the inference. This allows exploring the benefits and challenges of explicitly describing each source of uncertainty and of including additional prior information. The following schemes are considered in this study (Table 1):

1) The SLS scheme (Standard Least Squares) lumps the effects of all sources of errors into the remnant error term in eqn (8).

2) The OI scheme (Output-Input) explicitly accounts for rainfall and runoff uncertainty. Structural errors are handled entirely by the remnant error term. Vague priors are used for the terms $p(\boldsymbol{R})$ and $p(\Theta_R)$. However, prior information, derived from rating curve analysis, is used for the runoff error parameters Θ_Q.

3) The OI-CS scheme is an "enhanced" OI scheme, augmented using an informative prior for the term $p(\boldsymbol{R})$. This prior is derived using Conditional Simulation (CS) as described in Section 3.2.

4) The OIS scheme (Output-Input-Structural) explicitly accounts for rainfall and runoff uncertainty, and characterizes structural errors using a stochastic RR parameter. Note that it still uses the remnant error (8) to account for the inevitable imperfections of the uncertainty models.

5) The OIS-CS scheme is an "enhanced" OIS scheme, augmented using the CS prior for the term $p(\boldsymbol{R})$.

2.7 Quantification and decomposition of predictive uncertainty

2.7.1 "Total" predictive distributions

In Bayesian methods, the uncertainty in a quantity of interest (e.g. runoff Y) is usually quantified by means of the predictive distribution. Let Ξ denote the vector of all inferred quantities, and $p(\Xi \mid \tilde{\boldsymbol{D}})$ denote the posterior of parameters Ξ given observed data $\tilde{\boldsymbol{D}}$. By definition, the predictive distribution of Y is [*Gelman et al.*, 2004]:

$$p(Y \mid \tilde{\boldsymbol{D}}) = \int p(Y \mid \Xi, \tilde{\boldsymbol{D}}) p(\Xi \mid \tilde{\boldsymbol{D}}) \, \mathrm{d}\Xi \tag{11}$$

The "total" predictive distribution (TPD) in eqn (11) integrates over the posterior uncertainty in the parameters Ξ and can be obtained directly from the MCMC samples $(\Xi_{(i)})_{i=1\ldots N_{sim}}$. It is widely used

Author-produced version of the article published in
Water Resources Research (2011) vol. 47, doi : 10.1029/2011WR010643
The original publication is available at http://www.agu.org/journals/wr/

in hydrology, including flood forecasting [e.g. *Krzysztofowicz*, 1999; *Reggiani et al.*, 2009; *Todini*, 2008] and climate studies [*Rougier*, 2007].

2.7.2 "Partial" predictive distributions

In this study, the individual contributions of distinct sources of uncertainty are quantified by formulating "partial" predictive distributions (PPD's). The derivation of a PPD is illustrated using a simple 2-parameter model.

Let $p(\theta_1, \theta_2 \mid \tilde{\boldsymbol{D}})$ be the posterior of parameters θ_1 and θ_2. For example, θ_1 and θ_2 could be viewed as representing input and structural errors, which we are trying to disaggregate in this study. Now, consider the conditional distribution:

$$p(Y \mid \tilde{\boldsymbol{D}}, \theta_2^*) = \int p(Y \mid \theta_2^*, \theta_1)\, p(\theta_1 \mid \tilde{\boldsymbol{D}}, \theta_2^*) d\theta_1 , \tag{12}$$

where θ_2^* is a given conditioning value (e.g., the posterior mode).

Eqn (12) represents the uncertainty in Y contributed by the uncertainty in θ_1, conditional on θ_2^*. We hence refer to it as the "PPD of Y arising from the uncertainty in θ_1". The PPD $p(Y \mid \tilde{\boldsymbol{D}}, \theta_1^*)$, representing the uncertainty contributed by θ_2, can be defined in a similar manner.

Unlike the TPD, PPDs cannot in general be constructed directly from MCMC samples of the joint posterior distribution. Sampling from the conditional posterior distribution $p(\theta_1 \mid \tilde{\boldsymbol{D}}, \theta_2^*)$ in eqn (12) would, in general, require separate MCMC sampling. However, in the special case where the posteriors of θ_1 and θ_2 are independent, the conditional distribution $p(\theta_1 \mid \tilde{\boldsymbol{D}}, \theta_2^*)$ is equal to the marginal distribution $p(\theta_1 \mid \tilde{\boldsymbol{D}})$. The PPD then reduces to

$$p(Y \mid \tilde{\boldsymbol{D}}, \theta_2^*) = \int p(Y \mid \theta_2^*, \theta_1)\, p(\theta_1 \mid \tilde{\boldsymbol{D}}) d\theta_1 \tag{13}$$

Consequently, if the analysis of the full posterior suggests that θ_1 and θ_2 are near-independent, the PPD in eqn (13) can be approximated from the MCMC samples by generating a realization $Y^{(i)}$ for each parameter $(\theta_1^{(i)}, \theta_2^*)_{i=1\ldots N_{sim}}$. To the extent that θ_1 and θ_2 are independent, the sample $(Y^{(i)})_{i=1\ldots N_{sim}}$ is then an approximate realization from the PPD $p(Y \mid \tilde{\boldsymbol{D}}, \theta_2^*)$.

Author-produced version of the article published in
Water Resources Research (2011) vol. 47, doi : 10.1029/2011WR010643
The original publication is available at http://www.agu.org/journals/wr/

This study distinguishes between the following sources of errors: (i) rainfall errors; (ii) structural errors; and (iii) runoff + remnant errors. The corresponding PPDs are constructed from the MCMC samples generated during the inference by iterating the flowchart in Figure 1 for $i = 1{:}N_{sim}$.

3 Material and methods

3.1 Study area and hydrological model

3.1.1 The Yzeron catchment

The case study is based on the 129 km^2 Yzeron catchment in the Rhône-Alpes region of France, near Lyon (Figure 2a). Its regime is rainfall-dominated, with floods between autumn and spring, and extended periods of low flows in summer. The annual average rainfall and runoff are approximately 845 mm and 150 mm respectively, yielding an annual runoff coefficient of 0.18. The upstream elevations range from 400 to 917 m, with steep slopes often exceeding 10%.

Nearly 8 years of daily runoff (shown in Figure 2b) are used in this study. The last two years, 2007-2008 are used for calibration, while the preceding 6 years are used for validation.

Two separate sets of raingauges are used. The first set, denoted as R3D, comprises 3 raingauges in the lower areas of the catchment (squares in Figure 2a), with daily totals available for the whole period of study. The daily mean of the R3D raingauges provides an estimate of the daily areal rainfall (inverted bars in Figure 2b) that was used in the calibration and validation experiments.

The second set, R13H, comprises 13 raingauges located within the vicinity of the Yzeron catchment, shown as dots in Figure 2a. The spatial density of this network is quite high considering the moderate catchment size; moreover, it provides measurements at an hourly resolution. However, its observations are available only for the last 2 years of the study period. Consequently, the R13H data are used solely to investigate the error properties of the R3D estimates of the catchment-average rainfall. In particular, the high spatial density of the R13H gauges permits the spatial variability of rainfall to be described using conditional simulation (Section 3.2). The concurrent availability of the R3D and R13H data explains the use of the last 2 years of the study period for calibration, while only R3D data is used in the validation period.

3.1.2 The GR4J rainfall-runoff model

This study applies the widely-used GR4J model [*Perrin et al.*, 2003], which simulates catchment runoff using rainfall and potential evapotranspiration at a daily time step (Figure 2c). The model has two conceptual stores (production and routing), two unit-hydrograph elements and 4 calibration

parameters: the maximum production storage θ_1 [L, mm], the groundwater exchange parameter θ_2 [L/T, mm/day], the maximum routing storage θ_3 [L, mm] and the unit-hydrograph time-delay parameter θ_4 [T, days]. Further details can be found in *Perrin et al.* [2003].

3.2 Development of the rainfall error model

3.2.1 Conditional simulation (CS)

The uncertainty of areal rainfall estimates is generally dominated by sampling errors, i.e., errors due to the incomplete description of the rainfall spatial field using raingauges [*Moulin et al.*, 2009; *Severino and Alpuim*, 2005]. Conditional simulation (CS) is a geostatistical method that generates multiple replicates of the rainfall field based on the values measured at individual raingauges [e.g., *Vischel et al.*, 2009]. In most common CS methods, the replicates match the observed values at the raingauge locations, but differ elsewhere. The spatial variability of the replicates depends on the geostatistical properties (distribution, variogram, etc.) of the rainfall fields, which are estimated prior to generating the CS replicates.

CS provides a natural means to describe the uncertainty in the areal rainfall forcing and is therefore well suited for augmenting the statistical inference of hydrological models.

3.2.2 The Turning-Band-Method (TBM) rainfall generator for CS

The CS method used in this study was the Turning-Band-Method (TBM) rainfall generator. The main equations of the TBM geostatistical model are provided in Appendix A. Further details are provided by *Tompson et al.* [1989]. A summary of the main characteristics is provided below.

TBM generates three-dimensional fields that describe rainfall variability in two spatial (areal) dimensions and in the time dimension. Rainfall fields are constructed from the product of two independent fields: (i) a Boolean indicator field representing pixels with zero and non-zero rainfall; and (ii) a field of non-zero precipitation generated from a pre-specified distribution.

The TBM simulation depends on parameters describing the at-site rainfall distribution (e.g., mean and variance of a lognormal distribution) and the spatio-temporal properties of the observed rainfall fields (e.g., the spatio-temporal variogram). The simulated field is constructed to be consistent not only with the observed variogram of raw data (e.g., hourly rainfall), but also with the variograms of data aggregated over various durations (e.g., 2, 4, 6, 12, 24-hour intervals). This constraint is addressed using the integrative properties of random fields: given the variogram of the point

Author-produced version of the article published in
Water Resources Research (2011) vol. 47, doi : 10.1029/2011WR010643
The original publication is available at http://www.agu.org/journals/wr/

process that generates (unobserved) instantaneous rainfall, it is possible to derive the variograms of the aggregated fields. This operation is known as the regularization of the point variable to the aggregated variable [e.g., *Journel and Huijbregts*, 1978, chapter II]. Consequently, the inference of the variogram parameters of the (unobserved) point process is based on the (observed) variograms of observations aggregated over various durations. This allows the generated field to be consistent with the spatio-temporal properties of aggregated rainfall.

The TBM method generates Gaussian random fields, which are then transformed to obtain the indicator field and the non-zero precipitation field. This transformation is based on thresholding for the indicator field and on the transformation $F^{-1}\left(\Phi(\bullet)\right)$ for the non-zero precipitation field, where Φ is the cumulative density function (cdf) of the standard Gaussian distribution and F^{-1} is the inverse cdf of non-zero rainfalls. Care is needed at this step because these transformations alter the spatial correlations of the simulated random field. Therefore, empirical and analytical correction formulae are used to match the correlation structure of the final rainfall field to the observations (See Appendix A for details). Finally, Gibbs sampling is used to condition the simulations at the raingauge locations. *Onibon et al.* [2004] provide further details.

3.2.3 Derivation of the rainfall error model using CS data

Figure 3 depicts three representative CS replicates over 4 consecutive hourly steps. In all replicates, rainfall values match the observations at the conditioning gauge locations (empty squares, at the R13H locations), but differ elsewhere. For each replicate, the hourly rainfalls are aggregated to the daily scale and averaged over the catchment area. This yields the daily areal rainfall of the Yzeron catchment associated with a particular conditional replicate.

Figure 4a compares the time series of areal rainfall estimated from the R3D network to the distribution estimated from 34 CS replicates (conditioned on rainfall values from the R13H network). The limited number of replications is due to current computational constraints: a single CS of 2 years of hourly data over a 49 x 49 grid takes several hours on a standard desktop CPU. Improved computational strategies are beyond our scope and will be investigated in future work.

Figure 4a shows that the spread of the conditional replications is highly variable on a daily scale. For example, the individual replicates varied from 15 to 40 mm on day 119, while the R3D estimate was 25 mm. This suggests considerable uncertainty in the R3D areal rainfall estimates during this particular event. Conversely, the replicates ranged from 55 to 67 mm for day 134, with the R3D estimate of 62 mm, suggesting a markedly smaller uncertainty in the R3D data. Figure 4b

Author-produced version of the article published in
Water Resources Research (2011) vol. 47, doi : 10.1029/2011WR010643
The original publication is available at http://www.agu.org/journals/wr/

also shows that the standard deviation of CS replicates (computed for each day of the calibration period) has no clear relationship with the R3D-estimated areal rainfall. This implies that larger rainfall events are not necessarily subject to larger uncertainties.

Figure 4c compares the mean of the conditional replications $\hat{R}$ with the R3D-estimated values $\tilde{R}$ for each day of the calibration period. Overall, they are in acceptable agreement, suggesting the absence of strong systematic bias in the R3D estimates. However, a closer inspection reveals considerable discrepancies between the two estimates of the areal rainfall for small events. More precisely, on some days R3D estimates are zero even when CS suggests considerable precipitation (up to 20 mm). This suggests that significant rainfall events can be missed with only three raingauges, or that the CS is overestimating small events, or both.

The errors in the R3D estimates can be approximated as

$$\phi_t = \hat{R}_t / \tilde{R}_t \tag{14}$$

The multiplicative model was selected in an attempt to capture the heteroscedasticity of the rainfall errors, especially for large storm events. It has been used in several previous studies [e.g., *Kavetski et al.*, 2002; *Villarini et al.*, 2008; *Vrugt et al.*, 2008, and others].

