

A formal likelihood function for parameter and predictive inference of hydrologic models with correlated, heteroscedastic, and non-Gaussian errors

Gerrit Schoups¹ and Jasper A. Vrugt^{2,3,4}

Received 26 November 2009; revised 22 May 2010; accepted 10 June 2010; published 20 October 2010.

[1] Estimation of parameter and predictive uncertainty of hydrologic models has traditionally relied on several simplifying assumptions. Residual errors are often assumed to be independent and to be adequately described by a Gaussian probability distribution with a mean of zero and a constant variance. Here we investigate to what extent estimates of parameter and predictive uncertainty are affected when these assumptions are relaxed. A formal generalized likelihood function is presented, which extends the applicability of previously used likelihood functions to situations where residual errors are correlated, heteroscedastic, and non-Gaussian with varying degrees of kurtosis and skewness. The approach focuses on a correct statistical description of the data and the total model residuals, without separating out various error sources. Application to Bayesian uncertainty analysis of a conceptual rainfall-runoff model simultaneously identifies the hydrologic model parameters and the appropriate statistical distribution of the residual errors. When applied to daily rainfall-runoff data from a humid basin we find that (1) residual errors are much better described by a heteroscedastic, first-order, auto-correlated error model with a Laplacian distribution function characterized by heavier tails than a Gaussian distribution; and (2) compared to a standard least-squares approach, proper representation of the statistical distribution of residual errors yields tighter predictive uncertainty bands and different parameter uncertainty estimates that are less sensitive to the particular time period used for inference. Application to daily rainfall-runoff data from a semiarid basin with more significant residual errors and systematic underprediction of peak flows shows that (1) multiplicative bias factors can be used to compensate for some of the largest errors and (2) a skewed error distribution yields improved estimates of predictive uncertainty in this semiarid basin with near-zero flows. We conclude that the presented methodology provides improved estimates of parameter and total prediction uncertainty and should be useful for handling complex residual errors in other hydrologic regression models as well.

Citation: 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, W10531, doi:10.1029/2009WR008933.

1. Introduction

[2] Assessment of parameter and predictive uncertainty of hydrologic models is an essential part of any hydrologic study. Uncertainty analysis forms the basis for model comparison and selection [Schoups *et al.*, 2008], allows identification of robust water management strategies that take account of prediction uncertainties [Ajami *et al.*, 2008], and provides an impetus for targeted data collection aimed at

improving hydrologic predictions and water management [Feyen and Gorelick, 2004]. Furthermore, accurate parameter uncertainty estimation is often required for regionalization and extrapolation of hydrologic parameters to ungauged basins [Vrugt *et al.*, 2002; Zhang *et al.*, 2008].

[3] Uncertainty analysis is commonly based on a regression model, whereby observations are represented by the sum of a deterministic component, i.e., the hydrologic model, and a random component describing remaining errors or residuals. These residual errors typically consist of a combination of input, model structural, output, and parameter errors. Model parameter inferences are then based on a likelihood function quantifying the probability that the observed data were generated by a particular parameter set [Box and Tiao, 1992]. The mapping from parameter space to likelihood space results in the identification of a range of plausible parameter sets given the data and allows estimation of parameter and predictive uncertainty.

¹Department of Water Management, Delft University of Technology, Delft, Netherlands.

²Center for Nonlinear Studies, Los Alamos National Laboratory, Los Alamos, New Mexico, USA.

³Institute for Biodiversity and Ecosystem Dynamics, University of Amsterdam, Amsterdam, Netherlands.

⁴Department of Civil and Environmental Engineering, University of California, Irvine, California, USA.

[4] In recent years, much debate has focused on the use of either a formal or informal approach for specifying the likelihood function [Mantovan and Todini, 2006; Beven et al., 2008; Vrugt et al., 2008b; Stedinger et al., 2008; McMillan and Clark, 2009]. In the formal approach, one starts from an assumed statistical model for the residual errors, i.e., the functional form of the joint probability density function (pdf) of the residual errors is specified a priori. This statistical model is then used to derive the appropriate form for the likelihood function [Box and Tiao, 1992]. For example, assuming that the errors are independent and identically distributed according to a normal distribution with zero mean and a constant variance σ^2 , results in the standard least squares (SLS) approach for parameter estimation. An advantage of the formal approach is that error model assumptions are stated explicitly, and their validity can be verified a posteriori [e.g., Stedinger et al., 2008].

[5] The formal approach has been criticized for relying too strongly on residual error assumptions that do not hold in many applications [Beven et al., 2008]. In many cases, residuals errors are correlated, nonstationary, and non-Gaussian [Kuczera, 1983]. A common form of nonstationarity is heteroscedasticity, which in many studies is observed as an increase in error variance with streamflow discharge [Sorooshian and Dracup, 1980]. Violation of SLS assumptions may introduce bias in estimated parameter values and affect parameter and predictive uncertainty [Thyer et al., 2009]. Alternatively, informal likelihood functions have been proposed as a pragmatic approach to uncertainty estimation in the presence of complex residual error structures. A well-known example is the generalized likelihood uncertainty estimation methodology of Beven and Freer [2001]. Here the likelihood function is specified a priori without explicitly linking it to an underlying error model. The modeler has flexibility in specifying the form of the likelihood function, which makes the informal approach attractive in situations where traditional error assumptions are violated. For example, Beven et al. [2008] have advocated the use of a flat likelihood function to avoid overconditioning of the statistical error model on a single calibration data set. However, since the informal approach makes no explicit reference to the underlying error model, its assumptions are implicit and cannot be checked a posteriori. Further discussion and comparison of formal and informal approaches are given by Mantovan and Todini [2006], Beven et al. [2008], Vrugt et al. [2008b], Stedinger et al. [2008], and McMillan and Clark [2009].

[6] The main goal of this paper is to extend the applicability of the formal approach by deriving and applying a formal likelihood function based on a general error model that allows for model bias and for correlation, nonstationarity, and nonnormality of model residuals. As such, we preserve advantages of the formal approach (theoretical basis and possibility of diagnostic checking of error model assumptions), while gaining flexibility and reducing the need for unrealistic assumptions about the residual errors.

[7] We build on previous formal approaches that have been used to relax some of the SLS error assumptions [Sorooshian and Dracup, 1980; Kuczera, 1983; Thieman et al., 2001; Bates and Campbell, 2001]. In particular, we follow Bates and Campbell [2001] and account for serial dependence of residual errors using a general autoregressive (AR) time series model. The main contribution of our work lies in the

treatment of heteroscedasticity and nonnormality, whereas previous approaches have used data and model response transformations, e.g., Box-Cox transformations [Box and Tiao, 1992], to induce homoscedasticity (constant variance) and remove skewness, we instead rely on an explicit statistical model to account for heteroscedasticity and nonnormality. Error standard deviation is modeled as a linear function of simulated streamflow, and nonnormality is accounted for with a parametric error distribution that allows for separate control of kurtosis and skewness in the model residuals. As discussed below, our approach is both more flexible and more intuitive compared to the transformation method. Focus is on correct simulation and representation of total residual errors, i.e., measurement, model input, and model structural errors are treated in a lumped manner, as opposed to recent attempts at separating the various error sources in hydrologic modeling [Kavetski et al., 2006a; Kuczera et al., 2006; Gotzinger and Bardossy, 2008; Vrugt et al., 2008b; Reichert and Mieleitner, 2009; Thyer et al., 2009; Renard et al., 2010]. The lumped approach provides less insight into the sources of error but yields practical estimates of parameter and total prediction uncertainty.

[8] The next section presents the statistical modeling approach, derives a new likelihood function for parameter inference, and outlines a method for predictive simulation. The methodology is applied in section 3 to estimate parameter and predictive uncertainty of a spatially lumped rainfall-runoff model, using synthetic and real data from a humid and a semiarid basin. Following Thyer et al. [2009], we assess effects of assumptions in the error model on parameter and prediction uncertainty. Section 4 discusses and summarizes our findings.

2. Formal Likelihood Uncertainty Estimation

2.1. Model Formulation

[9] Our analysis is based on an additive nonlinear regression model of the form,

$$\mathbf{Y} = \mathbf{E} + \mathbf{e}, \quad (1)$$

where \mathbf{Y} is a vector of n streamflow observations; \mathbf{E} is a corresponding vector of expected values; and \mathbf{e} is a vector of zero-mean random errors or residuals, including measurement, model input, and model structural errors.

[10] Expected values are modeled using a mass balance-based hydrologic model h , yielding simulated values \mathbf{Y}_h as a function of an observed input \mathbf{X} and a vector of model parameters $\boldsymbol{\eta}_h$. Since errors in observations \mathbf{Y} , model input \mathbf{X} , and model structure may lead to systematic deviations or bias in hydrologic model predictions, multiplicative bias factors are introduced,

$$E_t = Y_{h,t}(\mathbf{X}|\boldsymbol{\eta}_h)\mu_t, \quad (2)$$

where mean flow E_t , hydrologically simulated flow $Y_{h,t}$, and bias factor μ_t all vary as a function of time t . The bias factor may be treated as a stochastic variable, similar to the rainfall multiplier approach introduced by Kavetski et al. [2006a, 2006b] to account for model input errors, or the time-variable model parameter approach of Kuczera et al. [2006] to account for model structural errors. A simpler

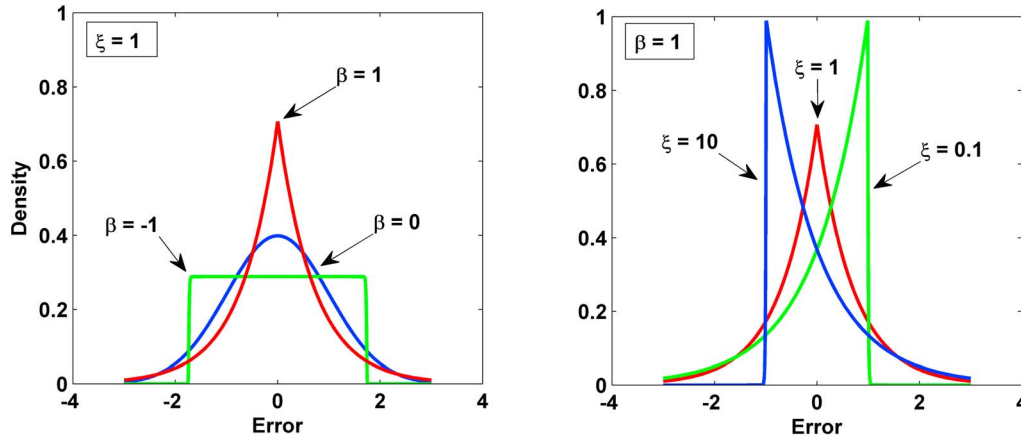


Figure 1. Densities of the skew exponential power (SEP) distribution with zero-mean and unit standard deviation for various values of the kurtosis (β) and skewness (ξ) parameters.

approach is attempted here by parameterizing bias factors as a function of simulated flow $Y_{h,t}$,

$$\mu_t = \exp(\mu_h Y_{h,t}) \quad (3)$$

where μ_h is a bias parameter to be inferred from the data. Equation (3) provides a simple way of amplifying the nonlinearity of the expected rainfall-runoff response. Other approaches, such as parameterizing μ_t as a function of (time-lagged) precipitation, could also be used.

[11] Residual errors \mathbf{e} are characterized by a joint probability density function (pdf) and a vector of parameters $\boldsymbol{\eta}_e$. A common approach is to assume that errors are independent and identically distributed (i.i.d.) according to a Gaussian density $N(0, \sigma^2)$. However, in hydrologic applications, residual errors usually violate these assumptions, as they exhibit temporal correlation, nonconstant variance (heteroscedasticity) and nonnormality. To deal with these nonideal situations, we propose the following model for the residual errors e_t ,

$$\Phi_p(B)e_t = \sigma_t a_t \text{ with } a_t \sim \text{SEP}(0, 1, \xi, \beta), \quad (4)$$

where $\Phi_p(B) = 1 - \sum_{i=1}^p \phi_i B^i$ is an autoregressive polynomial with p autoregressive parameters ϕ_i , B is the backshift operator ($B^i e_t = e_{t-i}$), σ_t is standard deviation at time t , and a_t is an i.i.d. random error with zero mean and unit standard deviation, described by a skew exponential power (SEP) density, to be defined below, with parameters ξ and β to account for nonnormality.

