Water Resources Research

RESEARCH ARTICLE
10.1029/2022WR032263

Key Points:

- We present a model-agnostic method for the generation of replicates of discharge records
- The nonparametric estimator provides an accurate estimate of random errors, which for the investigated discharge records are rather small
- The replicates portray accurately the assigned streamflow uncertainty and preserve statistical and hydrologic properties

Supporting Information:

Supporting Information may be found in the online version of this article.

Correspondence to:

J. A. Vrugt,
jasper@uci.edu

Citation:

de Oliveira, D. Y., & Vrugt, J. A. (2022). The treatment of uncertainty in hydrometric observations: A probabilistic description of streamflow records. *Water Resources Research*, *58*, e2022WR032263. https://doi.org/10.1029/2022WR032263

Received 17 MAR 2022
Accepted 31 OCT 2022

The Treatment of Uncertainty in Hydrometric Observations: A Probabilistic Description of Streamflow Records

Debora Y. de Oliveira[1] and Jasper A. Vrugt[1,2]

[1]Department of Civil and Environmental Engineering, University of California Irvine, Irvine, CA, USA, [2]Department of Earth System Science, University of California Irvine, Irvine, CA, USA

Abstract In this paper, we introduce a relatively simple data-driven method for the representation of the uncertainty in daily discharge records. The proposed method relies only on hourly discharge data and takes advantage of a nonparametric difference-based estimator in the characterization of random errors in discharge time series. We illustrate with corrupted streamflow data that the nonparametric estimator provides an accurate characterization of the nature (homoscedastic or heteroscedastic) and magnitude of these errors. In addition, we demonstrate the practical usefulness of the estimator using discharge time series of 500+ watersheds of the Catchment Attributes and MEteorology for Large-sample Studies data set. This analysis reveals that the magnitude of errors of aleatory nature in the investigated discharge records is rather small (less than 3% for 80% of the records). We then combine the effect of random errors and measurement frequency into a daily variance estimate, which serves as input to a streamflow generation approach. This procedure produces replicates of the discharge record which portray accurately the assigned streamflow uncertainty, preserve key statistical properties of the discharge record and are hydrologically realistic. The proposed method facilitates Bayesian analysis and supports tasks such as model diagnostics, data assimilation, uncertainty quantification and regionalization.

1. Introduction and Scope

River discharge serves many different purposes and is recorded automatically in many watersheds worldwide to support water supply, use and management decisions, hydroelectric generation and water quality assessment. In hydrology, streamflow records may be used to improve our understanding of hydrologic processes (Tetzlaff et al., 2017), regimes (Olden & Poff, 2003), extremes (Slater et al., 2021) and the changes thereof (Magilligan & Nislow, 2005), hydrologic alteration (Richter et al., 1996) and nonstationarity (Blöschl, 2017; Slater et al., 2021). Furthermore, the monitoring of river discharge is of crucial importance for (among others) drought and flood prediction (Brunner et al., 2021) and calibration of satellite-based streamgaging methods (Bjerklie et al., 2018).

Discharge records are subject to uncertainty and its estimation can reduce economic costs of water management decisions (H. McMillan et al., 2017) and allows robust research conclusions to be drawn, as demonstrated in data assimilation (Burgers et al., 1998), regionalization (Westerberg et al., 2016), and Bayesian model selection (Reuschen et al., 2021). Discharge time series are usually obtained by relating continuously measured river stage to discharge through so-called rating curves. Thus, uncertainty in discharge records arises due to measurement errors in the stage time series and also given that the rating curve used to transform stage into discharge is not uniquely defined and will vary over time at almost every gauging station in response to changing hydraulic characteristics of the stream channel and floodplain (see the review on sources of river discharge uncertainty by H. McMillan et al. (2012) and references therein).

Many methods were developed to estimate rating curve uncertainty (e.g., Coxon et al., 2015; H. K. McMillan & Westerberg, 2015; Le Coz et al., 2014, to name a few) and large differences between uncertainty estimates can be obtained depending on how each source of uncertainty is considered (Kiang et al., 2018). The propagation of errors in the stage time series into discharge records is often not accounted for. An exception is the work by Horner et al. (2018), which proposed an approach to evaluate the impact of stage errors, decomposed into aleatory and systematic errors, on streamflow uncertainty. Aleatory errors in stage records are due to waves and instrumental noise and systematic errors originate from instrument biases and calibration drifts over time (Horner et al., 2018).

Metadata on gauge information, stage-discharge measurements and rating curve estimation are essential in helping to contextualize and interpret streamflow records (e.g., Hannah et al., 2011), yet, such data are often lacking in large-sample hydrology data sets (Addor et al., 2020). This prevents users from assessing streamflow data reliability (see also Boldetti et al., 2010). In the absence of detailed information on the magnitude of the different sources of discharge uncertainty, at least some of them need to be specified based on expert knowledge or literature information. However, discharge uncertainty depends on the specific site considered (Coxon et al., 2015; Di Baldassarre & Montanari, 2009; Horner et al., 2018; Westerberg et al., 2016) and, as such, site-specific uncertainty estimates should be derived. Moreover, accurate error estimates are important to avoid compensation between the different sources (Horner et al., 2018).

The magnitude of aleatory errors in streamflow records (that originate from aleatory errors in stage time series) can be quantified using so-called error variance estimation methods. Such methods originate from nonparametric regression (Anderson, 1971; Hall et al., 1990; Zhou et al., 2015) and may be divided into two main classes. Estimators of the first class estimate the variance based on the sum of squared residuals from a nonparametric fit. This class includes, for example, spline smoothing (e.g., Silverman, 1985; Wahba, 1978) and kernel-based (e.g., Hall & Marron, 1990; Müller & Stadtmüller, 1987) estimators. The main drawback of this type of estimators is that they require the definition of a smoothing parameter. To overcome this issue, Garcia (2010) proposed a fully automated smoothing procedure, providing an efficient smoother for numerous applications in the area of data analysis. The second class consists of difference-based methods. Within this class of estimators, Rice (1984) introduced a first-order differencing estimator, Gasser et al. (1986) proposed a second-order difference-based estimator, and Hall et al. (1990) presented the kth-order difference-based estimator with k a fixed integer. The difference-based estimators of Rice (1984), Gasser et al. (1986), and Hall et al. (1990) consider a constant variance. Modifications to the estimator of Hall et al. (1990) to extend its applicability to nonconstant error variance were presented by Vrugt et al. (2005) and Zhou et al. (2015). Difference-based estimators do not require the estimation of a smoothing constant, yet, the user needs to specify the order of the difference-based estimator.

In this paper, we revisit the nonparametric estimator of Vrugt et al. (2005) and analyze its practical applicability to the estimation of random errors in streamflow records. We provide guidelines into the selection of the order of the difference-based estimator and a MATLAB implementation. We believe this will lower the threshold for others to consider implementing this approach. The estimates of random errors are combined with the effect of measurement frequency and serve as input to a relatively simple data-driven method for the generation of replicates of discharge records. Our aim is to generate discharge time series that are as plausible as the original record. The proposed method relies on hourly discharge time series only and produces replicates which portray accurately the assigned streamflow uncertainty, preserve key statistical properties of the discharge record and are hydrologically realistic.

2. Methods

2.1. Estimation of Random Errors in Discharge Records

Let's consider a n-record of discharge values, $\tilde{\mathbf{y}} = [\tilde{y}_1 \tilde{y}_2 \ \dots \ \tilde{y}_n]^\top$, in units of mm per time. If the discharge record is subject to random errors, the entries of this vector may be written as follows

$$\tilde{\mathbf{y}} = \mathcal{H}(\mathbf{t}) + \epsilon, \quad \epsilon \sim \mathcal{N}_n(\mathbf{0}, \mathbf{\Sigma}_\epsilon), \tag{1}$$

where $\mathcal{H}(t)$ is the data generating process of the actual streamflow at time $t \geq 0$ and the $n \times 1$ vector of errors, $\epsilon = [\epsilon_1 \epsilon_2 \ \dots \ \epsilon_n]^\top$, consists of independent variates with zero-mean and variance, $\sigma^2_{\epsilon_t}$, for all $t \in \mathbb{N}_+$. Without loss of generality, we do not specify a time unit, but in our application of Equation 1 we will consider hourly and daily discharge data. Streamflow errors is well-known to be heteroscedastic in nature (Sorooshian & Dracup, 1980), hence, we must specify n different error variances, $\boldsymbol{\sigma}^2_\epsilon = \left[\sigma^2_{\epsilon_1} \sigma^2_{\epsilon_2} \ \dots \ \sigma^2_{\epsilon_n}\right]^\top$. This results in the following formulation of the $n \times n$ covariance matrix of the random errors

$$\boldsymbol{\Sigma}_{\epsilon} = \mathbb{E}\left[\epsilon\epsilon^{\top}\right] = \begin{bmatrix} \sigma^2_{\epsilon_1} & 0 & \cdots & 0 \\ 0 & \sigma^2_{\epsilon_2} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sigma^2_{\epsilon_n} \end{bmatrix} = \mathrm{diag}\left(\boldsymbol{\sigma}^2_{\epsilon}\right), \tag{2}$$

where the function diag(**a**) produces a diagonal matrix with zeros everywhere and entries, $\mathbf{a} = [a_1\ a_2\ \ldots\ a_n]^{\top}$ on the main diagonal.