Figure 4d, which plots the multipliers ϕ versus the R3D rainfall estimates, reveals a complex distributional structure of rainfall errors. Multipliers associated with small-recorded rainfall values are predominantly larger than 1.0 (corresponding to the under-estimation reported earlier) and are highly variable. The discrepancies in small rainfall events have several possible explanations: (i) biases in the R3D areal averages due to insufficient spatial coverage; and/or (ii) biases in the CS of small rainfall events. Multipliers tend to stabilize around 1.0 for higher rainfall values, suggesting an absence of strong systematic biases and a limited heteroscedasticity. While the low heteroscedasticity of multipliers associated with larger events supports the multiplicative error model (14), the difficulty in describing errors in small rainfall suggests that simple models, such as Gaussian multipliers, may not be adequate over the entire rainfall range and need future refinement.

3.2.4 Diagnostic evaluation of CS predictions versus R3D gauges

To investigate the reliability of CS for small rainfall events, we evaluated the CS replicates against R3D raingauge values (as opposed to areal averages), by comparing the rainfall series from a given R3D raingauge g_k with the CS predictive distribution at the pixel containing g_k. The reliability of the CS predictive distribution is evaluated using the predictive QQ plot, which displays the *p*-

Author-produced version of the article published in
Water Resources Research (2011) vol. 47, doi : 10.1029/2011WR010643
The original publication is available at http://www.agu.org/journals/wr/

values of the observations within the predictive distribution against the quantiles of the uniform distribution. A statistically reliable predictive distribution leads to *p*-values close to the 1:1 line. Departures from the bisector have specific diagnostic interpretations (see Laio and Tamea [2007] and Thyer et al. [2009] for further details).

While Figure 5b suggests that the CS predictive distribution is reliable for daily rainfall exceeding 2 mm, Figure 5a suggests poorer reliability for small rainfalls. In particular, numerous observations have *p*-values of zero, suggesting a tendency of the CS to over-estimate the actual rainfall. The discrepancies in small rainfall events discussed in previous section 3.2.3 are therefore at least partly due to biases in the CS of small events.

Since the current analysis shows that CS reliably quantifies the uncertainty in the larger rainfall events, which are generally (though not always) of primary interest, it supports the use of CS as a tool to derive rainfall error estimates for hydrological applications. The investigation of the apparently poor CS performance for small rainfall is deferred to a future study.

3.2.5 Conditional simulation as a prior on the true rainfall in BATEA

A key advantage of the Bayesian paradigm is its ability to augment the inference with independent knowledge. In this study, we incorporate the information from the geostatistical analysis of the R13H network into a BATEA calibration of a hydrological model forced with the R3D rainfall. This is achieved by using the CS replicates to specify the term $p(\boldsymbol{R})$ in the BATEA posterior (10).

The prior $p(\boldsymbol{R})$ is described using independent Gamma distributions with time-varying parameters $\hat{\alpha}_t$ and $\hat{\beta}_t$, describing the rainfall at all time steps *t* where rainfall exceeds 2 mm,

$$p(\boldsymbol{R}) = \prod_{t=1, R_t > 2mm}^{N_t} p_{\Gamma}\left(R_t \mid \hat{\alpha}_t, \hat{\beta}_t\right) \tag{15}$$

The scale $\hat{\alpha}_t$ and shape $\hat{\beta}_t$ at step *t* are estimated by matching the moments of the Gamma distribution to the moments of the CS replicates described in Section 3.2.3. Note that the specification of the prior $p(\boldsymbol{R})$ is based solely on the R13H data (analyzed using CS) and does not use data from the R3D network. A posteriori, the R3D data is used indirectly in the exploratory analyses reported in sections 3.2.3 - 3.2.4.

Note that the exclusion of rainfalls below 2mm from the error model is used as a computational acceleration strategy to remove insensitive degrees of freedom from the inference. This

Author-produced version of the article published in
Water Resources Research (2011) vol. 47, doi : 10.1029/2011WR010643
The original publication is available at http://www.agu.org/journals/wr/

approximation has little effect on the inference results because the predicted runoff is largely insensitive to small rainfalls.

3.2.6 Rainfall error model

The likelihood of rainfall errors, $p\left(\tilde{\boldsymbol{R}} \mid \boldsymbol{R}, \boldsymbol{\Theta}_R\right)$ in eqn (10), is specified as follows

$$R_t = \tilde{R}_t \times \phi_t\,,\ \phi_t \sim TN(\mu_\phi, \sigma_\phi^2, 0)\,, \tag{16}$$

where $TN(a, b^2, 0)$ denotes a Gaussian distribution with mean a and standard deviation b, truncated at zero. Similarly to eqn (15), the error model (16) is applied only on days where $R_t > 2mm$. The advantages and limitations of rainfall models (15)-(16) are discussed in Section 5.4.

3.3 Development of the runoff error model

Runoff uncertainty was investigated by analyzing the rating curve and related stage-discharge gaugings. The Yzeron catchment can be considered well-gauged, with gaugings covering a large fraction of the flow duration curve.

Figure 6 shows the runoff measurement errors, defined as the difference between the runoff gaugings and the runoff predicted by the rating curve ("RCP runoff"). There is a clear trend of runoff measurement errors increasing with the RCP runoff.

In view of Figure 6, we hypothesized a heteroscedastic error model, where runoff uncertainty is Gaussian with a zero mean and a standard deviation σ_Q proportional to the RCP runoff,

$$Q = \tilde{Q} + \varepsilon_Q; \qquad \varepsilon_Q \sim N(0, \sigma_Q^2), \quad \sigma_Q = a + b\tilde{Q}, \tag{17}$$

where Q is the gauged runoff and $\tilde{Q}$ is the RCP runoff. In the context of eqn (3), $\boldsymbol{\Theta}_Q = (a, b)$.

Eqn (17) was fitted to the Yzeron runoff data (with vague priors on a and b) using the WINBUGS software [*Spiegelhalter et al.*, 2003]. The 90% predictive limits of the runoff measurement error model are shown in Figure 6. The fanning out of the uncertainty bounds for large runoff values is dominated by extrapolation from lower flows, where many more gaugings are available. This deficiency arises due to limited gauging data in the high flow range (a single measurement for flows exceeding 10 mm).

Author-produced version of the article published in
Water Resources Research (2011) vol. 47, doi : 10.1029/2011WR010643
The original publication is available at http://www.agu.org/journals/wr/

The posterior mean and standard deviation for the parameters of the rating curve error model (17) were $a = 0.0032 \pm 0.0015$ and $b = 0.096 \pm 0.014$. Since the precision of these estimates is relatively high, they were fixed at their posterior means during the subsequent BATEA calibration.

Note that eqn (17), in combination with the remnant error model (8), allows deriving the runoff likelihood term in eqn (10). Given the error models selected here, observed runoff is treated as a realization from a Gaussian distribution with mean $\hat{Q}_t$ and variance $\left(a + b\tilde{Q}\right)^2 + \sigma_\varepsilon^2$.

3.4 Representation of structural errors

The characterization of structural error of the GR4J model is explored using stochastic daily variation of its parameter θ_1. We also investigate a more traditional exogenous treatment of structural errors using the remnant error term (see also section 2.4).

When θ_1 is treated as stochastic, it is assumed to follow a truncated Gaussian distribution with unknown mean μ_{θ_1} and standard deviation σ_{θ_1},

$$\theta_1 \sim TN(\mu_{\theta_1}, \sigma_{\theta_1}^2, 0) \tag{18}$$

Note that θ_1 controls the maximum storage of the production store (Figure 2c). It may seem surprising, or even imprudent, to make this quantity time-dependent because the actual storage can then in principle exceed the maximum capacity. However, a separate sensitivity analysis [similar to Figure 5 of *Kuczera et al.*, 2006] indicated that this parameter, when made stochastic, had the largest impact on model predictions. Importantly, we examined the inferred stochastic variability of θ_1 to determine its effect on the storage values and long-term water balance (Section 4.4.2).

4 Inference results and posterior diagnostics

This section describes the application of the calibration schemes of Section 2.6 and Table 1 to the Yzeron data, using the input, output and structural error models constructed in Sections 3.2-3.4.

4.1 Well-posedness of the calibration schemes

The convergence of MCMC samples reflects the statistical characteristics of the target distribution. In particular, slowly convergent sampling is often indicative of ill-posed posteriors [*Renard et al.*, 2010]. Such posteriors arise when the data contains insufficient information to identify the quantities of interest and no prior information is available or used.

Author-produced version of the article published in
Water Resources Research (2011) vol. 47, doi : 10.1029/2011WR010643
The original publication is available at http://www.agu.org/journals/wr/

The MCMC convergence was assessed using the GR criterion [see *Cowles and Carlin*, 1996 for a broader review; *Gelman et al.*, 2004]. For all calibration schemes except OIS (see below), GR statistics were below 1.2 for all inferred quantities, suggesting a well-posed inference. As expected intuitively, Table 1 shows that convergence is faster for lower-dimensional inference schemes. Yet it also highlights the impact of the prior on the speed of MCMC convergence. Despite having exactly the same likelihood function and the same number of inferred quantities, scheme OI-CS converges ten times faster than OI because it uses the informative CS prior. This emphasizes that the computational cost of an inference depends more on its structure than just on its dimensionality. An in-depth discussion of dimensionality and computation in hierarchical models is provided by Spiegelhalter [2002]; see also the synthetic investigations in Renard et al. [2010].

The OIS scheme, which attempts to infer both rainfall and structural errors without using the CS prior, suffered from a prohibitively slow rate of MCMC convergence, with GR statistics for several inferred quantities (including hydrological parameters and latent variables) still exceeding 3.0 after 10^6 MCMC iterations. Inspection of the simulated values revealed strong negative correlations between the latent variables for input and structural errors (with some posterior cross-correlations exceeding -0.9). Moreover, the posterior standard deviations of inferred quantities were higher than in OIS-CS scheme by a factor of about 3 on average, but exceeding 10 for some latent variables. This computational behavior is symptomatic of ill-posedness [see detailed discussion in *Renard et al.*, 2010]. In practical terms, this means that rainfall and structural errors are not simultaneously identifiable solely from the given forcing-response time series with no associated error estimates.

The non-convergence of the OIS scheme, contrasted with the convergence of the OIS-CS scheme, supports a key conclusion of *Renard et al.* [2010], namely that the specification of informative priors for rainfall and runoff uncertainty is a necessary step to ensure well-posedness when both forcing and structural errors are modeled hierarchically using latent variables.

4.2 Reliability of total predictive uncertainty (all schemes)

This section examines the adequacy of the predictive distribution of runoff. Posterior scrutiny of the predictive distribution is important because violations of calibration assumptions can result in unreliable and misleading predictions [*Hall et al.*, 2007; *Thyer et al.*, 2009]. In addition to visual appraisals, which are of clear value to a hydrological expert, a more formal approach for evaluating the reliability of a predictive distribution is given by the predictive QQ plot (see Section 3.2.3). However, reliability alone is insufficient to demonstrate that a particular predictive method is superior to another [e.g. *Gneiting et al.*, 2007]. In particular, the precision of the predictive

Author-produced version of the article published in
Water Resources Research (2011) vol. 47, doi : 10.1029/2011WR010643
The original publication is available at http://www.agu.org/journals/wr/

distribution also needs to be assessed. Moreover, the reliability of the total predictive distribution does not prove that all individual error models are correctly specified – it is a necessary but insufficient condition. This topic is further discussed in section 5.2.

Figure 7 shows the predictive QQ plots constructed for the validation period. In addition, Figure 8 shows the total predictive distributions from schemes OI, OI-CS and OIS-CS for several flood events. Those figures allow a visual appraisal of the precision (i.e., sharpness or resolution) of the TPDs. Several important results can be noted:

1) The SLS scheme produces an unreliable predictive distribution. The shape of the QQ curve in Figure 7a suggests a general over-estimation of predictive uncertainty. However, when restricted to runoffs above 2 mm (65 days, Figure 7b), it shows that predictive uncertainty is actually severely under-estimated for large runoffs, with many observations outside of their predicted range.

2) The shape of the OI-curve in Figure 7 suggests a severe under-prediction of observations. This is confirmed by the second row of Figure 8, with the predicted runoff being consistently lower than the observed values.

3) The OI-CS scheme slightly under-estimates the predictive uncertainty, with about 1% of the observations lying outside of the predictive range (p-values of 0 and 1 by convention). On the other hand, Figure 8 shows that OI-CS yields markedly more precise predictions compared to other schemes.

4) The OIS-CS scheme yields a reliable estimation of the predictive uncertainty, for all runoff ranges (Figure 7a-b). However, Figure 8 shows that its predictive precision is the lowest, suggesting that, in this application, representing structural errors using a stochastic parameter has increased the predictive uncertainty compared to the OI-CS setup, where structural errors were represented as part of the additive remnant error term.

Further insights can be gained by examining the estimated parameters of the rainfall and structural error models, as listed in Table 2:

1) The standard deviation of the rainfall multipliers is estimated as 1.54 in the OI scheme, but reduces to 0.27 in the OIS-CS scheme. This occurs because the OI scheme lacks an adequate description of structural errors (the homoscedastic Gaussian remnant error term is poorly suited to this) and, by increasing its standard deviation, the rainfall error model can compensate for unaccounted structural errors. This compensation is detectable in this case study because of the availability of independent prior knowledge on rainfall errors.

Author-produced version of the article published in
Water Resources Research (2011) vol. 47, doi : 10.1029/2011WR010643
The original publication is available at http://www.agu.org/journals/wr/

2) Conversely, the estimated parameters of the rainfall error models are similar in schemes OI-CS and OIS-CS. This illustrates the constraint exerted by the CS prior, limiting the interactions between rainfall and structural errors. However, recall that removing the stochastic variability in θ_1 (OI-CS) leads to a slight under-estimation of the predictive uncertainty.

4.3 Decomposition of total predictive uncertainty into forcing, response and structural components (OIS-CS only)

The previous section showed that the BATEA methodology yields reliable estimates of predictive uncertainty when prior information on rainfall and runoff errors is available (scheme OIS-CS). It is then of practical significance and scientific interest to explore and evaluate the relative contributions of forcing and structural errors to the total predictive uncertainty.

