PUBLISHED VERSION

Evin, Guillaume; Kavetski, Dmitri; Thyer, Mark Andrew; Kuczera, George Alfred

[Pitfalls and improvements in the joint inference of heteroscedasticity and autocorrelation in hydrological model calibration](http://onlinelibrary.wiley.com/doi/10.1002/wrcr.20284/abstract)
Water Resources Research, 2013; 49(7):4518-4524

Originally Published at:
http://onlinelibrary.wiley.com/doi/10.1002/wrcr.20284/abstract

http://hdl.handle.net/2440/80732

Pitfalls and improvements in the joint inference of heteroscedasticity and autocorrelation in hydrological model calibration

Guillaume Evin,[1] Dmitri Kavetski,[1] Mark Thyer,[1] and George Kuczera[2]

[1] Residual errors of hydrological models are usually both heteroscedastic and autocorrelated. However, only a few studies have attempted to explicitly include these two statistical properties into the residual error model and jointly infer them with the hydrological model parameters. This technical note shows that applying autoregressive error models to raw heteroscedastic residuals, as done in some recent studies, can lead to unstable error models with poor predictive performance. This instability can be avoided by applying the autoregressive process to standardized residuals. The theoretical analysis is supported by empirical findings in three hydrologically distinct catchments. The case studies also highlight strong interactions between the parameters of autoregressive residual error models and the water balance parameters of the hydrological model.

1. Introduction

[2] The residual errors of hydrological models, which represent the combined effects of data and model errors, are usually both heteroscedastic and autocorrelated [e.g., *Sorooshian and Dracup*, 1980; *Kuczera*, 1983; *Bates and Campbell*, 2001]. Heteroscedasticity is related to larger errors being generally associated with larger rainfalls and streamflows [e.g., *Villarini and Krajewski*, 2008; *Thyer et al.*, 2009]. It can be represented by directly conditioning the variance of residual errors on explanatory variables such as runoff [e.g., *Sorooshian and Dracup*, 1980; *Thyer et al.*, 2009; *Pianosi and Raso*, 2012], or by applying Box-Cox and other transformations [e.g., *Kuczera*, 1983; *Bates and Campbell*, 2001; *Smith et al.*, 2010]. The direct conditioning approach is of particular interest because it allows exploiting additional information through explanatory variables. Autocorrelation is related to the "memory" of hydrological models, with storage errors propagating across multiple consecutive time steps [e.g., *Kavetski et al.*, 2003]. It can be represented using autoregressive (AR) models [*Kuczera*, 1983], typically under lag-1 [AR(1)] assumptions [*Schaefli et al.*, 2007; *Schoups and Vrugt*, 2010].

[3] This study pursues improved probabilistic descriptions of predictive and parametric uncertainties in hydrological modeling using residual error models [e.g., *Kuczera*, 1983; *Bates and Campbell*, 2001; *Gallagher and Doherty*, 2007; *Willems*, 2009; *Smith et al.*, 2010, and many others]. It focuses on the *joint* inference of heteroscedasticity, autocorrelation and hydrological parameters. Recent work in this direction includes *Schoups and Vrugt* [2010], where a heteroscedastic skewed exponential distribution was combined with an AR(1) model, and all statistical and hydrological parameters estimated jointly. In this note, we show that a seemingly straightforward combination of heteroscedasticity and autocorrelation can result in error models with poor statistical and computational properties. We then present an alternative conceptualization of the heteroscedastic AR(1) model with a notably more robust performance.

[4] The presentation is structured as follows. Section 2 derives the statistical properties of two alternative heteroscedastic AR(1) error models. Section 3 empirically compares the predictive reliability, precision, and parameter inference (including parameter interactions) for the two error models on three hydrologically distinct catchments. The note concludes with a summary of key findings and practical recommendations in section 4.

2. Heteroscedastic Autocorrelated Residual Error Models

[5] Most statistical calibration schemes are based on residual errors ε_t, defined as

$$\varepsilon_t = \tilde{Q}_t - Q_t(\boldsymbol{\theta}_H, \tilde{\boldsymbol{X}}_{1:t}, \boldsymbol{S}_0) \tag{1}$$

where $\tilde{Q}_t$ and Q_t are, respectively, the observed and simulated flows at time step t. As indicated, Q_t is a function of the forcing data $\tilde{X}$ (e.g., rainfall and evapotranspiration), the hydrological model parameters $\boldsymbol{\theta}_H$ and the initial conditions $\boldsymbol{S}_0$.

[1] School of Civil, Environmental and Mining Engineering, University of Adelaide, Adelaide, Australia.
[2] School of Engineering, University of Newcastle, Callaghan, New South Wales, Australia.