[12] The p th order autoregressive model, $\text{AR}(p)$, in equation (4) accounts for dependence and correlation between errors. The model could be further extended to more general autoregressive, moving average (ARMA) models, as was done by Kuczera [1983]. However, published literature and experience with simulating residual errors of rainfall-runoff models suggests that autoregressive (AR) models typically suffice. Furthermore, Bates and Campbell [2001] point out that approximation of ARMA models by higher-order AR models avoids problems with multiple local optima in ARMA models. Hence, for practical purposes we limit our presentation to AR models, but if necessary, the approach could be easily adapted to include ARMA models.

[13] Heteroscedasticity is explicitly accounted for by assuming that error standard deviation σ_t increases linearly with mean flow E_t ,

$$\sigma_t = \sigma_0 + \sigma_1 E_t, \quad (5)$$

where σ_0 and σ_1 are parameters to be inferred from the data. Error standard deviations typically increase as a function of flow, i.e., $\sigma_1 > 0$, for example, due to increasing uncertainty in the stage-discharge relationship at higher flows [Sorooshian and Dracup, 1980; Di Baldassarre and Montanari, 2009]. A similar heteroscedastic model was used by Thyer *et al.* [2009], although in their study, it was used as a streamflow measurement error model, with the σ_0 and σ_1 parameters estimated using rating curve data and fixed before hydrologic model calibration. Here equation (5) is used to represent heteroscedasticity of the total model residuals (including observation, input, and structural errors), and the σ_0 and σ_1 parameters are inferred simultaneously with the hydrologic model parameters.

[14] Finally, the $\text{SEP}(0, 1, \xi, \beta)$ density in equation (4) accounts for nonnormality of model residuals, with pdf expressed as (see Appendix A for details),

$$p(a_t | \xi, \beta) = \frac{2\sigma_\xi}{\xi + \xi^{-1}} \omega_\beta \exp\left\{-c_\beta |a_{\xi,t}|^{2/(1+\beta)}\right\}, \quad (6)$$

where $a_{\xi,t} = \xi^{-\text{sign}(\mu_\xi + \sigma_\xi a_t)} (\mu_\xi + \sigma_\xi a_t)$, and values for μ_ξ , σ_ξ , c_β , ω_β are computed as a function of skewness parameter ξ and kurtosis parameter β , as detailed in Appendix A. Kurtosis parameter β takes on values between -1 and $+1$ and determines the peakedness of the pdf, while skewness parameter ξ affects asymmetry ($\xi > 0$), as illustrated in Figure 1. The density is symmetric for $\xi = 1$ and positively (negatively) skewed for $\xi > 1$ ($\xi < 1$). In the case of a symmetric density, a uniform distribution results when $\beta = -1$, a Gaussian distribution when $\beta = 0$, and a Laplace or double-exponential distribution when $\beta = 1$. Hence, parameters ξ and β allow us to relax the assumption of Gaussian errors. In particular, values for $\beta > 0$ result in more peaked densities with heavier tails compared to a Gaussian pdf, which is useful for making parameter inference robust against outliers.

Table 1. Several Likelihood Functions Used in the Hydrologic Literature, Their Assumptions, and Relation to Equation (8) in This Paper^a

Likelihood Reference	Correlation	Heteroscedasticity	Noise Distribution	Implementation Using Equation (8)
Standard Least Squares (SLS)	Independent	Homoscedastic	Gaussian	$\phi_i = 0$ $\sigma_1 = 0$ $\xi = 1, \beta = 0$
<i>Sorooshian and Dracup</i> [1980, equation (26)]	Independent	Heteroscedastic	Gaussian	$\phi_i = 0$ $\xi = 1, \beta = 0$
<i>Sorooshian and Dracup</i> [1980, equation (20)]	AR(1)	Homoscedastic	Gaussian	$\phi_i = 0$ ($i > 1$) $\sigma_1 = 0$ $\xi = 1, \beta = 0$
<i>Kuczera</i> [1983]	ARMA(p, q)	Homoscedastic after Box-Cox transformation	Gaussian after Box-Cox transformation	$\beta = 0$
<i>Bates and Campbell</i> [2001]	AR(p)	Homoscedastic after Box-Cox transformation	Gaussian after Box-Cox transformation	$\beta = 0$
<i>Thiemann et al.</i> [2001]	Independent	Homoscedastic after log-transformation	Exponential power after log-transformation	$\phi_i = 0$

^aBias is never included, i.e., $\mu_1 = 0$ in equation (3).

2.2. Parameter Uncertainty

[15] The model formulated in the previous section contains a number of parameters $\eta = \{\eta_h, \eta_e\}$, including parameters of the hydrologic model η_h and the residual error model η_e . Parameter uncertainty after observing data \mathbf{Y} can be expressed by a posterior parameter pdf [Box and Tiao, 1992],

$$p(\eta|\mathbf{Y}) \propto \ell(\eta|\mathbf{Y})p(\eta), \quad (7)$$

where $p(\eta)$ is the prior pdf of the parameters, reflecting knowledge of the parameters before data \mathbf{Y} are available, and $\ell(\eta|\mathbf{Y})$ is the likelihood function. In Appendix B, an expression for the likelihood function is derived based on the error model defined in equations (4)–(6). The resulting expression for the log-likelihood function is,

$$L(\eta|\mathbf{Y}) = n \log \frac{2\sigma_\xi\omega_\beta}{\xi + \xi^{-1}} - \sum_{t=1}^n \log \sigma_t - c_\beta \sum_{t=1}^n |a_{\xi,t}|^{2/(1+\beta)}, \quad (8)$$

where errors $a_{\xi,t}$ and values for σ_t , σ_ξ , c_β , ω_β are computed as outlined in the previous section. The log-likelihood function in equation (8) relaxes common assumptions about residual errors and is therefore anticipated to be more applicable in hydrologic studies. This will be investigated in section 3, where the performance of the generalized log-likelihood function (“GL”) in equation (8) will be compared to the common standard least squares (“SLS”) approach. As discussed in Appendix B, equation (8) is a conditional log-likelihood function, valid for moderate to large sample sizes n typically available in rainfall-runoff applications. For example, *Sorooshian and Dracup* [1980] compared exact and conditional likelihood functions for the special case of an AR(1) error model with Gaussian innovations and found them to be in close agreement.

[16] Table 1 summarizes several previously used formal likelihood functions in rainfall-runoff modeling applications and shows how the log-likelihood function in equation (8) can be reduced to these by making specific assumptions about the residual errors. For example, for Gaussian errors ($\xi = 1$, $\beta = 0$) that are homoscedastic ($\sigma_1 = 0$) and independent ($\phi_i = 0$), equation (8) reduces to the SLS approach. Other previously used likelihood functions are also listed. *Sorooshian and Dracup* [1980] introduced a multivariate Gaussian error model and derived likelihood functions for cases of either heteroscedastic, i.e., nonconstant variance,

or first-order, auto-correlated errors. Their approach was generalized by *Kuczera* [1983], who considered a general ARMA model for the errors, in combination with a Box-Cox transformation of observed and simulated streamflow to account for heteroscedastic and skewed errors. A similar approach was adopted by *Bates and Campbell* [2001] but using AR rather than ARMA models to account for correlation. Finally, *Thiemann et al.* [2001] neglected error correlation but proposed the exponential power distribution [Box and Tiao, 1992] to model kurtosis, while using a log-transformation to account for heteroscedasticity and skewness. It should be clear from Table 1 that the log-likelihood function in equation (8) generalizes previous approaches and introduces additional flexibility to simultaneously account for correlated, heteroscedastic, and non-Gaussian residuals.

[17] With the specification of a prior parameter pdf, equation (8) can be used to calculate posterior parameter uncertainty using equation (7), e.g., by repeated Monte Carlo sampling of parameter sets from the prior parameter space. This is efficiently done using Monte Carlo Markov chain (MCMC) simulation [Bates and Campbell, 2001; Vrugt et al., 2003; Engeland et al., 2005; Vrugt et al., 2006; Kuczera and Parent, 1998; Smith and Marshall, 2008]. The MCMC algorithm used in this paper is called DREAM-ZS (DiffeRential Evolution Adaptive Metropolis algorithm) and was developed by Vrugt et al. [2009]. DREAM-ZS is based on the original DREAM algorithm [Vrugt et al., 2009] but uses sampling from an archive of past states to generate candidate points in each individual chain. Sampling from the past circumvents the need for a large number of parallel chains, designed to accelerate convergence for high-dimensional problems. Experience with DREAM-ZS suggests that only three parallel chains are needed to appropriately explore the posterior pdf, reducing time required for burn-in. Moreover, DREAM-ZS does not require outlier detection and removal, maintaining detailed balance at every single step in each of the parallel chains. Finally, DREAM-ZS contains a snooker update to generate jumps beyond parallel direction updates [ter Braak and Vrugt, 2008] and increase diversity of candidate points.

2.3. Predictive Uncertainty

[18] In addition to parameter uncertainty, we are also interested in predictive uncertainty of the model. Predic-

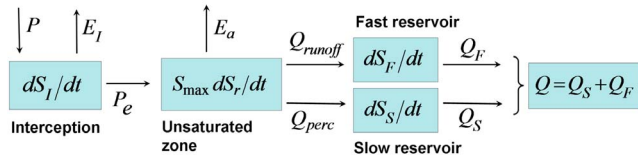


Figure 2. Model structure and hydrologic process parameterizations. Boxes represent water balance units with indicated water fluxes simulated as follows: evaporation $E_I = \min(E_p, S_{I,0})$, effective precipitation $P_e = P - (I_{\max} - S_{I,0})$, evaporation $E_a = (E_p - E_I)f(S_r|\alpha_E)$, runoff $Q_{\text{runoff}} = P_e f(S_r|\alpha_F)$, percolation $Q_{\text{perc}} = Q_{\text{smax}} f(S_r|\alpha_S)$, fast streamflow $Q_F = K_F S_F$, and base flow $Q_S = K_S S_S$. The functional relationship from fluxes from and storage in the unsaturated reservoir is parameterized as $f(S_r|\alpha) = \frac{1-e^{-\alpha S_r}}{1-e^{-\alpha}}$, where α is a process-specific parameter. All parameters are listed and explained in Table 2.

tive percentiles \mathbf{Y}_α , corresponding to a specified exceedance probability $1 - \alpha$, are obtained from the following relation,

$$P(\mathbf{Y} \leq \mathbf{Y}_\alpha | \mathbf{X}) = P([\mathbf{Y}_h(\mathbf{X}|\boldsymbol{\eta}_h)\boldsymbol{\mu} + \mathbf{e}(\boldsymbol{\eta}_e)]_{1..J} \leq \mathbf{Y}_\alpha | \mathbf{X}) = \alpha, \quad (9)$$

where J parameter sets $\boldsymbol{\eta} = \{\boldsymbol{\eta}_h, \boldsymbol{\eta}_e\}$ are randomly sampled from the posterior parameter pdf obtained with the MCMC algorithm and are used to generate J time series for model output \mathbf{Y}_h and errors \mathbf{e} . These J time series correspond to J model predictions at each time step, from which we can compute prediction percentiles \mathbf{Y}_α for each time step (e.g., taking $\alpha = 0.975$ and $\alpha = 0.025$ yields time series of the 97.5% and 2.5% prediction percentiles, which together constitute the 95% prediction uncertainty bands). For $\mathbf{e} = 0$, we obtain prediction percentiles \mathbf{Y}_α due to uncertainty in the hydrologic model parameter values. Estimation of total predictive uncertainty requires computing errors \mathbf{e} , which involves generating independent samples from a SEP distribution. This is done using the following algorithm based on the studies by Johnson [1979] and Würtz and Chalabi [2009]:

[19] 1. Generate a sample g_t from the γ distribution with shape parameter $(1 + \beta)/2$ and scale parameter 1.

[20] 2. Generate a random sign s_t (+1 or -1) with equal probability.

