Extending the Applicability of the Generalized Likelihood Function for Zero-Inflated Data Series

Debora Y. Oliveira[1] [ID], **Pedro L. B. Chaffe[2]** [ID], **and João H. M. Sá[1]** [ID]

[1]Graduate Program in Environmental Engineering, Federal University of Santa Catarina, Florianópolis, Brazil, [2]Department of Sanitary and Environmental Engineering, Federal University of Santa Catarina, Florianópolis, Brazil

Abstract Proper uncertainty estimation for data series with a high proportion of zero and near zero observations has been a challenge in hydrologic studies. This technical note proposes a modification to the Generalized Likelihood function that accounts for zero inflation of the error distribution (ZI-GL). We compare the performance of the proposed ZI-GL with the original Generalized Likelihood function using the entire data series (GL) and by simply suppressing zero observations ($GL^{Y>0}$). These approaches were applied to two interception modeling examples characterized by data series with a significant number of zeros. The ZI-GL produced better uncertainty ranges than the GL as measured by the precision, reliability and volumetric bias metrics. The comparison between ZI-GL and $GL^{Y>0}$ highlights the need for further improvement in the treatment of residuals from near zero simulations when a linear heteroscedastic error model is considered. Aside from the interception modeling examples illustrated herein, the proposed ZI-GL may be useful for other hydrologic studies, such as for the modeling of the runoff generation in hillslopes and ephemeral catchments.

1. Introduction

Bayesian inference has been widely used in hydrology for predictive uncertainty estimation. Within the Bayesian framework, two main approaches are used: (1) lumping all sources of uncertainty together and treating them as an additive term to the deterministic hydrological predictions (e.g., Schoups & Vrugt, 2010); (2) separately accounting for the contribution of different sources of uncertainty (Kavetski et al., 2006; Renard et al., 2010, 2011; Vrugt et al., 2008). Regardless of the method used for uncertainty estimation, it has been recognized that the correct characterization of the residual errors is essential for obtaining meaningful parameter values and reliable uncertainty ranges (Schoups & Vrugt, 2010; Smith et al., 2010; Thyer et al., 2009).

Different approaches have been used to deal with heteroscedasticity and correlation, typical characteristics of the hydrological modeling errors. Heteroscedasticity was considered by using a mixture of normal distributions for low and high flows (Schaefli et al., 2007), by modeling the residual standard deviation as a linear function of predicted values (Evin et al., 2014; Schoups & Vrugt, 2010; Westra et al., 2014; and many others), and by using Box-Cox transformations of data (Cheng et al., 2014; Smith et al., 2010, 2015). A first-order autoregressive model is the most frequently used formulation to deal with correlated residuals (Evin et al., 2013, 2014; Schaefli et al., 2007; and many others), although more generic frameworks are illustrated in some studies (e.g., Schoups & Vrugt, 2010).

The impact of zero and near zero observations has also been acknowledged. Despite recent studies highlighting, the need for special handling of zero and near zero observations in uncertainty estimation (Evin et al., 2013, 2014; McInerney et al., 2017), there are only a few methods developed for this purpose. Smith et al. (2010) introduced a mixture likelihood to separate the contribution of zero observations (with zero and nonzero residuals) and nonzero observations (with nonzero residuals). Wang and Robertson (2011) and Li et al. (2013) treated zero observations as censored data and considered their contribution to the likelihood as being the probability of the variable being equal or (hypothetically) below zero. Westra et al. (2014) censored the values below a certain threshold from the data series to avoid the negative impact of the corresponding residuals on the likelihood quantification.

In a recent paper, Schoups and Vrugt (2010) introduced a flexible likelihood function to deal with non-Gaussian, heteroscedastic, and correlated residuals. This likelihood function has been proved to be useful in

many hydrological modeling studies (e.g., Koskela et al., 2012; Schoups & Vrugt, 2010) and other studies suggested its application (e.g., Li et al., 2012). However, the presence of zero and near zero observations in the data series has been recognized as a difficulty in the proper quantification of the predictive uncertainty (Evin et al., 2013; Honti et al., 2013; Koskela et al., 2012; Schoups & Vrugt, 2010).

In this study, we present a modified version of the Generalized Likelihood function (GL) of Schoups and Vrugt (2010) that accounts for zero inflation of the error distribution. The proposed method is similar to the zero-inflation approach presented in Smith et al. (2010) except that the residual time series is divided into zero and nonzero states based on simulated instead of observed values. The proposed modification to the GL addresses the shortcomings discussed in previous publications, further extending its applicability. We illustrate the usefulness of the proposed likelihood function with a rainfall interception modeling example.

This paper is organized as follows. Section 2 presents the proposed zero-inflated version of the Generalized Likelihood function (ZI-GL). Section 3 empirically illustrates the usefulness of the ZI-GL with an interception modeling case study. This technical report ends with a summary of the main findings and recommendations for further studies in section 4.

2. Zero-Inflated Generalized Likelihood Function

2.1. Motivation

Schoups and Vrugt (2010) presented the Generalized Likelihood function to model non-Gaussian, heteroscedastic, and correlated residuals. Instead of the commonly assumed Gaussian distribution, residuals are modeled by a skewed exponential power (SEP) distribution, which has kurtosis (β) and skewness (ξ) parameters. The use of this flexible error distribution enables to some extent the handling of data series with zero observations, which can be modeled with zero error and therefore, may result in a more kurtotic distribution than a Gaussian distribution. However, for cases with prolonged periods of zero observations, increasing β may not be sufficient to accommodate the zero inflation of the error distribution (Koskela et al., 2012).

2.2. Formulation

In this work, we propose a modification of the Generalized Likelihood function to extend its applicability to situations where the data series has a high proportion of zero observations. We based our approach on the zero-inflation method presented by Smith et al. (2010), except that we split the residual time series according to simulated instead of observed values. Conditioning the residual error model on simulated values allows its use in prediction, where observations are not available. The raw residuals (e_t), defined as the difference between the observed (y_t) and the simulated ($\hat{y}_t$) variable at time t,

$$e_t = y_t - \hat{y}_t, \tag{1}$$

are separated in three classes: zero residuals from zero simulations (e_1); nonzero residuals from zero simulations (e_2); and nonzero residuals from nonzero simulations (e_3). The raw residuals are transformed to account for heteroscedasticity and autocorrelation (hereafter a refers to these transformed residuals) and are modeled in two stages. First, a binomial probability model is used for residuals corresponding to zero simulations ($a_{t,sim=0}$), where a_1 are associated with a probability ρ and a_2 are associated with a probability $1 - \rho$. Transformed residuals from class 2 are subsequently modeled by a SEP distribution with mean of 1, unit variance, skewness parameter ξ, and kurtosis parameter β. The residual model corresponding to the zero-simulation state is therefore,

$$a_{t,sim=0} \sim \begin{cases} 0 & \text{with probability } \rho \\ \text{SEP}(1,1,\xi,\beta) & \text{with probability } 1-\rho \end{cases}. \tag{2}$$

where $\rho = n_1/(n_1 + n_2)$ is the probability of a zero error given a zero simulation and is computed directly from the number of zero simulations with zero error n_1 and the number of zero simulations with nonzero error n_2. A standardized SEP distribution (zero mean and unit variance) is used to model transformed residuals from nonzero simulations ($a_{t,sim>0}$),

$$a_{t,sim>0} \sim \text{SEP}(0,1,\xi,\beta). \tag{3}$$

Heteroscedasticity is considered by assuming a linear model for the error variance, i.e.,

$$\sigma_t = \sigma_0 + \sigma_1 \hat{y}_t, \tag{4}$$

where σ_t is the standard deviation at time t, σ_0 is the heteroscedasticity intercept, and σ_1 is the heteroscedasticity slope. The standard deviation σ_t has different meanings depending on whether heteroscedasticity or autocorrelation is first considered. This issue will be discussed later in this section. Correlation between errors is accounted for by assuming a first-order autoregressive model, AR(1).