Figure 9 shows the TPD and PPD for the forcing, structural and response errors (see Section 2.7 for details). Under the hypotheses made in this case study (including the hydrological model and the data error models), predictive uncertainty in the runoff appears dominated by structural errors. Albeit significant, rainfall errors explain a smaller part of TPD, with runoff and remnant errors contributing even less. The identifiability of the parameters of the rainfall and structural error models (in particular, their near-independence, maximum absolute posterior correlation of about 0.12) provides confidence that the PPDs with respect to rainfall and structural errors can be interpreted as representing the individual contributions of these random variables to the TPD.

Note that this study uses partial predictive intervals in the decomposition of uncertainty. Since these correspond to conditional distributions (see section 2.7), the choice of the conditioning values may affect the decomposition of uncertainty. In the validation analyses presented here, we condition all latent variables on the modal estimate of their mean (μ_ϕ in eqn (16), see also Figure 1), because it represents the "most-likely" estimate of individual latent variables. Note that a PPD derived with such conditioning excludes the effects of random rainfall errors (as intended for a PPD reflecting structural uncertainty only), but includes the effects of systematic rainfall biases (since the posterior mean of the multipliers in general deviates from unity). Further design and interpretation of partial predictive limits will be carried out in a separate development.

These results suggest that, in this particular application, a greater reduction in predictive uncertainty can be achieved by improving the hydrological model rather than by improving the accuracy of the rainfall and runoff data. We also stress that insights such as those above could not have been obtained with approaches that do not attempt to isolate structural uncertainty and therefore motivate further research efforts on the decompositional approach.

Author-produced version of the article published in
Water Resources Research (2011) vol. 47, doi : 10.1029/2011WR010643
The original publication is available at http://www.agu.org/journals/wr/

4.4 Posterior scrutiny of error model hypotheses

4.4.1 Input errors (OIS-CS only)

Figure 10 shows diagnostic plots to scrutinize the rainfall error model in eqn (16). In particular, Figure 10a-b suggests that the assumption of independent rainfall multipliers from a truncated Gaussian distribution is plausible (however, note the considerable posterior uncertainty).

Figure 10c-d yields further insights into the identifiability of rainfall errors. It assesses the extent to which the posterior estimates of true rainfall differ from the prior, i.e., whether the rainfall-runoff data and hydrological model contain sufficient information to modify the prior estimates of the true rainfall estimated using CS. Figure 10c compares the 90% credibility intervals of the true rainfall arising from the prior and the posterior distributions. In most cases, these intervals are similar, suggesting that the information content of the calibration data only marginally modifies the prior CS-based estimates of true areal rainfall. A few exceptions can be observed: e.g., on day 136, the posterior is considerably tighter than the prior.

The contribution of the rainfall-runoff data to the refinement of the rainfall error estimates during the hydrological model calibration can be quantified using an "uncertainty reduction factor" (URF). In this work, the URF is defined as the ratio of the posterior and prior standard deviations of each inferred rainfall multiplier. It can be interpreted as follows: (i) URF $\approx$ 0 implies a significant reduction of uncertainty in the areal rainfall estimates (high information content in calibration data); (ii) URF > 1 indicates increased uncertainty (e.g., if the calibration data conflicts with the prior); (iii) URF $\approx$ 1 indicates that the calibration data has not refined the rainfall error model and the inference of the rainfall multipliers is governed by the prior. Note that case (iii) is "non-informative" solely with respect to the inference of rainfall errors, and does *not* imply that the inference of the hydrological model parameters is non-informative or governed by the priors.

Figure 10d plots the URFs versus the corresponding R3D rainfall values. Two points can be made:

1) For large rainfall values (>20 mm), the URFs are mainly between 0.8 and 1, indicating little reduction in uncertainty. This implies that the prior (rather than the data) controls the inference of rainfall errors affecting large events. This is the likely reason for the ill-posedness of scheme OIS, which does not use the prior information in eqn (15).

2) URFs for smaller rainfall events are highly variable, with some multipliers having a significant reduction in uncertainty after calibration. Although perhaps un-expected given the low sensitivity

Author-produced version of the article published in
Water Resources Research (2011) vol. 47, doi : 10.1029/2011WR010643
The original publication is available at http://www.agu.org/journals/wr/

of the hydrological model to small rainfalls, such reductions could be explained by the constraint exerted by the error model in eqn (16): during calibration, multipliers with a large prior variance will have their posterior variances tightened approximately to σ_ϕ^2. Inadequacies of the simple rainfall error model (16) and the CS replicates for small rainfalls (section 3.2.3) may also be responsible for the differences in the URF patterns.

4.4.2 Structural errors (OIS-CS only)

Similarly to rainfall errors, Figure 11a-b suggests that the structural error model based on stochastic variation of θ_1 at the storm time scale, is plausible. However, as noted in Section 3.4, it is important to check the evolution of storage with respect to the production store capacity because stochastic variations of parameter θ_1 may lead to a store content exceeding the store capacity.

Figure 11c-d shows the evolution of storage during the calibration period. While the store remained consistently below its full capacity, exceedances did occur on some rare occasions. A closer inspection of GR4J [*Perrin et al.*, 2003] suggests two possible problematic scenarios:

1) If rainfall exceeds PET, a part of the net rainfall fills the production store, with the remainder being routed through unit hydrographs. However, when the store exceeds its capacity, some water is subtracted from the production store. Note that this does not create a water balance error because this overflowing water is simply transferred to the routing components. Moreover, in the 2-year calibration period, overflows due to stochastic variations of θ_1 amounted to a total of <1.5 mm, which is minor in the overall context of a 2-year runoff volume of nearly 300 mm.

2) If PET exceeds rainfall, a part of the store content is evaporated. The actual evaporation is computed as a function of the net PET and the store level. While exceeding the store capacity could result in the actual ET exceeding PET, this never occurred in this study.

4.4.3 Residual diagnostics

Figure 12 shows distributional and autocorrelation diagnostics for the standardized residuals. Note that, for all schemes except SLS, the residuals combine runoff and remnant errors (section 2). For those schemes, standardization is therefore performed by dividing the raw residual at time step *t* by $\sqrt{\sigma_{Q(t)}^2 + \sigma_\varepsilon^2}$, where σ_Q is the standard deviation of runoff errors (which are heteroscedastic with respect to the runoff magnitude, as shown in Eqn (17)) and σ_ε is the standard deviation of remnant errors (which, in this case study, are homoscedastic, as shown in Eqn (8)).

Author-produced version of the article published in
Water Resources Research (2011) vol. 47, doi : 10.1029/2011WR010643
The original publication is available at http://www.agu.org/journals/wr/

Several comments can be made:

1) Accounting for data errors (schemes OI-CS, OI and OIS-CS) markedly reduces the skewness and excess kurtosis (Figure 12a-b) of the standardized residuals. However, skewness and kurtosis remained statistically significant for all calibration schemes, including OIS-CS. This further discredits the assumption of homoscedastic Gaussian remnant errors and needs to be addressed.

2) The autocorrelation tends to decrease when more sources of errors are represented explicitly in the inference scheme (Figure 12c). The amount of prior information also appears to be an important factor, with markedly higher autocorrelations for scheme OI-CS than for scheme OI. Nevertheless, given appreciable remaining autocorrelation, the remnant error model may need autoregressive components.

5 Discussion

5.1 Quantification of predictive uncertainty

Section 4.2 indicated that the predictive distribution of runoff was fairly reliable for schemes OI-CS and OIS-CS. It can be seen that scheme OI-CS slightly under-estimates predictive uncertainty (see Figure 7 and Section 4.2) while scheme OI yields significantly larger estimated input errors and predictive uncertainty. This suggests that the CS prior constrains the input error estimates and reduces their ability to interact with structural errors, and compensate for un-accounted errors.

Arguably the most reliable predictive distribution is obtained with the OIS-CS scheme (Section 4.2), which includes the CS prior and an explicit characterization of structural errors using stochastic-parameters. Importantly, the OIS scheme, which omits prior information on the rainfall errors, leads to an ill-posed inference (Section 4.1). This is consistent with previous findings that priors on rainfall and runoff uncertainty control the well-posedness of Bayesian hierarchical inferences in hydrology [*Renard et al.*, 2010]. However, further work is warranted to improve the predictive precision of the OIS-CS scheme. If, as it appears for this case study, structural uncertainty is the dominant uncertainty, improving the predictive precision will likely require tightening the characterization of structural errors, as well as improving the hydrological model.

5.2 Decomposition of predictive uncertainty

The empirical results in Section 4.3 suggest that decomposing predictive uncertainty into its contributing sources is possible when independent estimates of rainfall and runoff data uncertainty are available and used in the BATEA inference. The reliability of this decomposition can be examined by considering (i) the reliability of the total predictive distribution, in combination with

(ii) the reliability of the individual data and structural components. However, scrutinizing individual components of a predictive distribution is considerably more challenging than scrutinizing the full predictive uncertainty, as discussed next.

The reliability of this decomposition can be examined by considering (i) the reliability of the total predictive distribution, in combination with (ii) the reliability of the individual data and structural components. However, scrutinizing "partial" uncertainties is considerably more challenging than scrutinizing the full predictive uncertainty, as discussed next.

Direct scrutiny of the estimated contribution of rainfall uncertainty to the uncertainty in the predicted runoff requires accurate areal rainfall estimates. Since this is rarely available, the adequacy of the decomposition can be investigated *indirectly* by scrutinizing the inferred distribution of latent variables. In this study, this posterior diagnostic was carried out only partially, by comparing the inferred/predicted rainfall errors with those suggested by the R3D raingauge network. Due to the short length of the dense-gauged R13H rainfall time series for this catchment, it was used entirely to construct the rainfall error model for the calibration period, and was not used to check the rainfall PD in the validation period. In applications where longer periods of densely gauged rainfall are available, it could be partitioned between calibration and validation.

Future avenues for scrutinizing the rainfall component include comparing inferred rainfall errors with the errors suggested by other sources, such as radar. Although radar estimates are affected by complex measurement errors [e.g. *Kirstetter et al.*, 2010], they can provide spatial information that is not captured by sparse raingauge networks. For instance, comparing the location of the main mass of a rainstorm with the location of the raingauges may shed light at least on the sign of the error (i.e., whether the raingauge network has under- or over-estimated the areal rainfall).

Direct validation of the estimated structural uncertainty requires highly accurate forcing-response data, so that structural errors can be isolated. This is seldom achievable in practice, except in densely gauged experimental catchments. However, indirect strategies are possible. For instance, assessing the stability of structural error estimates when different rainfall and runoff data are used provides a useful measure of the interactions between data and structural errors.

5.3 On the treatment of structural error

The treatment of structural error remains a topic of active research [e.g., see the discussion by *Beven*, 2005; *Doherty and Welter*, 2010]. This study does not aim to compare or improve methods for representing structural errors. Instead, it uses two particular structural error methods as part of a

study pursuing error decomposition by exploiting independently derived data error models. We view this as a logical first step before structural error characterization is tackled.

Many other distinct strategies have been proposed to represent structural errors, including stochastic state errors [*Moradkhani et al.*, 2005a], model-averaging schemes [e.g., *Duan et al.*, 2007; *Marshall et al.*, 2007], multi-model frameworks [e.g., *Clark et al.*, 2008] and other approaches [e.g., *Bulygina and Gupta*, 2009; *Jacquin and Shamseldin*, 2007]. Which of these approaches – if any – provides an adequate description of structural errors remains an open question. In particular, some authors have argued that the epistemic nature of structural uncertainty makes it poorly suited to a statistical treatment [e.g., *Beven*, 2008]. Our view is that, in a particular modeling context, such as hydrological modeling, such proposition prove or refute a priori. Yet the extent to which a statistical scheme succeeds in representing structural error can be scrutinized a posteriori by inspecting total and partial predictive uncertainties, applying residual and other diagnostics, etc.

5.4 Limitations and future work

While we are optimistic with respect to the practical feasibility of the Bayesian approach in the context of hydrologic prediction, several significant challenges remain to be tackled.

First, immediate limitations with respect to data availability are noted. In particular, CS requires a distributed raingauge network to calibrate the CS parameters and variograms. Applications where no reliable information exists to inform the data error models are unlikely to be suitable for a decomposition of sources of error. This provides a strong argument in favor of continuing measurement and experimental campaigns, and improving operational networks.

Second, the geostatistical rainfall model used in this paper can be improved to overcome the lack of reliability for small rainfall (see Section 3.2.4). For example, an approximate classification of rainfall events into more homogenous rainfall types (e.g. localized convective storms vs. frontal rainfall events) could be performed, and the geostatistical properties (e.g. variograms) estimated separately for each type. Similarly, orographic effects could be included through a regression with respect to elevation.

Third, while the error model in eqn (16) is geared primarily towards characterizing the errors in the larger rainfall events, the limitations of the multiplicative error model are noted. It is unable to handle errors in zero rainfalls (i.e. for a localized storm not recorded by the raingauge network) and appears poorly suited for errors in small rainfalls (see Section 3.2.3).

Author-produced version of the article published in
Water Resources Research (2011) vol. 47, doi : 10.1029/2011WR010643
The original publication is available at http://www.agu.org/journals/wr/

Finally, applications at a subdaily scale would require additional development of the rainfall and runoff error models, in particular including autocorrelation [*McMillan et al.*, 2011].

Other areas in need of further research attention include:

1) Generalization of the rating curve error model to rigorously distinguish between random and systematic rating curve errors, and to account for their likely autocorrelation at short time-scales. Several options are emerging, including Bayesian approaches [e.g., *Moyeed and Clarke*, 2005], dynamic schemes [*Dottori et al.*, 2009] and other methods [e.g., *McMillan et al.*, 2010].