[21] 3. Compute $EP_t = s_t |g_t|^{(1+\beta)/2} \frac{\Gamma^{1/2}[(1+\beta)/2]}{\Gamma^{1/2}[3(1+\beta)/2]}$, which is a sample from the exponential power distribution, $EP(0, 1, \beta)$.

[22] 4. Generate a random sign w_t (+1 or -1) with probabilities $1 - \frac{\xi}{\xi + \xi^{-1}}$ and $\frac{\xi}{\xi + \xi^{-1}}$.

[23] 5. Compute $SEP_t = -w_t |EP_t| \xi^w$, which is a sample from the skew exponential power distribution, $SEP(\mu_\xi, \sigma_\xi, \xi, \beta)$, with μ_ξ and σ_ξ given by (A5) and (A6).

[24] 6. Normalize: $a_t = (SEP_t - \mu_\xi)/\sigma_\xi$

[25] The algorithm is repeated n times ($t = 1 \dots n$) to obtain n independent samples a_t from the skew exponential power density $SEP(0, 1, \xi, \beta)$. Corresponding heteroscedastic and correlated errors e_t are obtained using equation (4). A MATLAB function that implements this simulation algorithm, as well as the generalized log-likelihood function

of equation (8), is available upon request from the first author.

3. Application to Rainfall-Runoff Modeling

[26] We use daily data of mean areal precipitation, mean areal potential evaporation, and streamflow from two US basins, namely, the French Broad River basin at Asheville, NC, and the Guadalupe River basin at Spring Branch, TX. These are, respectively, the wettest and driest of the 12 MOPEX basins described in the study by Duan *et al.* [2006]. Daily records of precipitation and potential evaporation are input into a lumped conceptual rainfall-runoff model [Schoups *et al.*, 2010] based on the FLEX modeling system [Fenicia *et al.*, 2007] to simulate daily streamflow. The model considers interception, throughfall, evaporation, runoff generation, percolation, and surface and subsurface routing of water to the basin outlet. Runoff generation is assumed to be dominated by saturated overland flow and is simulated as a function of basin water storage without an explicit dependence on rainfall intensity. This assumption is typically valid for temperate climates but may be violated in the semiarid Guadalupe River basin. Snow accumulation and snowmelt are also not accounted for, although these processes occur in the French Broad River basin. The severity of these model structural errors will be evaluated in the case studies below. Model structure and hydrologic process parameterizations are shown in Figure 2. Note that our approach is similar to commonly used conceptual rainfall-runoff model structures: The nonlinear soil-moisture accounting store combined with parallel reservoirs for slow and fast hydrologic response was advocated by Jakeman and Hornberger [1993] and has been used in many other studies as well. The model includes a total of seven hydrologic parameters that need to be estimated, as summarized in Table 2. We assume the residuals to be described by a first-order auto-regressive error model, equation (4), with correlation coefficient ϕ_1 . This AR(1) model will be extended to higher-order AR models if dictated by the data.

[27] Table 2 summarizes all hydrologic and error model parameters and their prior uncertainty ranges. We assume uniform priors for all parameters, which is deemed acceptable here due to the large number of data points used ($n =$

Table 2. Prior Uncertainty Ranges of Hydrologic and Error Model Parameters^a

Parameter	Symbol	Minimum	Maximum	Units
Maximum interception	I_{\max}	0	10	mm
Soil water storage capacity	S_{\max}	10	1000	mm
Maximum percolation rate	Q_{smax}	0	100	mm/d
Evaporation parameter	α_E	0	100	-
Runoff parameter	α_F	-10	10	-
Time constant, fast reservoir	K_F	0	10	days
Time constant, slow reservoir	K_S	0	150	days
Heteroscedasticity intercept	σ_0	0	1	mm/d
Heteroscedasticity slope	σ_1	0	1	-
Autocorrelation coefficient	ϕ_1	0	1	-
Kurtosis parameter	β	-1	1	-
Skewness parameter	ξ	0.1	10	-
Bias parameter	μ_h	0	100	(mm/d) ⁻¹

^aPercolation parameter α_S is set to zero (i.e., percolation is assumed to be a linear function of soil storage, see Figure 2).

Table 3. Summary of True and Inferred Hydrologic and Error Model Parameters Using Synthetically Generated Daily Streamflow^a

Parameter	Cases ^b									
	1		2		3		4		5	
	True Value	GL	True Value	GL	True Value	GL	True Value	GL	True Value	GL
I_{\max}	2	2.3	0.8	0.6	2	1.7	0.3	0.6	2	2.7
		5.0	5.8	7.8		7.8	7.2	7.2		1.1
S_{\max}	100	99	93	105	100	104	83	88	100	105
		104	122	185		185	180	180		90
Q_{\max}	7	7.2	6.6	7.3	7	7.9	6.1	6.0	7	7.3
		7.9	8.6	14.7		14.7	13.2	13.2		158
α_E	50	66	37	52	50	63	30	27	50	64
		97	95	98		98	98	98		28
α_F	-0.5	-0.45	-0.70	-0.35	-0.5	-0.46	-1.1	-0.95	-0.5	-0.42
		-0.20	0.21	0.21		1.4	1.4	1.31		97
K_F	3	3.0	2.9	3.0	3	2.9	2.7	2.8	3	3.0
		3.1	3.2	3.2		3.3	3.3	3.3		0.75
K_S	60	61	56	59	60	60	47	48	60	60
		67	66	78		78	72	72		3.2
σ_0	0.5	0.49	0.46	0.3	0.3	0.29	0.26	0.23	0.3	0.30
		0.51	0.33	0.33		0.32	0.32	0.32		0.27
σ_1	0	0.0	0.00	0.10	0.1	0.10	0.09	0.11	0.1	0.10
		0.01	0.12	0.12		0.12	0.13	0.13		0.09
ϕ_1	0	0.01	-0.04	-0.01	0.7	0.69	0.67	0.70	0.7	0.69
		0.06	0.04	0.04		0.73	0.73	0.73		0.11
β	0	-0.06	-0.14	-0.03	0	0.03	-0.08	0.91	1	0.98
		0.06	0.08	0.08		0.14	0.14	0.77		0.71
ξ	1	1.0	0.96	1.0	1	1.0	0.96	0.99	2	2.0
		1.11	1.08	1.11		1.11	1.06	1.06		1.89
										2.12

^aSynthetic data are generated using the algorithm described in section 2.3, and inference is based on the corresponding generalized likelihood function “GL” in equation (8). Maximum likelihood values (under GL headings) and, next to it, 95% intervals are given. In each case, reported values are averages from 10 repeated 5 year synthetic experiments.

^bCases: 1 = homoscedastic, uncorrelated, Gaussian errors; 2 = heteroscedastic, uncorrelated, Gaussian errors; 3 = heteroscedastic, correlated, Gaussian errors; 4 = heteroscedastic, correlated, Laplacian errors; 5 = heteroscedastic, correlated, skewed Laplacian errors.

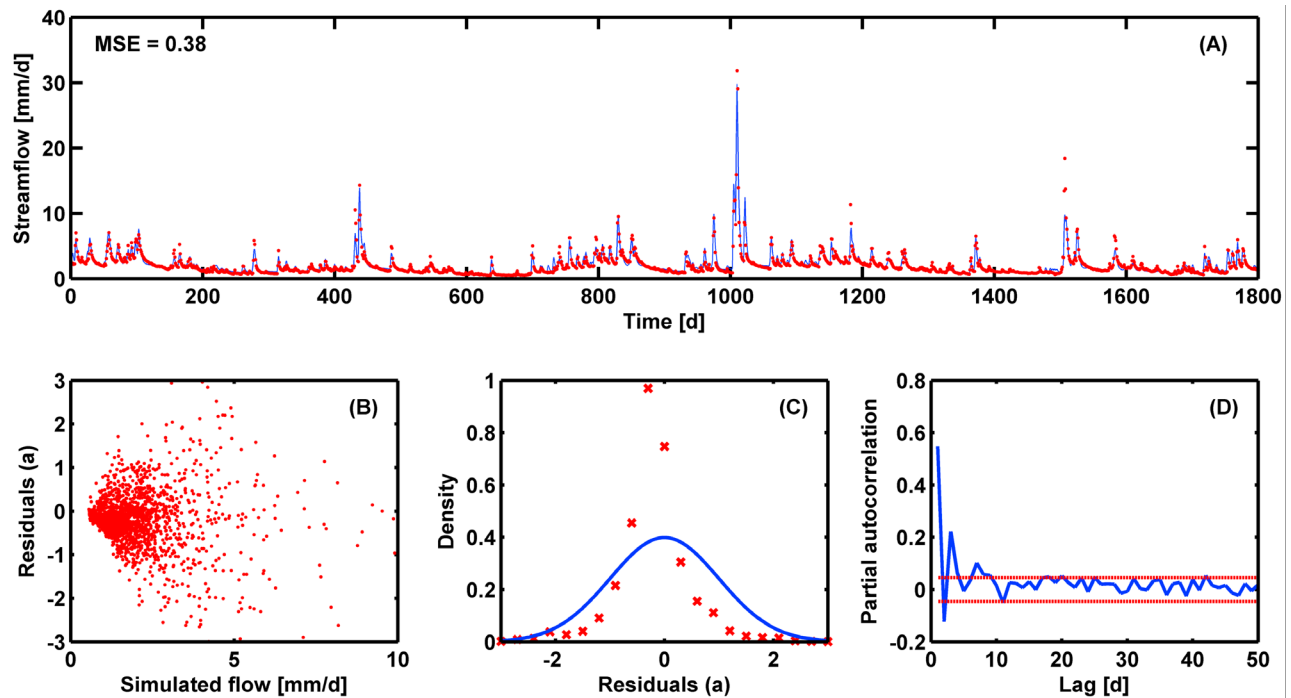


Figure 3. SLS calibration for the French Broad River basin: (a) time series of maximum-likelihood streamflow predictions (solid line) and observations (dots), (b) residuals a_t as a function of simulated flow, (c) assumed (solid line) and actual (crosses) pdf of residuals a_t , and (d) partial autocorrelation coefficients of residuals a_t with 95% significance levels.

1825). With smaller sample size, more attention should be paid to selection of priors and their effect on the posterior. In addition to the hydrologic parameters, a maximum of five error model parameters need to be estimated from measured discharge data. However, as will be discussed in the case studies below, the number of error model parameters varies depending on each case as dictated by the data. Discussion of the results in the following sections will focus on parameter and predictive uncertainty estimated using the generalized formal likelihood function derived in equation (8). Hereafter, these results are referred to as “GL” (generalized likelihood). To benchmark our results, we will include comparison against a traditional standard least squares approach (“SLS”) assuming independent, homoscedastic, Gaussian error residuals.

3.1. Synthetic Data: Verification of Estimation and Simulation Algorithm

[28] Before presenting our findings using measured streamflow data, the methodology is first tested using artificial discharge data. Time series of daily streamflow data were generated with the seven-parameter hydrologic model using observed daily precipitation and potential evaporation from the French Broad River basin. This synthetic discharge record was subsequently corrupted with artificial errors mathematically defined in equation (4) using the algorithm described in section 2.3. Hydrologic and error model parameters were inferred from the corrupted, synthetic streamflow data using the log-likelihood of equation (8) and the MCMC algorithm in section 2.2.

[29] Table 3 presents results for different cases, including heteroscedastic, correlated, and non-Gaussian errors. Our method is able to infer the underlying error structure and hydrologic model parameter values for all cases presented, with deviations between true and maximum-likelihood parameter values that are small compared to the 95% parameter uncertainty intervals obtained in each case. From the results in Table 3, it is therefore concluded that the MCMC algorithm is able to simultaneously infer the correct hydrologic and error model parameters. These results confirm that both the inference method in section 2.2, based on the error model of equation (4) and corresponding log-likelihood function of equation (8), and the simulation method in section 2.3, are correctly implemented. In the following sections we investigate whether the error model defined in equation (4) provides an accurate representation of residual errors encountered in rainfall-runoff applications.