If autocorrelation is treated before heteroscedasticity, the raw residuals are modeled following (Schoups & Vrugt, 2010)

$$e_t - \phi e_{t-1} = \sigma_t a_t \text{ with } a_t \sim \text{SEP}(\mu, 1, \xi, \beta), \tag{5}$$

where ϕ is the parameter of a first-order autoregressive polynomial, σ_t is the standard deviation of the autocorrelation-corrected residuals at time t, and a_t is an innovation from the corresponding SEP density, with mean μ ($\mu = 1$ for residual class 2 and $\mu = 0$ for residual class 3), unit variance, skewness parameter ξ, and kurtosis parameter β. Evin et al. (2013) demonstrated that applying an AR(1) model to the standardized residuals,

$$\eta_t = \frac{e_t}{\sigma_t}, \tag{6}$$

avoided the instability that may occur when the AR(1) model is applied to raw (heteroscedastic) residuals. Considering the reparameterization of this approach presented by Evin et al. (2014), equation (5) becomes

$$\eta_t - \phi \eta_{t-1} = a_t \text{ with } a_t \sim \text{SEP}(\mu, 1, \xi, \beta). \tag{7}$$

In this case, the variance of the standardized residuals is $1/(1 - \phi^2)$ and the standard deviation of equations (4) and (6) corresponds to $\sigma_t = \sqrt{1 - \phi^2} \sigma_e$, where σ_e is the standard deviation of raw residuals.

The modified log-likelihood function (L) that accounts for zero inflation, heteroscedasticity, autocorrelation, and non-Gaussian residuals is

$$
\begin{aligned}
L(\boldsymbol{\theta}|\mathbf{y}) = \ & n_1 \log \rho + n_2 \log(1-\rho) + n_2 \log \frac{2\sigma_{\xi_2}\omega_{\beta_2}}{\xi_2 + \xi_2^{-1}} - \sum_{t_2=1}^{n_2} \log \sigma_{t_2} - c_{\beta_2}\sum_{t_2=1}^{n_2} |a_{\xi_2, t_2}|^{2/(1+\beta_2)} + \\
& n_3 \log \frac{2\sigma_{\xi_3}\omega_{\beta_3}}{\xi_3 + \xi_3^{-1}} - \sum_{t_3=1}^{n_3} \log \sigma_{t_3} - c_{\beta_3}\sum_{t_3=1}^{n_3} |a_{\xi_3, t_3}|^{2/(1+\beta_3)},
\end{aligned} \tag{8}
$$

where $\boldsymbol{\theta} = \{\boldsymbol{\theta}_s, \boldsymbol{\theta}_e\}$ is the parameter set (composed of the deterministic model parameters $\boldsymbol{\theta}_s$ and the residual error model parameters $\boldsymbol{\theta}_e = \{\rho, \sigma_{t_2}, \sigma_{t_3}, \xi_2, \xi_3, \beta_2, \beta_3, \phi\}$), $\mathbf{y}$ is a vector with the observations (measurements of the observed system response), n_1 is the number of zero simulations with zero error, n_2 is the number of zero simulations with nonzero error, n_3 is the number of nonzero simulations with nonzero error, $a_{\xi_i, t_i} = \xi_i^{-sign(\mu_{\xi_i} + \sigma_{\xi_i} a_{t_i})} (\mu_{\xi_i} + \sigma_{\xi_i} a_{t_i})$ and $\omega_{\beta_i}, c_{\beta_i}, \mu_{\xi_i}$ and σ_{ξ_i} are calculated from β_i and ξ_i, where i is the index of the corresponding residual class, and are given in Appendix A of Schoups and Vrugt (2010; equations (A2), (A3), (A5) and (A6), respectively). The ZI-GL is reduced to GL as the number of zero simulations approaches zero.

Other heteroscedastic error models can be considered instead of the linear model presented herein. However, the implications of the transformation to the likelihood function should be carefully analyzed (see McInerney et al., 2017, for a comparison between heteroscedastic error models used in hydrologic studies).

2.3. Generation of Probabilistic Predictions

In a Bayesian framework, the posterior parameter distributions are used to create the predictive uncertainty that results from parameter uncertainty, i.e., samples from the posterior parameter distributions are used to run the deterministic model and generate the uncertainty related to parameter uncertainty. The probabilistic predictions are generated by adding to each deterministic simulation, at each time step, an innovation sampled from the inferred residual error model after "detransformation" to account for both heteroscedasticity and autocorrelation. For time steps with zero simulations, for each deterministic simulation, the binomial probability model (equation (2)) is used to determine whether the SEP density with parameters β_2 and ξ_2 should be used. The value of ρ calculated during model calibration is considered in both calibration and validation modes, since in predictive settings observations are not available to allow the computing of n_1

and n_2. For time steps with nonzero simulations, innovations are drawn from the standardized SEP density with parameters β_3 and ξ_3. The generation of samples from the corresponding SEP density is done following the simulation algorithm presented in Schoups and Vrugt (2010). When autocorrelation is considered before heteroscedasticity, innovations are multiplied by the standard deviation σ_t and subsequently equation (5) is used to create the residual time series. If heteroscedasticity is treated before autocorrelation, standardized residuals are generated by using equation (7) followed by the multiplication of the resulting time series by the standard deviation σ_t.

3. Empirical Case Study

3.1. Interception Model, Data, and Parameter Inference

We illustrate the applicability of the proposed zero-inflated Generalized Likelihood function with an interception modeling example. We used a simplified version of the Rutter model (Rutter et al., 1971) with a linear threshold model for the canopy drainage (Vrugt et al., 2003). This interception model requires as input gross rainfall and potential evaporation data and provides as output interception evaporation (I) and throughfall (Tf) estimates. The canopy water balance is calculated by

$$\frac{dS}{dt} = aP - D - I, \tag{9}$$

where S (mm) is the water storage in the canopy, a (dimensionless) is an interception efficiency parameter, P (mm h^{-1}) is the gross rainfall rate, D (mm h^{-1}) is the canopy drainage, and I (mm h^{-1}) is the interception evaporation rate. Canopy drainage is calculated as

$$D = \begin{cases} b(S-c) & \text{if } S > c \\ 0 & \text{otherwise} \end{cases}, \tag{10}$$

where b (h^{-1}) is a drainage parameter and c (mm) is the storage capacity. Throughfall rate is estimated as the sum of the canopy drainage and the proportion of rain falling directly on the ground,

$$Tf = (1-a)P + D, \tag{11}$$

and interception evaporation is assumed to increase linearly with the storage in the canopy,

$$I = dE_p \frac{S}{c}, \tag{12}$$

where d (dimensionless) is an evaporation efficiency parameter and E_p (mm h^{-1}) is the potential evaporation rate. The interception model has four parameters (a, b, c, and d). This model was implemented using a fixed-step implicit Euler scheme, with an integration time step of 1 min.