2) Further appraisal of the stochastic-parameter approach and the development of more informative structural error models. The work by *Reichert and Mieleitner* [2009], who used an Uhlenbeck process to characterize the time structure of stochastic parameters in lieu of the epoch-dependence assumption [*Kuczera et al.*, 2006], is an important advance in this direction.

3) A more flexible remnant error model, in particular, allowing for autocorrelation and heteroscedasticity [e.g., see the recent work by *Schoups and Vrugt*, 2010; *Smith et al.*, 2010].

4) A better understanding of structural errors. In particular, the use of structural errors to diagnose, compare or improve hydrological models remains an important area of future research [e.g. *Reichert and Mieleitner*, 2009; *Smith et al.*, 2008].

5) The computational implementation is an area of ongoing work [e.g., *Kuczera et al.*, 2010a, and others; *Vrugt et al.*, 2008]. Moreover, given the emerging evidence that in many cases the geometrical complexity of parameter distributions is an artifact of the numerical implementation of the hydrological model, the use of efficient gradient-based schemes for optimization and uncertainty analysis is of interest [e.g., *Kavetski and Clark*, 2010; *Kavetski et al.*, 2006d].

The utility of these developments should be scrutinized using stringent posterior diagnostics. In particular, the predictive QQ plot [e.g., *Laio and Tamea*, 2007; *Thyer et al.*, 2009] and similar reliability checks, in combination with appraisals of the predictive precision, provide an objective yardstick to empirically evaluate the practical performance of the inference, and make quantitative judgments on their suitability for operational purposes.

6 Conclusions

The application of the Bayesian framework in a real-data case study confirms earlier findings that prior information on data uncertainty is not merely beneficial, but essential for a meaningful and reliable quantification and decomposition of the predictive uncertainty. In particular:

1) Simultaneous inference of forcing and structural errors within the hierarchical framework is ill-posed unless informative priors on forcing and response uncertainties are specified;

2) Ignoring sources of error may lead to unreliable predictions. Conversely, including additional error models improved the reliability of the total uncertainty estimates. We stress that this improvement was demonstrated in the validation period and thus unlikely to be due to potential over-fitting;

3) Including informative priors on rainfall uncertainty demonstrably improves the reliability of runoff predictions (scrutinized in a validation time period) and paves the way for a quantitative decomposition of the total predictive uncertainty into its contributing causes.

In this study, where the GR4J model was calibrated to a 3-gauge daily rainfall observation network in the Yzeron catchment (France), structural uncertainty appears to dominate data uncertainty. This conclusion is likely to be catchment- and model- dependent. In addition, further work is needed to further develop and test techniques for analyzing and communicating partial uncertainties.

The use of rainfall conditional simulation as part of hydrological model calibration represents a significant advance in the treatment of rainfall uncertainty in hydrological calibration. Whereas earlier work, including data assimilation approaches, previous applications of BATEA and analogous hierarchical Bayesian methods, used largely heuristic “rule-of-thumb” considerations in the specification of rainfall uncertainty, this study demonstrates that conditional simulation can provide more reliable and precise estimates of the uncertainties in individual rainfall measurements, and how these uncertainties vary in time. This demonstrably improves the statistical reliability of the model predictions when compared against methods that disregard such information. Perhaps more importantly, approximate decompositions of predictive uncertainty become possible, including separate estimation of structural errors of the hydrological model.

More generally, this study takes an important step towards more reliable uncertainty quantification and decomposition, which would be beneficial for many key scientific and operational purposes in hydrological and environmental, including (i) improved probabilistic forecasts and predictions, (ii) meaningful hydrological model evaluations un-obscured by data errors, and (iii) more efficient research and operational resource allocation to reduce predictive uncertainty.

Given the manifest significance of a robust quantitative understanding of data and modeling uncertainties in environmental studies, further development and implementation of instrumental and statistical procedures is needed to estimate the accuracy and precision of environmental data at

the data-collection and post-processing stages. The Bayesian paradigm, with its philosophy of using and refining the knowledge of all uncertain quantities – be it model parameters, true forcings, or some error properties of the latter – provides a very appealing platform for the systematic integration of these insights into environmental model inference and prediction.

Acknowledgments

This work is supported by a FAST grant from the Department of Innovation, Industry, Science and Research (Australia), the Ministry of Higher Education and Research (France) and the Ministry of Foreign and European Affairs (France). The helpful comments by Jasper Vrugt, Keith Beven and three anonymous reviewers substantially improved this paper and are gratefully acknowledged.

References

Beven, K. J. (2005), On the concept of model structural error, *Water Sci. Technol.*, *52*(6), 167-175.

Beven, K. J. (2006), A manifesto for the equifinality thesis, *J. Hydrol.*, *320*, 18–36.

Beven, K. J. (2008), On doing better hydrological science, *Hydrol. Process.*, *22*(17), 3549-3553, doi: 10.1002/hyp.7108.

Bulygina, N., and H. Gupta (2009), Estimating the uncertain mathematical structure of a water balance model via Bayesian data assimilation, *Water Resour. Res.*, *45*(12), W00B13, doi: 10.1029/2007wr006749.

Clark, M. P., and A. G. Slater (2006), Probabilistic quantitative precipitation estimation in complex terrain, *Journal of Hydrometeorology*, *7*(1), 3-22.

Clark, M. P., A. G. Slater, D. E. Rupp, R. A. Woods, J. A. Vrugt, H. V. Gupta, T. Wagener, and L. E. Hay (2008), Framework for Understanding Structural Errors (FUSE): A modular framework to diagnose differences between hydrological models, *Water Resources Research*, *44*.

Cowles, M. K., and B. P. Carlin (1996), Markov chain Monte Carlo convergence diagnostics: A comparative review, *J. Am. Stat. Assoc.*, *91*(434), 883-904.

Cressie, N., C. A. Calder, J. S. Clark, J. M. V. Hoef, and C. K. Wikle (2009), Accounting for uncertainty in ecological analysis: the strengths and limitations of hierarchical statistical modeling, *Ecological Applications*, *19*(3), 553-570.

Di Baldassarre, G., and A. Montanari (2009), Uncertainty in river discharge observations: a quantitative analysis, *Hydrol. Earth Syst. Sci.*, *13*(6), 913-921.

Doherty, J., and D. Welter (2010), A short exploration of structural noise, *Water Resour. Res.*, *46*(5), W05525.

Dottori, F., M. L. V. Martina, and E. Todini (2009), A dynamic rating curve approach to indirect discharge measurement, *Hydrol. Earth Syst. Sci. Discuss.*, *6*(1), 859-896.

Duan, Q., N. K. Ajami, X. Gao, and S. Sorooshian (2007), Multi-model ensemble hydrologic prediction using Bayesian model averaging, *Adv. Water Resour.*, *30*(5), 1371-1386.

Evensen, G. (1994), Sequential Data Assimilation with a Nonlinear Quasi-Geostrophic Model Using Monte-Carlo Methods to Forecast Error Statistics, *Journal of Geophysical Research-Oceans*, *99*(C5), 10143-10162.

Gelman, A., J. B. Carlin, H. S. Stern, and D. B. Rubin (2004), *Bayesian data analysis*, 2 ed., 696 pp., Texts in Statistical Science, Chapman & Hall

Gneiting, T., F. Balabdaoui, and A. E. Raftery (2007), Probabilistic forecasts, calibration and sharpness, *J. R. Stat. Soc. Ser. B-Stat. Methodol.*, *69*, 243-268.

Goldstein, M., and J. Rougier (2009), Reified Bayesian modelling and inference for physical systems, *J. Stat. Plan. Infer.*, *139*(3), 1221-1239.

Goovaerts, P. (2000), Geostatistical approaches for incorporating elevation into the spatial interpolation of rainfall, *J. Hydrol.*, *228*(1-2), 113-129.

Gotzinger, J., and A. Bardossy (2008), Generic error model for calibration and uncertainty estimation of hydrological models, *Water Resources Research*, *44*.

Hall, J., E. O'Connell, and J. Ewen (2007), On not undermining the science: coherence, validation and expertise. Discussion of Invited Commentary by Keith Beven Hydrological Processes, 20, 3141-3146 (2006), *Hydrol. Process.*, *21*(7), 985-988.

Herschy, R. (1994), The Analysis of Uncertainties in the Stage Discharge Relation, *Flow Meas Instrum*, *5*(3), 188-190.

Huard, D., and A. Mailhot (2008), Calibration of hydrological model GR2M using Bayesian uncertainty analysis, *Water Resources Research*, *44*.

Jacquin, A., and A. Y. Shamseldin (2007), Development of a possibilistic method for the evaluation of predictive uncertainty in rainfall-runoff modeling, *Water Resources Research*, *43*.

Journel, A. G., and C. J. Huijbregts (1978), *Mining geostatistics*, 600 pp., Academic Press, London

Kavetski, D., and M. P. Clark (2010), Ancient numerical daemons of conceptual hydrological modeling: 2. Impact of time stepping schemes on model analysis and prediction, *Water Resour. Res.*, *46*(10), W10511.

Kavetski, D., S. Franks, and G. Kuczera (2002), Confronting Input Uncertainty in Environmental Modelling in Calibration of Watershed Models, in *Water Science and Application Series 6*, edited by Q. Y. Duan, H. V. Gupta, S. Sorooshian, A. Rousseau and R. Tourcotte, pp. 49-68, American Geophysical Union, Washington DC.

Kavetski, D., G. Kuczera, and S. W. Franks (2006a), Calibration of conceptual hydrological models revisited: 1. Overcoming numerical artefacts, *J. Hydrol.*, *320*(1-2), 173-186.

Kavetski, D., G. Kuczera, and S. W. Franks (2006b), Bayesian analysis of input uncertainty in hydrological modeling: 1. Theory, *Water Resources Research*, *42*(3).

Kavetski, D., G. Kuczera, and S. W. Franks (2006c), Bayesian analysis of input uncertainty in hydrological modeling: 2. Application, *Water Resources Research*, *42*(3).

Kavetski, D., G. Kuczera, and S. W. Franks (2006d), Calibration of conceptual hydrological models revisited: 2. Improving optimisation and analysis, *J. Hydrol.*, *320*(1-2), 187-201.

Kennedy, M. C., and A. O'Hagan (2001), Bayesian Calibration of Computer Models, *Journal of the Royal Statistical Society. Series B (Statistical Methodology)*, *63*(3), 425-464.

Kirstetter, P. E., G. Delrieu, B. Boudevillain, and C. Obled (2010), Toward an error model for radar quantitative precipitation estimation in the Cevennes-Vivarais region, France, *J. Hydrol.*, *394*(1-2), 28-41, doi: 10.1016/j.jhydrol.2010.01.009.

Krajewski, W. F., G. J. Ciach, and E. Habib (2003), An analysis of small-scale rainfall variability in different climatic regimes, *Hydrol. Sci. J.-J. Sci. Hydrol.*, *48*(2), 151-162.

Krzysztofowicz, R. (1999), Bayesian theory of probabilistic forecasting via deterministic hydrologic model, *Water Resources Research*, *35*(9), 2739-2750.

Krzysztofowicz, R. (2002), Bayesian system for probabilistic river stage forecasting, *J. Hydrol.*, *268*(1-4), 16-40.

Kuczera, G., and B. J. Williams (1992), Effect of Rainfall Errors on Accuracy of Design Flood Estimates, *Water Resources Research*, *28*(4), 1145-1153.

Kuczera, G., D. Kavetski, S. Franks, and M. Thyer (2006), Towards a Bayesian total error analysis of conceptual rainfall-runoff models: Characterising model error using storm-dependent parameters, *J. Hydrol.*, *331*(1-2), 161-177.

Kuczera, G., D. Kavetski, B. Renard, and M. Thyer (2010a), A limited-memory acceleration strategy for MCMC sampling in hierarchical Bayesian calibration of hydrological models, *Water Resources Research.*, *46*.

Kuczera, G., B. Renard, M. Thyer, and D. Kavetski (2010b), There are no hydrological monsters, just models and observations with large uncertainties!, *Hydrological sciences Journal*, *55*(6).

Laio, F., and S. Tamea (2007), Verification tools for probabilistic forecasts of continuous hydrological variables, *Hydrol. Earth Syst. Sci.*, *11*(4), 1267-1277.

Lang, M., K. Pobanz, B. Renard, E. Renouf, and E. Sauquet (2010), Extrapolation of rating curves by hydraulic modelling and its relative impact on flood frequency analysis with historical data, *Hydrological sciences Journal.*, *55*(6).

Marshall, L., D. Nott, and A. Sharma (2007), Towards dynamic catchment modelling: a Bayesian hierarchical mixtures of experts framework, *Hydrol. Process.*, *21*(7), 847-861.

McMillan, H., J. Freer, F. Pappenberger, T. Krueger, and M. Clark (2010), Impacts of uncertain river flow data on rainfall-runoff model calibration and discharge predictions, *Hydrol. Process.*, *24*(10), 1270-1284.

McMillan, H., B. Jackson, M. Clark, D. Kavetski, and R. Woods (2011), Rainfall uncertainty in hydrological modelling: An evaluation of multiplicative error models, *J. Hydrol.*, *400*(1-2), 83-94.

Molini, A., L. G. Lanza, and P. La Barbera (2005), The impact of tipping-bucket raingauge measurement errors on design rainfall for urban-scale applications, *Hydrol. Process.*, *19*(5), 1073-1088.

Moradkhani, H., and S. Sorooshian (2008), General Review of Rainfall-Runoff Modeling: Model Calibration, Data Assimilation, and Uncertainty Analysis, in *Hydrological Modelling and the Water Cycle: Coupling the Atmosheric and Hydrological Models*, edited by S. Sorooshian, K. L. Hsu, E. Coppola, B. Tomassetti, M. Verdecchia and G. Visconti, pp. 1-24.