Corresponding author: D. Kavetski, School of Civil, Environmental and Mining Engineering, University of Adelaide, Adelaide, SA 5005, Australia. (dmitri.kavetski@adelaide.edu.au)

[6] The Bayesian posterior distribution corresponding to such residual error models is

$$p(\boldsymbol{\theta}_H, \boldsymbol{\theta}_\varepsilon | \tilde{\mathbf{Q}}, \tilde{\boldsymbol{X}}) \propto p(\tilde{\mathbf{Q}} | \boldsymbol{\theta}_H, \boldsymbol{\theta}_\varepsilon, \tilde{\boldsymbol{X}}) p(\boldsymbol{\theta}_H, \boldsymbol{\theta}_\varepsilon) \tag{2}$$

where $\boldsymbol{\theta}_\varepsilon$ denotes the parameters of the residual error model (such as standard deviation, etc.). The likelihood function in equation (2) is given by the joint pdf of the residuals

$$p(\tilde{\mathbf{Q}} | \boldsymbol{\theta}_H, \boldsymbol{\theta}_\varepsilon, \tilde{\boldsymbol{X}}) = p(\boldsymbol{\varepsilon}[\boldsymbol{\theta}_H, \tilde{\boldsymbol{X}}, \tilde{\mathbf{Q}}] | \boldsymbol{\theta}_\varepsilon) \tag{3}$$

where $\boldsymbol{\varepsilon}[\boldsymbol{\theta}_H, \tilde{\boldsymbol{X}}, \tilde{\mathbf{Q}}]$ represents the vector of residual errors, computed over the calibration period $t = 1, \ldots, N_t$. A warm-up period $t = -N_w, \ldots, 0$ is used prior to the calibration period.

2.1. Approach 1: AR(1) Model Applied to Raw Residuals

2.1.1. Formulation and Basic Properties

[7] Let the raw residual errors ε_t be described by

$$\varepsilon_t = \phi_\varepsilon \varepsilon_{t-1} + z_t; \qquad z_t \sim N\left(0, \sigma_{z(t)}^2\right) \tag{4}$$

where the innovations z_t are independent zero-mean Gaussian deviates with a time-varying standard deviation $\sigma_{z(t)}$, and $\phi_\varepsilon \in (-1, 1)$ is the lag-1 autoregressive parameter.

[8] To account for heteroscedasticity, we use a linear model for the standard deviation $\sigma_{z(t)}$, conditioned on the simulated streamflow Q_t, with parameters a_z and b_z,

$$\sigma_{z(t)} = a_z + b_z Q_t \tag{5}$$

[9] This approach in equations (4) and (5) is analogous to the scheme used by *Schoups and Vrugt* [2010], except here, in order to focus more directly on autocorrelation and heteroscedasticity, we use Gaussian assumptions and do not introduce bias, skew, and kurtosis parameters.

[10] From equation (4), the residual errors ε_t can be expressed in terms of the innovations z_t

$$\begin{aligned} \varepsilon_t &= z_t + \phi_\varepsilon(z_{t-1} + \phi_\varepsilon \varepsilon_{t-2}) \\ &= z_t + \phi_\varepsilon z_{t-1} + \phi_\varepsilon^2 z_{t-2} + \ldots + \phi_\varepsilon^{t+N_w+1} \varepsilon_{-N_w-1} \approx \sum_{i=0}^{t+N_w} \phi_\varepsilon^i z_{t-i} \end{aligned} \tag{6}$$

where, as $N_w \to \infty$, the term $\phi_\varepsilon^{t+N_w+1} \varepsilon_{-N_w-1}$ becomes negligible provided $|\phi_\varepsilon| < 1$.

[11] Equation (6) shows that the marginal distribution of an AR(1) process with heteroscedastic Gaussian innovations is also Gaussian: ϕ is a constant and equation (6) is hence a sum of independent scaled Gaussian deviates.

[12] Equation (6) also shows that the process has zero mean ($E[\varepsilon_t] = 0$) because $E[z_t] = 0$. Its variance can be obtained by considering that the innovations z_t are defined as mutually independent,

$$\text{var}(\varepsilon_t) = E\left[(\varepsilon_t - E[\varepsilon_t])^2\right] = E\left[\left(\sum_{i=0}^{t+N_w} \phi_\varepsilon^i z_{t-i}\right)^2\right] = \sum_{i=0}^{t+N_w} \phi_\varepsilon^{2i} \sigma_{z(t-i)}^2 \tag{7}$$

[13] In general, $\text{var}(\varepsilon_t)$ does not have a closed form expression. When $\sigma_{z(t)} = \sigma_z$ (i.e., constant variance in time), it simplifies to $\text{var}(\varepsilon_t) = \sigma_z^2 / (1 - \phi_\varepsilon^2)$, which corresponds to the marginal variance of a homoscedastic AR(1) process. Note that when the autoregressive coefficient ϕ_ε approaches 1, the AR(1) process becomes nonstationary [*Box and Jenkins*, 1976]. In particular, when $\phi_\varepsilon = 1$, the term $\phi_\varepsilon^{t+N_w+1} \varepsilon_{-N_w-1}$ in equation (6) does not decay as $N_w \to \infty$, and hence $\text{var}(\varepsilon_t) = \sum_{i=0}^{\infty} \sigma_{z(t-i)}^2$ becomes infinite. As will be shown in section 3, this has major implications on residual error model behavior.