Hourly throughfall data from two interception plots were used for the identification of the interception model parameters. One plot is covered by a secondary native Atlantic forest (Mixed Ombrophilous Forest) and the other is a *Pinus taeda* plantation. Both plots are located in Southern Brazil. Hourly gross rainfall was measured in a nearby meteorological station and used as forcing in the interception model. Potential evaporation was assumed as being equal to 4 mm d^{-1} which was converted into hourly values by using a sinusoidal variation between 6 A.M. and 6 P.M. and constant rate in night hours. The available data series correspond to the period of 26 February 2014 to 6 October 2014 for the native forest plot and 23 August 2008 to 17 November 2008 for the pine plantation plot.

The data series were separated in rainfall events considering a rainless period preceding and following the event of 12 h. Rain events of less than 0.25 mm depth—which correspond to a single tip of the gauge—were discarded. Events that resulted in throughfall greater than gross precipitation were discarded as well. This procedure resulted in 59 and 35 events for the native forest and pine plantation plots, respectively. The events were sorted in ascending order of gross rainfall and the ones in even positions were assigned to the calibration data series and the remaining events produced the validation data series. The calibration and validation data series contain 51 and 40% (native forest) and 59 and 56% (pine plantation) of zero observations, respectively. Without considering zero observations, the proportion of data equal to measurement

Table 1
Assumptions of Each Residual Error Model Considered in This Study and Corresponding Number of Calibrated Parameters (D)

Formulation	Likelihood	Distribution	Heteroscedasticity	Correlation	Fixed parameters	D
Generalized	L1	Gaussian	Homoscedastic	Independent	$\sigma_1 = 0, \beta = 0, \xi = 1, \phi = 0$	5
Likelihood function	L2	Gaussian	Heteroscedastic	Independent	$\beta = 0, \xi = 1, \phi = 0$	6
	L3	Skewed exponential power	Heteroscedastic	Independent	$\xi = 1, \phi = 0$	7
	L4/L5[a]	Skewed exponential power	Heteroscedastic	Correlated	$\xi = 1$	8
Zero-inflated	L1	Skewed exponential power	Homoscedastic	Independent	$\sigma_1 = 0, \beta = 0, \phi = 0$	7
Generalized	L2	Skewed exponential power	Heteroscedastic	Independent	$\beta = 0, \phi = 0$	8
Likelihood function[b]	L3	Skewed exponential power	Heteroscedastic	Independent	$\phi = 0$	9
	L4/L5[a]	Skewed exponential power	Heteroscedastic	Correlated		10

[a]For L4 the AR(1) model is applied to raw residuals (e_t) and for L5 it is applied to standardized residuals ($\eta_t = e_t/\sigma_t$). [b]For the ZI-GL approach, the presented assumptions correspond to residuals from class 3. Residuals from class 2 are considered homoscedastic and are modeled with a skewed exponential power distribution with fixed parameters $\beta = 1$ and $\xi = 10$.

resolution is 7 and 4% (native forest) and 44 and 29% (pine plantation) for the calibration and validation data series, respectively.

The differential evolution adaptive Metropolis (DREAM) algorithm (Vrugt, 2016; Vrugt et al., 2008, 2009) was used for parameter inference and uncertainty estimation. DREAM requires the setup of some parameters that depend on the case study: the dimension of the problem D, i.e., the number of parameters to be estimated (interception model parameters + error model parameters); the number of Markov chains N; and the number of generations T. In this study, we set $N = \max(10; 2D)$ and $T = 3,000$. When necessary, T was increased until convergence to a stationary distribution was achieved. Additionally, the user must choose a likelihood function to evaluate the model residuals and define a prior distribution for each parameter. The proposed modification of the Generalized Likelihood function was tested against its original formulation introduced by Schoups and Vrugt (2010), considering different levels of complexity for the residual error model (Table 1). The error models included in this analysis were chosen in an iterative manner, where the increase in complexity from one error model to another was done after verifying the inability of the preceding model to meet one or more assumptions made about the model residuals. The use of both the original Generalized Likelihood function formulation (GL) and its corresponding zero-inflated version (ZI-GL) with each error model diverge from the likelihood selection framework presented by Smith et al. (2015). Following that framework, the potential for zero inflation should be verified as a first step and would likely conduct the likelihood selection process toward a zero-inflated likelihood. However, maintaining the two