If the errors are independent and identically distributed random variables with zero-mean and nonconstant variance, then we can resort to the nonparametric estimator of Vrugt et al. (2005) to estimate $\boldsymbol{\sigma}^2_{\epsilon}$ from the discharge record. This estimator belongs to the class of difference-based variance estimation methods (Hall et al., 1990) and differences the hourly discharge time series, $\tilde{\mathbf{y}}_{\mathrm{h}} = [\,\tilde{y}_{1\mathrm{h}}\ \tilde{y}_{2\mathrm{h}}\ \ldots\ \tilde{y}_{n\mathrm{h}}\,]^{\top}$, k consecutive times to yield a local estimate, $\widehat{\sigma}^2_{\mathrm{h}}$, of the hourly error variance

$$\widehat{\sigma}^2_{\mathrm{h}} = \binom{2k}{k}^{-1} \left(\Delta^k\left(\tilde{y}_{t\mathrm{h}}\right)\right)^2, \tag{3}$$

where $\Delta^1\left(\tilde{y}_{t\mathrm{h}}\right) = \tilde{y}_{t\mathrm{h}} - \tilde{y}_{t-1\mathrm{h}}$, $\Delta^2\left(\tilde{y}_{t\mathrm{h}}\right) = \tilde{y}_{t\mathrm{h}} - 2\tilde{y}_{t-1\mathrm{h}} + \tilde{y}_{t-2\mathrm{h}}$, $\Delta^3\left(\tilde{y}_{t\mathrm{h}}\right) = \tilde{y}_{t\mathrm{h}} - 3\tilde{y}_{t-1\mathrm{h}} + 3\tilde{y}_{t-2\mathrm{h}} - \tilde{y}_{t-3\mathrm{h}}$, and so forth, and $\binom{a}{b} = \frac{a!}{b!(a-b)!}$ is the binomial coefficient. The $k+1$ coefficients of the kth-order difference operator, $\Delta^k\left(\tilde{y}_{t\mathrm{h}}\right)$, honor the Pascal triangle. The derivation of Equation 3 for $k = 1$ is provided in Text S1 in Supporting Information S1. An alternative implementation of this nonparametric estimator is provided in Text S2 in Supporting Information S1. Nonparametric estimators of this kind have a long history in time series analysis (Anderson, 1971; Hall et al., 1990; Rice, 1984) and will provide an accurate characterization of random errors if (a) the data-generating process, $\mathcal{H}(t)$, is sufficiently smooth, and (b) the measurement frequency is high compared to the typical timescale of $\mathcal{H}(t)$. We will demonstrate that the hourly streamflow records of the Catchment Attributes and MEteorology for Large-sample Studies (CAMELS) watersheds satisfy both conditions. A scatter plot of $\widehat{\sigma}^2_{\mathrm{h}}$ and the corresponding discharge will now reveal the nature of the random errors of the discharge record. Note that each $\widehat{\sigma}^2_{\mathrm{h}}$ individually is rather meaningless, since its value is estimated using only $k+1$ data points. In the estimator of Hall et al. (1990), valid for homoscedastic errors only, the variance estimate is computed as the arithmetic mean of all $\widehat{\sigma}^2_{\mathrm{h}}$ values (see Equation 13 in Supporting Information S1). A similar approach must be pursued for heteroscedastic errors. In this paper, we assume the following error function to describe the heteroscedasticity of the random errors,

$$\sigma^2_{t\mathrm{h}} = \left(\alpha y_{t\mathrm{h}} + \beta m_{\mathrm{h}}\right)^2, \tag{4}$$

where $m_{\mathrm{h}} = \frac{1}{n}\sum_{t=1}^{n} \tilde{y}_{t\mathrm{h}}$ (mm/d) is the arithmetic mean of the hourly discharge data, α, $\beta \geq 0$ are non-negative dimensionless coefficients that determine the nature and magnitude of the errors and $t = (1, 2, \ldots, n)$. The use of the mean discharge in the intercept, βm_{h}, of the error function of Equation 4 warrants the application of a common β value to catchments with widely different flow magnitudes. For $\alpha = 0$, the n-vector of random errors will have a constant variance and for $\alpha > 0$ the magnitude of the random errors will increase with discharge magnitude. Heteroscedasticity of discharge errors is expected since a power function (or similar) is used to transform stage measurements into discharge estimates. The use of a linear heteroscedastic model is justified by visual inspection of the relationship between $\widehat{\sigma}_{\mathrm{h}}$ and discharge magnitude (as will be shown later) and it is also supported by previous hydrologic studies (e.g., Evin et al., 2013, 2014; Schoups & Vrugt, 2010; Thyer et al., 2009). The application of the nonparametric estimator directly to stage time series along with rating curve information may reveal a more accurate description of the relationship between $\widehat{\sigma}_{\mathrm{h}}$ and discharge magnitude than the linear heteroscedastic model used herein.

Algorithm S1 in Supporting Information S1 presents a MATLAB implementation of the nonparametric error estimator. This function, called `error_estimation`, requires as input argument the $n \times 1$ vector $\tilde{\mathbf{y}}_{\mathrm{h}}$ of hourly discharge values and returns the estimates $\hat{\alpha}$ and $\hat{\beta}$ of the error function (Equation 4). These estimates may be obtained by two methods: (a) using all data; and (b) from a moving average of the error variance computed from

a window of m data pairs on either side of the data point when the data pairs are sorted by $\overline{y}_h$ (see definition in Equation 17 in Supporting Information S1). This moving average substantially reduces data dispersion and provides a more robust characterization of the $(\overline{y}_h - \widehat{\sigma}_h)$ relationship. The estimation method may be specified as optional input argument. In case method 2 is selected, the value of m may also be specified, otherwise a default value of $m = 100$ is used. Investigations with corrupted simulated streamflow data showed that the choice $m = 100$ works well with hourly discharge time series. Optional outputs include the $(n - k) \times 1$ vectors $\widehat{\boldsymbol{\sigma}}_h^2$ and $\overline{\mathbf{y}}_h$ of discharge error variances and corresponding discharge estimates, respectively. These two vectors characterize the error relationship of the watershed under investigation. The value of k may be specified as optional input argument, otherwise a default value of $k = 3$ is assigned. As optional input, the user may also specify a tolerance value below which the estimated error will be removed from the analysis.

2.1.1. Case Study I: Simulated Streamflow Data

To demonstrate that Equation 3 provides an accurate characterization of the nature and magnitude of random errors in streamflow records, we benchmark our method by application to error corrupted streamflow record. Table S1 in Supporting Information S1 summarizes main characteristics of the five watersheds selected for this first case study.

Corrupted streamflow data is generated by adding uncorrelated normal variates, with variance that follows the error function specified in Equation 4, to hourly simulated discharge time series. Text S3 in Supporting Information S1 provides a more detailed explanation about the generation of the simulated and corrupted streamflow time series. The simulated discharge record of each watershed is displayed in Figure S1 in Supporting Information S1. The $n \times 1$ vector of error corrupted hourly discharge is referred to as pseudo streamflow record in the remainder of this paper. This sets these fabricated time series apart from the discharge records of the CAMELS watersheds. We test the nonparametric error estimator of Equation 3 for two different types of error. In the case of homoscedastic errors, $\alpha = 0$, and we set β equal to 0.001, 0.01, 0.1, 1.0, and 10. This results in five pseudo discharge records for each watershed with increasing levels of homoscedastic error. In the case of heteroscedastic errors, we set $\beta = 0$ and consider separately, $\alpha = 0.001$, 0.01, 0.05, 0.1, and 0.3. This brings the total to 10 pseudo discharge records for each watershed.

