

# Water Resources Research®



## REPLY

10.1029/2023WR036550

### Key Points:

- We address the concerns of Knoben and Clark (2023) on the use of the nonparametric estimator with hourly discharge data
- We provide further clarification on the interpretation of the variance estimates obtained with the nonparametric estimator

### Correspondence to:

J. A. Vrugt,  
jasper@uci.edu

### Citation:

de Oliveira, D. Y., & Vrugt, J. A. (2024). Reply to comment by W. Knoben and M. Clark on “the treatment of uncertainty in hydrometric observations: A probabilistic description of streamflow records”. *Water Resources Research*, 60, e2023WR036550. <https://doi.org/10.1029/2023WR036550>

Received 23 OCT 2023

Accepted 5 JAN 2024

## Reply to Comment by W. Knoben and M. Clark on “The Treatment of Uncertainty in Hydrometric Observations: A Probabilistic Description of Streamflow Records”

Debora Y. de Oliveira<sup>1</sup>  and Jasper A. Vrugt<sup>1</sup> 

<sup>1</sup>Department of Civil and Environmental Engineering, University of California, Irvine, CA, USA

**Abstract** In this Reply, we address the concerns of Knoben and Clark (2023, <https://doi.org/10.1029/2022WR034294>) or KC23 that “the assumptions needed to effectively use difference-based variance estimation methods are not always met by hourly streamflow records.” There should be little doubt that the assumptions of our difference-based estimator will sometimes be violated in hourly streamflow records but the results from de Oliveira and Vrugt (2022, <https://doi.org/10.1029/2022wr032263>) and confirmed by the findings in our Reply show that such violations are sporadic enough not to affect much the error variance estimates. Snowmelt as pointed out by KC23 (<https://doi.org/10.1029/2022WR034294>) may not have received sufficient attention in our paper, yet their 365-day record is simply not long enough to demonstrate bias of our discharge error variance estimates (and their dependence on flow level). This would require analysis of a much longer, multi-year, streamflow record of a snowmelt-driven watershed. The snowmelt catchment analyzed in dOV22 (<https://doi.org/10.1029/2022wr032263>) did not demonstrate bias in the discharge error variance estimates. We also provide additional clarification on the interpretation of the variance estimates obtained with the nonparametric estimator, and discuss the main issues in the test case presented in Knoben and Clark (2023, <https://doi.org/10.1029/2022WR034294>).