3.2. First Case Study: French Broad Basin

3.2.1. Evaluation of Error Models

[30] Five years of observed daily forcing (precipitation, potential evaporation) and observed daily streamflow were used to identify hydrologic and error model parameters. Figure 3 shows results for the fitted model using SLS with seven hydrologic parameters and one error parameter (constant error variance). Note that the model mimics the data quite well, reproducing most minor and major flow events. Nevertheless, closer inspection of the model residuals in Figure 3 reveals that the SLS assumptions do not hold: (1) the error variance increases as a function of simulated flow, suggesting heteroscedasticity; (2) the error histogram is

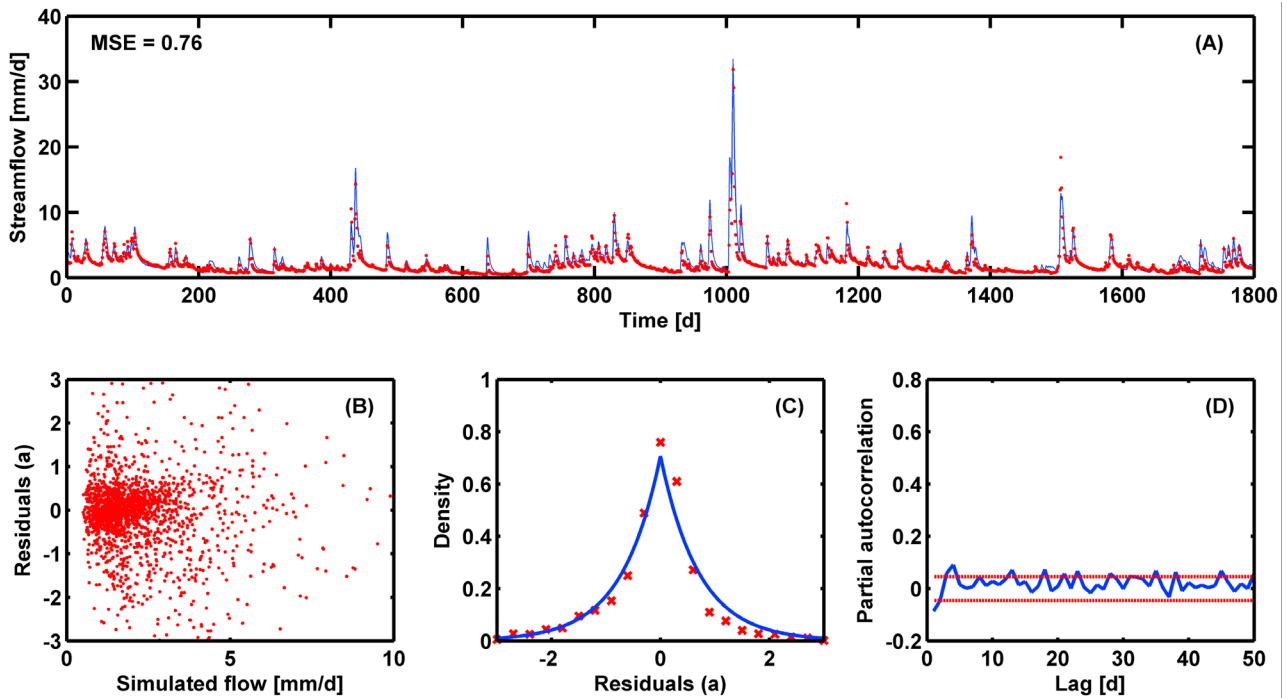


Figure 4. GL calibration for the French Broad River basin: (a) time series of maximum-likelihood streamflow predictions (solid line) and observations (dots), (b) residuals a_t as a function of simulated flow, (c) assumed (solid line) and actual (crosses) pdf of residuals a_t , and (d) partial autocorrelation coefficients of residuals a_t with 95% significance levels.

more peaked than the assumed Gaussian pdf, and (3) errors are significantly correlated at a lag of one, violating the independence assumption of SLS. Such violations have been reported in other rainfall-runoff studies as well [e.g., Kuczera, 1983; Feyen *et al.*, 2007; Thyer *et al.*, 2009]. These errors may be due to a combination of measurement, model input, and model structural errors. For example, neglecting snow accumulation and melt processes may be one reason for correlation in residual errors, especially during winter and spring, but eventually throughout the year as parameter values partially compensate for structural errors during calibration.

[31] To account for deviations from SLS assumptions, a second calibration was performed in which heteroscedasticity, nonnormality, and error correlations were explicitly accounted for using the generalized likelihood function (“GL”) of equation (8). The hydrologic parameters are now augmented with two variance parameters (σ_1 and σ_0), one shape parameter (β), and a first-order autocorrelation coefficient (ϕ_1).

This results in a model that again fits the data quite well, albeit with a mean-square-error (MSE) that is twice as large as under SLS (Figure 4). One explanation is that GL puts less emphasis on fitting peak flows because of heteroscedastic errors. Moreover, since SLS minimizes the MSE, it will always yield smaller MSE values than other methods. Comparing models based on MSE assumes that errors satisfy SLS assumptions. As is clear from Figure 3, these assumptions do not hold here. A better way of evaluating the appropriateness of the GL error model relative to the SLS error model is to compare their respective maximum log-likelihood values (or posterior densities). For the 5 year calibration data set, we find that the GL error model has a much larger log-likelihood than the SLS error model, i.e., 540 versus -1690 . This is perhaps not that surprising as the GL error model contains three additional parameters for data fitting. However, corresponding maximum log-likelihood values during an independent 20 year evaluation period (1970 for GL versus -6140 for SLS), as well as values

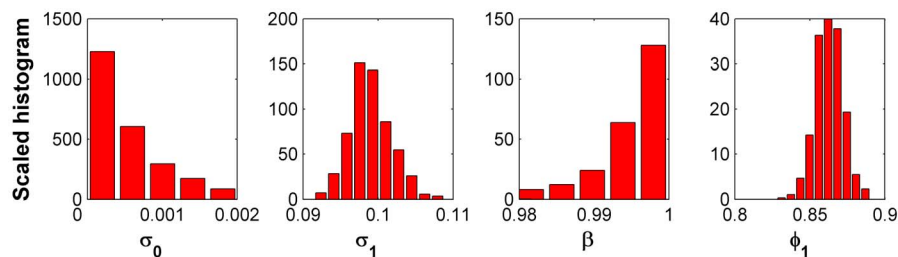


Figure 5. Posterior histograms of error model parameters using GL on the French Broad River basin.

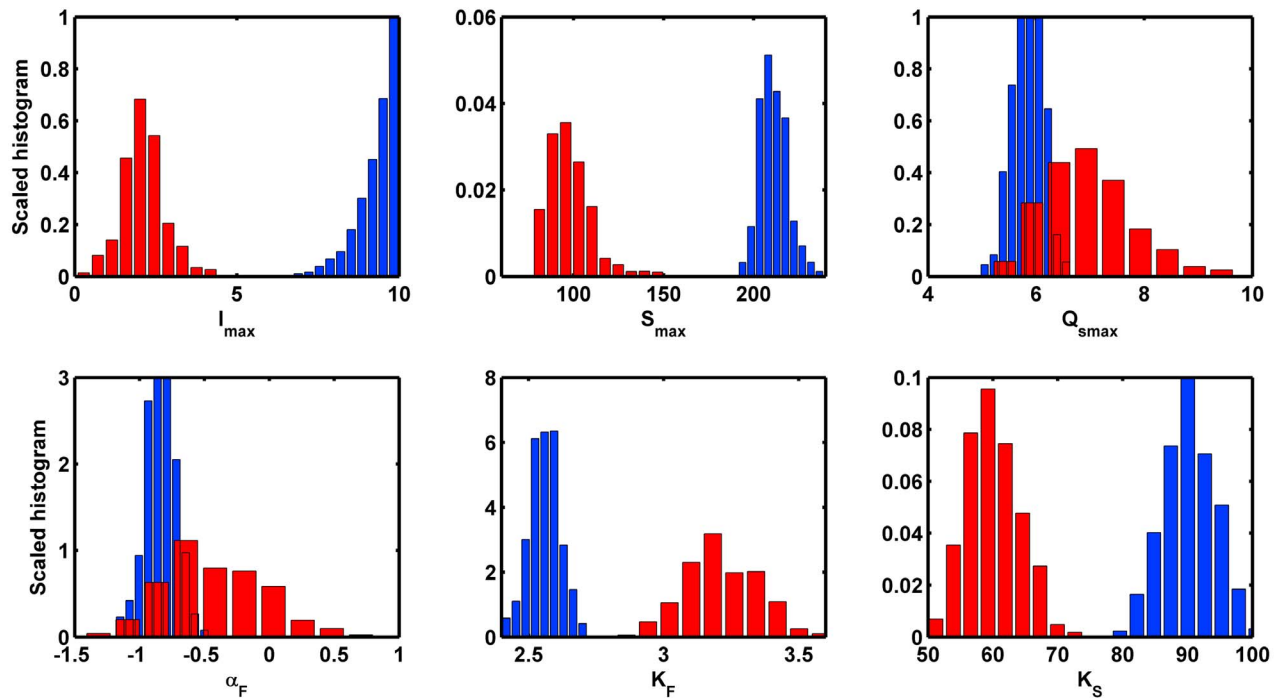


Figure 6. Posterior histograms of hydrologic model parameters using SLS (black) and GL (gray) on the French Broad River basin.

for the Bayesian Information Criterion (to be minimized) [Marshall *et al.*, 2005] (−997 for GL versus 3440 for SLS) both confirm superiority of the GL error model.

[32] In addition, parameter inference using GL is consistent with the a priori assumptions, as shown by the diagnostic plots in Figure 4. Heteroscedasticity and correlation between errors have been removed, and the inferred error distribution closely matches the empirical distribution of the model residuals. The corresponding posterior histograms of the error parameters are shown in Figure 5, indicating that all four parameters are well identified. Note that kurtosis parameter β approaches a value of 1, which means that the errors follow a Laplace or double-exponential distribution. As shown in Figure 4, the Laplace distribution is more peaked than the Gaussian distribution, and also has heavier tails, which makes it robust against outliers. Our results show strong evidence for heavy-tailed errors.

3.2.2. Parameter Uncertainty

[33] So far, the results indicate that residual errors are better represented by an error model that explicitly accounts for heteroscedasticity, correlation, and nonnormality, compared to the simplifying assumptions inherent in SLS. The next question we wish to address is what effect violation of SLS assumptions has on estimates of hydrologic parameter and predictive uncertainty.

[34] Figure 6 presents posterior parameter histograms for all seven hydrologic parameters based on SLS and GL inference strategies. Two important findings can be deduced from these plots. First, parameter inference based on invalid error assumptions (SLS) yields parameter estimates that deviate significantly from those obtained with a more appropriate error model (GL). This is most notably the case for parameter S_{\max} , the soil water storage capacity, which is a key parameter in the model for separating effective rainfall

into runoff, evaporation, and percolation. Parameter estimates under SLS may even lead to nonphysical values, as is the case for parameter I_{\max} , which represents vegetation interception capacity. Using SLS, this parameter sits at its upper bound of 10 mm, whereas under GL more realistic values around 2 mm are obtained [Breuer *et al.*, 2003].

[35] The second finding in Figure 6 is that parameter uncertainty, as measured by the width or spread of the posterior parameter histograms, is underestimated by SLS compared to GL. This is, for example, clearly the case for parameters Q_{\max} and α_F . Greater uncertainty under GL results from a combination of factors (1) accounting for error correlation reduces the information content of the data, (2) the heteroscedastic error model assigns greater uncertainty to peak flows, and (3) the inferred Laplace distribution has heavier tails than the Gaussian distribution. Underestimation of parameter uncertainty using SLS was also found by Thyer *et al.* [2009].