2.1.2. Likelihood Function

[14] Approach 1 introduces error model parameters $\boldsymbol{\theta}_\varepsilon^{(1)} = (\phi_\varepsilon, a_z, b_z)$ and results in the likelihood

$$p(\boldsymbol{\varepsilon}) = p(\varepsilon_1, \ldots, \varepsilon_n) = p(\varepsilon_1) \prod_{t=2}^{N_t} p(\varepsilon_t | \varepsilon_{t-1}) \tag{8}$$

$$p(\varepsilon_t | \varepsilon_{t-1}) = N\left(\varepsilon_t | \phi_\varepsilon \varepsilon_{t-1}, \sigma_{z(t)}^2\right) = N\left(\varepsilon_t - \phi_\varepsilon \varepsilon_{t-1} | 0, \sigma_{z(t)}^2\right) \tag{9}$$

where $N(x|\mu, \sigma^2)$ denotes the pdf of a scalar Gaussian random deviate x with mean μ and standard deviation σ.

[15] These equations correspond exactly to the scheme of *Schoups and Vrugt* [2010] when the latter is applied with no bias, skew, or kurtosis. For the first time step, we use the marginal distribution $p(\varepsilon_1)$, which as shown above is Gaussian with zero mean and variance given by equation (7).

2.1.3. Lag-1 Autocorrelation of the Residuals

[16] The lag-1 autocorrelation of the raw residuals can be derived as

$$\begin{aligned} \rho(\varepsilon_t, \varepsilon_{t-1}) &= \frac{E(\varepsilon_t \varepsilon_{t-1})}{\sqrt{\text{var}(\varepsilon_t)\text{var}(\varepsilon_{t-1})}} = \frac{E(z_t \varepsilon_{t-1}) + E(\phi_\varepsilon \varepsilon_{t-1}^2)}{\sqrt{\text{var}(\varepsilon_t)\text{var}(\varepsilon_{t-1})}} \\ &= \frac{\phi_\varepsilon \text{var}(\varepsilon_{t-1})}{\sqrt{\text{var}(\varepsilon_t)\text{var}(\varepsilon_{t-1})}} = \phi_\varepsilon \sqrt{\frac{\text{var}(\varepsilon_{t-1})}{\text{var}(\varepsilon_t)}} \end{aligned} \tag{10}$$

[17] Equation (10) shows that in Approach 1 the lag-1 autocorrelation of the residuals depends on both the autoregressive parameter ϕ_ε and on the form of heteroscedasticity.

2.2. Approach 2: AR(1) Model Applied to Standardized Residuals

[18] This section considers an alternative heteroscedastic AR(1) model, where the AR(1) assumptions are applied to standardized residuals, rather than to the raw residuals.

2.2.1. Formulation and Basic Properties

[19] Let us define a standardized residual error η_t as follows:

$$\eta_t = \frac{\varepsilon_t}{\sigma_{\varepsilon(t)}} \tag{11}$$

where $\sigma_{\varepsilon(t)}$ is the standard deviation of the raw residual at time t.

[20] Analogously to Approach 1, we can assume linear heteroscedasticity,

$$\sigma_{\varepsilon(t)} = a_\varepsilon + b_\varepsilon Q_t \tag{12}$$

[21] We now apply the AR(1) process on the standardized residual errors, defining innovations y_t as

$$\eta_t = \phi_\eta \ \eta_{t-1} + y_t; \quad y_t \sim N\left(0, \sigma_y^2\right) \tag{13}$$

[22] Equation (13) corresponds to a standard homoscedastic AR(1) process and, since var $[\eta_t] = 1$ due to the standardization in equation (11), it follows that $\sigma_y^2 = 1 - \phi_\eta^2$.

2.2.2. Likelihood Function

[23] The error model parameters in Approach 2 are $\theta_\varepsilon^{(2)} = \left(\phi_\eta, a_\varepsilon, b_\varepsilon\right)$. The likelihood function, given by the joint pdf of the residuals, $p(\varepsilon)$, must account for the transformation in equation (11)

$$p(\boldsymbol{\varepsilon}) = \det\left|\frac{\partial \boldsymbol{\eta}}{\partial \boldsymbol{\varepsilon}}\right| p(\boldsymbol{\eta}) \tag{14}$$

where det $|\partial\boldsymbol{\eta}/\partial\boldsymbol{\varepsilon}|$ is the absolute value of the determinant of the Jacobian matrix of the transformation in equation (11) and can be derived to be $\prod_{t=1}^{N_t} \partial\eta_t/\partial\varepsilon_t = \prod_{t=1}^{N_t} 1/\sigma_{\varepsilon(t)}$.

[24] The likelihood function is then

$$p(\boldsymbol{\varepsilon}) = p(\varepsilon_1, ..., \varepsilon_n) = \frac{p(\eta_1)}{\sigma_{\varepsilon(1)}} \prod_{t=2}^{N_t} \frac{p(\eta_t|\eta_{t-1})}{\sigma_{\varepsilon(t)}} \tag{15}$$

[25] Since we assumed Gaussian innovations y_t in equation (13), we have

$$p(\eta_t|\eta_{t-1}) = N\left(\eta_t|\mu = \phi_\eta\eta_{t-1}, \sigma^2 = 1 - \phi_\eta^2\right)$$
$$= N\left(\eta_t - \phi_\eta\eta_{t-1}|\mu = 0, \sigma^2 = 1 - \phi_\eta^2\right) \tag{16}$$

[26] In addition, from section 2.2.1, the marginal distribution at $t = 1$ is $p(\eta_1) = N\left(0, 1^2\right)$.