Table 2
Interception and Error Model Parameters Specifications

| Parameter		Description	Minimum	Maximum	Unit
Interception model	a	Interception efficiency	0.1	1	
	b	Drainage parameter	1.0	1,000	d^{-1}
	c	Storage capacity	0.1	10	mm
	d	Evaporation efficiency	0.1	5	
Generalized	σ_0	Heteroscedasticity intercept	0	50	mm d^{-1}
likelihood function	σ_1	Heteroscedasticity slope	0	1	
	β	Kurtosis parameter	-1	1	
	ϕ	Autocorrelation coefficient	0	1	
Zero-inflated generalized	$\sigma_{0,2}$	Standard deviation for residual class 2	0	50	mm d^{-1}
likelihood function	$\sigma_{0,3}$	Heteroscedasticity intercept for residual class 3	0	50	mm d^{-1}
	$\sigma_{1,3}$	Heteroscedasticity slope for residual class 3	0	1	
	β_3	Kurtosis parameter for residual class 3	-1	1	
	ξ_3	Skewness parameter for residual class 3	0.1	10	
	ϕ	Autocorrelation coefficient	0	1	
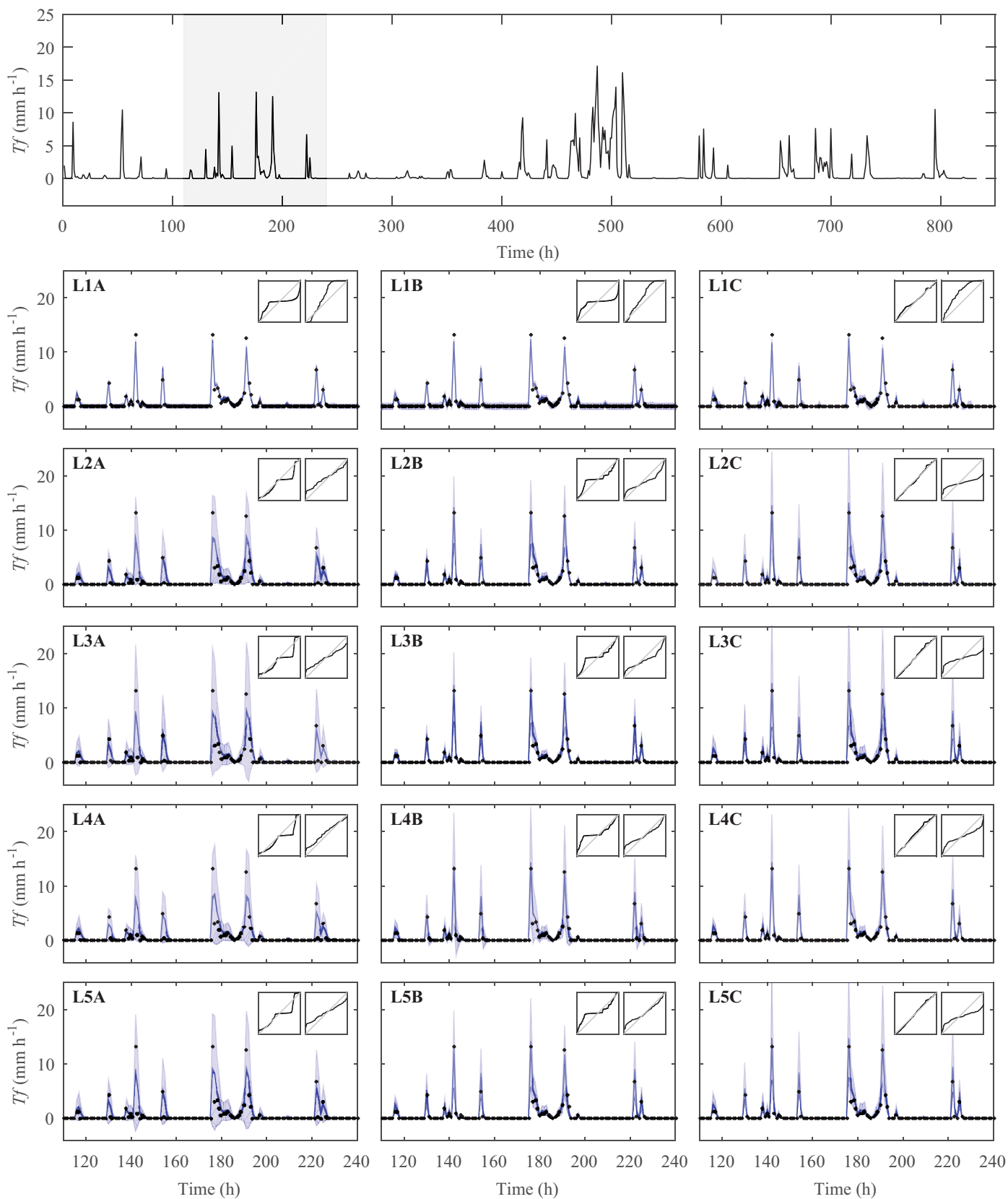
Figure 1. (continued)

formulations allowed us to separate the effects resulting from increasing residual model complexity from those caused by accounting for zero inflation of the error distribution. We also compared the ZI-GL with the results obtained by simply suppressing zero observations from the data series when using the original GL formulation (GL$^{y>0}$). In this case, we simply omitted all residuals (zero and nonzero) from zero observations from the calculation of the likelihood function value. Heteroscedasticity was considered by assuming a linear model for the error variance and a first-order autoregressive model AR(1) was used to account for residual autocorrelation. The AR(1) model was applied to the entire series. The ZI-GL formulation requires the specification of two SEP densities. For the residual class 2 (nonzero residuals from zero simulations), we fixed β_2 to a value of 1 and ξ_2 to a value of 10. For the residual class 3 (nonzero residuals from nonzero simulations), the SEP parameters β_3 and ξ_3 were calibrated with the interception model parameters and the remaining error model parameters. In the ZI-GL approach, heteroscedasticity was considered only for residuals from class 3. A uniform prior for each parameter was used with ranges specified in Table 2.

The last 3,000 parameter sets generated with the DREAM algorithm were used to represent the parameter uncertainty and to create the probabilistic predictions as described in section 2.3. The performance of each error model was assessed using different metrics. The reliability of the probabilistic predictions was evaluated by visual inspection of the predictive QQ plot (Thyer et al., 2009), which compares the empirical cumulative distribution function (cdf) of the sample of p values $F_{\hat{y}(t)}(y_t)$ with the cdf of a uniform distribution, and quantified using the reliability metric (Evin et al., 2014; McInerney et al., 2017),

$$\text{Reliability}[\hat{\mathbf{y}}, \mathbf{y}] = \frac{2}{N} \sum_{t=1}^{N} \left| F_U \left[F_{\hat{y}(t)}(y_t) \right] - F_\Omega \left[F_{\hat{y}(t)}(y_t) \right] \right|, \tag{13}$$

where $F_{\hat{y}(t)}$ is the cdf of the predictive distribution at time t and y_t is the observation, F_U is the cdf of the uniform distribution $U(0,1)$, and F_Ω is the empirical cdf. The cdf of the predictive distribution is discontinuous at $y_t = 0$. For that reason, the p values corresponding to zero observations were considered as a random value between the p value just before $y_t = 0$ and the p value just after $y_t = 0$. If the observations are samples of the predictive distribution, the predictive QQ plot follows the 1:1 line and the reliability metric is zero. The precision is related to the width of the probabilistic predictions and was measured by the precision metric (McInerney et al., 2017),

$$\text{Precision}[\hat{\mathbf{y}}, \mathbf{y}] = \frac{1}{N} \sum_{t=1}^{N} \text{sdev } \hat{\mathbf{y}}_t \bigg/ \frac{1}{N} \sum_{t=1}^{N} y_t, \tag{14}$$

where sdev $\hat{\mathbf{y}}_t$ is the standard deviation of the probabilistic predictions at time step t. The overall capability of the model to simulate the water balance is quantified by the volumetric bias metric (McInerney et al., 2017),

$$\text{Bias}[\hat{\mathbf{y}}, \mathbf{y}] = \left| \frac{\sum_{t=1}^{N} y_t - \sum_{t=1}^{N} \hat{y}_{t,\text{mean}}}{\sum_{t=1}^{N} y_t} \right|, \tag{15}$$

where $\hat{y}_{t,\text{mean}}$ is the mean of the predictions at time step t. For all three metrics, zero indicates perfect performance. For a more detailed interpretation of the above metrics, the reader is encouraged to refer to the cited references. Similar metrics were also used in other hydrological modeling studies (e.g., Renard et al., 2010; Thyer et al., 2009).