[36] Further evaluation and comparison of parameter estimates by SLS and GL was done by using 5 year calibration periods with different starting points. Each calibration period was shifted by 1 year, such that subsequent periods have 4 years of data in common. Ten different calibration periods were considered, and for each data set, parameters were inferred using both SLS and GL. Experience suggests that 5 years of daily streamflow data contains enough information about the parameters of conceptual rainfall-runoff models, and therefore, no significant variations in parameter estimates between calibration data sets are anticipated. Resulting estimates of parameter uncertainty for each of the seven hydrologic parameters are shown in Figure 7. Parameter robustness was quantified by computing the variance of the posterior parameter means, divided by total variance of the MCMC posterior samples over all

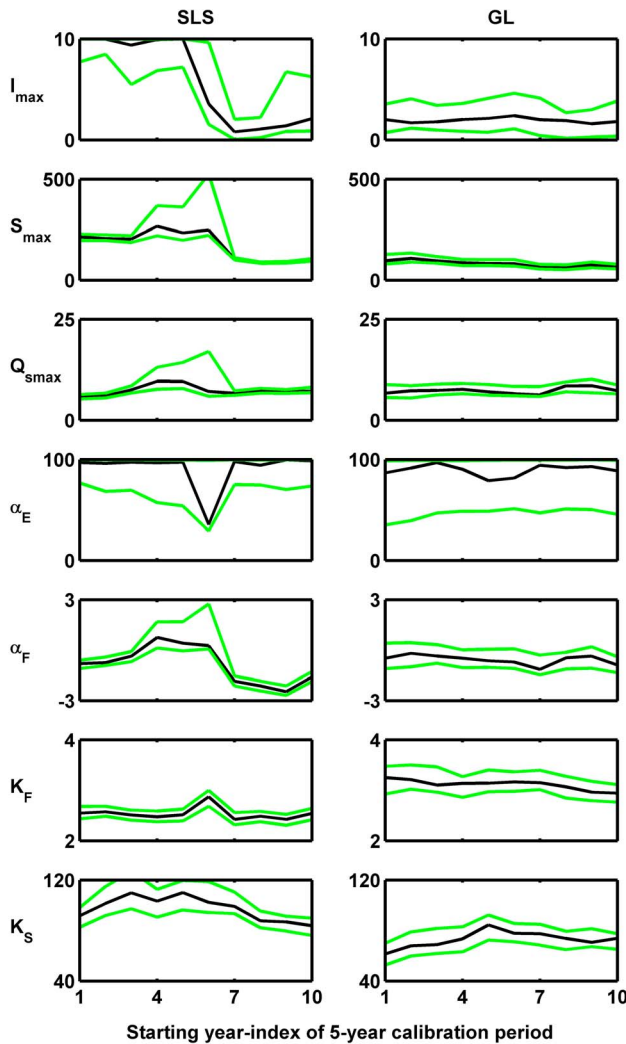


Figure 7. Maximum-likelihood parameter values (black) and 95% uncertainty bands (gray) using SLS and GL on the French Broad River basin.

calibration periods. For each parameter, this ratio was smaller for GL (range, 0.04–0.84) than for SLS (range, 0.22–0.99). Together with Figure 7, these results indicate that GL parameter estimates are consistent between calibration data sets, whereas SLS parameter estimates are more sensitive to variations in the calibration data. The latter confirms results reported by *Thyer et al.* [2009]. Robustness of the GL inference results can be attributed to three factors: (1) by accounting for heteroscedasticity less weight is given to high flows, making the inference less sensitive to large flow events in different calibration data sets; (2) long tails of the Laplace distribution allow for a larger number of large errors, which again induces robustness against outliers and random variations in large flow events; (3) accounting for autocorrelation in the residual errors filters out measurement, model input, and model structural errors, resulting in less biased and more consistent parameter estimates [*Vrugt et al.*, 2005].

3.2.3. Predictive Uncertainty

[37] In addition to parameter uncertainty, it is expected that assumptions about the residual errors have direct

consequences for estimates of prediction uncertainty. Figure 8 shows time series and quantile–quantile (QQ) plots for flow predictions using SLS assumptions. It is obvious from these results that SLS yields inadequate streamflow predictions. The problem with ignoring heteroscedasticity is clearly visible in the time series plots: use of a constant (average) error variance results in an overestimation of prediction uncertainty for low flows and an underestimation for high flows. As the streamflow record is dominated by low flows, interspersed with high-flow events, the dominating feature of the calibration and validation QQ plots is their S-shaped curvature, indicative of overestimation of prediction uncertainty [*Thyer et al.*, 2009]. Such overestimation may even lead to prediction uncertainty bands that become negative, as evident in the time series plot of Figure 8.

[38] By contrast, predictive uncertainty computed without making simplifying assumptions of SLS results in uncertainty bands that, although not perfect, are more realistic (Figure 9). Now, prediction bands at low flows are narrower and more closely bracket observed flows. In addition, larger uncertainty for high flows is accounted for through the use of an error variance that increases as a function of simulated streamflow. The QQ plots in Figure 9 confirm that predictions under GL better represent observed flows, as QQ plots approach the 1:1 line, especially for the validation data set, whereas some systematic over-prediction is apparent for the calibration data set.

3.3. Second Case Study: Guadalupe River Basin

[39] In the second case study, we study a much drier basin. Streamflow is characterized by extended periods of near-zero flows, alternated with strongly nonlinear peak flows following rainfall events. It is expected that the simple rainfall-runoff model used here may not account for all relevant processes in this basin, such as infiltration-excess runoff generation and flash-flooding [*Clark et al.*, 2008]. Therefore, residual errors are anticipated to be more complex than in the first case study, thereby representing a greater challenge to a formal uncertainty analysis. We follow a similar approach as in the first case study, in that we gradually increase complexity of the error model, starting from the simplest model, namely SLS, until satisfactory results are obtained by posterior checks and analysis of predictive uncertainty bands.

[40] Model predictions and corresponding diagnostic plots using SLS are shown in Figure 10. As before, SLS assumptions of constant variance and normally distributed residuals are clearly violated. In addition, residuals are slightly correlated at small lags but less so than in the first case study. Results in Figure 10 show that the SLS error model yields unrealistic prediction uncertainty bands. The situation is especially problematic in this case with near-zero flows, where SLS results in negative lower prediction bounds for most of the flow record.

[41] Next, the analysis was repeated with additional error parameters for heteroscedasticity, autocorrelation, and non-normality (kurtosis), following the same approach as in the first case study. Results (not shown) revealed that, although error assumptions are fulfilled, prediction uncertainty bands are large and meaningless. This is caused by the large inferred value for the autocorrelation coefficient, which is

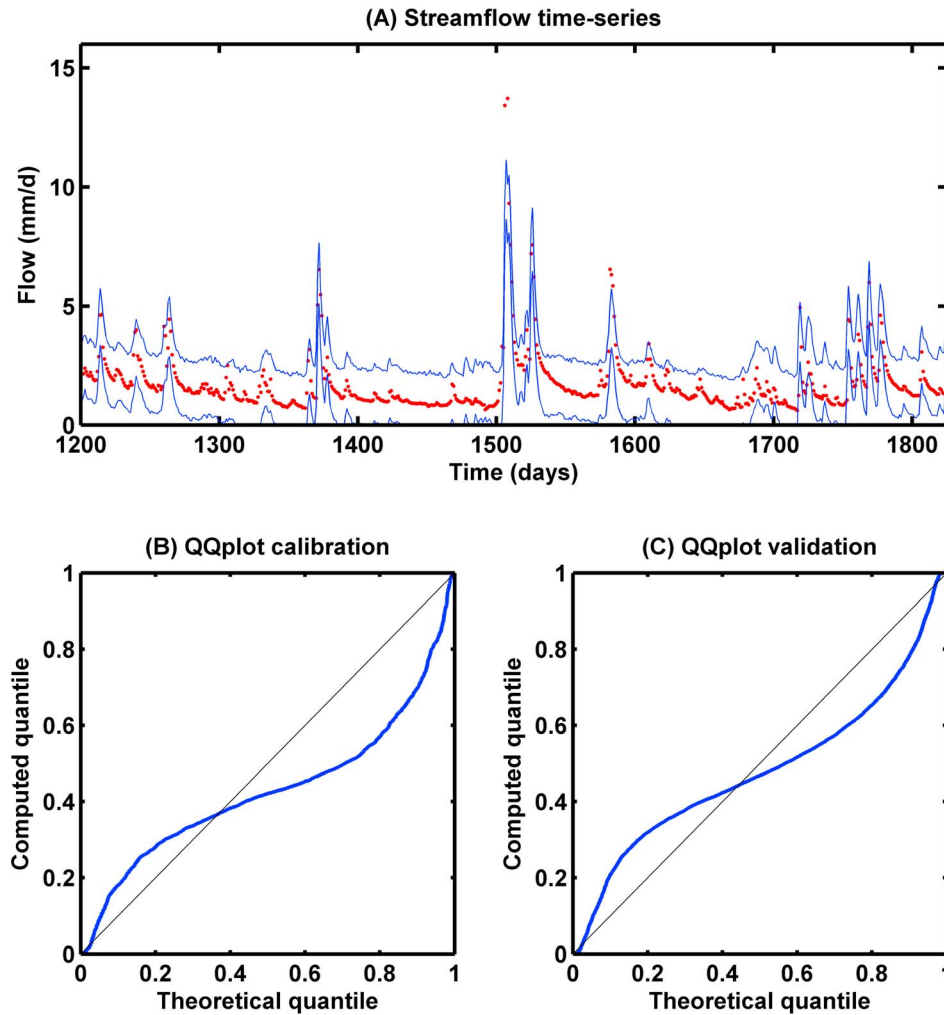


Figure 8. Predictive uncertainty using SLS on the French Broad River basin: (a) time series of observations (dots) and 95% total prediction uncertainty bands (solid lines); (b and c) QQ plots for calibration and validation periods.

close to 1 in this case. A value for ϕ_1 near 1 amounts to a random walk, resulting in large random errors. A possible explanation for this result is that the residual errors are too severe to be accounted for by a simple AR(1) model. Therefore, higher-order AR models were also considered, specifically AR(2) and AR(4). However, this resulted in similarly unrealistic error bounds as the ones for an AR(1) model.

[42] As an alternative to accounting for error correlations, improvements in prediction of expected streamflow values were sought by introducing time-variable model bias factors μ_t , as in equation (2), to compensate for large residual errors. Therefore, bias parameter μ_h in equation (3) was inferred together with the hydrologic model parameters and error model parameters for heteroscedasticity (σ_0 , σ_1) and kurtosis (β). Magnitude of the bias factors in equation (2) was limited to a value of 10. While improving the results, prediction uncertainty bands and diagnostic plots in this case (not shown) indicated the need for two additional corrections. First, significant correlation remained between residual errors at lag 1, suggesting the need to include an AR model. Second, lower-prediction uncertainty bands became

negative, because of the general low flows in this basin, combined with the use of a symmetric distribution for the residual errors. Therefore, a final run included the following error model parameters: bias parameter μ_h , parameters for heteroscedasticity (σ_0 , σ_1), kurtosis (β), and skew (ξ), and a first-order autocorrelation coefficient ϕ_1 . To avoid problems described above, the correlation coefficient was not automatically inferred but was instead fixed to a value of 0.4, based on inspection of diagnostic plots.