2.2.3. Lag-1 Autocorrelation

[27] The lag-1 autocorrelation of residual errors in Approach 2 is

$$\rho(\varepsilon_t, \varepsilon_{t-1}) = \frac{\mathrm{E}(\varepsilon_t\varepsilon_{t-1})}{\sqrt{\mathrm{var}\,(\varepsilon_t)\mathrm{Var}\,(\varepsilon_{t-1})}} = \frac{\sigma_{\varepsilon(t)}\sigma_{\varepsilon(t-1)}\phi_\eta}{\sigma_{\varepsilon(t)}\sigma_{\varepsilon(t-1)}} = \phi_\eta \tag{17}$$

where we used the result

$$\mathrm{E}(\varepsilon_t\varepsilon_{t-1}) = \sigma_{\varepsilon(t)}\sigma_{\varepsilon(t-1)}\mathrm{E}(\eta_t\eta_{t-1}) = \sigma_{\varepsilon(t)}\sigma_{\varepsilon(t-1)}\phi_\eta \tag{18}$$

[28] Equation (17) shows that in Approach 2 the lag-1 autocorrelation $\rho(\varepsilon_t, \varepsilon_{t-1})$ is constant in time and corresponds exactly to the autoregressive parameter ϕ_η. This can be contrasted with equation (10) for Approach 1, where the lag-1 autocorrelation has a more complex structure.

2.3. Similarities and Differences Between Approaches 1 and 2

[29] Although Approaches 1 and 2 represent heteroscedasticity and autocorrelation using similar equations, the order in which these properties are treated is different (Table 1). In Approach 1, an AR(1) model is applied to the raw residuals, followed by the application of a heteroscedastic model to its innovations (by time-varying the conditional variance of the AR(1) process). In contrast, Approach 2 applies a heteroscedastic model to standardize the raw residuals, followed by the application of a homoscedastic AR(1) model to the standardized residuals (i.e., after the heteroscedasticity has been removed, or at least substantially reduced).

[30] These structural differences lead to important differences in the mathematical behavior of the two approaches. Both approaches include autoregressive equations that accumulate the errors from previous innovations. However, in Approach 1 the innovations are heteroscedastic and can result in particularly large accumulated errors. For example, equation (7) shows that large innovations associated with peak streamflows propagate into the predictive uncertainty of subsequent recession time steps. This propagation is particularly strong when $\phi_\varepsilon \approx 1$. Conversely, in Approach 2 the innovations of the AR(1) process are homoscedastic, and the heteroscedasticity in equation (12) is applied after (rather than before) the innovations are accumulated. The practical implications of these differences on the behavior of the error models will be investigated in an empirical case study (section 3).

[31] Approach 2 can be viewed as a particular case of the common "variance-stabilizing" strategy of transforming residual errors and then specifying a (possibly correlated) *homoscedastic* probability distribution of the transformed residuals. For example, see equations (A(4)) and (A(5)) in *Bates and Campbell* [2001], where a homoscedastic AR(1) process is applied to Box-Cox-transformed residuals. In Approach 2, the homoscedastic AR(1) process is applied to standardized residuals. Unlike the Box-Cox transformation, the explicit standardization in equations (11) and (12) directly exploits the conditioning on simulated streamflow as an explanatory variable.

[32] In contrast, Approach 1 corresponds to applying a joint "decorrelation" transformation $z_t = e_t - \phi \ e_{t-1}$ followed by specifying a *heteroscedastic* distribution of the transformed residuals z. It can hence be viewed as a special case of an "autocorrelation-reduction" strategy where the aim of the transformation is to reduce autocorrelation rather than heteroscedasticity.

3. Empirical Case Study

3.1. Hydrological Data, Models, and Methodology

[33] Approaches 1 and 2 are compared in three catchments: Lacmalac and Tinderry (South-East of Australia),

Table 1. Summary of Equations for Approaches 1 and 2

Approach 1	Approach 2
$\varepsilon_t = \phi_\varepsilon\varepsilon_{t-1} + z_t$; $z_t \sim N(0, \sigma_{z(t)}^2)$ $\sigma_{z(t)} = a_z + b_z Q_t$	$\varepsilon_t = \sigma_{\varepsilon(t)}\eta_t$ $\sigma_{\varepsilon(t)} = a_\varepsilon + b_\varepsilon Q_t$ $\eta_t = \phi_\eta\eta_{t-1} + y_t;$ $y_t \sim N(0, \sigma_y^2)$