3.2. Predictive Uncertainty

Compared to the original Generalized Likelihood function (GL), the proposed zero-inflated version of the GL (ZI-GL) improved the quality of the predictive uncertainty for all error models as measured by the precision,

Figure 1. Predictive uncertainty derived using each of the five error models (L1–L5) for the native forest case study. The observed throughfall series of the entire validation period is presented in the top plot. The predictions using the original Generalized Likelihood function formulation (GL), by simply suppressing residuals from zero observations from original Generalized Likelihood function (GL$^{y>0}$) and using the zero-inflated version of the Generalized Likelihood function (ZI-GL) are presented in the left, middle, and right columns, respectively. In order to make the figure clearer, only a portion of the validation period is shown (gray shaded area of the top plot). The dots correspond to the throughfall observations. The light and dark shaded areas indicate the 90% total predictive uncertainty and the uncertainty that results from parameter uncertainty, respectively. The performance metrics were calculated for the entire validation period. The insets in the upper right corner of each plot displays the predictive QQ plot for observed throughfall lower (left) and greater (right) than 2 mm h^{-1}. If the predictive uncertainty is appropriate, then the black line should be close to the 1:1 gray line.

reliability, and volumetric bias metrics (which can be observed by comparing the first and third columns of Figure 1 and the corresponding metrics in Figures 2a and 2c). The poor performance obtained for the GL is evidenced by visual inspection of the predictive uncertainty, which assumes even negative values (without any physical meaning) (first column of Figure 1). Other studies also encountered this same problem when modeling the hydrological behavior of dry basins (Evin et al., 2013; Schoups & Vrugt, 2010). Considering a skewed density when using GL avoided these negative predictive bounds; however, the predictive uncertainty remained large and therefore meaningless (results not shown). Omitting the residuals from zero observations ($GL^{y>0}$) resulted in better behaved uncertainty bounds compared to the GL approach (first and second columns of Figure 1). When the $GL^{y>0}$ was used, only the error model L4 resulted in significant
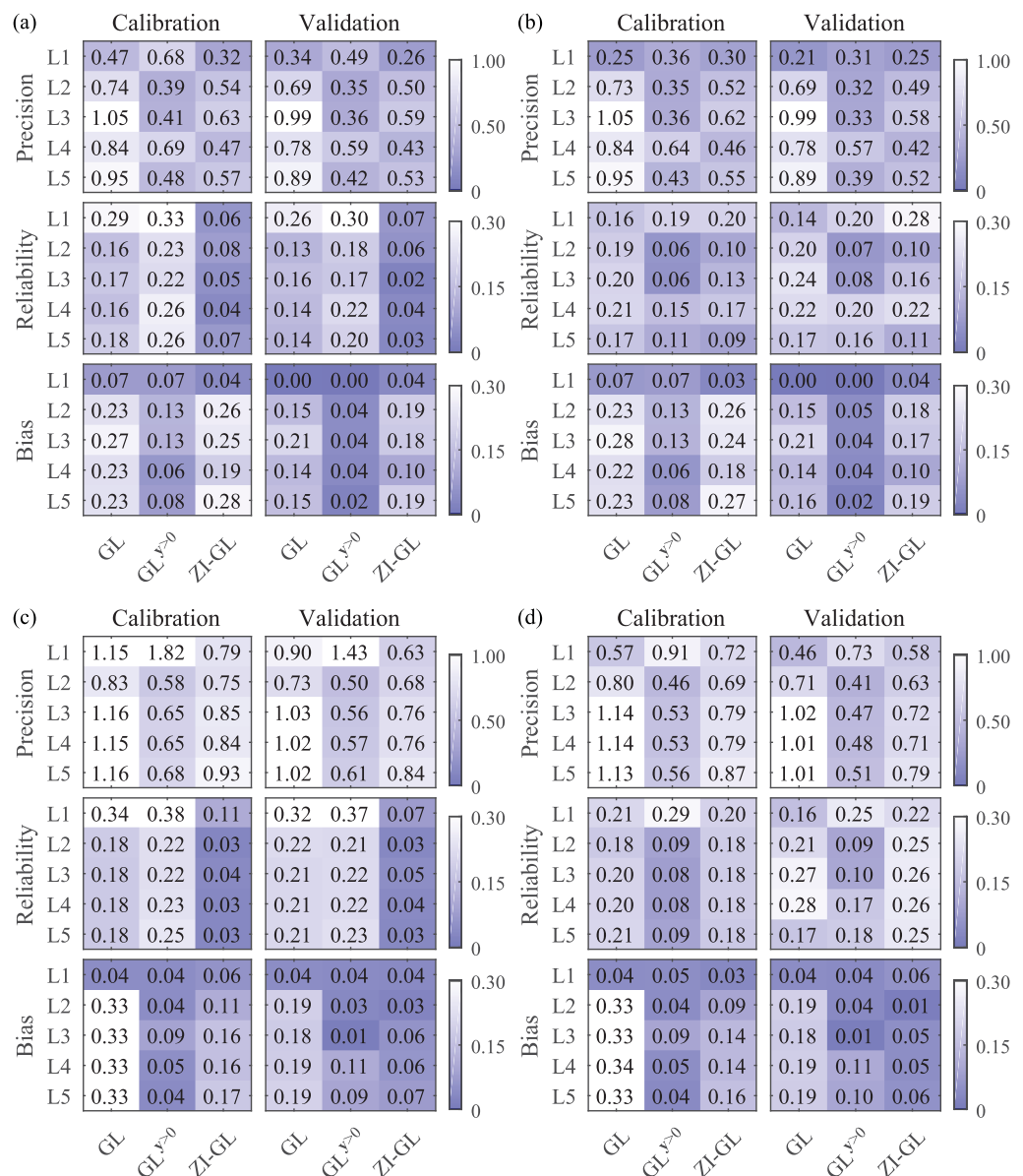

Figure 2. Performance metrics for uncertainty ranges derived using each of the five error models (L1–L5) in the original Generalized Likelihood function (GL), by simply suppressing zero observations from the data series ($GL^{y>0}$) and using the zero-inflated version of the Generalized Likelihood function (ZI-GL). The metrics are presented for (a) the native forest case study considering the entire time series, (b) the native forest case study considering only time steps with nonzero residuals, (c) the pine plantation case study considering the entire time series, and (d) the pine plantation case study considering only time steps with nonzero residuals. For all three metrics, zero indicates perfect performance.

negative values for the predictive uncertainty. The large uncertainty bound obtained with that error model was a result of the relatively high inferred value for the autoregressive model parameter (maximum likelihood value for ϕ of 0.36 for the native forest case study) combined with the use of raw residuals in the AR(1) model. As demonstrated in *Evin* et al. (2013), removing autocorrelation of raw instead of standardized residuals may result in a wide and poor predictive uncertainty. Removing heteroscedasticity before autocorrelation (error model L5), as suggested in the aforementioned study, avoided the occurrence of negative bounds, which can be observed by comparing the predictive uncertainty for L4 and L5 in the second column of Figure 1. For the native forest (and pine plantation), the use of the error model L5 with the GL (GL$^{y>0}$) accounted for the high autocorrelation between standardized residuals, with maximum likelihood value for ϕ greater than 0.30. As a consequence, larger differences in the performance metrics values between L4 and L5 were obtained in these situations.

The poor performance of the GL for the error models L2 to L5 may be a consequence of the negative impact of residuals from near zero simulations on the likelihood quantification when the linear heteroscedastic error model is considered, which is acknowledged in many hydrological modeling studies (e.g., Evin et al., 2013; Westra et al., 2014). The improvement in the quality of the predictive uncertainty obtained when the ZI-GL approach is used illustrates the ability of this method to at least partially handle this problem. In the ZI-GL, nonzero residuals from zero simulations are modeled separately and, therefore, these residuals are removed from the estimation of the linear heteroscedastic error model. Since GL$^{y>0}$ simply omit the residuals from zero observations from the computation of the likelihood value, a fewer number of residuals from near zero simulations is considered in this method. As a result, except for L4 whose poor predictive bounds was discussed before, the precision and volumetric bias metrics obtained with GL$^{y>0}$ were lower than using the ZI-GL approach (darker blue for GL$^{y>0}$ compared to ZI-GL in Figure 2a). By avoiding the increase in the contribution to the likelihood of residuals from near zero simulations, which can assume overly large values after their standardization using the heteroscedastic error model, greater weight is given to residuals from large simulated values, which have a higher impact on the quantification of the precision and bias metrics. The ZI-GL provided better results than the GL$^{y>0}$ considering the reliability metric (darker blue for ZI-GL compared to GL$^{y>0}$ in Figures 2a and 2c), since the incorporation of a zero-inflation approach results in a better estimate of the predictive uncertainty for zero residuals. However, when only time steps with nonzero residuals are considered, the reliability metric is lower for the GL$^{y>0}$ method (darker blue for GL$^{y>0}$ compared to ZI-GL in Figures 2b and 2d).