[43] Figure 11 shows resulting diagnostic plots and flow predictions using the GL approach. Compared to the SLS results in Figure 10, prediction uncertainty bands are much improved, providing a better description of observed values and remaining positive throughout the flow record. In addition, the inferred error distribution more closely matches the assumed distribution. The inferred value for skewness parameter ξ in this case is about 1.3 (Figure 12), indicating a positively skewed error distribution, as can also be noted in Figure 11c. Diagnostic plots in Figure 11 further indicate that the GL error model removed residual heteroscedasticity and first-order correlation, with some minor correlation

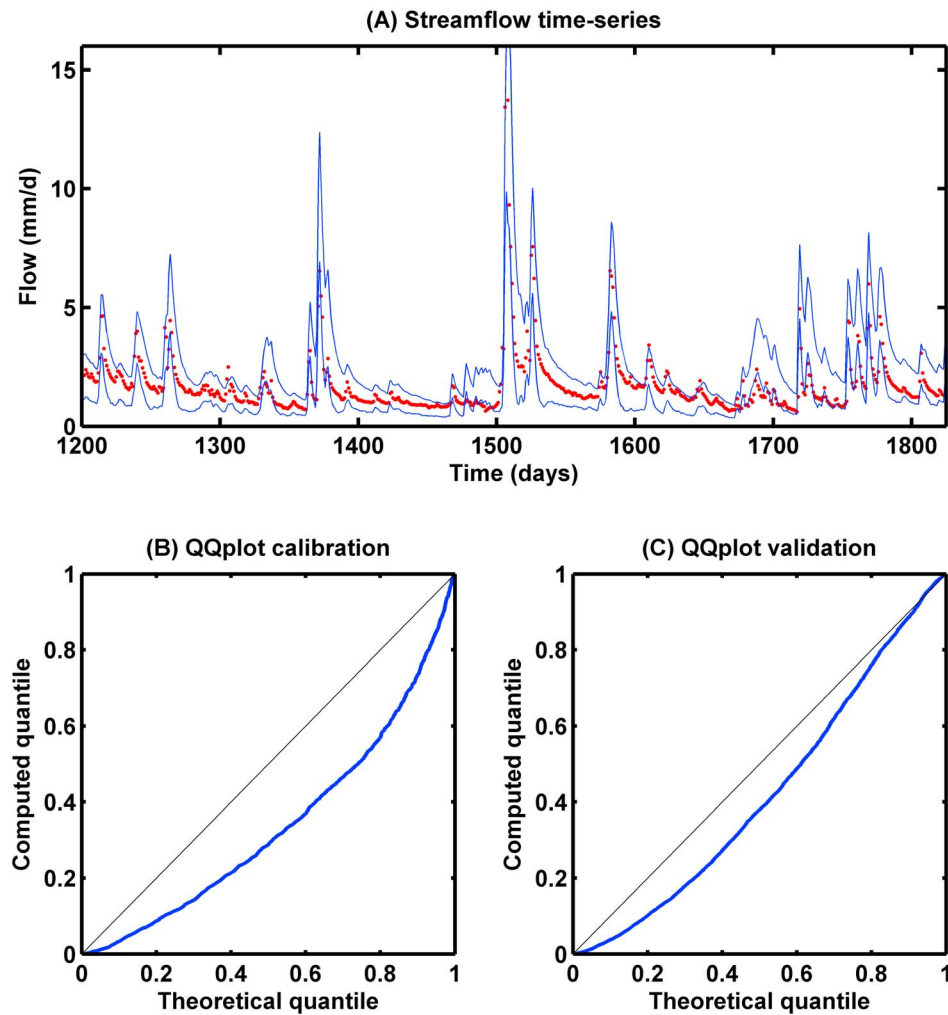


Figure 9. Predictive uncertainty using GL on the French Broad River basin: (a) time series of observations (dots), and 95% total prediction uncertainty bands (solid lines); (b and c) QQ plots for calibration and validation periods.

remaining at greater lags. Figure 12 summarizes posterior histograms for all hydrologic model and error model parameters using the GL approach. Most parameters are fairly well identified, except for two parameters (S_{\max} and K_S) that reach their upper bound, indicating that even larger values for these parameters are preferred. A physical interpretation could be that these large inferred values for S_{\max} and K_S are indicative of relatively dry and deep vadose zones in semiarid basins, resulting in large storage effects and slow response times for subsurface flow. Finally, Table 4 shows somewhat elevated posterior correlations around 0.9 between parameters $Q_{s\max}$ and both α_E and α_F , suggesting that one of these could be fixed in this semiarid basin.

4. Discussion and Conclusions

[44] Results for both case studies confirm previous work [e.g., Kuczera, 1983; Vrugt *et al.*, 2008a; Thyer *et al.*, 2009] that accurate estimation of parameter and prediction uncertainty depends on an adequate statistical representation of the residual errors. In both case studies, traditional SLS assumptions of residual independence, homoscedasticity,

and normality were clearly violated. Our approach relaxes these assumptions and relies on a statistical residual error model, with corresponding likelihood function, which explicitly accounts for correlation, heteroscedasticity, and non-normality of model residual errors.

[45] Application of the approach to rainfall-runoff modeling in the wet basin (French Broad River) resulted in encouraging results. The error model appears to be sufficiently flexible to describe the errors in this case. A first-order autoregressive model, combined with a linear heteroscedasticity model and a Laplace distribution for the residual errors, worked quite well, as revealed by posterior error diagnostic plots. In addition, this error model, and corresponding likelihood function, generated parameter estimates that are robust to the specific data record used for inference, and yielded improved estimates of predictive uncertainty compared to SLS. Nonrobustness of SLS was also exposed by the work of Thyer *et al.* [2009]. Parameter robustness using GL is attributed to the use of a more accurate residual error model.

[46] For the semiarid basin, residual errors are more complex and multiplicative bias factors were introduced to compensate for some of the larger residual errors. These

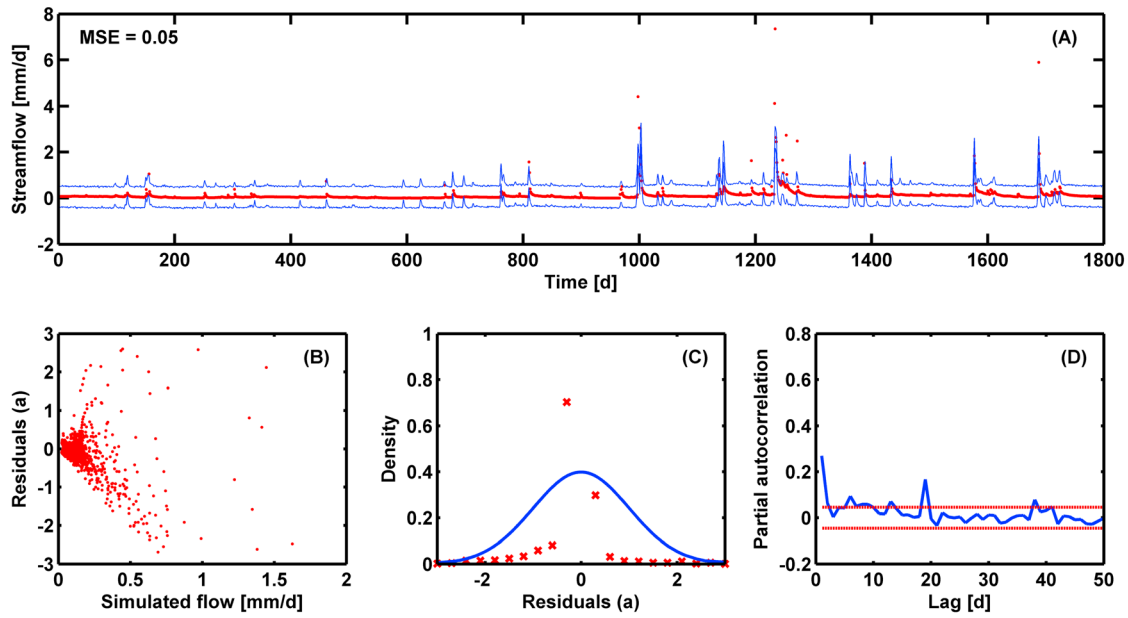


Figure 10. Predictive uncertainty and diagnostic plots using SLS on the Guadalupe river basin: (a) time series of observations (dots) and 95% total prediction uncertainty bands (solid lines), (b) residuals a_t as a function of simulated flow, (c) assumed (solid line) and actual (crosses) pdf of residuals a_t , and (d) partial autocorrelation coefficients of residuals a_t with 95% significance levels.

errors are caused by a combination of measurement errors, rainfall volume/timing errors, and model structural errors. For example, it could be hypothesized that runoff in this basin is dominated by the infiltration-excess mechanism (Hortonian overland flow). This would imply that the runoff

coefficient not only depends on soil wetness, as accounted for in our hydrologic model, but also on rainfall intensity. As rainfall intensities are expected to vary quite significantly throughout a single day, daily average rainfall rates used in this study cannot capture these dynamics. Hence, a switch to

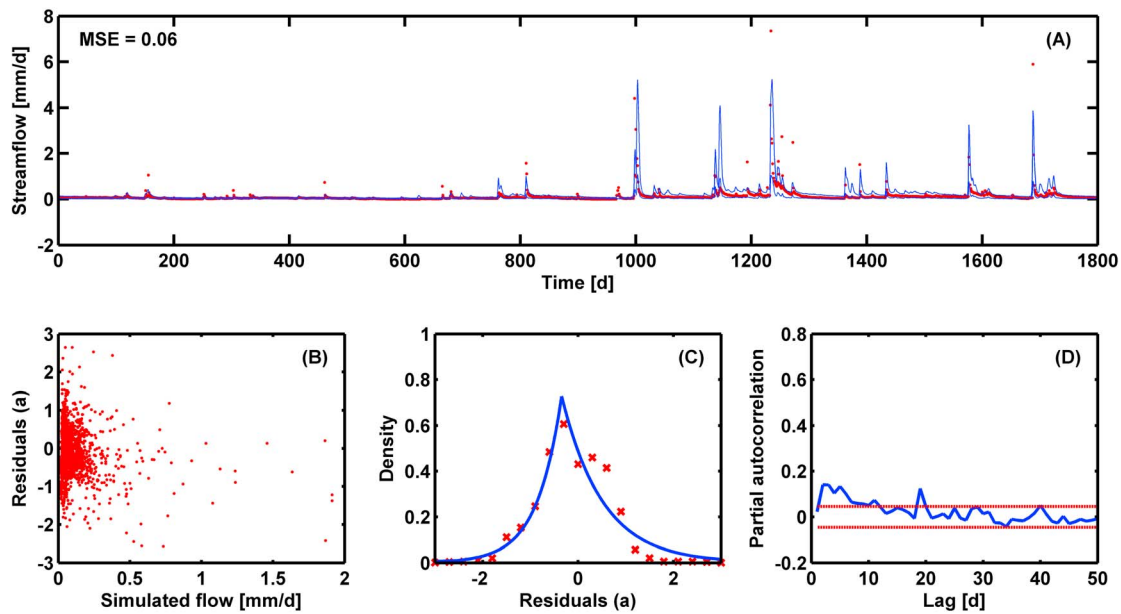


Figure 11. Predictive uncertainty and diagnostic plots using GL on the Guadalupe river basin: (a) time series of observations (dots) and 95% total prediction uncertainty bands (solid lines), (b) residuals a_t as a function of simulated flow, (c) assumed (solid line) and actual (crosses) pdf of residuals a_t , and (d) partial autocorrelation coefficients of residuals a_t with 95% significance levels.