2.1.2. Case Study II: Hourly Error Estimates for the CAMELS Catchments

We demonstrate the practical usefulness of the estimator by application to catchments from the CAMELS data set (Addor et al., 2017a; Newman et al., 2015). Preliminary analysis revealed that a daily time step can be too coarse to provide an accurate estimation of the random errors, especially for low levels of errors (results not shown). Therefore, we resort to the hourly streamflow time series made available by Gauch et al. (2020). We only included in our analysis catchments without any missing daily streamflow data within the period available in the CAMELS data set, and for which the corresponding hourly time series from Gauch et al. (2020) were also available. These two criteria combined resulted in a total of 504 catchments being included in our analysis. More details of the experimental data can be found in Text S4 in Supporting Information S1.

2.2. Effect of Measurement Frequency on Discharge Uncertainty

We can turn the hourly error variances, σ_{th}^2, of Equation 4 into daily estimates, $\widehat{\sigma}_{td}^2$ (mm^2/d^2), as follows (derivation in Text S5 in Supporting Information S1)

$$\begin{aligned}\widehat{\sigma}_{td}^2 &= \frac{1}{u}\frac{1}{u-1}\sum_{i=1}^{u}\left(\tilde{y}_{ih} - \frac{1}{u}\sum_{j=1}^{u}\tilde{y}_{jh}\right)^2 + \frac{1}{u}\frac{1}{u}\sum_{i=1}^{u}\sigma_{ih}^2 \\ &= \frac{1}{u}\sigma_{\tilde{y}_h}^2 + \frac{1}{u^2}\sum_{i=1}^{u}(\alpha\tilde{y}_{ih} + \beta\tilde{m}_h)^2,\end{aligned} \tag{5}$$

where $\sigma_{\tilde{y}_h}^2$ (mm^2/d^2) denotes the variance of the $u = 24$ hourly discharge values and $\tilde{m}_h = \frac{1}{n}\sum_{t=1}^{n}\tilde{y}_{th}$ (mm/d) is the arithmetic mean of the hourly discharge record. The first term on the right-hand side is simply a multiple of $1/u$ of the variance of hourly discharge values and expresses the effect of measurement frequency on daily estimates of the error variance. This term will decrease if we increase u, that is, if we use data at a sub-hourly resolution. The second term on the right-hand side is a multiple of $1/u$ of the mean hourly error variance. Now, in analogy

to Equation 4, we use a linear relationship to describe the functional dependence of $\hat{\sigma}_{\mathrm{td}}$ and the daily discharge, $\tilde{y}_{\mathrm{td}}$ (mm/d), as follows

$$\sigma_{\mathrm{td}} = a\tilde{y}_{\mathrm{td}} + b\tilde{m}_{\mathrm{d}} \tag{6}$$

where a (−) and b (−) signify the daily slope and intercept, respectively, and $\tilde{m}_{\mathrm{d}} = \frac{1}{n}\sum_{t=1}^{n} \tilde{y}_{\mathrm{td}}$ (mm/d) is the arithmetic mean of the daily discharge data. The (linear) model is needed since the hourly discharge time series does not cover the entire period we are interested in (of the daily data). If there were no missing data in the hourly discharge time series, the linear model of Equation 6 would not be required and the estimates provided by Equation 5 could be used directly in the subsequent steps of the proposed method.

2.3. Generation of Replicates of Streamflow Records

Now that the daily error variance is estimated, we can create discharge replicates by adding to the record, $\tilde{\mathbf{y}}$, a n-vector of errors, $\epsilon^* \sim \mathcal{N}_n(\mathbf{0}, \boldsymbol{\Sigma}_\epsilon)$, which is drawn at random from the n-variate normal distribution with zero mean and covariance matrix, $\boldsymbol{\Sigma}_\epsilon$, of Equation 2. The symbol $\boldsymbol{\epsilon}^*$ differentiates the sampled errors from the "true" but unknown random errors, $\boldsymbol{\epsilon}$, of Equation 1. Since the data generating process, $\mathcal{H}(\mathbf{t})$, is unknown, we do not add $\boldsymbol{\epsilon}^*$ to $\mathcal{H}(\mathbf{t})$ (as would be expected from Equation 1) but to the discharge record, $\tilde{\mathbf{y}}$, which already incorporate errors. The replicates, thus, constitute perturbations of the discharge record. This approach will create replicates that are less smooth than the original record and express unrealistically large and/or sudden changes in day-to-day streamflow. This undesired behavior is particularly noticeable in long recession periods.

We must rectify these anomalies and make sure that the discharge replicates are consistent with the discharge record. We can simply revise our procedure and induce autocorrelation between the n entries of a perturbation time series, p, that added to the discharge record will result in replicates that (among others) preserve the smoothness of the original time series and portray the inferred heteroscedastic relationship. For this purpose, we replace the covariance matrix, $\boldsymbol{\Sigma}_\epsilon$, of Equation 2 with a covariance matrix, $\boldsymbol{\Sigma}_p$, written as a product of the $n \times n$ correlation matrix of an AR(k) process, $\mathbf{R}_{\tilde{y}}$, and the $n \times n$ diagonal matrix of the daily error variances (see Text S6 in Supporting Information S1). We can now generate realizations of the discharge record by drawing perturbation time series, $\mathbf{p} \sim \mathcal{N}_n(\mathbf{0}, \boldsymbol{\Sigma}_p)$, with covariance matrix, $\boldsymbol{\Sigma}_p$. There are several ways in which we can do this. We use Cholesky factorization and decompose the covariance matrix, $\boldsymbol{\Sigma}_p$, into a lower triangular matrix, $\mathbf{L}$, with positive diagonal entries so that $\boldsymbol{\Sigma}_p = \mathbf{L}\mathbf{L}^\top$ (Stewart, 1998). Next, we can generate replicates of the discharge time series, $\tilde{\mathbf{y}}_{\mathrm{r}}$, using, $\tilde{\mathbf{y}}_{\mathrm{r}} = \tilde{\mathbf{y}} + \mathbf{L}\boldsymbol{v}^*$, where $\boldsymbol{v}^* \in \mathbb{R}^n$ is a $n \times 1$ vector of variates drawn from the standard normal distribution (see e.g., Gentle (2006), for the generation of a random vector from a multivariate normal distribution). This method is CPU-efficient and minimizes computational costs for long data records.

Algorithm S2 in Supporting Information S1 presents a MATLAB implementation of the replicate generation procedure. This function, called `replicate_generation`, requires as input arguments the $n \times 1$ vector $\tilde{\mathbf{y}}$ of daily discharge values, the number N of replicates that should be generated, and the $n \times 1$ vector $\boldsymbol{\sigma}_d$ of standard deviations of the daily discharge record. The code returns a $n \times N$ matrix that contains one replicate $\mathbf{y}_{\mathrm{r}}$ in each column.

3. Results

3.1. Estimation of Random Errors in Discharge Records

3.1.1. Case Study I: Simulated Streamflow Data

Figure 1 presents the results of the nonparametric estimator of Equation 3 using second-order differencing for the 10 pseudo discharge records of the Leaf River (USGS 02472000), an example of catchment with a *strong winter regime* according to the functional classification of Brunner et al. (2020). The top and bottom panels correspond to the homoscedastic and heteroscedastic error cases, respectively, and present scatter plots of the $(\overline{y}_{\mathrm{h}}, \hat{\sigma}_{\mathrm{h}})$ data points for $\alpha = 0$ and (a) $\beta = 0.001$, (b) $\beta = 0.01$, (c) $\beta = 0.1$, (d) $\beta = 1.0$, and (e) $\beta = 10$ and in the heteroscedastic error case with $\beta = 0$ and (f) $\alpha = 0.001$, (g) $\alpha = 0.01$, (h) $\alpha = 0.05$, (i) $\alpha = 0.1$, and (j) $\alpha = 0.3$. The blue squares in each graph correspond to a moving average of the error variances obtained from a window of 100 data pairs on either side of the data point. The red line displays the error function of Equation 4 used to corrupt the