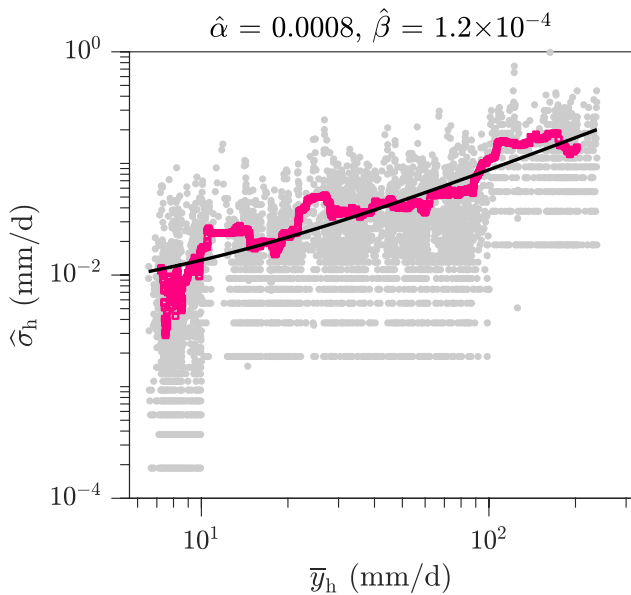
### 1. Introduction

Knoben and Clark (2023, hereafter KC23) expressed some concerns about the method proposed in Vrugt et al. (2005, hereafter V05) and de Oliveira and Vrugt (2022, hereafter dOV22) to estimate the magnitude of aleatory errors in discharge time series. KC23 stated that “the assumptions needed to effectively use the statistical approach of V05; dOV22 are not always met by hourly streamflow records.” We certainly do not rule out the possibility that the assumptions of our difference-based estimator will be violated at times in hourly streamflow records but the results from de Oliveira and Vrugt (2022) and confirmed by the findings in this Reply, demonstrate that such anomalous events do not affect much the error variance estimates. The difference-based variance estimator will provide a robust characterization of the magnitude of aleatory errors and its dependence on flow level if the discharge record is long enough. The 1-year streamflow record of KC23 does not satisfy this length requirement. Furthermore, the suggestion by KC23 that their anomalous event (snowmelt) is expected to repeat every year does not demonstrate bias of our estimator as it disregards the premise of difference-based estimation. Differencing the discharge time series only once will eliminate (or reduce) trends and seasonality. Each additional differencing step removes more memory from the time series ultimately leaving us with a record of random perturbations provided that (a) the data generating process  $H(t)$  is sufficiently smooth and (b) the measurement frequency is high compared to the typical timescale of  $H(t)$ . Thus, even if an apparently anomalous event (snowmelt) repeats every year, annual discharge records will differ and yield different error variance estimates. Only a multi-year streamflow record of a snowmelt-driven watershed will tell us whether the rapidly rising hydrographs due to snowmelt systematically bias the discharge error variance estimates. The snowmelt catchment analyzed in dOV22 did not show such bias (see Figure S5 in dOV22), but, if as KC23 suggest this is the case, then one could follow the suggestion of dOV22 for intermittent catchments and difference the discharge time series additional times.

The main concern expressed in KC23 is further separated into two issues, associated with (a) the appropriateness of local variance estimates,  $\hat{\sigma}_n^2$ , provided by the nonparametric estimator, and (b) the effectiveness of the smoothing procedure. We argue and demonstrate that these perceived issues originate from a misinterpretation and misuse of the nonparametric estimator presented in dOV22, specifically, KC23 inappropriate interpretation

© 2024. The Authors.

This is an open access article under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits use, distribution and reproduction in any medium, provided the original work is properly cited.



**Figure 1.** Scatter plot of hourly discharge  $\bar{y}_h$  (mm/d) and its estimated error standard deviation  $\hat{\sigma}_h$  for the Bow River catchment. Each gray dot signifies a different data pair. The pink squares portray the central moving average of  $\hat{\sigma}_h$  using a window of 100 data pairs. The solid black line is the regression line. Note that we are still using a logarithmic scale in the Figures presented in this Reply given the distribution of the variables under analysis.

of individual  $\hat{\sigma}_h^2$  values as variance estimates and gross overestimation of the aleatory errors. KC23 also questioned whether the magnitude of aleatory errors could be effectively estimated from a streamflow time series without accounting simultaneously for other sources of error.

This Reply is organized along the main comments of KC23. In Section 2 we discuss the robustness of the nonparametric estimator when applied to hourly discharge records. Then, in Section 3 we provide further clarification on the interpretation of the variance estimates provided by the nonparametric estimator. Lastly, Section 4 discusses the use of the proposed nonparametric estimator to estimate aleatory errors in the presence of other (and potentially larger) sources of uncertainty in discharge records.

## 2. Robustness of the Nonparametric Estimator With Hourly Discharge Records

Through a series of synthetic case studies involving catchments with contrasting hydrologic conditions (Figure S1 in dOV22) and varying degrees of homoscedastic and heteroscedastic errors, dOV22 showed that the nonparametric error estimator provides unbiased estimates of the magnitude of aleatory errors of hourly discharge records. The estimator distinguishes random errors from the underlying signal (Section 3.1.1 in dOV22), hence returns aleatory variance estimates of near zero when applied directly to error-free discharge data (not shown). This merely confirms that hourly discharge data satisfy assumptions (a) and (b) of the estimator listed above.