Table 2. Catchments Properties and Calibration and Validation Periods

Catchment Name and Number	Area (km^2)	Mean Annual Rainfall (mm)	Mean Annual Runoff (mm)	Runoff Coefficient	Percentage of Days with Runoff <1 mm	Calibration Period	Validation Period
Lacmalac 410057	673	1182	397	0.33	65.0	28 Oct 1976–2 Oct 1978	3 Oct 1978–6 Sep 1983
Tinderry 410734	490	808	106	0.13	95.3	22 Jun 1980–8 Sep 1988	9 Sep 1988–19 Apr 1995
French Broad 3451500	2448	1413	800	0.57	16.3	8 Sep 1973–25 Nov 1981	26 Nov 1981–30 Apr 1998

and the French Broad River (Asheville, Texas). The climatology of these catchments and the daily data periods used in the analysis are listed in Table 2. These catchments provide a range of different climatologies and hydrological regimes: Tinderry is dry (ephemeral), Lacmalac is wet and French Broad River is particularly wet. Simulated streamflow is obtained using GR4J, a lumped rainfall-runoff model with 4 fitted parameters [*Perrin et al.*, 2003]. Uniform priors are used on all inferred quantities, with parameter ranges specified in Table 3. The posterior distributions are optimized using a quasi-Newton method [*Kavetski and*

Clark, 2010] and sampled using a multistage Metropolis algorithm [*Thyer et al.*, 2009].

[34] Several performance metrics are used. Statistical reliability is evaluated using the predictive QQ plot [*Thyer et al.*, 2009], precision is quantified using the average (in time) coefficient of variation of the predictive distribution. The adequacy of the AR(1) approximation is assessed using the autocorrelation function (ACF) of the standardized innovations, i.e., $z_t/\sigma_{z(t)}$ for Approach 1 (see equation (4)) and $y_t/\sqrt{1 - \phi_\eta^2}$ for Approach 2 (see equation (13)). The ACF should be as close as possible to 0. We appraise the extent to which the standardized innovations are homoscedastic by plotting them against the quantile of the simulated flow, and the extent to which they are Gaussian by comparing their empirical marginal density to a Gaussian distribution. Finally, we inspect the posterior parameter distributions and the interactions between hydrological and error model parameters.

Table 3. Parameter Specifications

	Parameter	Description	Minimum	Maximum
θ_H	θ_1 (mm)	Maximum capacity of the production store	100	20,000
	θ_2 (mm)	Groundwater exchange coefficient	−500	500
	θ_3 (mm)	Maximum capacity of the routing store	1	500
	θ_4 (days)	Time base of unit hydrograph	0.5	10
θ_ε	a (mm)	Heteroscedasticity intercept	0.0001	100
	b	Heteroscedasticity slope	0.0001	10
	ϕ	Autoregressive coefficient	−0.999	0.999

3.2. Visual Assessment of Predictive Performance

[35] As shown in Figure 1, Approach 1 produces very vague predictions, with exceedingly wide and poorly behaved predictive bounds, especially just after the recessions. This behavior is particularly pronounced in the
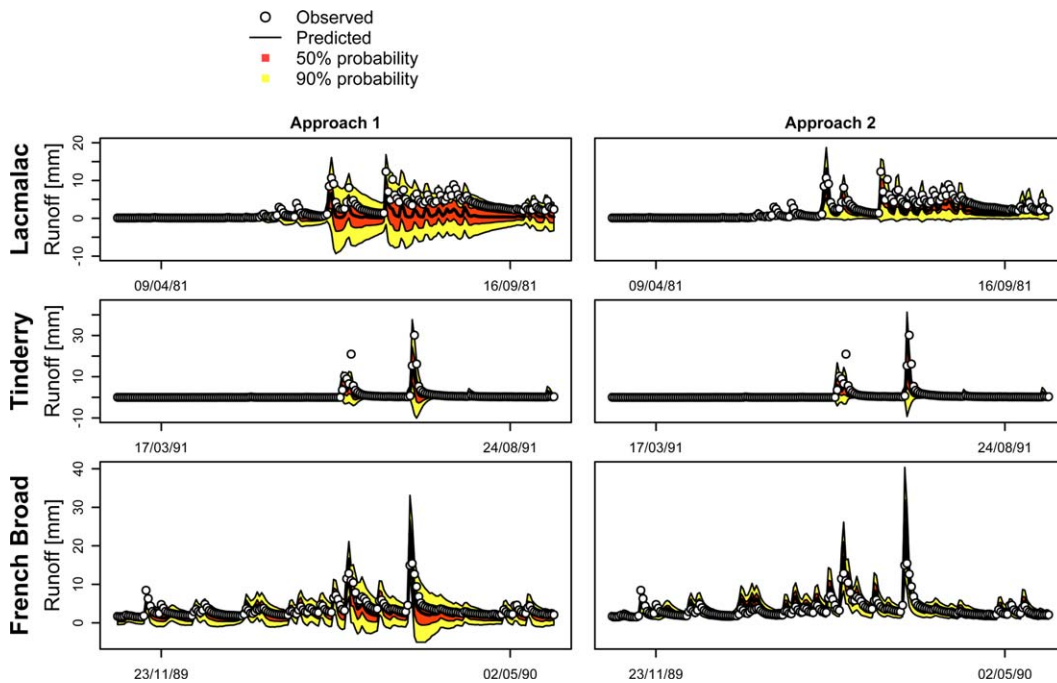

Figure 1. Predictive distributions of streamflow during representative portions of the validation period.