Moradkhani, H., K. L. Hsu, H. Gupta, and S. Sorooshian (2005a), Uncertainty assessment of hydrologic model states and parameters: Sequential data assimilation using the particle filter, *Water Resources Research*, *41*(5).

Moradkhani, H., S. Sorooshian, H. V. Gupta, and P. R. Houser (2005b), Dual state–parameter estimation of hydrological models using ensemble Kalman filter, *Adv. Water Resour.*, *28*, 135-147.

Moradkhani, H., K. Hsu, Y. Hong, and S. Sorooshian (2006), Investigating the impact of remotely sensed precipitation and hydrologic model uncertainties on the ensemble streamflow forecasting, *Geophys. Res. Lett.*, *33*(12).

Moulin, L., E. Gaume, and C. Obled (2009), Uncertainties on mean areal precipitation: assessment and impact on streamflow simulations, *Hydrol. Earth Syst. Sci.*, *13*(2), 99-114.

Moyeed, R. A., and R. T. Clarke (2005), The use of Bayesian methods for fitting rating curves, with case studies., *Adv. Water Resour.*, *28*, 807-818.

Onibon, H., T. Lebel, A. Afouda, and G. Guillot (2004), Gibbs sampling for conditional spatial disaggregation of rain fields, *Water Resources Research*, *40*(8).

Oudin, L., C. Perrin, T. Mathevet, V. Andreassian, and C. Michel (2006), Impact of biased and randomly corrupted inputs on the efficiency and the parameters of watershed models, *J. Hydrol.*, *320*(1-2), 62-83.

Perrin, C., C. Michel, and V. Andreassian (2003), Improvement of a parsimonious model for streamflow simulation, *J. Hydrol.*, *279*(1-4), 275-289.

Rajaram, H., and K. P. Georgakakos (1989), Recursive Parameter-Estimation of Hydrologic-Models, *Water Resources Research*, *25*(2), 281-294.

Reggiani, P., M. Renner, A. H. Weerts, and P. van Gelder (2009), Uncertainty assessment via Bayesian revision of ensemble streamflow predictions in the operational river Rhine forecasting system, *Water Resources Research*, *45*.

Reichert, P., and J. Mieleitner (2009), Analyzing input and structural uncertainty of nonlinear dynamic models with stochastic, time-dependent parameters, *Water Resources Research*, *45*.

Reichle, R. H. (2008), Data assimilation methods in the Earth sciences, *Adv. Water Resour.*, *31*(11), 1411-1418, doi: 10.1016/j.advwatres.2008.01.001.

Reichle, R. H., D. B. McLaughlin, and D. Entekhabi (2002), Hydrologic data assimilation with the ensemble Kalman filter, *Monthly Weather Review*, *130*(1), 103-114.

Reitan, T., and A. Petersen-Overleir (2009), Bayesian methods for estimating multi-segment discharge rating curves, *Stochastic Environmental Research and Risk Assessment. In press.*

Renard, B., D. Kavetski, M. Thyer, G. Kuczera, and S. W. Franks (2010), Understanding predictive uncertainty in hydrologic modeling: The challenge of identifying input and structural errors, *Water Resources Research*, *46*.

Rougier, J. C. (2007), Probabilistic Inference for Future Climate Using an Ensemble of Climate Model Evaluations, *Clim. Change*, *81*, 247-264.

Salamon, P., and L. Feyen (2009), Assessing parameter, precipitation, and predictive uncertainty in a distributed hydrological model using sequential data assimilation with the particle filter, *J. Hydrol.*, *376*(3-4), 428-442.

Salamon, P., and L. Feyen (2010), Disentangling uncertainties in distributed hydrological modeling using multiplicative error models and sequential data assimilation, *Water Resources Research. in press.*

Schoups, G., and J. A. Vrugt (2010), A formal likelihood function for parameter and predictive inference of hydrologic models with correlated, heteroscedastic, and non-Gaussian errors, *Water Resour. Res.*, *46*(10), W10531, doi: 10.1029/2009wr008933.

Seo, D.-J., H. D. Herr, and J. C. Schaake (2006), A statistical post-processor for accounting of hydrologic uncertainty in short-range ensemble streamflow prediction, *Hydrology and Earth System Sciences - Discussion*, *3*, 1987-2035.

Severino, E., and T. Alpuim (2005), Spatiotemporal models in the estimation of area precipitation, *Environmetrics*, *16*, 773-802.

Smith, P. J., K. J. Beven, and J. A. Tawn (2008), Detection of structural inadequacy in process-based hydrological models: A particle-filtering approach, *Water Resources Research*, *44*(1), doi: 10.1029/2006WR005205.

Smith, T., A. Sharma, L. Marshall, R. Mehrotra, and S. Sisson (2010), Development of a formal likelihood function for improved Bayesian inference of ephemeral catchments, *Water Resources Research*, *46*, doi: W12551

10.1029/2010wr009514.

Spiegelhalter, D. J., A. Thomas, and N. G. Best (2003), *WinBugs, Version 1.4, User Manual*, Institute of Public Health, Cambridge, UK

Spiegelhalter, D. J., N. G. Best, B. R. Carlin, and A. van der Linde (2002), Bayesian measures of model complexity and fit, *J. R. Stat. Soc. Ser. B-Stat. Methodol.*, *64*, 583-616.

Thyer, M., B. Renard, D. Kavetski, G. Kuczera, S. W. Franks, and S. Srikanthan (2009), Critical evaluation of parameter consistency and predictive uncertainty in hydrological modelling: a case study using bayesian total error analysis, *Water Resources Research*, *45*.

Todini, E. (2008), A model conditional processor to assess predictive uncertainty in flood forecasting, *International Journal of River Basin Management 6*(2), 123-137.

Tompson, A. F. B., R. Ababou, and L. W. Gelhar (1989), Implementation of the 3-Dimensional Turning Bands Random Field Generator, *Water Resources Research*, *25*(10), 2227-2243.

Villarini, G., P. V. Mandapaka, W. F. Krajewski, and R. J. Moore (2008), Rainfall and sampling uncertainties: A rain gauge perspective, *J. Geophys. Res.-Atmos.*, *113*(D11).

Vischel, T., T. Lebel, S. Massuel, and B. Cappelaere (2009), Conditional simulation schemes of rain fields and their application to rainfall-runoff modeling studies in the Sahel, *J. Hydrol.*, *375*(1-2), 273-286.

Vrugt, J. A., C. G. H. Diks, H. V. Gupta, W. Bouten, and J. M. Verstraten (2005), Improved treatment of uncertainty in hydrologic modeling: Combining the strengths of global optimization and data assimilation, *Water Resources Research*, *41*(1).

Vrugt, J. A., C. J. F. ter Braak, M. P. Clark, J. M. Hyman, and B. A. Robinson (2008), Treatment of input uncertainty in hydrologic modeling: Doing hydrology backward with Markov chain Monte Carlo simulation, *Water Resour. Res.*, *44*.

Weerts, A. H., and G. Y. H. El Serafy (2006), Particle filtering and ensemble Kalman filtering for state updating with hydrological conceptual rainfall-runoff models, *Water Resources Research*, *42*(9).

Willems, P. (2001), Stochastic description of the rainfall input errors in lumped hydrological models, *Stoch. Environ. Res. Risk Assess.*, *15*(2), 132-152.

Young, P. (1998), Data-based mechanistic modelling of environmental, ecological, economic and engineering systems, *Environmental Modelling & Software*, *13*(2), 105-122.

Young, P., P. McKenna, and J. Bruun (2001), Identification of non-linear stochastic systems by state dependent parameter estimation, *International Journal of Control*, *74*(18), 1837-1857.

Author-produced version of the article published in
Water Resources Research (2011) vol. 47, doi : 10.1029/2011WR010643
The original publication is available at http://www.agu.org/journals/wr/

Appendix A. Details of the TBM rainfall generator

The models given in Table A 1 apply to the point-process generating (unobserved) instantaneous rainfall. Moreover, the variograms are those used to generate Gaussian random fields, and may differ from the empirical variograms of observed data. The following steps are necessary:

Step 1: Pass from Gaussian to real space

A Gaussian field $\boldsymbol{U} = U(x,y,t)$ generated using the variograms in Table A 1 is transformed into a real-space random field $\boldsymbol{R} = R(x,y,t)$ using the following transformations:

- For the indicator field: $R(x,y,t) = \begin{cases} 0 \text{ if } U(x,y,t) < \Phi^{-1}(\zeta_I) \\ 1 \text{ otherwise} \end{cases}$
- For the non-zero field: $R(x,y,t) = F_W^{-1}(\Phi(U(x,y,t)))$

where Φ is the standard Gaussian cdf and F_W is the cdf of the at-site distribution in Table A 1.

These transformations alter the correlations of the transformed field, which will not match those of the Gaussian field. The variograms γ_I and γ_W in real space are therefore derived as follows:

- For the indicator field, an exact formula can be used,

$$\gamma_I(d) = \int_0^{\arcsin\left(\sqrt{\gamma_I^*(d)/2}\right)} \exp\left(-\frac{\Phi^{-1}(\zeta_I)}{1+\cos t}\right) dt$$

- For the non-zero field, the transformation is assumed to affect only the sill of the variogram,

$$\gamma_W(d) = \mathrm{Var}[W] \times \gamma_W^*(d),$$

where $\mathrm{Var}[W]$ is the variance of the at-site distribution (Table A 1). Simulation studies suggest that this is a reasonable hypothesis when the coefficient of variation of the at-site distribution is moderate.

Author-produced version of the article published in
Water Resources Research (2011) vol. 47, doi : 10.1029/2011WR010643
The original publication is available at http://www.agu.org/journals/wr/

Step 2: Convert partial variograms into the total variogram

Given the variograms of the indicator field *I* and the non-zero field *W*, the variogram $\gamma(d)$ of the rainfall field $Z = I*W$ can be derived as follows [*Lepioufle*, 2009]:

$$\gamma(d) = \left[1 - \zeta_I - \gamma_I(d)\right]\gamma_W(d) + 2\gamma_I(d)\gamma_{Tr}(d)$$

where $\gamma_{Tr}(d)$ is the transition variogram between zero and non-zero rainfall. If *I* and *W* are independent, this variogram does not depend on the distance *d* and is equal to [*Lepioufle*, 2009]:

$$\gamma_{Tr}(d) = \frac{\boldsymbol{\mu}_W^2 + \mathrm{Var}[W]}{2}$$

Step 3: Convert simultaneous rainfall into cumulated rainfall

The variogram $\gamma(d)$ describes the instantaneous rainfall field, yet the observed data are rainfall cumulated over a given duration (e.g., one hour). It is therefore necessary to derive the variogram of the cumulated rainfall field.

Let (x_1, y_1) and (x_2, y_2) be two pixels in the simulation domain, and τ_1 and τ_2 two time points. $h = \sqrt{d_S^2\left[(x_1, y_1);(x_2, y_2)\right]} = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2}$ is the spatial distance between (x_1, y_1) and (x_2, y_2). The spatial variogram of the rainfall cumulated over a duration *D* can be derived as follows [e.g., *Journel and Huijbregts*, 1978]:

$$\gamma_D(h) = \frac{1}{D^2}\iint_{[0;D]} \gamma(d_{S-T}\left[(x_1, y_1, \tau_1);(x_2, y_2, \tau_2)\right])d\tau_1 d\tau_2 - \frac{1}{D^2}\iint_{[0;D]} \gamma(d_{S-T}\left[(0,0,\tau_1);(0,0,\tau_2)\right])d\tau_1 d\tau_2$$

Step 4: Estimation

$\gamma_D(h)$ is the spatial variogram of observed data cumulated over a duration *D*. The parameters in Table A 1 can therefore be estimated by fitting the variograms $\gamma_{D_1}(h), ..., \gamma_{D_k}(h)$ to the empirical variograms of observed data cumulated over durations $D_1 \ldots D_k$. A simple least-square fitting criterion is used in this case study, with durations 1, 3, 6, 12 and 24 hours. The inferred parameters are given in Table A 2.

Author-produced version of the article published in
Water Resources Research (2011) vol. 47, doi : 10.1029/2011WR010643
The original publication is available at http://www.agu.org/journals/wr/

Figure 1. Derivation of partial predictive uncertainties from the MCMC analysis of the joint BATEA posterior. Here, $(\boldsymbol{\theta}^{(i)}, \boldsymbol{\phi}^{(i)}, \mu_{\phi}^{(i)}, \sigma_{\phi}^{(i)}, \boldsymbol{\theta}_1^{(i)}, \mu_{\theta_1}^{(i)}, \sigma_{\theta_1}^{(i)}, \sigma_{\varepsilon}^{(i)})$ is the ith MCMC sample, $i = 1{:}N_{sim}$ and $(\breve{\boldsymbol{\theta}}, \breve{\boldsymbol{\phi}}, \breve{\mu}_{\phi}, \breve{\sigma}_{\phi}, \breve{\boldsymbol{\theta}}_1, \breve{\mu}_{\theta_1}, \breve{\sigma}_{\theta_1}, \breve{\sigma}_{\varepsilon})$ is a selected point-estimate of inferred quantities on which the PPDs are conditioned (in this study, the modal values are used).

Figure 2. Data and model used in the case study. (a) Map of the Yzeron catchment; (b) rainfall (R3D) and runoff time series; (c) schematic of GR4J [adapted from *Perrin et al.*, 2003].

Figure 3. Examples of conditional rainfall replicates ("CS replicates") over the Yzeron catchment. Here, 3 replicates comprising 4 consecutive hourly steps are shown.

Figure 4. Comparison of the CS replicates to catchment-average R3D rainfall estimates. (a) Time series, dots = R3D, lines = replicates; (b) standard deviation of replicates; (c) mean of replicates; (d) rainfall multiplier.

Figure 5. Evaluation of the predictive distribution of daily rainfall obtained using CS against the three R3D raingauges (each symbol corresponding to a specific raingauge). P-values on the bisector line indicate statistically reliable predictions. (a) non-zero rainfalls smaller than 2 mm; (b) rainfalls larger than 2 mm.