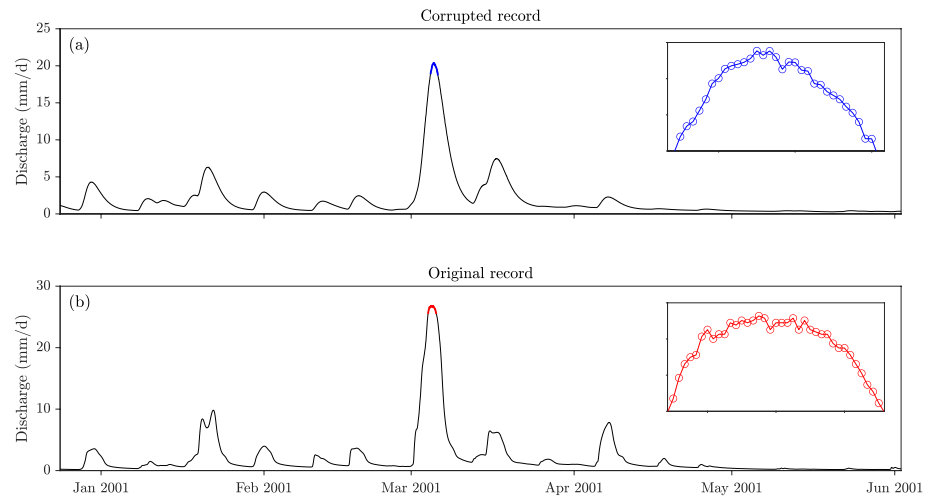
Despite our comprehensive testing of the nonparametric error estimator in dOV22, KC23 is not convinced by the results of our methodology and

argue that there is a “*lack of compelling evidence that these methods can be applied to hourly discharge observations.*” KC23 argue that “*the uncorrelated random nature of the synthetic errors creates time series that are not as smooth as real streamflow records.*” KC23 illustrate this issue by corrupting the discharge record from the Bow River at Banff (Alberta, Canada) with aleatory errors whose standard deviation is more than one order of magnitude larger than its actual (estimated) value. Specifically, KC23 assumed a slope  $\alpha = 0.01$  of the error function (Equation 4 in dOV22), whereas the estimated slope,  $\hat{\alpha}$ , of this record is less than 0.001, more than 10 times smaller! (see Figure 1). The consequences of this gross overestimation of  $\alpha$  are evident in Figure 1 of KC23. The aleatory errors are unrealistically large and the corrupted discharge time series is overly jagged. If  $\hat{\alpha} = 0.0008$  of Figure 1 were used instead, the aleatory errors would be much smaller on average and the perturbed discharge time series would be in much better agreement with the streamflow record. Unlike what KC23 have done, we do not visualize the perturbed ( $\hat{\alpha} = 0.0008$ ) and measured hydrographs as we do not expect the two records to be similar. We will address this issue in the next paragraph.

There is a second problem with the method used by KC23 to generate corrupted streamflow records. This issue is important but less evident. KC23 adds the aleatory errors directly to the discharge record,  $\tilde{y}$ , but this time series already contains errors. Then, additional perturbations are only expected to deteriorate further record smoothness (among others) even if the induced errors are consistent with the aleatory errors expressed by the streamflow records. Instead, we should only add the random errors to a discharge time series that is free of aleatory errors.

In dOV22, different magnitudes of homoscedastic and heteroscedastic errors were tested and used to corrupt a simulated time series (free of aleatory errors) of watersheds with contrasting hydrologic regimes. In Figure 2a, we present such an error-corrupted time series for the Leaf River catchment.

Random errors were added to the simulated discharge record by sampling from a normal distribution with zero mean and heteroscedastic variance  $\hat{\sigma}_h^2$  expressed by the error function of this watershed (we set  $\hat{\alpha} = 0.004$ —see Figure 2 in dOV22). In Figure 2b, we also present the hourly streamflow record. At first glance the two records appear similar. Discrepancies between the time series reflect the combined effect of measurement and modeling errors. When we zoom in on the largest storm event in early March 2001, we observe a strong resemblance between the peaks of the corrupted and original records. Both records express a similar smoothness. Thus, the



**Figure 2.** (a) Error corrupted time series and (b) original discharge record of the Leaf River near Collins, MS (USGS 02472000). The corrupted time series was generated by adding to the simulated hourly discharges  $y_h$  aleatory errors sampled from a normal distribution with zero mean and variance equal to  $(\alpha y_h + \beta m_h)^2$ , where  $m_h$  is the arithmetic mean of the  $y_h$ 's. Coefficients  $\alpha$  and  $\beta$  are derived from the discharge record using the nonparametric estimator. Further information is found in the Supporting Information (SI) of [dOV22](#).

claim of [KC23](#) that “the uncorrelated random nature of the synthetic errors creates time series that are not as smooth as real streamflow records” is not supported by our results.