Lacmalac and French Broad River catchments, which are both humid. It appears related to the error accumulation behavior discussed in section 2.3. Another undesirable property of Approach 1 is its exceedingly high proportion of large negative predicted flows. Even if the predictive distribution was truncated at zero to avoid negative flows, the resulting distribution would remain poor. Conversely, Approach 2 does not suffer from these problems and has much better-behaved prediction limits with a few small negative flows that can be safely truncated.

[36] For the Tinderry catchment, the predictive bounds obtained with Approaches 1 and 2 are very similar, and contain an exceedingly large fraction of negative flows. Given the ephemeral nature of this catchment, avoiding this degeneracy is likely to require a specialized treatment of low and zero flows, including the development of truncated likelihood functions [*Smith et al.*, 2010].

3.3. Statistical Metrics of Predictive Performance

[37] Figure 2 compares Approaches 1 and 2 using a range of diagnostics. Consistently with the hydrographs in Figure 1, Approach 2 clearly outperforms Approach 1 in the Lacmalac and French Broad catchments, especially in terms of predictive precision (row 2 of Figure 2). Approach 2 also produces near-perfect reliability in the French Broad River, whereas Approach 1 systematically under-estimates streamflow (row 1 of Figure 2).

[38] The error heteroscedasticity appears adequately captured by equations (5) and (12) in the (wet) Lacmalac and French Broad catchments, with the standardized innovations being nearly homoscedastic as required (row 3 of Figure 2). This is not the case in the (ephemeral) Tinderry basin, where there is a clear remaining trend in the relationship between the standardized innovations and streamflow. The curvature in this trend suggests a nonlinear heteroscedastic model warrants investigation.

[39] Density plots of the standardized innovations (row 4 of Figure 2) reveal that they are generally symmetric but kurtotic, which is consistent with earlier studies [*Schoups and Vrugt*, 2010]. The kurtosis is particularly strong in the Tinderry catchment, where it is combined with a slight asymmetry. The Gaussian assumptions are hence questionable in this case. The standardized innovations obtained with Approach 1 exhibit a slight positive bias for the French Broad catchment, which is not the case when Approach 2 is used.

[40] The autocorrelation structure of the residuals appears well approximated by the AR(1) assumption, with the ACF of the standardized innovations being close to 0 for both Approaches 1 and 2, in all three catchments (row 5 of Figure 2). This clearly contrasts with the large autocorrelations when the AR(1) error model component is omitted, and highlights the need for autocorrelated error models.

[41] In general, Approaches 1 and 2 produce very similar, and generally poor, results in the Tinderry catchment (Approach 1 yields a small gain in reliability while Approach 2 produces slightly higher precision). This can be attributed to the ephemerality of the Tinderry catchment and reinforces the need to improve the hydrological model (e.g., to capture the wetting-up thresholds) and employ a specialized treatment of zero and near-zero flows in the likelihood function [*Smith et al.*, 2010].

3.4. Parameter Inference

[42] Figure 3 shows that the posterior parameter distributions are generally well behaved with both Approaches 1 and 2; most parameters appear well identified. However, the inferences are different and some problematic features are evident. In Approach 1, the GR4J water balance parameter θ_2 is highly negative (indicating export of groundwater from the catchment) and is negatively correlated with the heteroscedasticity slope b_z. In Approach 2, θ_2 is close to 0, though it remains correlated with b_ε. Approach 2 also has a strong positive correlation between ϕ_η and b_ε, which appears absent in Approach 1. Similar results (not shown here) were found in the other two catchments.

4. Conclusions and Recommendations

[43] This technical note illustrates the challenges of fitting hydrological model parameters jointly with the autocorrelation and heteroscedasticity parameters of residual error models. An empirical case study was undertaken, based on three catchments with diverse hydrological dynamics and the widely used GR4J hydrological model. Two distinct residual error models based on Gaussian AR(1) processes were compared using a range of diagnostics. The main conclusions are as follows:

[44] (1) When jointly inferring heteroscedasticity and autocorrelation parameters, applying Gaussian AR(1) models directly to the raw residual errors can produce a poorly behaved error model with grossly exaggerated predictive uncertainty. This instability can be avoided by applying the Gaussian AR(1) process to standardized residuals. This empirical finding appears consistent with analytical insights into the error accumulation properties of the two statistical error models.

[45] (2) Hydrological parameters directly controlling the water balance can interact strongly when fitted jointly with heteroscedasticity and autocorrelation parameters. Since the current study employed a single hydrological model, further testing is needed to ascertain its relation to conclusion 1 above.

[46] (3) Applying the Gaussian AR(1) process to standardized residuals results in strong interactions between autoregressive and heteroscedastic parameters. Further research is required to understand the origin of these interactions and whether they can be eliminated or at least reduced (e.g., using alternative autocorrelation structures, parameter transformations or reparameterizations).

[47] (4) Ephemeral catchments remain particularly hard to model, and in addition to a robust representation of heteroscedasticity and autocorrelation, require a more robust handling of skew, near-zero flows, seasonality and other aspects (confirming previous studies on this topic). Moreover, in all three basins, there is evidence of excess kurtosis in the standardized innovations.