In order to reduce the negative impact of residuals from zero and near zero simulations on the likelihood calculation, we set the heteroscedasticity intercept parameter to a fixed value (= 0.12 mm h^{-1}) and performed the same analysis of the predictive uncertainty presented herein (the results are presented as supporting information). When the heteroscedasticity intercept parameter is fixed, the performance metrics for the three approaches are much more similar. Considering the entire time series, lower reliability values are obtained for the ZI-GL, since this approach is the only one that do not overestimate the predictive uncertainty for zero residuals. The precision metric is also better for ZI-GL. For GL and GL$^{y>0}$, the residual standard deviation for all residuals from zero simulations is equal to the fixed value of the heteroscedasticity intercept, which results in an overestimated uncertainty for these time steps. The worst reliability for ZI-GL when only time steps with nonzero residuals are considered is a consequence of the use of the "inflated SEP" for all residuals from zero simulations, which ultimately results in p values greater than ρ for time steps corresponding to residuals from class 2 (nonzero residuals from zero simulations).

The error model that assumes homoscedastic, Gaussian, and independent residuals (L1) resulted in better estimates of the total simulated volume (lower volumetric bias metric for the mean predictive throughfall values—darker blue for L1 in Figures 2a and 2c). Homoscedastic error model gives equal weight to all raw residuals, which are in general larger for higher simulated values. Since the peaks correspond to the largest proportion of simulated throughfall, the total volume is expected to be better simulated when a likelihood that produces a better fit for higher values is used. The same result is expected to occur in very responsive catchments which respond rapidly to precipitation. For example, in the hydrological modeling of a semiarid Australian catchment where a high proportion of the annual flow occurs in a short period of the year, Westra et al. (2014) found a correspondence between models ranked based on the Nash-Sutcliffe coefficient of efficiency and the total annual flow error. However, this error model resulted in an overestimated uncertainty for
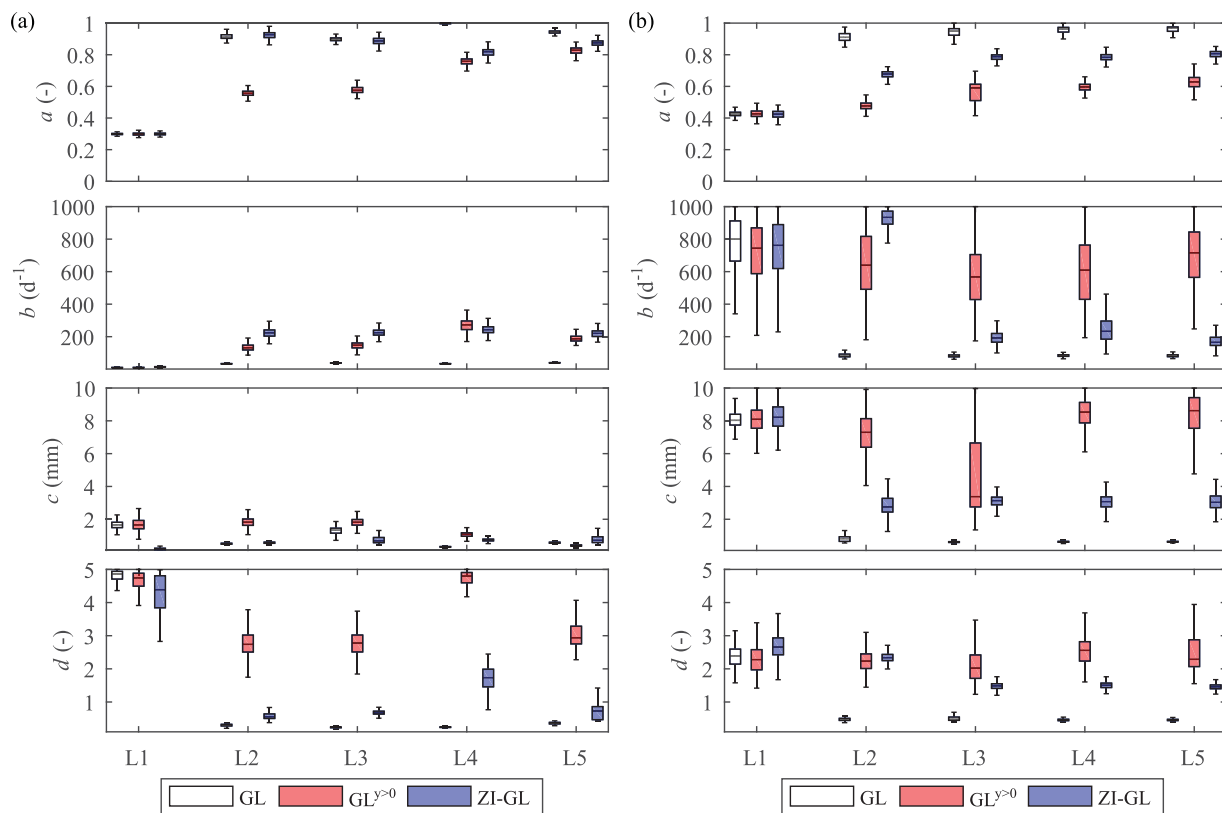
Figure 3. Box plots of the posterior parameter distributions derived using each of the five error models (L1–L5) using the Generalized Likelihood function (GL), by simply suppressing zero observations from the data series (GL$^{y>0}$) and using the zero-inflated version of the Generalized Likelihood function (ZI-GL) for (a) the native forest case study and (b) the pine plantation case study. The y axis limits were set to the prior range of each parameter.

low throughfall values and a systematic underprediction of higher values, as a consequence of the narrow estimated predictive uncertainty (predictive QQ plots corresponding to L1 on the insets of the Figure 1).

3.3. Parameter Inference

The choice of the error model had a large influence on the posterior parameter distribution (Figure 3). For the first error model (L1), the posterior parameter estimates were very similar for the original GL (GL), by simply suppressing zero observations from the data series (GL$^{y>0}$) and using the zero-inflated version of the Generalized Likelihood function (ZI-GL). Two main shifts in parameter values are evident: (1) when changing from homoscedastic (L1) to heteroscedastic model (L2–L5) (compare boxplots for L1 with L2–L5 in Figure 3); and (2) when changing from the GL to the GL$^{y>0}$ and to the ZI-GL when heteroscedasticity is considered (compare white with red and blue boxplots in Figure 3). Both shifts are a consequence of the different relative contribution of residuals from near zero simulations to the likelihood value. Therefore, distinct weights are given to processes and corresponding parameters related to these low values depending on the choice of the error model. When the heteroscedasticity intercept parameter was set to a fixed value of 0.12 mm h^{-1}, the differences between the posterior parameter estimates obtained for GL, GL$^{y>0}$ and ZI-GL were greatly reduced (supporting information Figure S3).