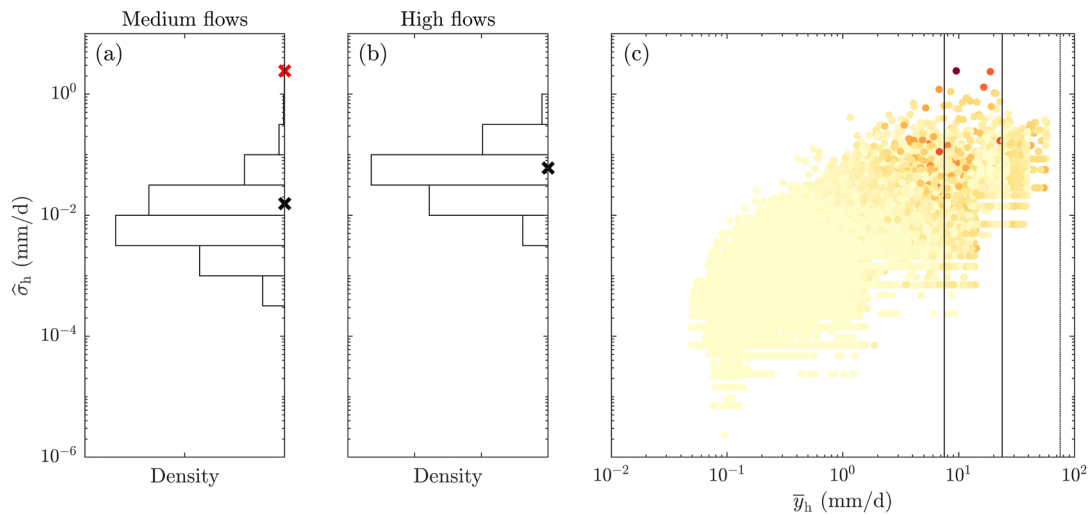
### 3. Clarification on the Interpretation of the Variance Estimates Provided by the Nonparametric Estimator

In this section, we address the alleged drawbacks of our method, that is, (a) the appropriateness of local variance estimates,  $\hat{\sigma}_h^2$ , provided by the nonparametric estimator, and (b) the effectiveness of the smoothing procedure, and argue that these issues originate from a misinterpretation and misuse of the nonparametric estimator.

First, [KC23](#) seems to be looking at each individual value or a few values of  $\hat{\sigma}_h^2$  as an estimate of the variance for a specific flow level. As discussed in [dOV22](#), “each  $\hat{\sigma}_h^2$  individually is rather meaningless, since its value is estimated using only  $k + 1$  points.” Text S1 (SI) of [dOV22](#) presents the derivation of the nonparametric estimator equation for  $k = 1$  and should help clarify that the variance estimate is obtained from all  $\hat{\sigma}_h^2$  values. For homoscedastic errors, this can be accomplished by computing the arithmetic mean of all  $\hat{\sigma}_h^2$  values (Hall et al., 1990). For heteroscedastic errors, we apply the nonparametric estimator locally in the time series after which we summarize the resulting  $(\hat{y}_h, \hat{\sigma}_h)$  data pairs in [dOV22](#) by fitting a linear regression function to the point cloud (e.g., see Figure 1 of this Reply). Thus, the statement made by [KC23](#) that “the estimated standard deviation of the error,  $\hat{\sigma}_\tau$ , tends to be particularly high when the hydrograph changes rapidly, and not, as the error model used by [V05](#) and [dOV22](#) suggests it should be (see Equation 3), when the largest flow magnitudes occur” is misleading, since it implies that we can actually get a robust variance estimate for each (or a few) discharge value(s) separately. We illustrate this point further in Figure 3 using the hourly discharge record of the Leaf River catchment.

While an individual  $(\bar{y}_h, \hat{\sigma}_h)$  data pair in the “medium” flow range (denoted with a red cross) corresponds to the largest value of  $\hat{\sigma}_h$ , there is an overall tendency of  $\hat{\sigma}_h$  to increase with flow level. While it may be true that we yield some of the largest  $\hat{\sigma}_h$ -values when the hydrograph changes rapidly, the so-obtained  $(\bar{y}_h, \hat{\sigma}_h)$  data pairs make up only a small fraction of the point cloud. As a result, these “outliers” have a negligible impact on the overall  $\bar{y}_h - \hat{\sigma}_h$  relationship (see also Figure 1).