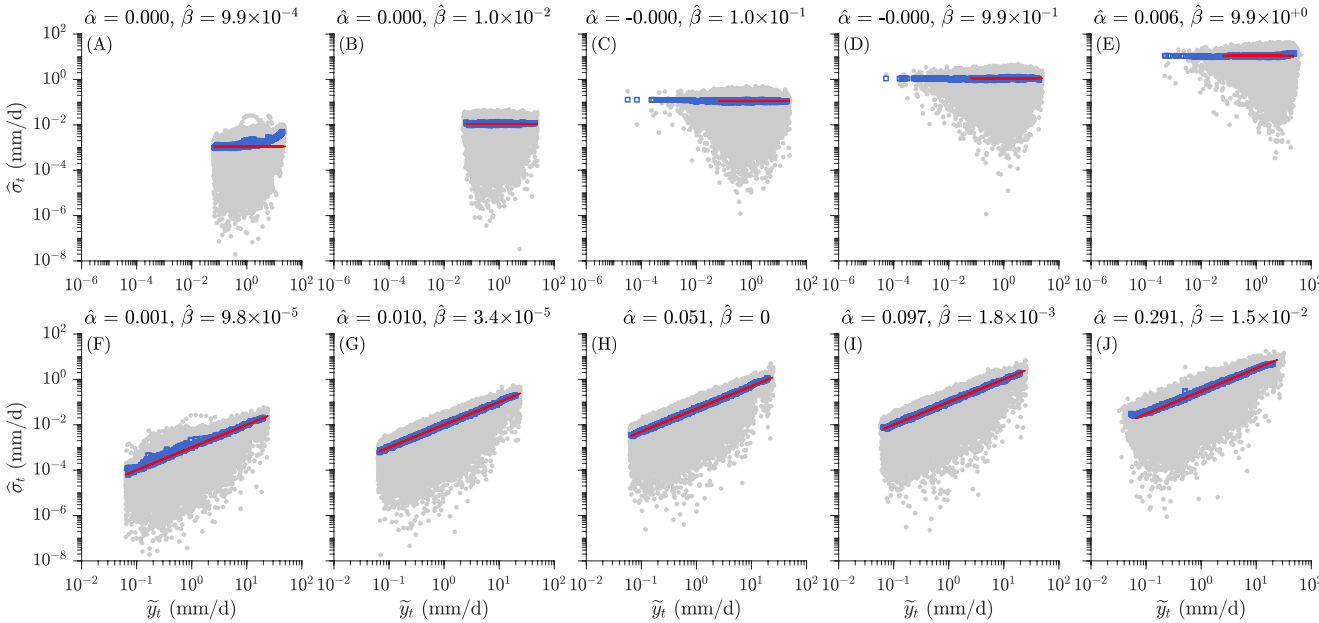

Figure 1. Relationship between the error deviations $\widehat{\sigma}_t$ and corresponding hourly discharge values obtained using second-order differencing ($k = 2$) for the 10 pseudo discharge records of the Leaf River near Collins, MS (USGS 02472000), an example of catchment with a *strong winter regime*. The top and bottom panels correspond to the homoscedastic and heteroscedastic error cases, respectively, and present scatter plots of the $\left(\overline{y}_{\mathrm{h}}, \widehat{\boldsymbol{\sigma}}_{\mathrm{h}}\right)$ data points for $\alpha = 0$ and (a) $\beta = 0.001$, (b) $\beta = 0.01$, (c) $\beta = 0.1$, (d) $\beta = 1.0$ and (e) $\beta = 10$ and in the heteroscedastic error case with $\beta = 0$ and (f) $\alpha = 0.001$, (g) $\alpha = 0.01$, (h) $\alpha = 0.05$, (i) $\alpha = 0.1$, and (j) $\alpha = 0.3$. Each gray dot signifies a different data pair. The blue squares portray the moving average of the error deviation computed from a window of 100 data pairs on either side of the data point. The red line displays the error function of Equation 4 used to corrupt the simulated discharge record. The least squares values of the coefficients $\hat{\alpha}$ and $\hat{\beta}$ of the linear regression model, $\alpha\overline{y}_{\mathrm{rh}} + \beta m_{y_{\mathrm{h}}}$, which is fitted to the blue squares are listed in each graph.

simulated discharge record. The least squares values of the coefficients, $\hat{\alpha}$ and $\hat{\beta}$, of the linear regression model, $\alpha\overline{y}_{\mathrm{rh}} + \beta m_{y_{\mathrm{h}}}$ fitted to the blue squares are listed in each graph.

The results of this analysis highlight several important findings. In the first place, note in Figure 1a the mismatch between the error standard deviations (blue squares) and the error function of Equation 4 used to corrupt the simulated discharge record (red line) for larger values of $\overline{y}_{\mathrm{h}}$. Thus, we present in Figure S2 in Supporting Information S1 the results obtained using third-order differencing. It is clear that, for very low level of errors, $k = 3$ works best (Figure 1a and Figure S2a in Supporting Information S1). Second, the nonparametric estimator using third-order differencing provides an accurate characterization of the nature of the hourly discharge random errors for all error levels (Figure S2 in Supporting Information S1). The estimator returns a constant variance across flow levels in the homoscedastic error case, and produces a nonconstant error standard deviation for the heteroscedastic error case. Third, the blue squares provide an accurate characterization of the actual error function (red line) that was used to generate the pseudo discharge records. It should therefore not be a surprise that the least squares estimates, $\hat{\alpha}$ and $\hat{\beta}$, match almost perfectly their counterparts, α and β of Equation 4. Lastly, the nonparametric estimator returns $\hat{\alpha} \to 0$ and $\hat{\beta} \to 0$ when applied directly to error-free discharge data (not shown). This is an encouraging result and highlights that the estimator can differentiate the errors from the underlying signal (data generating process). This testifies to the smoothness and measurement resolution of hourly discharge data and inspires confidence that the two stipulated requirements of variance-based difference estimation are met. Figures S3–S6 in Supporting Information S1 present our results for the other four watersheds with contrasting hydrologic regimes. Based on the findings for the Leaf River basin, we fix $k = 3$ and use third-order differencing in our application of Equation 3 in these catchments. The results for these watersheds are qualitatively similar to those presented in Figure 1 for the Leaf River. Note the same pattern observed previously in Figure 1a for $\beta = 0.001$ is repeated for the catchment with the *intermittent regime* (Figure S3a in Supporting Information S1), but now for $k = 3$. Following the same reasoning, increasing the value of k would help to improve the estimation in this case (not shown).

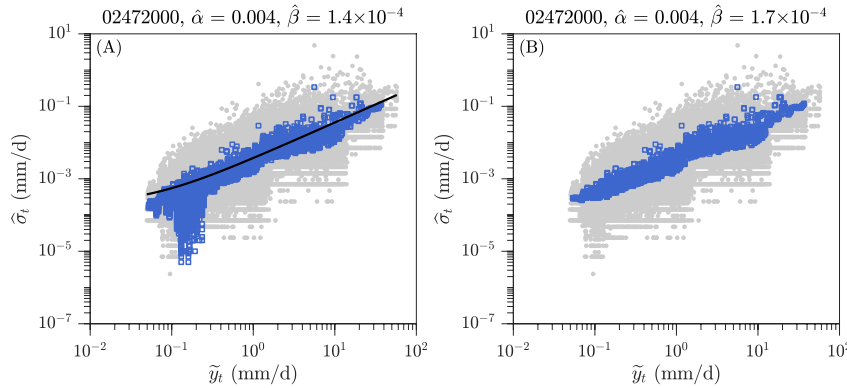

Figure 2. Relationship between the error deviations $\hat{\sigma}_t$ and corresponding hourly discharge values for the Leaf River near Collins, MS (USGS 02472000), an example of catchment with a *strong winter regime*: (a) using all data, and (b) removing from the analysis $\hat{\sigma}_{th}$ values smaller than 10^{-10}. Each gray dot signifies a different data pair. The blue squares portray the moving average of the error deviation computed from a window of 100 data pairs on either side of the data point. The solid black line signifies the regression line.

To check whether α and β are consistently estimated, we repeated the simulation 1,000 times for each setting. We report in Figures S7–S11 in Supporting Information S1 the distribution of $\hat{\alpha}$ and $\hat{\beta}$ for the homoscedastic error case and in Figures S12–S16 in Supporting Information S1 for the heteroscedastic error case, using the pseudo discharge records of the five selected watersheds. It can be seen that in general the mean $\hat{\alpha}$ and $\hat{\beta}$ values derived from the nonparametric estimator of Equation 3 are in excellent agreement with their true values used to corrupt the simulated discharge record. For the catchment with the *intermittent regime*, α and β are consistently overestimated for $\beta = 0.001$ (Figures S8a and S8f in Supporting Information S1) and, to a lesser extent, for $\beta = 0.01$ (Figures S8b and S8g in Supporting Information S1). The same result holds true for heteroscedastic errors, for which α and β are consistently overestimated for $\alpha = 0.001$ (Figures S13a and S13f in Supporting Information S1). These results confirm our earlier conclusion that, in this catchment, a larger value of k would be needed to correctly characterize very low levels of errors.