Figure 6. Heteroscedastic model of runoff measurement errors estimated from rating curve analysis of the Yzeron catchment.

Figure 7. Evaluation of the predictive distributions of runoff estimated using different BATEA schemes. The PQQ plot for observed runoff in validation is shown for: (a) all runoffs; and (b) runoffs exceeding 2 mm.

Figure 8. Total predictive uncertainty for the five largest events of the validation period. Each column depicts a different storm event. Shaded areas represent 50 and 90% predictive intervals (from darkest to lightest), the black line represents the predictive median. Observed runoff values are shown with dots.

Figure 9. Decomposition of the predictive uncertainty estimated using the OIS-CS scheme. Observed runoff values are shown with dots. Shaded areas represent 50 and 90% predictive intervals (from darkest to lightest) for the total predictive distribution, thick lines represent 90% predictive intervals for the partial predictive distribution. The figure shows the contribution of

Author-produced version of the article published in
Water Resources Research (2011) vol. 47, doi : 10.1029/2011WR010643
The original publication is available at http://www.agu.org/journals/wr/

output and remnant errors (first row), structural errors (second row), and input errors (third row). In this case study, structural uncertainty appears dominant.

Figure 10. Posterior diagnostics for the rainfall error model in the OIS-CS scheme. (a) QQ plot of estimated rainfall multipliers (posterior mode). The red line represents the truncated Gaussian distribution, grey bars represent 90% posterior intervals for each multiplier; (b) autocorrelation function of estimated rainfall multipliers; (c) prior vs. posterior estimates of true rainfall (90% intervals are shown); (d) uncertainty reduction factor (URF), defined as the ratio of posterior and prior standard deviations of individual rainfall multipliers.

Figure 11. Posterior diagnostics for the structural error model in the OIS-CS scheme. (a) QQ plot of estimated θ_1 values (posterior mode). The red line represents the truncated Gaussian distribution, grey bars represent 90% posterior intervals for each θ_1 value; (b) autocorrelation function of estimated θ_1 values; (c) content (thin black line) and stochastic capacity (thick red line) of the production store; (d) content (thin black line) and capacity (thick red line) of the routing store.

Figure 12. Diagnostics of standardized residuals (which represent combined runoff and remnant errors). Dashed lines indicate significance limits (level $\alpha = 5\%$).

Author-produced version of the article published in
Water Resources Research (2011) vol. 47, doi : 10.1029/2011WR010643
The original publication is available at http://www.agu.org/journals/wr/

Table 1. Summary of calibration schemes and MCMC convergence results.

Name	Rainfall errors			Runoff errors		Stochastic parameter		Remnant errors		RR model parameters	Case study details		
	Model	Prior $p(\boldsymbol{R})$	Prior $p(\Theta_R)$	Model	Prior $p(\Theta_Q)$	Model	Prior $p(\Theta_\Lambda)$	Model	Prior $p(\Theta_\varepsilon)$	Prior $p(\boldsymbol{\theta})$	Number of inferred quantities	Convergence: Iteration (x 10^3)	Convergence: CPU time (hour)[$]
SLS	n/a	n/a	n/a	n/a	n/a	n/a	n/a	Eq. (8)	Vague	Vague	5	0.6	0.01
OI	Eq. (16)	Vague	Vague	Eq. (17)	Dirac[£]	n/a	n/a	Eq. (8)	Vague	Vague	173	259.5	3.28
OI-CS	Eq. (16)	Eq. (15)	Vague	Eq. (17)	Dirac[£]	n/a	n/a	Eq. (8)	Vague	Vague	173	10.5	0.23
OIS	Eq. (16)	Vague	Vague	Eq. (17)	Dirac[£]	Eq. (18)	Vague	Eq. (8)	Vague	Vague	268	∞	∞
OIS-CS	Eq. (16)	Eq. (15)	Vague	Eq. (17)	Dirac[£]	Eq. (18)	Vague	Eq. (8)	Vague	Vague	268	647.0	16.55

[£] Parameters *a* and *b* of eqn (17) are fixed at their estimated values.

[$] 2.4 GHz desktop CPU

Author-produced version of the article published in
Water Resources Research (2011) vol. 47, doi : 10.1029/2011WR010643
The original publication is available at http://www.agu.org/journals/wr/

Table 2. Rainfall and structural uncertainty estimated as part of the hydrological model inference using BATEA. The posterior medians are reported, followed by the corresponding posterior standard deviations.

	Rainfall multipliers		Stochastic θ_1	
	Mean	Standard deviation	Mean	Shtandard deviation
OI	0.20 ± 0.17	1.54 ± 0.15	-	-
OI-CS	1.18 ± 0.03	0.30 ± 0.03	-	-
OIS-CS	1.15 ± 0.03	0.27 ± 0.02	218 ± 23	84 ± 17

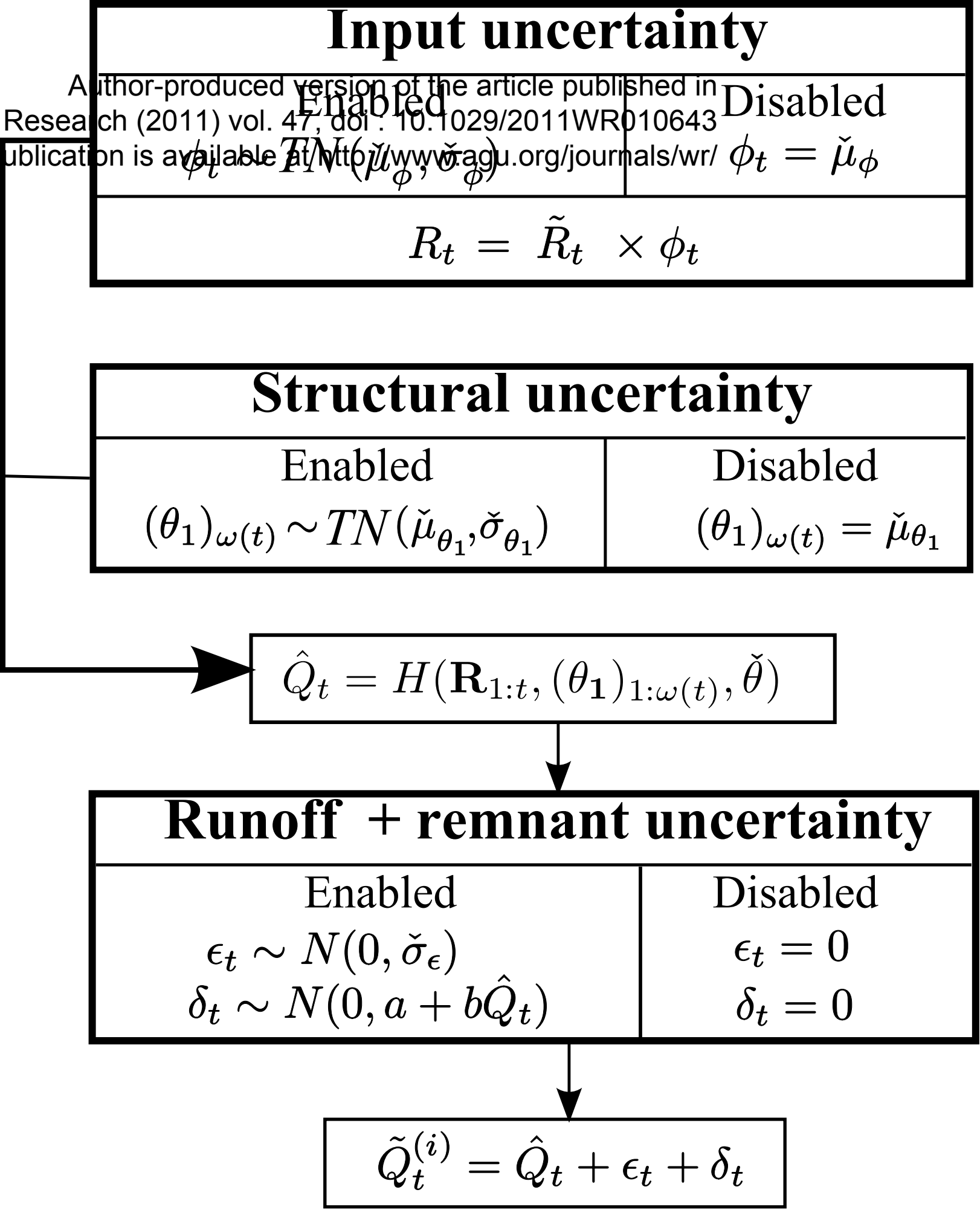
Author-produced version of the article published in
Water Resources Research (2011) vol. 47, doi : 10.1029/2011WR010643
The original publication is available at http://www.agu.org/journals/wr/

Table A 1. Geostatistical rainfall models used in this case study. The parameters are indicated in bold font.

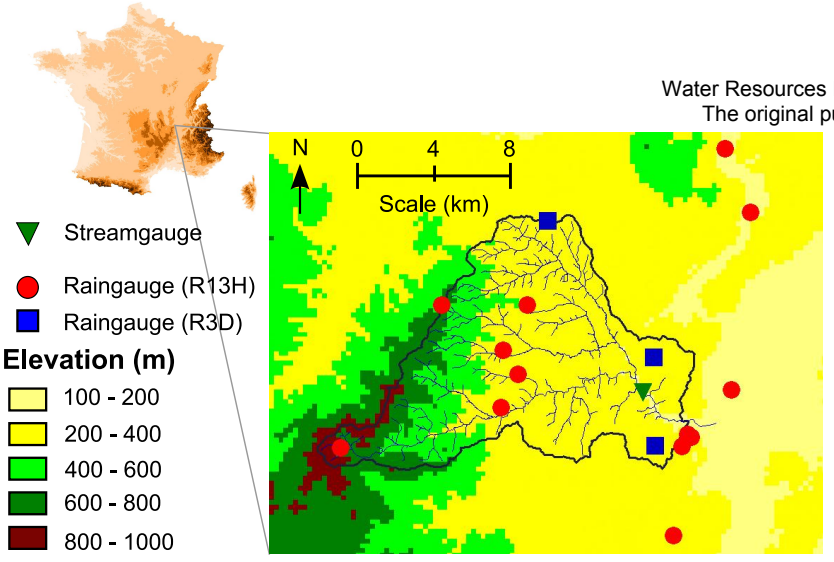
	Indicator field $I(x,y,t)$	Non-zero rainfall field $W(x,y,t)$
At-site distribution	Binomial, $\Pr(I(x,y,t)=0)=\boldsymbol{\zeta}_I$	Inverse Gaussian, $p_W(w)=\sqrt{\frac{\boldsymbol{\lambda}_W}{2\pi w^3}}\exp\left[-\frac{\boldsymbol{\lambda}_W(w-\boldsymbol{\mu}_W)^2}{2\boldsymbol{\mu}_W^2 w}\right]$
Spatio-temporal distance	$d_{S-T}\left[(x_1,y_1,t_1);(x_2,y_2,t_2)\right]=\sqrt{d_S^2\left[(x_1,y_1);(x_2,y_2)\right]+\boldsymbol{\alpha}_I^2 d_T^2\left[t_1;t_2\right]}$ $=\sqrt{(x_2-x_1)^2+(y_2-y_1)^2+\boldsymbol{\alpha}_I^2(t_2-t_1)^2}$	$d_{S-T}\left[(x_1,y_1,t_1);(x_2,y_2,t_2)\right]=\sqrt{d_S^2\left[(x_1,y_1);(x_2,y_2)\right]+\boldsymbol{\alpha}_W^2 d_T^2\left[t_1;t_2\right]}$ $=\sqrt{(x_2-x_1)^2+(y_2-y_1)^2+\boldsymbol{\alpha}_W^2(t_2-t_1)^2}$
Variogram (in Gaussian space)	Spherical, $\gamma_I^*(d)=\begin{cases}\frac{3d}{2\boldsymbol{\rho}_I}-\frac{d^3}{2\boldsymbol{\rho}_I^3} & \text{if } d<\boldsymbol{\rho}_I\\ 1 & \text{otherwise}\end{cases}$	Spherical, $\gamma_W^*(d)=\begin{cases}\frac{3d}{2\boldsymbol{\rho}_W}-\frac{d^3}{2\boldsymbol{\rho}_W^3} & \text{if } d<\boldsymbol{\rho}_W\\ 1 & \text{otherwise}\end{cases}$

Table A 2. Estimated parameters and simulation grid for the geostatistical models.