The above focus on just a few members of the point cloud distracts from what is truly important in determining the error variance of streamflow records. Only a long data record guarantees a large enough sample of  $(\bar{y}_h, \hat{\sigma}_h)$  data pairs at medium and high flows needed for a robust characterization of the  $\bar{y}_h - \hat{\sigma}_h$  relationship. [dOV22](#) used more than 10 years of streamflow data for each catchment (median of 28 years). This length of the hourly discharge records “was judged [long] enough to infer the error model of each watershed.” To our surprise, [KC23](#)

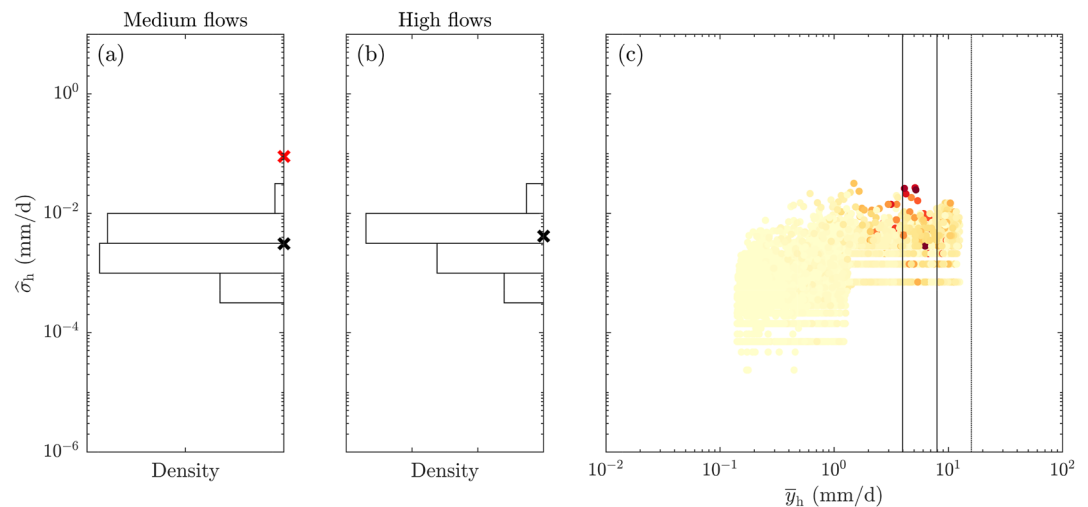


**Figure 3.** Application of the nonparametric estimator to the *entire available discharge record* of the Leaf River near Collins, MS (USGS 02472000): (a) Distribution of  $\hat{\sigma}_h$  for medium flows, (b) distribution of  $\hat{\sigma}_h$  for high flows, (c) relationship between hourly discharge and the error standard deviation  $\hat{\sigma}_h$ . Each dot signifies a different data pair. The red cross in panel (a) corresponds to the largest value of  $\hat{\sigma}_h$  and the black crosses in panels (a) and (b) denote the mean error standard deviations of both flow levels.

used a discharge record that is less than 365 days! This cannot be expected to yield an accurate characterization of the  $\bar{y}_h - \hat{\sigma}_h$  relationship. The  $(\bar{y}_h, \hat{\sigma}_h)$  point cloud is just too small (compare Figure 1 and Figure 2 of [dOV22](#)) and the fitting of the regression line becomes overly sensitive to extremal  $\hat{\sigma}_h$  data points such as the red dots displayed in Figure 3.

Figure 4 illustrates the consequences of using a much too short discharge record. We present the results of the nonparametric estimator using only 1 year of hourly streamflow values from the Leaf River. The value of  $\hat{\sigma}_h$  for “medium” flows is comparable to that of “high” flows. This contradicts our earlier findings in Figure 3, which supports a heteroscedastic  $\bar{y}_h - \hat{\sigma}_h$  relationship.