Altogether, the results of the synthetic case study confirm that application of Equation 3 with hourly data provides unbiased estimates of the error variance for catchments with contrasting hydrologic conditions. Our results are in agreement with Dette et al. (1998) and illustrate that higher order estimators are needed in case of substantial increase in the variation of the function subject to differencing, for example, for the intermittent catchment, and/or a substantial decrease in the error variance, for example, for very low level of error ($\beta = 0.001$). These numerical experiments also demonstrate that third-order differencing, $k = 3$, suffices for practical application.

3.1.2. Case Study II: Hourly Error Estimates for the CAMELS Catchments

To provide insights into the estimated random errors of real discharge records, please consider Figure 2 which presents a scatter plot of the $(\overline{y}_h, \hat{\sigma}_h)$ data pairs derived from Equation 3 using hourly discharge time series from Gauch et al. (2020) of the Leaf River near Collins, MS (USGS 02472000). This is only one watershed of a much larger cohort of 500+ catchments of the CAMELS data set (Addor et al., 2017a; Newman et al., 2015) investigated in this study. Each gray dot signifies a different data pair. The blue squares correspond to a moving average of the error deviations obtained from a window of 100 data pairs on either side of the data point. As is evident from the raw data (gray dots), the error standard deviation increases substantially with flow level. This confirms the heteroscedastic nature of random errors in streamflow records. The $(\overline{y}_h, \hat{\sigma}_h)$-data scatter (blue squares) is reasonably approximated with a line ($R^2 = 0.45$), justifying the common assumption that the standard deviation of the discharge errors increases linearly with flow magnitude (e.g., Evin et al., 2013, 2014; Schoups & Vrugt, 2010; Thyer et al., 2009). Indeed, the solid black line.

$$\sigma_{th} = \alpha \tilde{y}_{th} + \beta \tilde{m}_h, \tag{7}$$

with slope, $\alpha = 4 \times 10^{-3}$ (−), and intercept, $\beta = 1.4 \times 10^{-4}$ (−), provides a reasonable description of the relationship between the hourly error standard deviations, $\hat{\sigma}_{th}$, and corresponding discharge values, $\tilde{y}_{th}$. This relationship results from the application of Equation 3 and can either be homoscedastic or heteroscedastic. Alternative

models, such as an exponential function, were also tested but their use was not justified by an overall increase in model fit.

Figures S17–S20 in Supporting Information S1 present our findings for the four other catchments with contrasting functional regimes according to the classification scheme of Brunner et al. (2020). Figures S17–S20 in Supporting Information S1 confirm that error heteroscedasticity is universal across discharge records of contrasting hydrologic regimes. Figure S21 in Supporting Information S1 visualizes the values of the slope of the hourly discharge error model for the 500+ watersheds of the CAMELS data set and reveals that the magnitude of random errors in the investigated discharge records is rather small (less than 3% for 80% of the records). Figure S22 in Supporting Information S1 presents the coefficient of determination, R^2, of the linear error model of Equation 7 for each CAMELS catchment. The relatively high median R^2 of 0.77 provides support for the use of a linear $(\bar{y}_h, \hat{\sigma}_h)$-relationship.

3.2. Daily Uncertainty Estimates for the CAMELS Catchments

Figure S23 in Supporting Information S1 visualizes the values of the slope of the daily discharge error model for the 500+ watersheds of the CAMELS data set. The ranges 0.03–0.08 and 0.01–0.22 encompass 50% and 95% of the slope values, respectively. The median slope of 0.05 for all CAMELS watersheds is substantially smaller than common values of 0.10–0.20 documented and/or used in the hydrologic literature (e.g., H. K. McMillan & Westerberg, 2015; H. McMillan et al., 2012; Thyer et al., 2009). This is an expected result since our method only combines the effect of random errors and measurement frequency into the daily variance estimate. Thus, this difference must be systematic uncertainty (which our method does not account for). Figure S24 in Supporting Information S1 presents the coefficient of determination, R^2, of the linear error model of Equation 6 for each CAMELS catchment.

3.3. Replicates of Streamflow Records for the CAMELS Catchments

We illustrate our streamflow generation approach by application to the daily discharge record of the Leaf River watershed. Figure 3a presents the 95% confidence interval of the $N = 1{,}000$ replicates of the discharge record for a representative portion of the data set. The discharge data are separately indicated with a solid blue line. The confidence intervals cannot be used to judge the hydrologic reasonableness of the different realizations, since they only convey information on the ensemble spread. Therefore, in the bottom panel, we zoom in on three separate periods with (b) low, (c) median, and (d) high flows, respectively. This allows us to plot a few individual replicates of the discharge record. Figures S25–S28 in Supporting Information S1 present our findings for the other four watersheds. The 95% streamflow confidence regions center on the discharge record and increase in width with flow magnitude. This heteroscedasticity is readily visible in the bottom panel, wherein higher streamflows display more dispersion among a selected sample of the replicates. The graphs also confirm that the replicates are smooth and exhibit a similar temporal persistence as the original discharge record. This is easy to verify in long recession periods with low flows, but is much more difficult to depict for median and high flows as their time windows are insufficient to illustrate this behavior. Our next analysis will provide further support for these claims. Figure 4 benchmarks the (a) spread, (b) sample autocorrelation function (ACF), (c) mean absolute discharge difference (MAD),

$$\mathrm{MAD} = \frac{1}{n-1}\sum_{t=2}^{n} |\tilde{y}_{t\mathrm{dr}} - \tilde{y}_{t-1\mathrm{dr}}|, \tag{8}$$

and (d) volume error (VE, in %)

$$\mathrm{VE} = 100\frac{\sum_{t=1}^{n}(\tilde{y}_{t\mathrm{d}} - \tilde{y}_{t\mathrm{dr}})}{\sum_{t=1}^{n}\tilde{y}_{t\mathrm{d}}}, \tag{9}$$

of the 1,000 replicates of the discharge record against their counterparts of the original discharge record of the Leaf River. These results confirm that the standard deviation of the replicates increases linearly with discharge magnitude. The data pairs (gray circles) lie almost perfectly on the error model (solid black line) of the Leaf River. The ACFs of the replicates (gray lines) are in excellent agreement with their counterpart computed from