[48] Finding 1 can help hydrologists improve the calibration of hydrological models by avoiding fundamental statistical problems, while Findings 2–4 highlight areas for further research, in particular including the analysis of these issues within more complex error models.
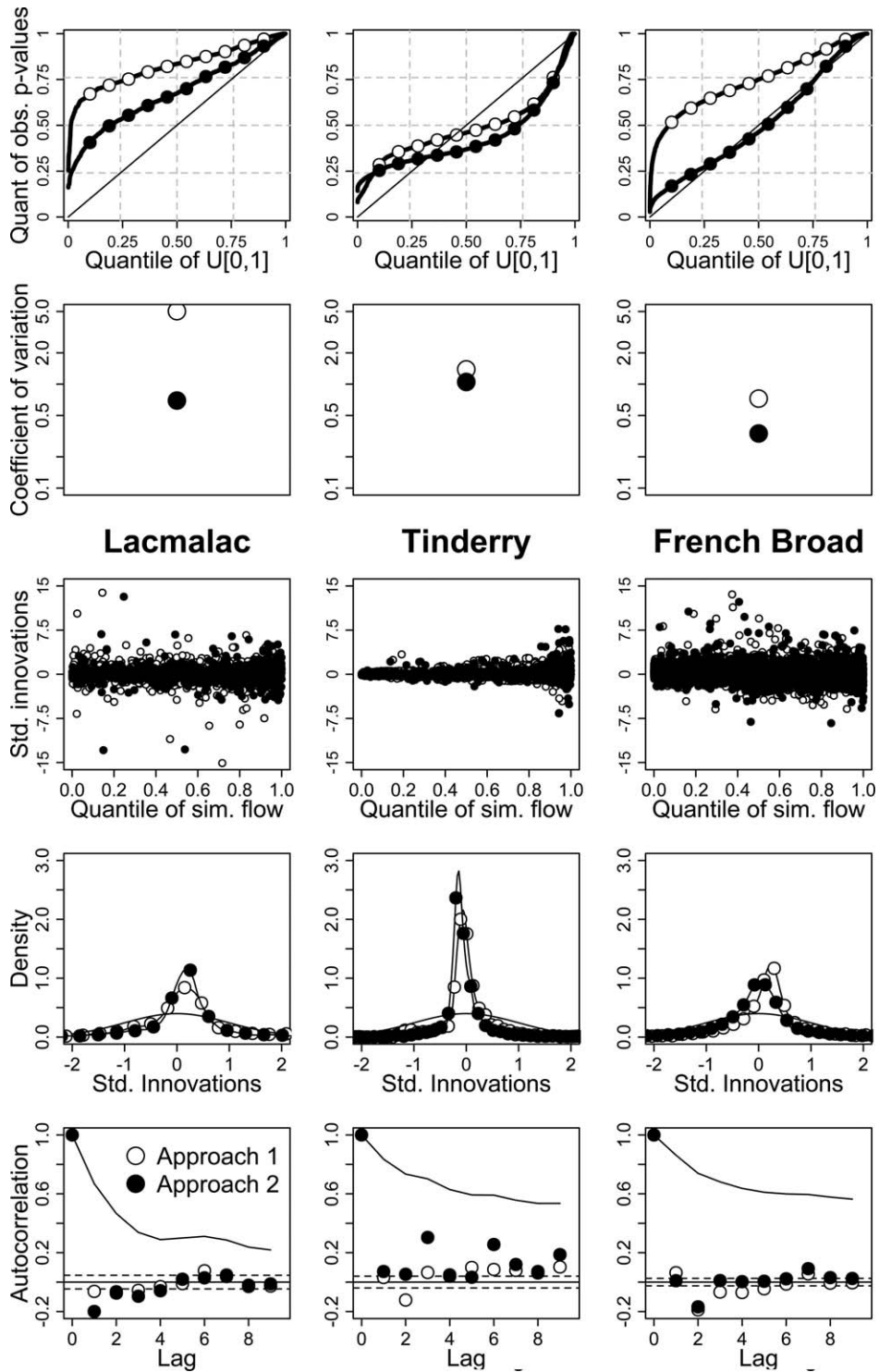
Figure 2. Comparison of predictive performances of Approach 1 (white circles) and Approach 2 (black circles) over the validation period using a range of diagnostics. Statistical reliability (Row 1): predictive QQ-plots (for ease of visualization, the symbols are used to distinguish between the curves rather than to denote individual data points). Precision (Row 2): average (in time) coefficient of variation of the predictive distribution. Heteroscedasticity (Row 3): standardized innovations as a function of quantiles of simulated flow. Distributional check (Row 4): probability density plots of standardized innovations (the black line illustrates the error model assumptions—a Gaussian pdf with zero mean and unit standard deviation). Autocorrelation (Row 5): autocorrelation function (ACF) plots (with 95% confidence intervals indicated with dotted lines). The residual autocorrelations obtained when the AR(1) error model component is omitted (i.e., Approach 1 with $\phi_\varepsilon = 0$) are shown with a black line.
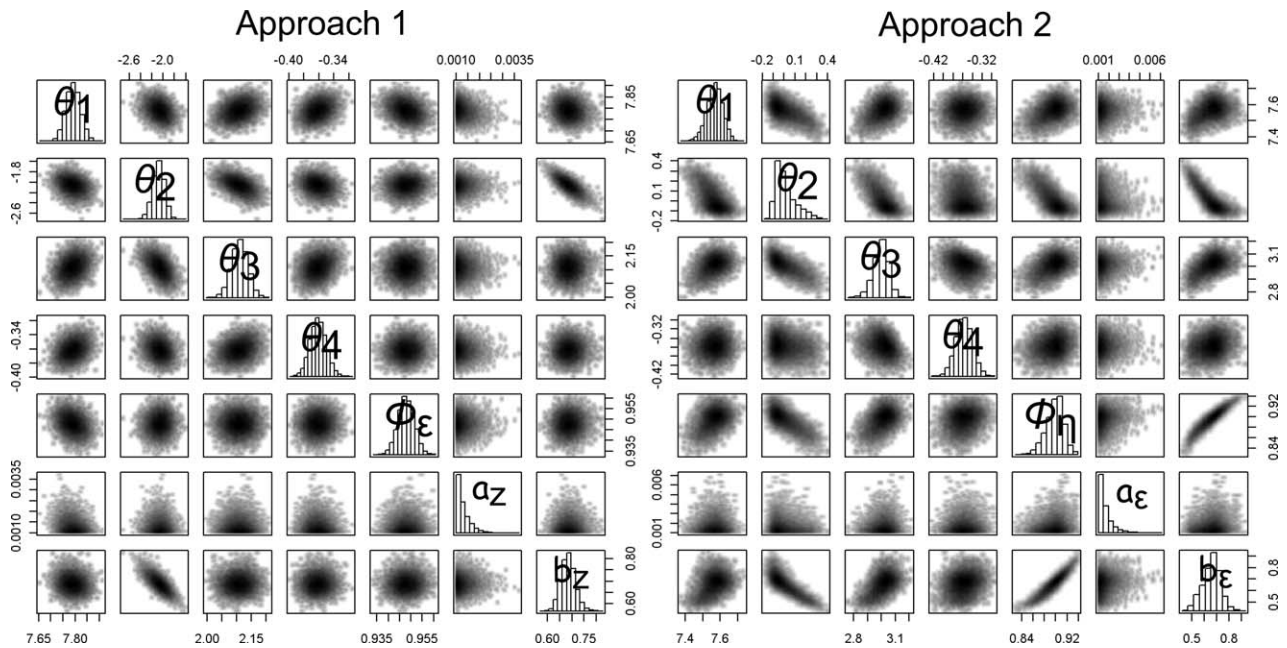
Figure 3. Posterior parameter distributions for the Lacmalac catchment.