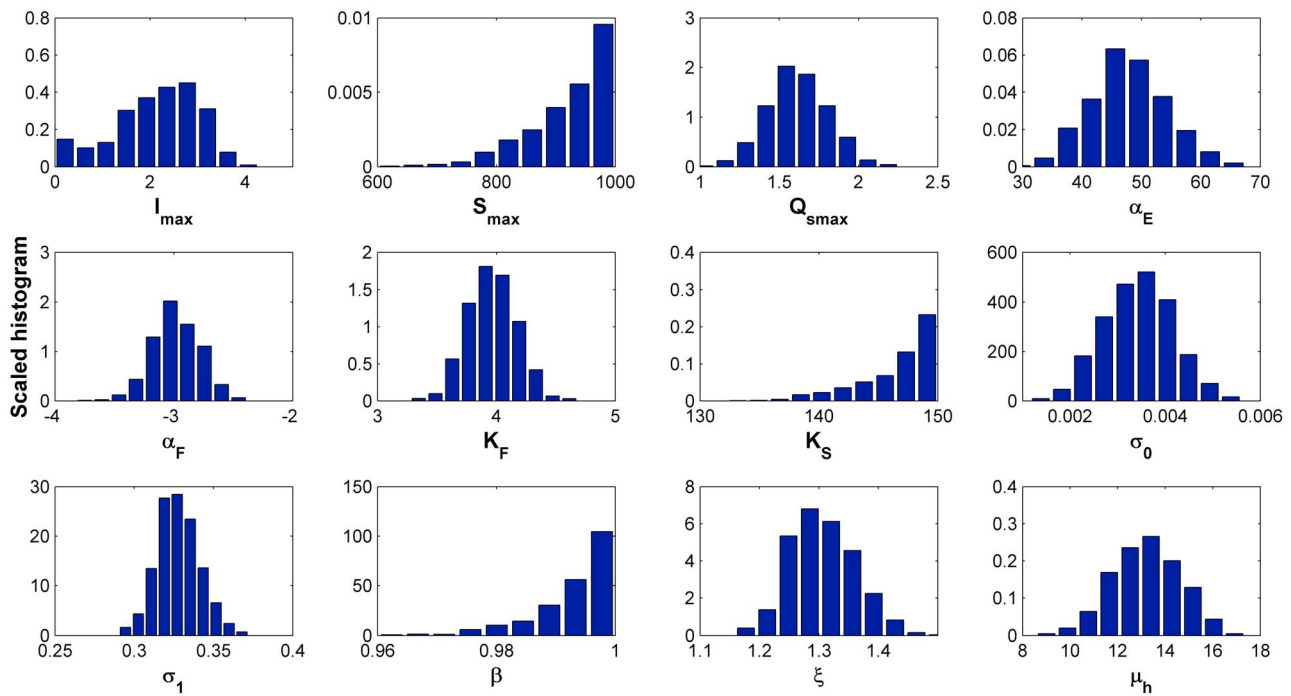


Figure 12. Posterior histograms of hydrologic and error model parameters using GL on the Guadalupe river basin.

hourly rainfall data and addition of a Hortonian overland flow mechanism to the model may be required to adequately simulate streamflow dynamics in the semiarid basin. Here multiplicative bias factors were used to account for some of these deficiencies. Although not perfect, this approach yields a simple and parsimonious method (one extra parameter) to account for model bias. Results for the semiarid basin also highlight the importance of accounting for skewness. Use of a skewed distribution yields more realistic prediction uncertainty bands, in the sense that negative lower prediction bounds, as obtained with a symmetric distribution, are avoided.

[47] We further note that both basins provided strong evidence in favor of a Laplace distribution (symmetric or skewed) for the residuals. The Laplace distribution may prove to be a good choice in many other cases as well, as heavy tailed residuals have been reported in several other studies [e.g., *Bates and Campbell, 2001; Schaeffli et al.,*

2007; *Yang et al., 2007*]. Inference with an additive Laplacian error model amounts to median or L_1 regression, which has long been advocated for its robustness against outliers [*Koenker and Bassett, 1978*].

[48] The modeling and inference approach presented here deviates from previous methods that have mostly relied on Box-Cox transformations to handle heteroscedastic and non-Gaussian residuals in hydrologic modeling (Table 1). Although we do not provide a rigorous comparison, we can still point out two advantages of our method. First, Box-Cox transformations are useful for removing heteroscedasticity and skewness; however, they typically do not account for heavy-tailed residuals, as shown by the case studies of *Bates and Campbell [2001]* and *Yang et al. [2007]*. Heavy-tailed residuals appear to be quite common in hydrologic modeling and are accounted for using our approach. Second, explicit modeling of the statistical error distribution as done here is more intuitive than the transformation method, in that

Table 4. Posterior Parameter Correlation Coefficient Matrix of the Hydrologic and Error Model Parameters Using GL on the Guadalupe River Basin^a

	I_{\max}	S_{\max}	Q_{smax}	α_E	α_F	K_F	K_S	σ_0	σ_1	β	ξ	μ_h
I_{\max}	1.00	-0.05	0.18	-0.10	0.38	0.01	-0.12	0.21	-0.21	-0.05	0.25	-0.27
S_{\max}		1.00	0.57	0.56	0.47	0.11	-0.03	-0.01	0.06	-0.05	0.16	0.16
Q_{smax}			1.00	<u>0.91</u>	<u>0.87</u>	-0.16	0.09	-0.20	0.08	-0.03	-0.09	-0.57
α_E				1.00	<u>0.69</u>	-0.20	0.25	-0.31	0.16	0.01	-0.22	-0.32
α_F					1.00	0.11	0.04	-0.16	0.01	-0.09	-0.18	<u>-0.71</u>
K_F						1.00	-0.03	0.10	-0.06	0.02	0.03	0.05
K_S							1.00	-0.40	0.26	0.16	-0.45	0.05
σ_0								1.00	<u>-0.79</u>	-0.01	0.57	0.19
σ_1									1.00	-0.05	-0.38	-0.01
β										1.00	-0.03	0.06
ξ											1.00	0.33
μ_h												1.00

^aCorrelations greater than 0.6 are underlined.

error model parameters have a direct relation to observable error statistics. This is illustrated by diagnostic plots in Figure 3 which suggest reasonable values for the error model parameters. In contrast, Box-Cox transformation parameter values cannot be easily deduced from such diagnostic plots.

[49] Our approach also deviates from other studies that have focused on separating the various error contributions that make up the model residuals, including measurement, model input, and model structural errors [e.g., *Kuczera et al.*, 2006; *Reichert and Mieleitner*, 2009; *Renard et al.*, 2010]. Such studies are important for testing hypotheses about possible causes for deviations between model predictions and data. However, complete disentangling of the various error sources, especially separation of rainfall and model structural errors, can be quite challenging, as pointed out by a recent study of *Renard et al.* [2010]. Our approach is less ambitious and focuses instead on a correct statistical description of the data and the total model residuals, without separating out various error sources. This results in a pragmatic method for estimating parameter and prediction uncertainty of hydrologic models without the need for explicit assumptions about various error contributions.

[50] In conclusion, the methodology proposed in this paper provides increased flexibility for describing residual errors in rainfall-runoff applications using a formal statistical approach. This flexibility translates into improved estimates of parameter and total prediction uncertainty, compared to traditional approaches that rely on unrealistic assumptions of independent, homoscedastic, and normally distributed model residuals. Although our application focused on streamflow simulation, the presented methodology is entirely general and may be useful for dealing with complex error residuals in other hydrologic regression models as well.

Appendix A

[51] The appendix shows how the standardized Skew Exponential Power (SEP) pdf in equation (6) can be obtained from the exponential power (EP) pdf of *Box and Tiao* [1992] using the method of *Fernandez and Steel* [1998]. Our development follows *Würtz and Chalabi* [2009], who implemented the SEP as part of the fGarch package in R. The standardized EP pdf with zero mean and unit standard deviation can be expressed as,

$$f(a|\beta) = \omega_\beta \exp\left\{-c_\beta |a|^{2/(1+\beta)}\right\}, \quad (A1)$$

where β is a kurtosis parameter ($-1 < \beta \leq 1$), and ω_β and c_β are given by [*Box and Tiao*, 1992; p. 157],

$$\omega_\beta = \frac{\Gamma^{1/2}[3(1+\beta)/2]}{(1+\beta)\Gamma^{3/2}[(1+\beta)/2]}, \quad (A2)$$

$$c_\beta = \left(\frac{\Gamma[3(1+\beta)/2]}{\Gamma[(1+\beta)/2]}\right)^{1/(1+\beta)}, \quad (A3)$$

where $\Gamma[x]$ is the γ function evaluated at x . *Fernandez and Steel* [1998] developed a general method for introducing skew in a symmetric density,

$$p(\varepsilon|\xi) = \frac{2}{\xi + \xi^{-1}} f\left(\xi^{-\text{sign}(\varepsilon)} \varepsilon\right), \quad (A4)$$

where ξ is a skewness parameter ($\xi > 0$) and f denotes a symmetric density, in our case, the standardized EP pdf in (A1). Mean and standard deviation of ε can be derived from equation (5) in *Fernandez and Steel* [1998],

$$\mu_\xi = M_1(\xi - \xi^{-1}) \quad (A5)$$

$$\sigma_\xi = \sqrt{(M_2 - M_1^2)(\xi^2 + \xi^{-2}) + 2M_1^2 - M_2}, \quad (A6)$$

where M_r is the r th absolute moment of the symmetric density f ,

$$M_r = 2 \int_0^\infty s^r f(s) ds. \quad (A7)$$

[52] For the standardized EP pdf in (A1), one obtains the following expressions for M_1 and M_2 ,

$$M_1 = \frac{\Gamma[1+\beta]}{\Gamma^{1/2}[3(1+\beta)/2]\Gamma^{1/2}[(1+\beta)/2]} \quad (A8)$$

$$M_2 = 1, \quad (A9)$$

which allows us to compute mean μ_ξ and standard deviation σ_ξ in (A5) and (A6). To obtain a standardized SEP density, with zero mean and unit standard deviation, the pdf in (A4) is scaled by standard deviation σ_ξ and ε is replaced by $\mu_\xi + \sigma_\xi a$,

$$p(a|\xi) = \frac{2\sigma_\xi}{\xi + \xi^{-1}} f\left(\xi^{-\text{sign}(\mu_\xi + \sigma_\xi a)} (\mu_\xi + \sigma_\xi a)\right). \quad (A10)$$

Substitution of (A1) into (A10) yields the standardized SEP density of equation (6).

Appendix B

[53] We derive the likelihood function in equation (8) from the assumed error model in equations (4)–(6). Likelihood $\ell(\boldsymbol{\eta}|\mathbf{Y})$ of the hydrologic and error model parameters $\boldsymbol{\eta}$ is defined as the joint pdf of the observations \mathbf{Y} for given parameters $\boldsymbol{\eta}$,

$$\ell(\boldsymbol{\eta}|\mathbf{Y}) \equiv p(\mathbf{Y}|\boldsymbol{\eta}) = p(\mathbf{e}|\boldsymbol{\eta}), \quad (B1)$$

where $\mathbf{Y} = (Y_1 \dots Y_n)'$ and $\mathbf{e} = (e_1 \dots e_n)'$. Splitting \mathbf{e} into two subsets, $\mathbf{e}_{1:p} = (e_1 \dots e_p)'$ and $\mathbf{e}_{p+1:n} = (e_{p+1} \dots e_n)'$, the joint pdf can be written as the product of marginal and conditional densities,

$$p(\mathbf{e}|\boldsymbol{\eta}) = p(\mathbf{e}_{1:p}|\boldsymbol{\eta}) p(\mathbf{e}_{p+1:n}|\mathbf{e}_{1:p}, \boldsymbol{\eta}). \quad (B2)$$

[54] The conditional pdf in (B2) can be recursively expanded to yield,

$$p(\mathbf{e}|\boldsymbol{\eta}) = p(\mathbf{e}_{1:p}|\boldsymbol{\eta}) \prod_{t=p+1}^n p(e_t|\mathbf{e}_{1:t-1}, \boldsymbol{\eta}), \quad (B3)$$

which, using the error model in equation (4), results in the following expression,

$$p(\mathbf{e}|\boldsymbol{\eta}) = p(\mathbf{e}_{1:p}|\boldsymbol{\eta}) \prod_{t=p+1}^n \sigma_t^{-1} p(a_t|\boldsymbol{\eta}) \quad (B4)$$

with density $p(a_t|\boldsymbol{\eta})$ given by equation (6).

[55] For Gaussian innovations a_t , the marginal pdf $p(\mathbf{e}_{1:p}|\boldsymbol{\eta})$ is also Gaussian, resulting in closed-form expressions for the exact likelihood function (see Newbold [1974] for ARMA models, and equation (17) in the study by Sorooshian and Dracup [1980] for AR(1) models with Gaussian innovations). However, with non-Gaussian innovations, the marginal pdf is typically quite complicated (see, e.g., Damsleth and El-Shaarawi [1989] for ARMA models with Laplace innovations). A common approach, valid for moderate to large sample sizes n typically encountered in rainfall-runoff modeling, approximates the marginal pdf $p(\mathbf{e}_{1:p}|\boldsymbol{\eta})$ by conditioning on unobserved residuals e_t ($t < 1$): $p(\mathbf{e}_{1:p}|\boldsymbol{\eta}) \simeq p(\mathbf{e}_{1:p}|\mathbf{e}_{1-p:0}, \boldsymbol{\eta}) = \prod_{t=1}^p \sigma_t^{-1} p(a_t|\boldsymbol{\eta})$. Inserting this approximation into (B4) yields a conditional likelihood function,

$$\ell(\boldsymbol{\eta}|\mathbf{Y}) \simeq \prod_{t=1}^n \sigma_t^{-1} p(a_t|\boldsymbol{\eta}), \quad (\text{B5})$$

which, with equation (6), can be written as,

$$\ell(\boldsymbol{\eta}|\mathbf{Y}) \simeq \prod_{t=1}^n \sigma_t^{-1} \frac{2\sigma_\xi}{\xi + \xi^{-1}} \omega_\beta \exp\left\{-c_\beta |a_{\xi,t}|^{2/(1+\beta)}\right\} \quad (\text{B6})$$

[56] Taking the log-transform of this expression yields the (conditional) log-likelihood function in equation (8). In (B6) and equation (8), unobserved residuals e_t ($t < 1$) are assumed to be equal to zero.