10.1029/2022WR032263

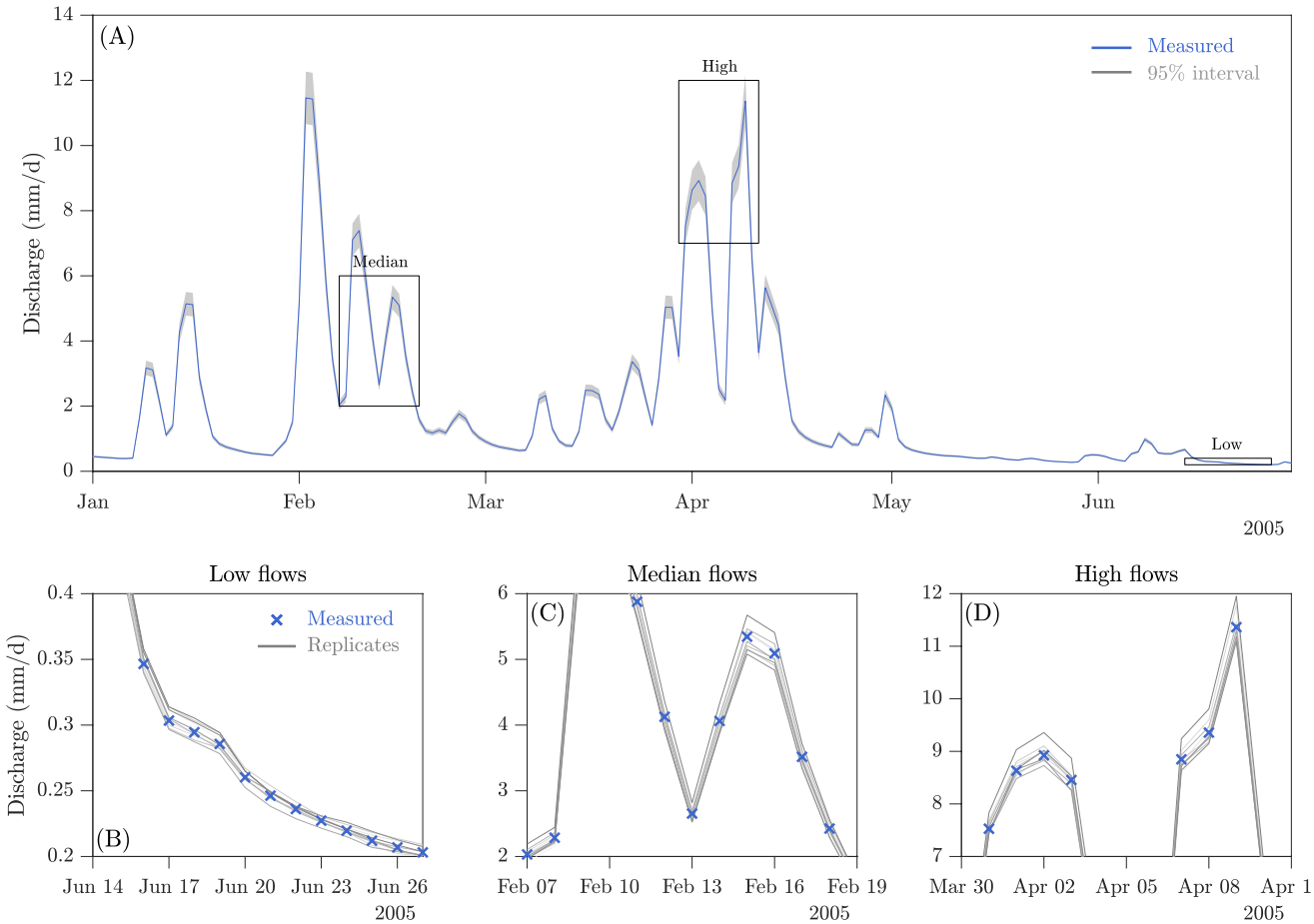

Figure 3. Illustration of the streamflow replicates generated using daily streamflow data from the Leaf River near Collins, MS (USGS 02472000), an example of catchment with a *strong winter regime*. (a) 99% confidence intervals (gray region) of the $N = 1{,}000$ replicates of the discharge record for a representative portion of the 34-year data set. The discharge data are separately indicated with a solid blue line. The top panel only visualizes percentiles of the discharge uncertainty without recourse to the underlying replicates. Therefore, the bottom panel displays a selection of the replicates for small excerpts of the discharge record with (b) low, (c) median, and (d) high flows, respectively.

the original discharge record. In fact, the traces of the replicates are so similar that their ACFs have collapsed to a line. The replicates not only preserve essential statistical properties of the discharge data, but also honor key hydrologic properties. The marginal distributions of the MAD and VE metrics center on their values computed from the daily discharge record. Figures S29–S32 in Supporting Information S1 confirm that the proposed method also works well for contrasting hydrologic regimes. The ensemble spread is governed by the magnitude of the estimated discharge errors. The larger the variance of the estimated errors, the larger the variation among the replicates, and, thus, the more dispersed the metrics will be around the values computed from the original discharge record. Figure S33 in Supporting Information S1 visualizes the values of the MAD metric for the 500+ watersheds of the CAMELS data set and confirms that the smoothness of the replicates are consistent with the smoothness of the original discharge record: the measured value is not inside the sampled distribution for three catchments only–these catchments correspond to the black circles in Figure S33 in Supporting Information S1.

4. Limitations and Future Work

The nonparametric estimator of Equation 3 assumes that the discharge errors are of aleatory nature only. We have confirmed with the synthetic case study (Section 3.1.1) that the estimator provides an accurate description of the nature and magnitude of such random errors. However, discharge time series are also subject to systematic errors as (among others) the rating curve used to convert water height into discharge (volume) will vary over time as a result of the variant hydraulic characteristics of the stream channel and floodplain (e.g., Coxon et al., 2015; H. K.

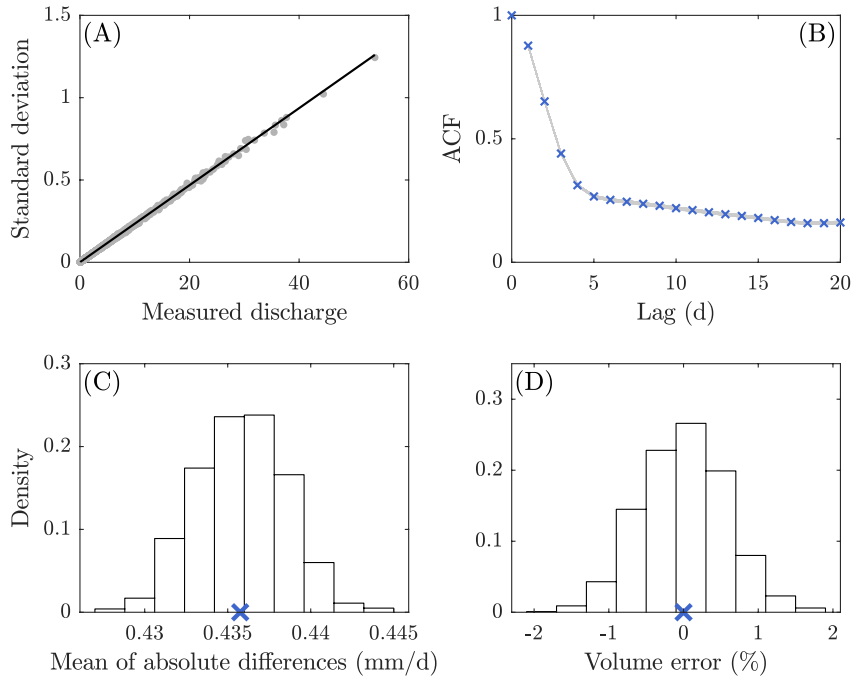

Figure 4. Characteristics of the streamflow replicates for the Leaf River near Collins, MS (USGS 02472000), an example of catchment with a *strong winter regime*. (a) Standard deviation of the $N = 1{,}000$ replicates as a function of discharge. Each gray dot signifies a different data pair. The solid black line signifies the heteroscedastic error model that was used to create the discharge replicates. (b) The autocorrelation function (ACF) of the replicates (gray lines) and the original record (blue crosses). (c and d) Frequency distributions of the mean absolute discharge differences and volume error of the 1,000 replicates. The blue crosses highlight the values computed from the original discharge record.

McMillan & Westerberg, 2015; Mansanarez et al., 2019; Westerberg & McMillan, 2015). To negate systematic errors and promote the accuracy and reliability of discharge estimates, the rating curves of most USGS gauges are re-calibrated every 6–8 weeks.

The proposed replicate generation procedure can accommodate expert knowledge and/or other information about discharge precision and/or accuracy. This may make obsolete the uncertainty estimates (Section 2.2), but leaves invariant the methodology for the generation of streamflow replicates (Section 2.3). In other words, if updated estimates of the uncertainty of daily discharge time series become available, a different covariance matrix, $\mathbf{\Sigma}_p$, could be specified. In that case, the new variance estimates would populate the main diagonal of $\mathbf{\Sigma}_p$, and are used to create discharge replicates using the method of Section 2.3. As a simple example, please consider Figures S34–S38 in Supporting Information S1 that present discharge replicates for the five selected catchments using $\sigma_{rd} = 0.20\tilde{y}_{rd}$ instead of the values presented in Section 3.2. The replicates now exhibit a larger spread but remain plausible, as shown in Figures S39–S43 in Supporting Information S1.

5. Summary and Conclusions

In this paper, we introduce a relatively simple data-driven method for the representation of uncertainty of daily discharge records. We account for two sources of uncertainty, namely, the effect of random errors (Section 2.1) and of measurement frequency (Section 2.2).

We demonstrated how the nonparametric estimator presented in Vrugt et al. (2005) can be applied to estimate the variance of random errors, using both fabricated and real discharge time series of 500+ watersheds of the CAMELS data set. We provide guidelines into the selection of the order of the difference-based estimator and a MATLAB implementation. We showed that (1) the nonparametric estimator provides an accurate characterization of the nature and magnitude of the random errors in discharge records (Section 3.1.1), (2) third-order differencing suffices for practical application of the estimator to hourly time series (Section 3.1.1), and (3) the magnitude of random errors in the investigated discharge records are rather small (Section 3.1.2). This third result suggests that

future work should consider a more complete treatment of the uncertainty in discharge records. The uncertainty estimates could be updated as more information (e.g., stage-discharge gagings) are included in the analysis.

The nonparametric estimator could also be applied directly to stage time series. A probabilistic method that generates an ensemble of rating curves could then be used to propagate these errors into discharge uncertainty. For example, the estimates of random errors in stage records could feed the method proposed by Horner et al. (2018), which in the absence of additional information specified the variance of random errors based on expert knowledge.