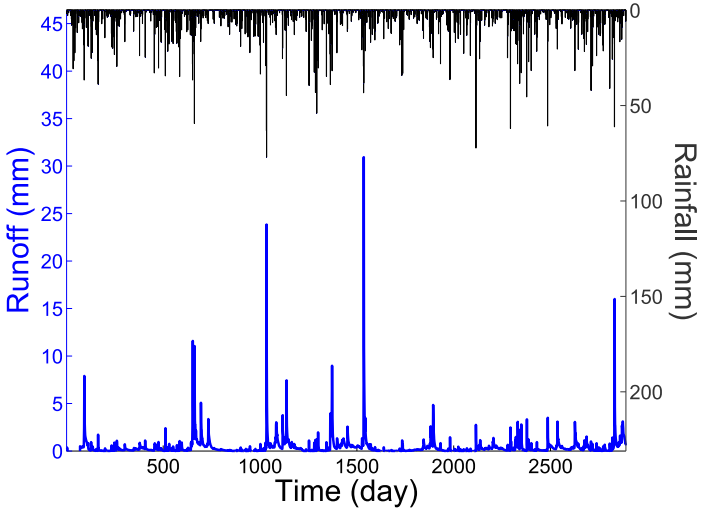
	Indicator field $I(x,y,t)$	Non-zero field $W(x,y,t)$
At-site distribution	$\boldsymbol{\zeta}_I=0.58$	$\boldsymbol{\mu}_W=1.13;\boldsymbol{\lambda}_W=0.36$
Spatio-temporal distance	$\boldsymbol{\alpha}_I=10\,\text{km.h}^{-1}$	$\boldsymbol{\alpha}_W=10\ \text{km.h}^{-1}$
Variogram (in Gaussian space)	$\boldsymbol{\rho}_I=30$ km	$\boldsymbol{\rho}_W=30$ km
Simulation grid	$\boldsymbol{\delta}_x=$ 500 m; $\boldsymbol{\delta}_y=$ 500 m; $\boldsymbol{\delta}_t=$ 1 hour	



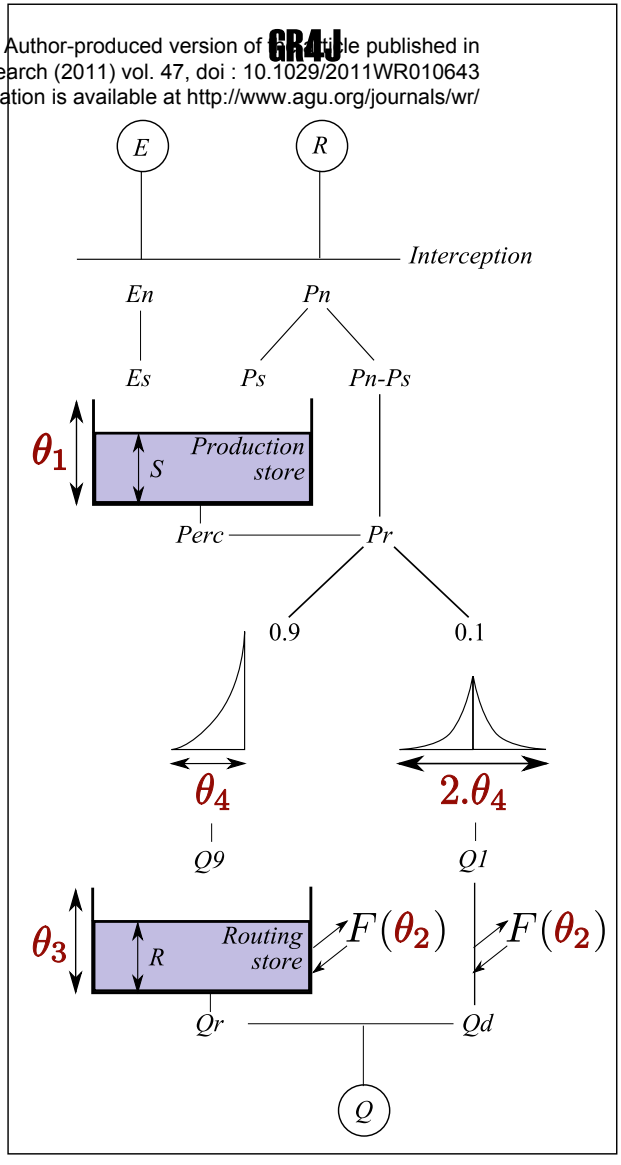
Author-produced version of the article published in
Water Resources Research (2011) vol. 47, doi : 10.1029/2011WR010643
The original publication is available at http://www.agu.org/journals/wr/



(a)

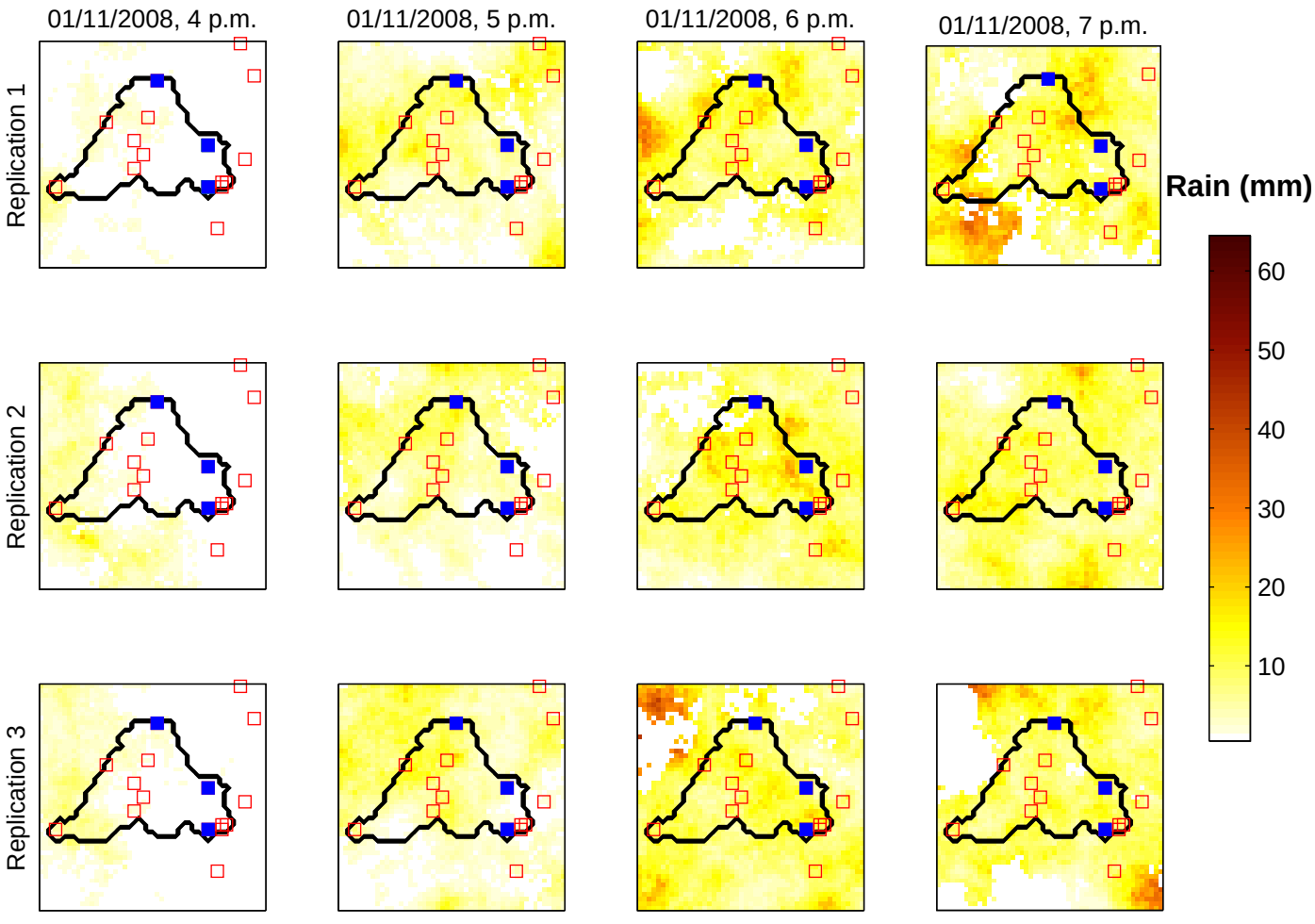


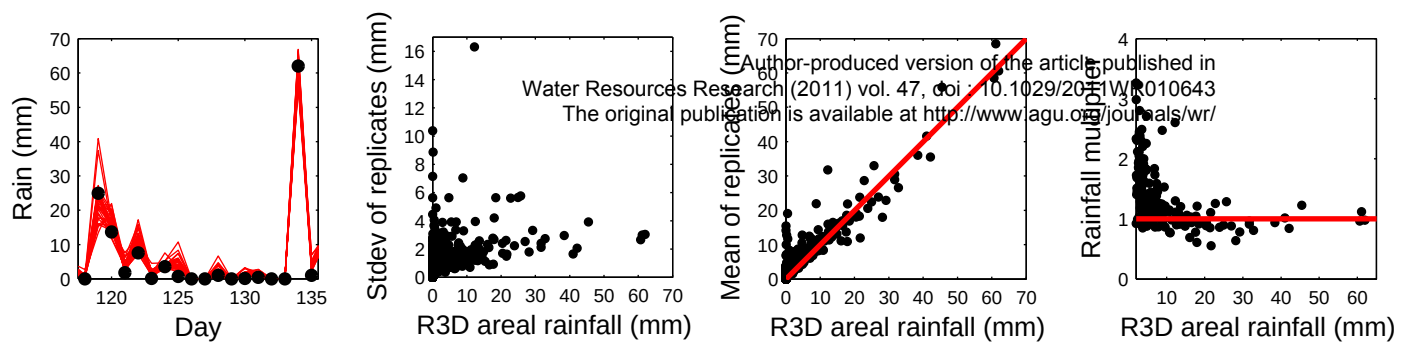
(b)



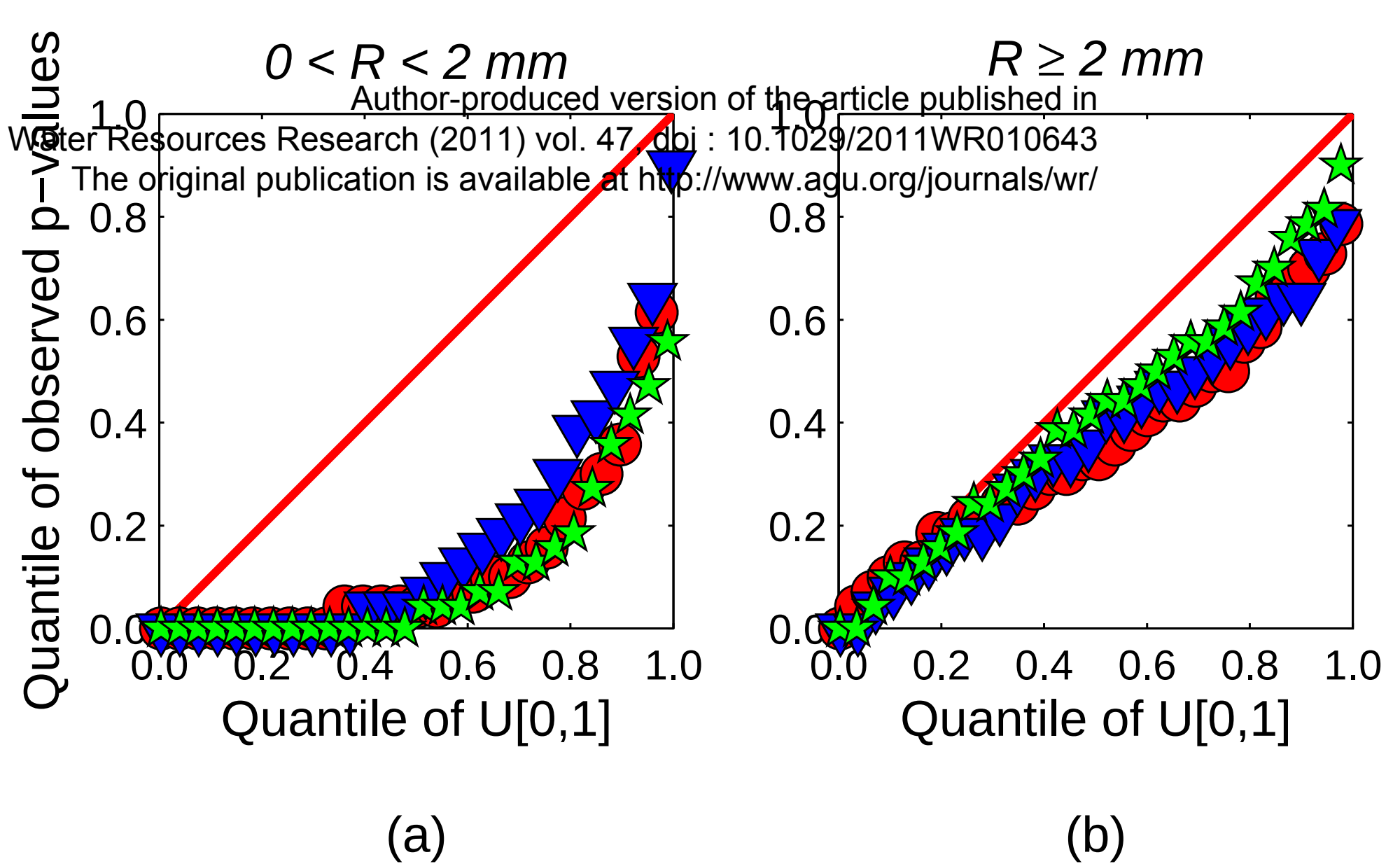
(c)

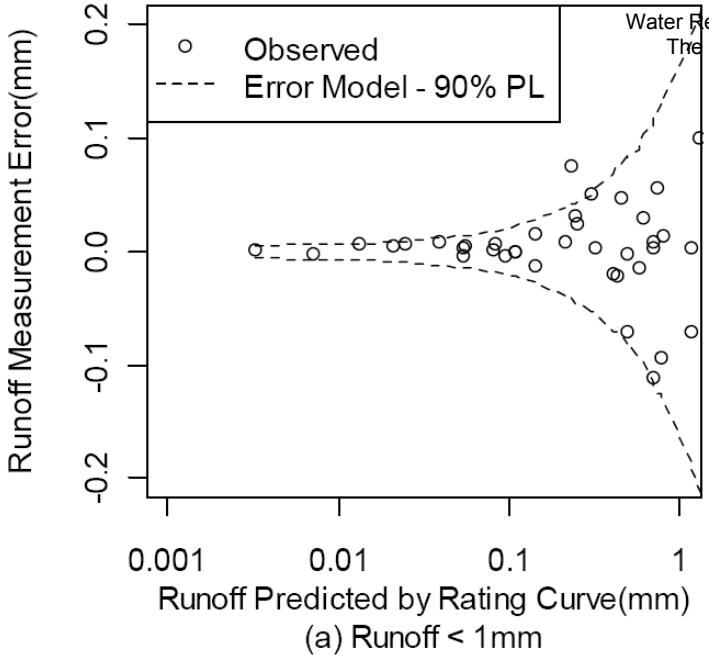
Author-produced version of the article published in
Water Resources Research (2011) vol. 47, doi : 10.1029/2011WR010643
The original publication is available at http://www.agu.org/journals/wr/