3.4. Residual Error Diagnostics

The transformed residual errors (innovations) were better represented by the assumed distribution with the consideration of a SEP density, autocorrelation, and when zero residuals were separately handled (GL$^{y>0}$ or ZI-GL) (L4 and L5 of GL$^{y>0}$ and ZI-GL in Figure 4a). For the pine plantation case study, the transformed residual errors deviate more strongly from the assumed distribution (results not shown). We hypothesize that this is a limitation caused by the data measurement resolution, which was 10 times greater than for the native forest case study. The data series collected in the pine plantation has a high proportion of throughfall data equal to the measurement resolution (section 3.1).
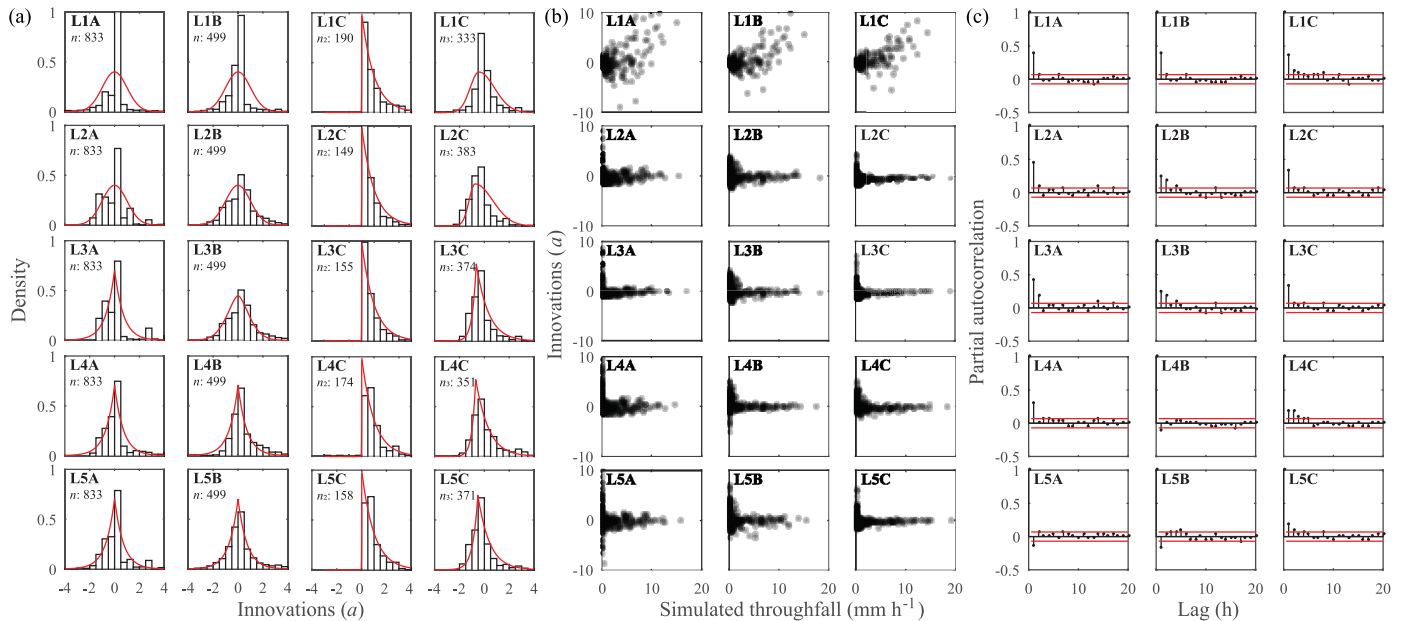
Figure 4. Model error diagnostic for each of the five error models (L1–L5) considering the original Generalized Likelihood function formulation (A), by simply suppressing zero observations from the data series (B) and using the zero-inflated version of the Generalized Likelihood function (C) for the native forest case study (validation period only). This diagnostic was performed using the maximum likelihood parameter set. From left to right: (a) histograms of the innovations (a_t), where the red line indicates the theoretical distribution; (b) innovations (a_t) as a function of simulated throughfall; and (c) the partial autocorrelation function of the innovations (a_t), where the red lines indicate the 95% significance levels. In the histograms of the ZI-GL approach, the distribution of residuals from class 2 (nonzero residuals from zero simulations) is presented on the left and the distribution of residuals from class 3 (nonzero residuals from nonzero simulations) is presented on the right. Partial autocorrelation was calculated for the entire data series.

The hypothesis of homoscedastic residuals was clearly violated (first row of Figure 4b). However, the consideration of a linear model for the error variance (L2–L5) resulted in large values for the transformed residuals of near zero simulations, which could have had a deteriorating impact on the likelihood quantification (as discussed in the section 3.2). The use of even more complex residual models, with the incorporation of autocorrelation (L4 and L5), exacerbated this problem. Furthermore, although reducing the lag 1 autocorrelation as compared to L2 and L3, the inclusion of an AR(1) model was not enough to remove residual correlation (plots for L4 and L5 in Figure 4c). Setting the heteroscedasticity intercept parameter to a fixed value of 0.12 mm h^{-1} prevented the occurrence of overly large transformed residuals from near zero simulations, especially for the GL and ZI-GL formulation (supporting information Figure S4b).

4. Conclusions and Recommendations

In this study, we introduced a modification of the Generalized Likelihood function of Schoups and Vrugt (2010) that accounts for zero inflation of the error distribution (ZI-GL). We follow the zero-inflation approach of Smith et al. (2010) except that the residual time series is divided into zero and nonzero states based on simulated instead of observed values. Conditioning the residual error model on simulated values allows its use in prediction, where observations are not available.

The use of the ZI-GL provided better uncertainty estimates relative to the GL as measured by the precision, reliability, and volumetric bias metrics. These results are consistent with those presented by Smith et al. (2010), which illustrated the positive effects of using a zero-inflated error model for parameter inference and predictive uncertainty estimate in the hydrological modeling of ephemeral catchments. We hypothesize that the poor performance of GL was a consequence of the negative impact on the likelihood quantification of residuals from near zero simulations, which assume overly large values after transformation using the heteroscedastic error model. This problem was also evident by visual inspection of the transformed residuals as a function of simulated values, which indicates that the linear heteroscedastic error model was not well performing for near zero observations. The need for special handling of these low values was also

acknowledged in other studies (e.g., Evin et al., 2013, 2014; Westra et al., 2014). The better performance of ZI-GL compared to GL illustrates the ability of the proposed likelihood to at least partially deal with this problem. However, the lower precision and bias metric values for the $GL^{y>0}$ method, in which the residuals from zero observations are omitted from the likelihood function, indicates that the residuals from near zero simulations are still a problem and requires further investigation.