In the second part of this paper, we presented theory and rationales of a method for the generation of equally plausible streamflow replicates (Section 2.3). The proposed method relies on hourly discharge time series only and (4) produces replicates which portray accurately the assigned streamflow uncertainty, preserve key statistical properties of the discharge record and are hydrologically realistic (Section 3.3).

The proposed replicate generation procedure has potential use in model diagnostics, facilitates Bayesian analysis and supports tasks such as data assimilation, uncertainty quantification and regionalization. In data assimilation, it has been shown that the observations must be perturbed at the analysis steps otherwise the variance of the ensembles will be too low (Burgers et al., 1998). The use of replicates of discharge records allows the temporal structure of the discharge time series to be preserved, which is not achieved when each perturbation is sampled independently. The impact of using replicates of discharge records in data assimilation will be investigated in a future study. In regionalization, streamflow uncertainty estimates could be used to assign weights to each discharge time series and to help discard gauges with inadequate hydrometric data quality.

Data Availability Statement

The Catchment Attributes and MEteorology for Large-sample Studies (CAMELS) data set is described in Newman et al. (2015) and can be downloaded from https://dx.doi.org/10.5065/D6MW2F4D (Newman et al., 2014). The hourly streamflow data of Gauch et al. (2020) are available at https://doi.org/10.5281/zenodo.4072700. The attributes of the CAMELS watersheds are described in Addor et al. (2017a) and can be downloaded from https://doi.org/10.5065/D6G73C3Q (Addor et al., 2017b). The regime classes of Brunner et al. (2020) are available from https://doi.org/10.4211/hs.069f552f96ef4e638f4bec281c5016ad (Brunner, 2020).

Acknowledgments
The authors highly appreciate the feedback received from Jack Eggleston and Terry Kenney concerning our questions about the USGS discharge data. The authors also want to acknowledge the comments provided by Francesco Serinaldi, Thorsten Wagener, Wouter Knoben, and two anonymous reviewers, which have helped to improve the quality of this paper. D. Y. de Oliveira gratefully acknowledges the financial support received from the Brazilian Federal Agency for Support and Evaluation of Graduate Education (Coordenação de Aperfeiçoamento de Pessoal de Nível Superior, CAPES), Grant 88881.174456/2018-01. This paper is in memory of Prof. Scott James from Baylor University, TX, whose sudden passing has left a big hole in our community. Many of us talk about interdisciplinary research. He actually did it.

References

Addor, N., Do, H. X., Alvarez-Garreton, C., Coxon, G., Fowler, K., & Mendoza, P. A. (2020). Large-sample hydrology: Recent progress, guidelines for new datasets and grand challenges. *Hydrological Sciences Journal*, *65*(5), 712–725. https://doi.org/10.1080/02626667.2019.1683182

Addor, N., Newman, A. J., Mizukami, N., & Clark, M. P. (2017a). The CAMELS data set: Catchment attributes and meteorology for large-sample studies. *Hydrology and Earth System Sciences*, *21*(10), 5293–5313. https://doi.org/10.5194/hess-21-5293-2017

Addor, N., Newman, A. J., Mizukami, N., & Clark, M. P. (2017b). *Catchment attributes for large-sample studies* [Dataset]. UCAR/NCAR. https://doi.org/10.5065/D6G73C3Q

Anderson, T. W. (1971). *The statistical analysis of time series*. Wiley.

Bjerklie, D. M., Birkett, C. M., Jones, J. W., Carabajal, C., Rover, J. A., Fulton, J. W., & Garambois, P.-A. (2018). Satellite remote sensing estimation of river discharge: Application to the Yukon River Alaska. *Journal of Hydrology*, *561*, 1000–1018. https://doi.org/10.1016/j.jhydrol.2018.04.005

Blöschl, G. (2017). Debates—Hypothesis testing in hydrology: Introduction. *Water Resources Research*, *53*(3), 1767–1769. https://doi.org/10.1002/2017WR020584

Boldetti, G., Riffard, M., Andréassian, V., & Oudin, L. (2010). Data-set cleansing practices and hydrological regionalization: Is there any valuable information among outliers? *Hydrological Sciences Journal*, *55*(6), 941–951. https://doi.org/10.1080/02626667.2010.505171

Brunner, M. (2020). Streamflow regimes for CAMELS dataset [Dataset]. HydroShare. https://doi.org/10.4211/hs.069f552f96ef4e638f4bec281c5016ad

Brunner, M. I., Melsen, L. A., Newman, A. J., Wood, A. W., & Clark, M. P. (2020). Future streamflow regime changes in the United States: Assessment using functional classification. *Hydrology and Earth System Sciences*, *24*(8), 3951–3966. https://doi.org/10.5194/hess-24-3951-2020

Brunner, M. I., Slater, L., Tallaksen, L. M., & Clark, M. (2021). Challenges in modeling and predicting floods and droughts: A review. *WIREs Water*, *8*(3), e1520. https://doi.org/10.1002/wat2.1520

Burgers, G., van Leeuwen, P. J., & Evensen, G. (1998). Analysis scheme in the ensemble Kalman filter. *Monthly Weather Review*, *126*(6), 1719–1724. https://doi.org/10.1175/1520-0493(1998)126<1719:ASITEK>2.0.CO;2

Coxon, G., Freer, J., Westerberg, I. K., Wagener, T., Woods, R., & Smith, P. J. (2015). A novel framework for discharge uncertainty quantification applied to 500 UK gauging stations. *Water Resources Research*, *51*(7), 5531–5546. https://doi.org/10.1002/2014WR016532

Dette, H., Munk, A., & Wagner, T. (1998). Estimating the variance in nonparametric regression—what is a reasonable choice? *Journal of the Royal Statistical Society: Series B*, *60*(4), 751–764. https://doi.org/10.1111/1467-9868.00152

Di Baldassarre, G., & Montanari, A. (2009). Uncertainty in river discharge observations: A quantitative analysis. *Hydrology and Earth System Sciences*, *13*(6), 913–921. https://doi.org/10.5194/hess-13-913-2009

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*(7), 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*(3), 2350–2375. https://doi.org/10.1002/2013WR014185

Garcia, D. (2010). Robust smoothing of gridded data in one and higher dimensions with missing values. *Computational Statistics & Data Analysis*, *54*(4), 1167–1178. https://doi.org/10.1016/j.csda.2009.09.020

Gasser, T., Sroka, L., & Jennen-Steinmetz, C. (1986). Residual variance and residual pattern in nonlinear regression. *Biometrika*, *73*(3), 625–633. https://doi.org/10.1093/biomet/73.3.625

Gauch, M., Kratzert, F., Klotz, D., Nearing, G., Lin, J., & Hochreiter, S. (2020). Data for "rainfall-runoff prediction at multiple timescales with a single long short-term memory network" [Dataset]. Zenodo. https://doi.org/10.5281/zenodo.4072701

Gentle, J. (2006). *Random number generation and Monte Carlo methods*. Springer.

Hall, P., & Marron, J. S. (1990). On variance estimation in nonparametric regression. *Biometrika*, *77*(2), 415–419. https://doi.org/10.1093/biomet/77.2.415

Hall, P., Kay, J. W., & Titterington, D. M. (1990). Asymptotically optimal difference-based estimation of variance in nonparametric regression. *Biometrika*, *77*(3), 521–528. https://doi.org/10.2307/2336990

Hannah, D. M., Demuth, S., van Lanen, H. A. J., Looser, U., Prudhomme, C., Rees, G., et al. (2011). Large-scale river flow archives: Importance, current status and future needs. *Hydrological Processes*, *25*(7), 1191–1200. https://doi.org/10.1002/hyp.7794

Horner, I., Renard, B., Le Coz, J., Branger, F., McMillan, H. K., & Pierrefeu, G. (2018). Impact of stage measurement errors on streamflow uncertainty. *Water Resources Research*, *54*(3), 1952–1976. https://doi.org/10.1002/2017WR022039

Kiang, J. E., Gazoorian, C., McMillan, H., Coxon, G., Le Coz, J., Westerberg, I. K., et al. (2018). A comparison of methods for streamflow uncertainty estimation. *Water Resources Research*, *54*(10), 7149–7176. https://doi.org/10.1029/2018WR022708

Le Coz, J., Renard, B., Bonnifait, L., Branger, F., & Le Boursicaud, R. (2014). Combining hydraulic knowledge and uncertain gaugings in the estimation of hydrometric rating curves: A Bayesian approach. *Journal of Hydrology*, *509*, 573–587. https://doi.org/10.1016/j.jhydrol.2013.11.016