We do not discount the possibility that for some specific discharge records, the individual  $\hat{\sigma}_h$  values deviate systematically from the assumed linear  $\bar{y}_h - \hat{\sigma}_h$ -relationship. But even if this is true, the method presented in [dOV22](#) is still plausible as “heteroscedasticity of discharge errors is expected since a power function (or



**Figure 4.** Application of the nonparametric estimator to a 1 year extract of the discharge record of the Leaf River: (a) distribution of  $\hat{\sigma}_h$  for medium flows, (b) distribution of  $\hat{\sigma}_h$  for high flows, (c) relationship between hourly discharge  $\bar{y}_h$  and the error standard deviation  $\hat{\sigma}_h$ . Each dot signifies a different data pair. The red cross in panel (a) corresponds to the largest value of  $\hat{\sigma}_h$  and the black crosses in panels (a) and (b) portray the mean error standard deviations of both flow levels.

*similar*) is used to transform stage measurements into discharge estimates.” The linear model should be able to capture the main features of this relationship. If so desired one can use more flexible parametric expressions for the  $(\bar{y}_h, \hat{\sigma}_h)$ -relationship. Model selection methods may then be used to judge which expression is most supported by the point cloud. Moreover, one could think of applying the nonparametric estimator directly to stage records instead. Along with other information about the uncertainty in discharge records (e.g., temporal changes in the rating curve), this should provide a better description of the variance of aleatory errors, as recognized in dOV22: “*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  $\hat{\sigma}_h$  and discharge magnitude than the linear heteroscedastic model used herein.*”

#### 4. On the Use of the Nonparametric Estimator to Estimate the Variance of Aleatory Errors Given Other Sources of Uncertainty in Streamflow Records

In this last section, we address the comment made in KC23 that “*the difference-based variance estimation method is divorced from current understanding of the sources of errors in streamflow records.*” As stated in dOV22 and recognized by KC23, we are only concerned with the estimation of aleatory uncertainty in discharge records. We contend that such aleatory errors can only be inferred in isolation from all other sources of modeling error. Simply by looking in detail at the discharge record. In essence, we are seeking the discharge time series of the data generating process. Whatever differences remain between this unknown signal and the discharge record are random errors and cannot be explained unless we use an overparameterized model. But this overfitting will show during an independent evaluation period. The nonparametric estimator builds on decades of research in the statistical literature and has a strong theoretical underpinning. This alone should inspire confidence in the ability of the nonparametric estimator to differentiate random errors from the underlying signal. This was confirmed by the synthetic case studies of dOV22.

We do not dispute that other sources of uncertainty may have a larger contribution to the total uncertainty of discharge records. This was recognized in dOV22: “*The nonparametric estimator assumes that the discharge errors are of aleatory nature only. We have confirmed with the synthetic case study 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.*”

#### Data Availability Statement

A MATLAB implementation of the NonParametric aleatory Error Estimator (NoPEE) is available from GitHub at <https://github.com/jaspervrugt/NoPEE> licensed under MIT. The daily discharge record of the Leaf River catchment was obtained from the CAMELS data set, which is described in Newman et al. (2015) and can be downloaded from <https://dx.doi.org/10.5065/D6MW2F4D>. The hourly streamflow data of Gauch et al. (2020) is available at <https://doi.org/10.5281/zenodo.4072700>. The hourly streamflow record from the Bow River catchment can be downloaded from <https://zenodo.org/records/8106091>.

#### References

- 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(11), e2022WR032. <https://doi.org/10.1029/2022WR032263>
- 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”. *Zenodo*. <https://doi.org/10.5281/zenodo.4072701>
- Hall, P., Kay, J. W., & Titterton, D. M. (1990). Asymptotically optimal difference-based estimation of variance in nonparametric regression. *Biometrika*, 77(3), 521–528. <https://doi.org/10.2307/2336990>
- Knoben, W. J. M., & Clark, M. (2023). Comment on “The treatment of uncertainty in hydrometric observations: A probabilistic description of streamflow records” by de Oliveira and Vrugt. *Water Resources Research*, 58(11), e2022WR032263. <https://doi.org/10.1029/2022WR034294>
- 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 hydro-meteorological 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>
- 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>

#### Acknowledgments

The first author gratefully acknowledges the financial support received from the Brazilian Federal Agency for Support and Evaluation of Graduate Education (CAPES), Grant 88881.174456/2018-01.