Setting the heteroscedasticity intercept parameter to a fixed value (higher compared to the one obtained when this parameter was calibrated) has proven to be an easy alternative to reduce the weight of residuals from near zero simulations on the likelihood quantification. In this case, the three formulations (GL, $GL^{y>0}$, and ZI-GL) presented very similar results with the ZI-GL being able to provide a better predictive uncertainty for lower values.

The GL has flexibility to deal with a wide range of residual errors characteristics and, we expect the results presented in this study to increase its applicability to zero-inflated data series. Aside from the interception case studies presented in this paper, the ZI-GL may be useful for other hydrological and environmental modeling studies dealing with zero-inflated error distributions. Application of the ZI-GL to other situations is therefore highly desirable to confirm its usefulness.

References

Cheng, Q. B., Chen, X., Xu, C. Y., Reinhardt-Imjela, C., & Schulte, A. (2014). Improvement and comparison of likelihood functions for model calibration and parameter uncertainty analysis within a Markov chain Monte Carlo scheme. *Journal of Hydrology*, *519*, 2202–2214. https://doi.org/10.1016/j.jhydrol.2014.10.008

Evin, G., Kavetski, D., Thyer, M., & Kuczera, G. (2013). Pitfalls and improvements in the joint inference of heteroscedasticity and autocorrelation in hydrological model calibration. *Water Resources Research*, *49*, 4518–4524. https://doi.org/10.1002/wrcr.20284

Evin, G., Thyer, M., Kavetski, D., McInerney, D., & Kuczera, G. (2014). Comparison of joint versus postprocessor approaches for hydrological uncertainty estimation accounting for error autocorrelation and heteroscedasticity. *Water Resources Research*, *50*, 2350–2375. https://doi.org/10.1002/2013WR014185

Honti, M., Stamm, C., & Reichert, P. (2013). Integrated uncertainty assessment of discharge predictions with a statistical error model. *Water Resources Research*, *49*, 4866–4884. https://doi.org/10.1002/wrcr.20374

Kavetski, D., Kuczera, G., & Franks, S. W. (2006). Bayesian analysis of input uncertainty in hydrological modeling: 1. Theory. *Water Resources Research*, *42*, W03407. https://doi.org/10.1029/2005WR004368

Koskela, J. J., Croke, B. W. F., Koivusalo, H., Jakeman, A. J., & Kokkonen, T. (2012). Bayesian inference of uncertainties in precipitation-streamflow modeling in a snow affected catchment. *Water Resources Research*, *48*, W11513. https://doi.org/10.1029/2011WR011773

Li, M., Wang, Q. J., & Bennett, J. (2013). Accounting for seasonal dependence in hydrological model errors and prediction uncertainty. *Water Resources Research*, *49*, 5913–5929. https://doi.org/10.1002/wrcr.20445

Li, M., Yang, D., Chen, J., & Hubbard, S. S. (2012). Calibration of a distributed flood forecasting model with input uncertainty using a Bayesian framework. *Water Resources Research*, *48*, W08510. https://doi.org/10.1029/2010WR010062

McInerney, D., Thyer, M., Kavetski, D., Lerat, J., & Kuczera, G. (2017). Improving probabilistic prediction of daily streamflow by identifying Pareto optimal approaches for modeling heteroscedastic residual errors. *Water Resources Research*, *53*, 2199–2239. https://doi.org/10.1002/2013WR014979

Renard, B., Kavetski, D., Kuczera, G., Thyer, M., & Franks, S. W. (2010). Understanding predictive uncertainty in hydrologic modeling: The challenge of identifying input and structural errors. *Water Resources Research*, *46*, W05521. https://doi.org/10.1029/2009WR008328

Renard, B., Kavetski, D., Leblois, E., Thyer, M., Kuczera, G., & Franks, S. W. (2011). Toward a reliable decomposition of predictive uncertainty in hydrological modeling: Characterizing rainfall errors using conditional simulation. *Water Resources Research*, *47*, W11516. https://doi.org/10.1029/2011WR010643

Rutter, A. J., Kershaw, K. A., Robins, R. C., & Morton, A. J. (1971). A predictive model of rainfall interception in forests. 1. Derivation of the model from observations in a plantation of Corsican pine. *Agricultural Meteorology*, *9*, 367–384.

Schaefli, B., Talamba, D. B., & Musy, A. (2007). Quantifying hydrological modeling errors through a mixture of normal distributions. *Journal of Hydrology*, *332*(3–4), 303–315. https://doi.org/10.1016/j.jhydrol.2006.07.005

Schoups, G., & Vrugt, J. A. (2010). A formal likelihood function for parameter and predictive inference of hydrologic models with correlated, heteroscedastic, and non-Gaussian errors. *Water Resources Research*, *46*, W10531. https://doi.org/10.1029/2009WR008933

Smith, T., Marshall, L., & Sharma, A. (2015). Modeling residual hydrologic errors with Bayesian inference. *Journal of Hydrology*, *528*, 29–37. https://doi.org/10.1016/j.jhydrol.2015.05.051

Smith, T., Sharma, A., Marshall, L., Mehrotra, R., & Sisson, S. (2010). Development of a formal likelihood function for improved Bayesian inference of ephemeral catchments. *Water Resources Research*, *46*, W12551. https://doi.org/10.1029/2010WR009514

Thyer, M., Renard, B., Kavetski, D., Kuczera, G., Franks, S. W., & Srikanthan, S. (2009). Critical evaluation of parameter consistency and predictive uncertainty in hydrological modeling: A case study using Bayesian total error analysis. *Water Resources Research*, *45*, W00B14. https://doi.org/10.1029/2008WR006825

Vrugt, J. A. (2016). Markov chain Monte Carlo simulation using the DREAM software package: Theory, concepts, and MATLAB implementation. *Environmental Modelling & Software*, *75*, 273–316. https://doi.org/10.1016/j.envsoft.2015.08.013

Vrugt, J. A., Dekker, S. C., & Bouten, W. (2003). Identification of rainfall interception model parameters from measurements of throughfall and forest canopy storage. *Water Resources Research*, *39*(9), 1251. https://doi.org/10.1029/2003WR002013

Vrugt, J. A., ter Braak, C. J. F., Clark, M. P., Hyman, J. M., & Robinson, B. A. (2008). Treatment of input uncertainty in hydrologic modeling: Doing hydrology backward with Markov chain Monte Carlo simulation. *Water Resources Research*, *44*, W00B09. https://doi.org/10.1029/2007WR006720

Vrugt, J. A., ter Braak, C. J. F., Diks, C. G. H., Robinson, B. A., Hyman, J. M., & Higdon, D. (2009). Accelerating Markov chain Monte Carlo simulation by differential evolution with self-adaptive randomized subspace sampling. *International Journal of Nonlinear Sciences and Numerical Simulation*, *10*(3), 273–290. https://doi.org/10.1515/IJNSNS.2009.10.3.273

Wang, Q. J., & Robertson, D. E. (2011). Multisite probabilistic forecasting of seasonal flows for streams with zero value occurrences. *Water Resources Research*, *47*, W02546. https://doi.org/10.1029/2010WR009333

Westra, S., Thyer, M., Leonard, M., Kavetski, D., & Lambert, M. (2014). A strategy for diagnosing and interpreting hydrological model nonstationarity. *Water Resources Research*, *50*, 5090–5113. https://doi.org/10.1002/2013WR014719