Magilligan, F. J., & Nislow, K. H. (2005). Changes in hydrologic regime by dams. *Geomorphology*, *71*(1), 61–78. https://doi.org/10.1016/j.geomorph.2004.08.017

Mansanarez, V., Renard, B., Coz, J. L., Lang, M., & Darienzo, M. (2019). Shift happens! Adjusting stage-discharge rating curves to morphological changes at known times. *Water Resources Research*, *55*(4), 2876–2899. https://doi.org/10.1029/2018WR023389

McMillan, H. K., & Westerberg, I. K. (2015). Rating curve estimation under epistemic uncertainty. *Hydrological Processes*, *29*(7), 1873–1882. https://doi.org/10.1002/hyp.10419

McMillan, H., Krueger, T., & Freer, J. (2012). Benchmarking observational uncertainties for hydrology: Rainfall, river discharge and water quality. *Hydrological Processes*, *26*(26), 4078–4111. https://doi.org/10.1002/hyp.9384

McMillan, H., Westerberg, I., & Branger, F. (2017). Five guidelines for selecting hydrological signatures. *Hydrological Processes*, *31*(26), 4757–4761. https://doi.org/10.1002/hyp.11300

Müller, H.-G., & Stadtmüller, U. (1987). Estimation of heteroscedasticity in regression analysis. *Annals of Statistics*, *15*(2), 610–625. https://doi.org/10.1214/aos/1176350364

Newman, A. J., Clark, M. P., Sampson, K., Wood, A., Hay, L. E., Bock, A., et al. (2015). Development of a large-sample watershed-scale hydrometeorological data set for the contiguous USA: Data set characteristics and assessment of regional variability in hydrologic model performance. *Hydrology and Earth System Sciences*, *19*(1), 209–223. https://doi.org/10.5194/hess-19-209-2015

Newman, A., Sampson, K., Clark, M. P., Bock, A., Viger, R. J., & Blodgett, D. (2014). A large-sample watershed-scale hydrometeorological dataset for the contiguous USA [Dataset]. UCAR/NCAR. https://doi.org/10.5065/D6MW2F4D

Olden, J. D., & Poff, N. L. (2003). Redundancy and the choice of hydrologic indices for characterizing streamflow regimes. *River Research and Applications*, *19*(2), 101–121. https://doi.org/10.1002/rra.700

Reuschen, S., Nowak, W., & Guthke, A. (2021). The four ways to consider measurement noise in Bayesian model selection – And which one to choose. *Water Resources Research*, *57*(11), e2021WR030391. https://doi.org/10.1029/2021WR030391

Rice, J. (1984). Bandwidth choice for nonparametric regression. *Annals of Statistics*, *12*(4), 1215–1230. https://doi.org/10.1214/aos/1176346788

Richter, B. D., Baumgartner, J. V., Powell, J., & Braun, D. P. (1996). A method for assessing hydrologic alteration within ecosystems. *Conservation Biology*, *10*(4), 1163–1174. https://doi.org/10.1046/j.1523-1739.1996.10041163.x

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*(10), 2009WR008933. https://doi.org/10.1029/2009WR008933

Silverman, B. W. (1985). Some aspects of the spline smoothing approach to non-parametric regression curve fitting. *Journal of the Royal Statistical Society: Series B*, *47*(1), 1–52. https://doi.org/10.1111/j.2517-6161.1985.tb01327.x

Slater, L. J., Anderson, B., Buechel, M., Dadson, S., Han, S., Harrigan, S., et al. (2021). Nonstationary weather and water extremes: A review of methods for their detection, attribution, and management. *Hydrology and Earth System Sciences*, *25*(7), 3897–3935. https://doi.org/10.5194/hess-25-3897-2021

Slater, L., Villarini, G., Archfield, S., Faulkner, D., Lamb, R., Khouakhi, A., & Yin, J. (2021). Global changes in 20-year, 50-year, and 100-year river floods. *Geophysical Research Letters*, *48*(6), e2020GL091824. https://doi.org/10.1029/2020GL091824

Sorooshian, S., & Dracup, J. A. (1980). Stochastic parameter estimation procedures for hydrologie rainfall-runoff models: Correlated and heteroscedastic error cases. *Water Resources Research*, *16*(2), 430–442. https://doi.org/10.1029/WR016i002p00430

Stewart, G. (1998). *Matrix algorithms: Volume 1: Basic decompositions*. Society for Industrial and Applied Mathematics.

Tetzlaff, D., Carey, S. K., McNamara, J. P., Laudon, H., & Soulsby, C. (2017). The essential value of long-term experimental data for hydrology and water management. *Water Resources Research*, *53*(4), 2598–2604. https://doi.org/10.1002/2017WR020838

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*(12), W00B14. https://doi.org/10.1029/2008WR006825

Vrugt, J. A., Diks, C. G. H., Gupta, H. V., Bouten, W., & Verstraten, J. M. (2005). Improved treatment of uncertainty in hydrologic modeling: Combining the strengths of global optimization and data assimilation. *Water Resources Research*, *41*(1), W01017. https://doi.org/10.1029/2004WR003059

Wahba, G. (1978). Improper priors, spline smoothing and the problem of guarding against model errors in regression. *Journal of the Royal Statistical Society: Series B*, *40*(3), 364–372. https://doi.org/10.1111/j.2517-6161.1978.tb01050.x

Westerberg, I. K., & McMillan, H. K. (2015). Uncertainty in hydrological signatures. *Hydrology and Earth System Sciences*, *19*(9), 3951–3968. https://doi.org/10.5194/hess-19-3951-2015

Westerberg, I. K., Wagener, T., Coxon, G., McMillan, H. K., Castellarin, A., Montanari, A., & Freer, J. (2016). Uncertainty in hydrological signatures for gauged and ungauged catchments. *Water Resources Research*, *52*(3), 1847–1865. https://doi.org/10.1002/2015WR017635

Zhou, Y., Cheng, Y., Wang, L., & Tong, T. (2015). Optimal difference-based variance estimation in heteroscedastic nonparametric regression. *Statistica Sinica*, *25*, 1377–1397. https://doi.org/10.5705/ss.2013.010

References From the Supporting Information

Box, G., Jenkins, G., Reinsel, G., & Ljung, G. (2015). *Time series analysis: Forecasting and control*. Wiley. Retrieved from https://books.google.com/books?id=rNt5CgAAQBAJ

Clark, M. P., & Kavetski, D. (2010). Ancient numerical daemons of conceptual hydrological modeling: 1. Fidelity and efficiency of time stepping schemes. *Water Resources Research*, *46*(10), 2009WR008894. https://doi.org/10.1029/2009WR008894

Clark, M. P., Slater, A. G., Rupp, D. E., Woods, R. A., Vrugt, J. A., Gupta, H. V., et al. (2008). Framework for Understanding Structural Errors (FUSE): A modular framework to diagnose differences between hydrological models. *Water Resources Research*, *44*(12), W00B02. https://doi.org/10.1029/2007WR006735

Falcone, J. A. (2011). *GAGES-II: Geospatial attributes of Gages for evaluating streamflow* (Tech. Rep.). USGS. https://doi.org/10.3133/70046617

Kavetski, D., & Clark, M. P. (2010). Ancient numerical daemons of conceptual hydrological modeling: 2. Impact of time stepping schemes on model analysis and prediction. *Water Resources Research*, *46*(10), 2009WR008896. https://doi.org/10.1029/2009WR008896

Oudin, L., Hervieu, F., Michel, C., Perrin, C., Andréassian, V., Anctil, F., & Loumagne, C. (2005). Which potential evapotranspiration input for a lumped rainfall–runoff model?: Part 2—Towards a simple and efficient potential evapotranspiration model for rainfall–runoff modelling. *Journal of Hydrology*, *303*(1), 290–306. https://doi.org/10.1016/j.jhydrol.2004.08.026

Raftery, A. E., Gneiting, T., Balabdaoui, F., & Polakowski, M. (2005). Using Bayesian model averaging to calibrate forecast ensembles. *Monthly Weather Review*, *133*(5), 1155–1174. https://doi.org/10.1175/MWR2906.1

Schoups, G., Vrugt, J. A., Fenicia, F., & van de Giesen, N. C. (2010). Corruption of accuracy and efficiency of Markov chain Monte Carlo simulation by inaccurate numerical implementation of conceptual hydrologic models. *Water Resources Research*, *46*(10), 2009WR008648. https://doi.org/10.1029/2009WR008648