References

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.

Box, G. E. P., and G. M. Jenkins (1976), *Time Series Analysis: Forecasting and Control,* Holden-Day, San Francisco, Calif.

Gallagher, M., and J. Doherty (2007), Parameter estimation and uncertainty analysis for a watershed model, *Environ. Modell. Software*, *22*(7), 1000–1020.

Kavetski, D., and M. P. Clark (2010), Ancient numerical daemons of conceptual hydrological modeling. Part 2: Impact of time stepping scheme on model analysis and prediction, *Water Resour. Res.*, *46*, W10511, doi:10510.11029/12009WR008896.

Kavetski, D., S. W. Franks, and G. Kuczera (2003), Confronting Input Uncertainty in Environmental Modelling, in *Calibration of Watershed Models*, edited by Q. Duan, H. V. Gupta, S. Sorooshian, A. N. Rousseau, R. Turcotte, pp. 49–68, AGU, Washington D. C.

Kuczera, G. (1983), Improved parameter inference in catchment models: 1. Evaluating parameter uncertainty, *Water Resour. Res.*, *19(5)*, 1151–1162, 10.1029/WR019i005p01151.

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

Pianosi, F., and L. Raso (2012), Dynamic modeling of predictive uncertainty by regression on absolute errors, *Water Resour. Res.*, *48*(3), W03516, doi:10.1029/2011WR010603.

Schaefli, B., D. B. Talamba, and A. Musy (2007), Quantifying hydrological modeling errors through a mixture of normal distributions, *J. Hydrol.*, *332*, 303–315, doi:10.1016/j.jhydrol.2006.07.005.

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:10510.11029/12009WR008933.

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 Resour. Res.*, *46*, W12551, doi:12510.11029/12010WR009514.

Sorooshian, S., and J. A. Dracup (1980), Stochastic parameter estimation procedures for hydrological rainfall-runoff models: Correlated and heteroscedastic error cases, *Water Resour. Res.*, *16*(2), 430–442.

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 Resour. Res.*, *45*, W00B14, doi:10.1029/2008WR006825.

Villarini, G., and W. F. Krajewski (2008), Empirically based modeling of spatial sampling uncertainties associated with rainfall measurements by rain gauges, *J. Hydrol.*, *31*, 1015–1023.

Willems, P. (2009), A time series tool to support the multi-criteria performance evaluation of rainfall-runoff models, *Environ. Modell. Software*, *24*(3), 311–321.