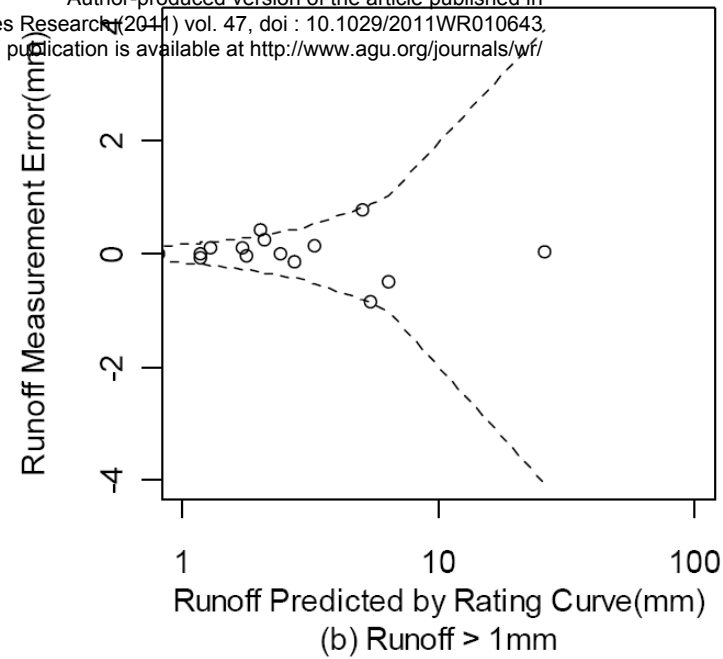




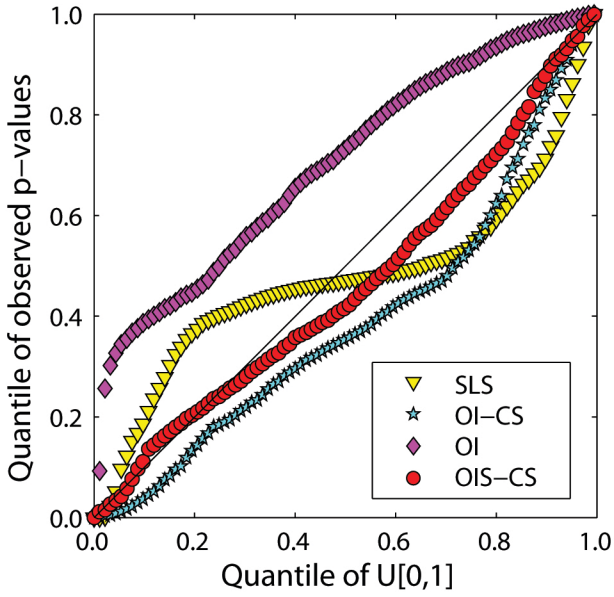
(a) (b) (c) (d)



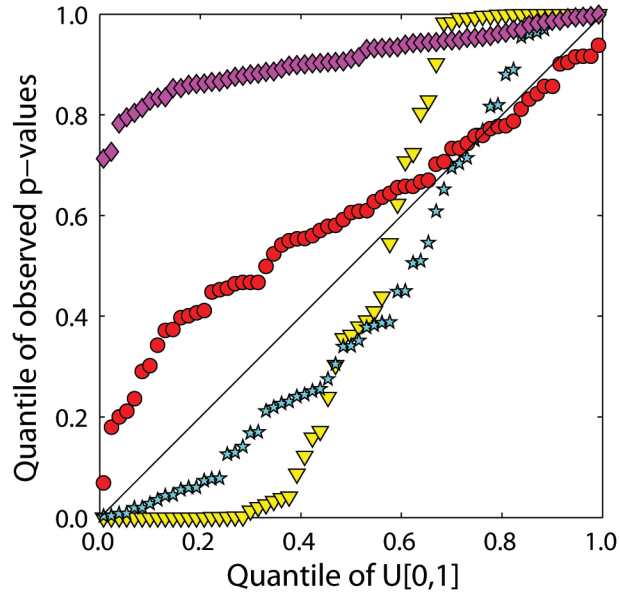




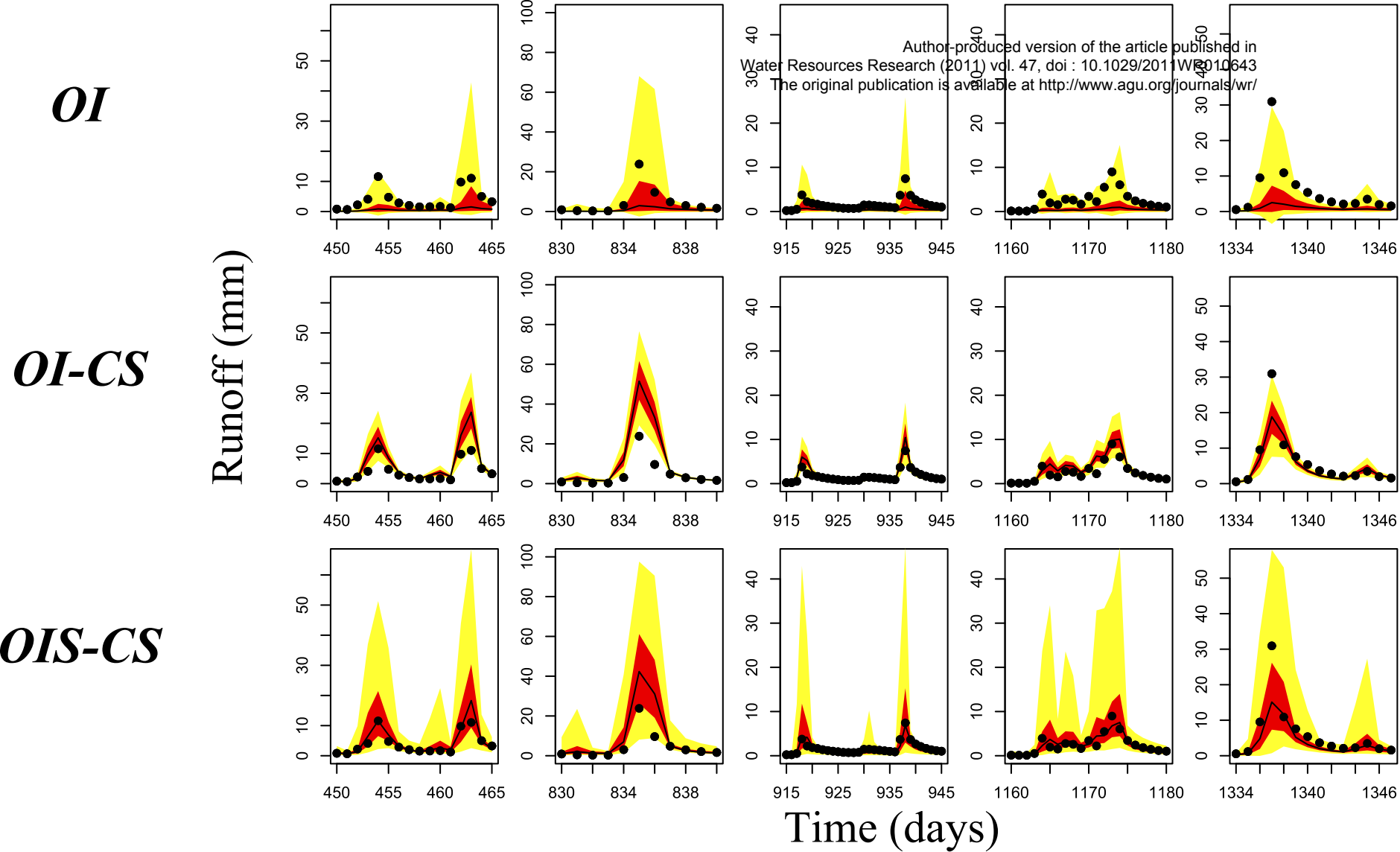
Author-produced version of the article published in
Water Resources Research (2011) vol. 47, doi : 10.1029/2011WR010643
The original publication is available at http://www.agu.org/journals/wr/

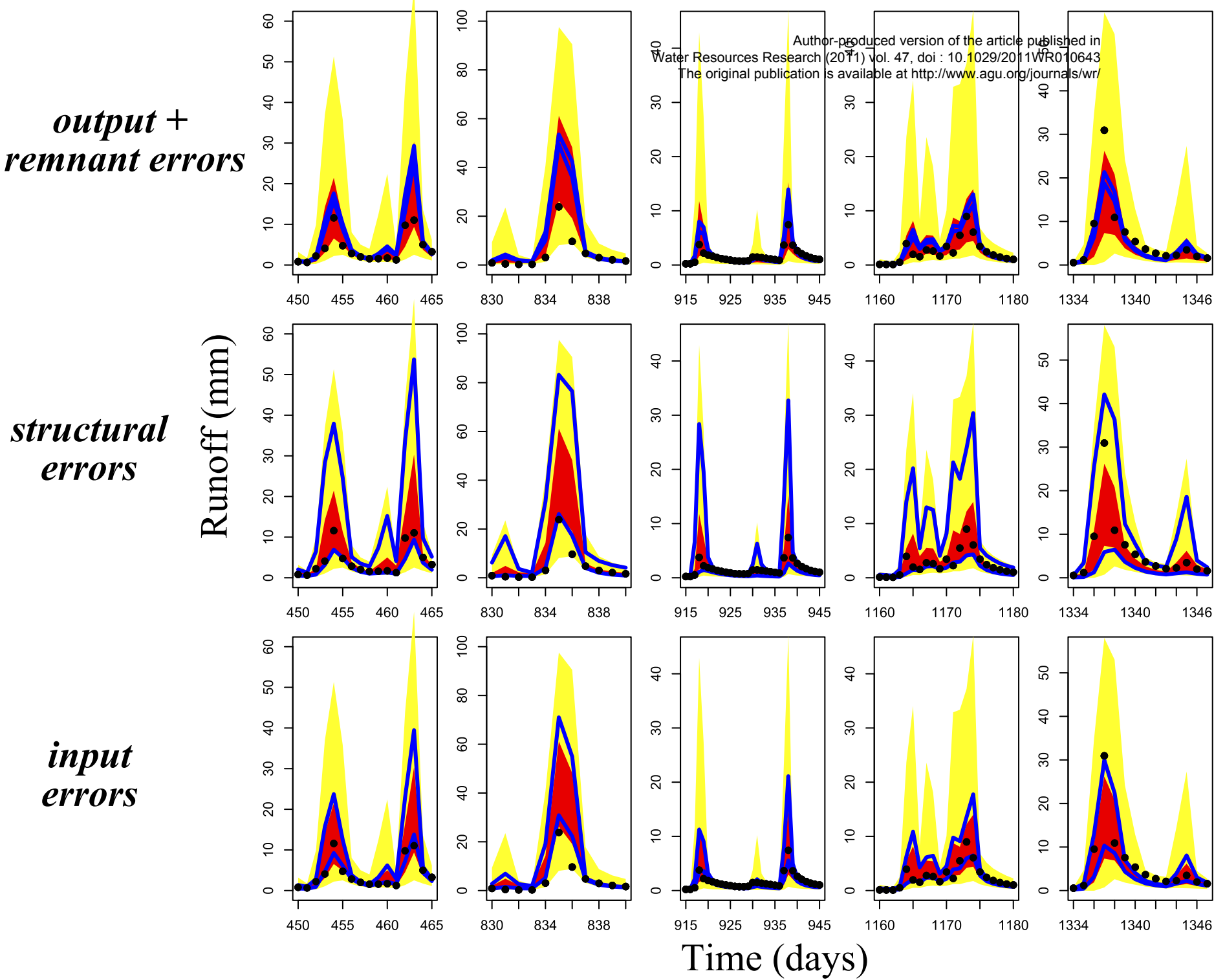


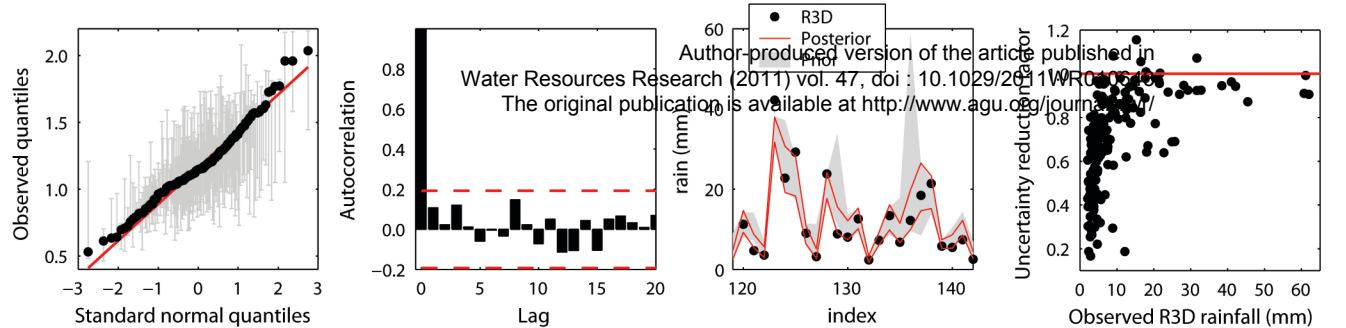
(a)



(b)







Author-produced version of the article published in
Water Resources Research (2011) vol. 47, doi : 10.1029/2011WR010643
The original publication is available at http://www.agu.org/journals/wr/