[57] **Acknowledgments.** We would like to thank three anonymous reviewers for their constructive criticism and Steven Weijs, Rolf Hut, Ronald van Noijen, and Nick van de Giesen for fruitful discussions. The second author is supported by a J. Robert Oppenheimer Fellowship from the LANL postdoctoral program.

References

- Ajami, N. K., G. M. Hornberger, and D. L. Sunding (2008), Sustainable water resource management under hydrological uncertainty, *Water Resour. Res.*, 44, W11406, doi:10.1029/2007WR006736.
- Bates, B. C., and E. P. Campbell (2001), A Markov chain Monte Carlo scheme for parameter estimation and inference in conceptual rainfall-runoff modeling, *Water Resour. Res.*, 37(4), 937–947.
- Beven, K., and J. Freer (2001), Equifinality, data assimilation, and uncertainty estimation in mechanistic modeling of complex environmental systems using the GLUE methodology, *J. Hydrol.*, 249, 11–29.
- Beven, K., P. J. Smith, and J. E. Freer (2008), So just why would a modeler choose to be incoherent, *J. Hydrol.*, 354, 15–32.
- Box, G. E. P., and G. C. Tiao (1992), *Bayesian Inference in Statistical Analysis*, 588 pp., John Wiley, New York.
- Breuer, L., K. Eckhardt, and H.-G. Frede (2003), Plant parameter values for models in temperate climates, *Ecol. Model.*, 169(2–3), 237–293.
- 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 Resour. Res.*, 44, W00B02, doi:10.1029/2007WR006735.
- Damsleth, E., and A. H. El-Shaarawi (1989), ARMA models with double-exponentially distributed noise, *J. R. Stat. Soc., Ser. B*, 51(1), 61–69.
- Di Baldassarre, G., and A. Montanari (2009), Uncertainty in river discharge observations: A quantitative analysis, *Hydrol. Earth Syst. Sci.*, 13, 913–921.
- Duan, Q., et al. (2006), Model Parameter Estimation Experiment (MOPEX): An overview of science strategy and major results from the second and third workshops, *J. Hydrol.*, 320, 3–17.
- Engeland, K., C. Xu, and L. Gottschalk (2005), Assessing uncertainties in a conceptual water balance model using Bayesian methodology, *Hydrol. Sci. J.*, 50, 45–63.
- Fenicia, F., H. H. G. Savenije, P. Matgen, and L. Pfister (2007), A comparison of alternative multiobjective calibration strategies for hydrological modeling, *Water Resour. Res.*, 43, W03434, doi:10.1029/2006WR005098.
- Fernandez, C., and M. J. F. Steel (1998), On Bayesian modeling of fat tails and skewness, *J. Am. Stat. Assoc.*, 93, 359–371.
- Feyen, L., and S. M. Gorelick (2004), Framework to evaluate the worth of hydraulic conductivity data for optimal groundwater resources management in ecologically sensitive areas, *Water Resour. Res.*, 41, W03019, doi:10.1029/2003WR002901.
- Feyen, L., J. A. Vrugt, B. O. Nuallain, J. van der Knijff, and A. de Roo (2007), Parameter optimization and uncertainty assessment for large-scale stream flow forecasting, *J. Hydrol.*, 332, 276–289.
- Gotzinger, J., and A. Bardossy (2008), Generic error model for calibration and uncertainty estimation of hydrological models, *Water Resour. Res.*, 44, W00B07, doi:10.1029/2007WR006691.
- Jakeman, A. J., and G. M. Hornberger (1993), How much complexity is warranted in a rainfall-runoff model?, *Water Resour. Res.*, 29(8), 2637–2649, doi:10.1029/93WR00877.
- Johnson, M. (1979), Computer generation of the exponential power distribution, *J. Stat. Comput. Sim.*, 9(3), 239–240.
- Kavetski, D., G. Kuczera, and S. W. Franks (2006a), Bayesian analysis of input uncertainty in hydrological modeling: 1. Theory, *Water Resour. Res.*, 42, W03407, doi:10.1029/2005WR004368.
- Kavetski, D., G. Kuczera, and S. W. Franks (2006b), Bayesian analysis of input uncertainty in hydrological modeling: 2. Application, *Water Resour. Res.*, 42, W03408, doi:10.1029/2005WR004376.
- Koenker, R., and G. Bassett Jr. (1978), Regression quantiles, *Econometrica*, 46(1), 33–50.
- Kuczera, G. (1983), Improved parameter inference in catchment models. 1. Evaluating parameter uncertainty, *Water Resour. Res.*, 19(5), 1151–1162, doi:10.1029/WR019i005p01151.
- Kuczera, G., and E. Parent (1998), Monte Carlo assessment of parameter uncertainty in conceptual catchment models: The Metropolis algorithm, *J. Hydrol.*, 211(1–4), 69–85.
- Kuczera, G., D. Kavetski, S. Franks, and M. Thyer (2006), Towards a Bayesian total error analysis of conceptual rainfall-runoff models: Characterizing model error using storm-dependent parameters, *J. Hydrol.*, 331, 161–177, doi:10.1016/j.jhydrol.2006.05.010.
- Mantovan, P., and E. Todini (2006), Hydrological forecasting uncertainty assessment: Incoherence of the GLUE methodology, *J. Hydrol.*, 330, 368–381.
- Marshall, L., D. Nott, and A. Sharma (2005), Hydrological model selection: A Bayesian alternative, *Water Resour. Res.*, 41, W10422, doi:10.1029/2004WR003719.
- McMillan, H., and M. Clark (2009), Rainfall-runoff model calibration using informal likelihood measures within a Markov chain Monte Carlo sampling scheme, *Water Resour. Res.*, 45, W04418, doi:10.1029/2008WR007288.
- Newbold, P. (1974), The exact likelihood function for a mixed autoregressive-moving average process, *Biometrika*, 61(3), 423–426.
- Reichert, P., and J. Mieleitner (2009), Analyzing input and structural uncertainty of nonlinear dynamic models with stochastic, time-dependent parameters, *Water Resour. Res.*, 45, W10402, doi:10.1029/2009WR007814.
- Renard, B., D. Kavetski, G. Kuczera, M. A. Thyer, and S. W. Franks (2010), Understanding predictive uncertainty in hydrologic modeling: The challenge of identifying input and structural errors, *Water Resour. Res.*, 46, W05521, doi:10.1029/2009WR008328.
- Schaefli, B., D. B. Talamba, and A. Musy (2007), Quantifying hydrological modeling errors through a mixture of normal distributions, *J. Hydrol.*, 332, 303–315.
- Schoups, G., J. A. Vrugt, F. Fenicia, and N. C. van de Giesen (2010), Inaccurate numerical implementation of conceptual hydrologic models corrupts accuracy and efficiency of MCMC simulation, *Water Resour. Res.*, doi:10.1029/2009WR008648, in press.
- Schoups, G., N. C. van de Giesen, and H. H. G. Savenije (2008), Model complexity control for hydrologic prediction, *Water Resour. Res.*, 44, W00B03, doi:10.1029/2008WR006836.
- Smith, T. J., and L. A. Marshall (2008), Bayesian methods in hydrologic modeling: A study of recent advancements in Markov chain Monte Carlo techniques, *Water Resour. Res.*, 44, W00B05, doi:10.1029/2007WR006705.
- Sorooshian, S., and J. A. Dracup (1980), Stochastic parameter estimation procedures for hydrologic rainfall-runoff models—correlated and heteroscedastic error cases, *Water Resour. Res.*, 16(2), 430–442, doi:10.1029/WR016i002p00430.
- Stedinger, J. R., R. M. Vogel, S. U. Lee, and R. Batchelder (2008), Appraisal of the generalized likelihood uncertainty estimation (GLUE) method, *Water Resour. Res.*, 44, W00B06, doi:10.1029/2008WR006822.

- ter Braak, C. J. F., and J. A. Vrugt (2008), Differential evolution Markov chain with snooker updater and fewer chains, *Stat. Comput.*, *18*, 435–446.
- Thiemann, M., M. Trosset, H. V. Gupta, and S. Sorooshian (2001), Bayesian recursive parameter estimation for hydrologic models, *Water Resour. Res.*, *37*(10), 2521–2535, doi:10.1029/2000WR900405.
- 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 modeling: A case study using Bayesian total error analysis, *Water Resour. Res.*, *45*, W00B14, doi:10.1029/2008WR006825.
- Vrugt, J. A., W. Bouten, H. V. Gupta, and S. Sorooshian (2002), Toward improved identifiability of hydrologic model parameters: The information content of experimental data, *Water Resour. Res.*, *38*(12), 1312, doi:10.1029/2001WR001118.
- Vrugt, J. A., H. V. Gupta, W. Bouten, and S. Sorooshian (2003), A Shuffled Complex Evolution Metropolis algorithm for optimization and uncertainty assessment of hydrologic model parameters, *Water Resour. Res.*, *39*(8), 1201, doi:10.1029/2002WR001642.
- Vrugt, J. A., B. A. Robinson, and V. V. Vesselinov (2005), Improved inverse modeling of flow and transport in subsurface media: Combined parameter and state estimation, *Geophys. Res. Lett.*, *32*, L18408, doi:10.1029/2005GL023940.
- Vrugt, J. A., H. V. Gupta, S. C. Dekker, S. Sorooshian, T. Wagener, and W. Bouten (2006), Application of stochastic parameter optimization to the Sacramento Soil Moisture Accounting Model, *J. Hydrol.*, *325*, 288–307.
- Vrugt, J. A., C. J. F. ter Braak, M. P. Clark, J. M. Hyman, and B. A. Robinson (2008a), Treatment of input uncertainty in hydrologic modeling: Doing hydrology backward with Markov chain Monte Carlo simulation, *Water Resour. Res.*, *44*, W00B09, doi:10.1029/2007WR006720.
- Vrugt, J. A., C. J. F. ter Braak, H. V. Gupta, and B. A. Robinson (2008b), Equifinality of formal (DREAM) and informal (GLUE) Bayesian approaches to hydrologic modeling, *Stoch. Env. Res. Risk A.*, *23*, 1059–1060, doi:10.1007/s00477-008-0274-y.
- Vrugt, J. A., C. J. F. ter Braak, C. G. H. Diks, B. A. Robinson, J. M. Hyman, and D. Higdon (2009), Accelerating Markov chain Monte Carlo simulation by differential evolution with self-adaptive randomized subspace sampling, *Int. J. Nonlinear Sci.*, *10*, 273–290.
- Würtl, D., and Y. Chalabi (2009), Rmetrics—autoregressive conditional heteroscedastic modelling, <http://cran.r-project.org/web/packages/fGarch/index.html>.
- Yang, J., P. Reichert, and K. C. Abbaspour (2007), Bayesian uncertainty analysis in distributed hydrologic modeling: A case study in the Thur River basin (Switzerland), *Water Resour. Res.*, *43*, W10401, doi:10.1029/2006WR005497.
- Zhang, Z., T. Wagener, P. Reed, and R. Bhushan (2008), Reducing uncertainty in predictions in ungauged basins by combining hydrologic indices regionalization and multiobjective optimization, *Water Resour. Res.*, *44*, W00B04, doi:10.1029/2008WR006833.

G. Schoups, Department of Water Management, Delft University of Technology, Stevinweg 1, PO Box 5048, 2600 GA Delft, Netherlands. (g.h.w.schoups@tudelft.nl)

J. A. Vrugt, Department of Civil and Environmental Engineering, University of California, Irvine, CA 92697, USA.