A log-sinh transformation for data normalization and variance stabilization

Q. J. Wang,[1] D. L. Shrestha,[1] D. E. Robertson,[1] and P. Pokhrel[1]

Received 26 May 2011; revised 12 March 2012; accepted 23 March 2012; published 8 May 2012.

[1] When quantifying model prediction uncertainty, it is statistically convenient to represent model errors that are normally distributed with a constant variance. The Box-Cox transformation is the most widely used technique to normalize data and stabilize variance, but it is not without limitations. In this paper, a log-sinh transformation is derived based on a pattern of errors commonly seen in hydrological model predictions. It is suited to applications where prediction variables are positively skewed and the spread of errors is seen to first increase rapidly, then slowly, and eventually approach a constant as the prediction variable becomes greater. The log-sinh transformation is applied in two case studies, and the results are compared with one- and two-parameter Box-Cox transformations.

Citation: Wang, Q. J., D. L. Shrestha, D. E. Robertson, and P. Pokhrel (2012), A log-sinh transformation for data normalization and variance stabilization, *Water Resour. Res.*, *48*, W05514, doi:10.1029/2011WR010973.

1. Introduction

[2] In many hydrological applications, it is highly desirable to quantify the uncertainty of hydrological model predictions. Hydrological model prediction variables are often highly skewed and their errors are typically not normal and display nonconstant variance (heteroscedasticity) [*Schoups and Vrugt*, 2010]. Various heteroscedastic error models have been proposed for hydrological applications [e.g., *Schoups and Vrugt*, 2010; *Thyer et al.*, 2009]. However, the most common treatment is to transform the prediction variable to normalize errors and stabilize their variance in the transformed space. This approach is conceptually attractive as it is statistically convenient to deal with errors that are normally distributed with constant variance.

[3] The most widely used transformation for such a purpose is the Box-Cox transformation with one or two parameters [*Box and Cox*, 1964]. The Box-Cox transformation has proven to be highly successful for many applications, including hydrological applications [e.g., *Kuczera*, 1983; *Bates and Campbell*, 2001; *Thyer et al.*, 2002; *Yang et al.*, 2007; *Engeland et al.*, 2010]. However, it is not without limitations [*Sakia*, 1992]. As will be demonstrated later in this paper, the one-parameter Box-Cox transformation is not always able to achieve variable normalization and variance stabilization. As a result, the probability distribution quantifying prediction uncertainty is sometimes unreliable. The two-parameter Box-Cox transformation is more flexible and can achieve better variable normalization and variance stabilization over most of the data range. However, this flexibility may lead to the assignment of unrealistically large uncertainty to predictions of large events.

[4] The nonparametric method of normal quantile transformation (NQT) is also widely used in hydrology because of its ability to deal with nonstandard frequency distribution shapes [e.g., *Montanari and Brath*, 2004]. However, it has a number of limitations. While the method aims to normalize model predictions and observations, there is no guarantee that model prediction errors after transformation are also normally distributed and have stable variance. Extrapolation beyond the available data range is also problematic.

[5] In this paper, we derive a new parametric transformation based on a pattern of errors commonly seen in hydrological model predictions. We compare the transformation with the one- and two-parameter Box-Cox transformations through two case studies.

2. New Transformation

[6] Figure 1a shows the errors of simulated daily flows of the GR4J daily rainfall-runoff model [*Perrin et al.*, 2003] applied to the Hurdle Creek catchment in Victoria, Australia. Error is defined here as the difference between observed and model simulated flows. The errors are displayed against the simulated flows. The spread of the errors is seen to grow with the simulated flows, but the growth slows down and eventually tapers off as the simulated flows become greater.

[7] A similar pattern of errors can be seen in Figure 1b. Here the WAPABA monthly water balance model [*Wang et al.*, 2011] was applied to the Lake Hume catchment located in southeastern Australia. Monthly flow volumes were simulated. This pattern of errors is also seen in the work of *Schoups and Vrugt* [2010] and indeed in many other hydrological modeling applications we have encountered.

[8] Denote the model simulation of a variable y as y_{sim}. Assume that any simulation bias that may exist has already been corrected. The expectation and variance of y are, respectively,

$$\mathrm{E}[y] = y_{\mathrm{sim}}, \tag{1}$$

$$\mathrm{Var}[y] = [s(y_{\mathrm{sim}})]^2. \tag{2}$$

[1]CSIRO Land and Water, Highett, Victoria, Australia.

Published in 2012 by the American Geophysical Union

19447973, 2012, 5, Downloaded from https://agupubs.onlinelibrary.wiley.com/doi/10.1029/2011WR010973 by Readcube (Labtiva Inc.), Wiley Online Library on [22/12/2022]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

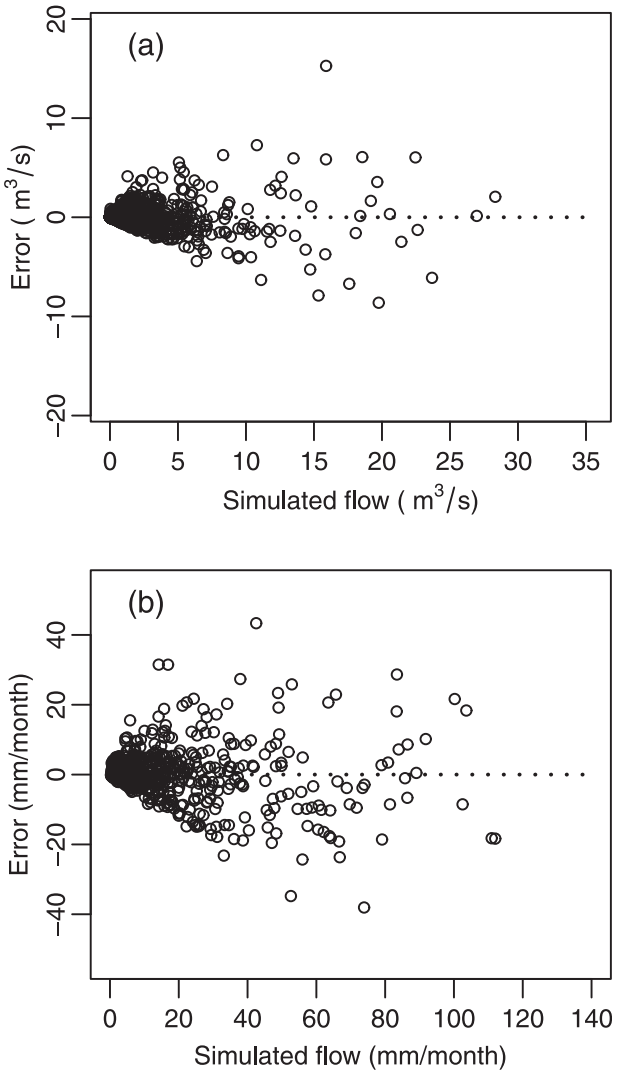

Figure 1. Model error spread of (a) GR4J simulated daily flows for the Hurdle Creek catchment in Victoria, Australia. (b) WAPABA simulated monthly flows for the Lake Hume catchment located across the border between Victoria and New South Wales, Australia.

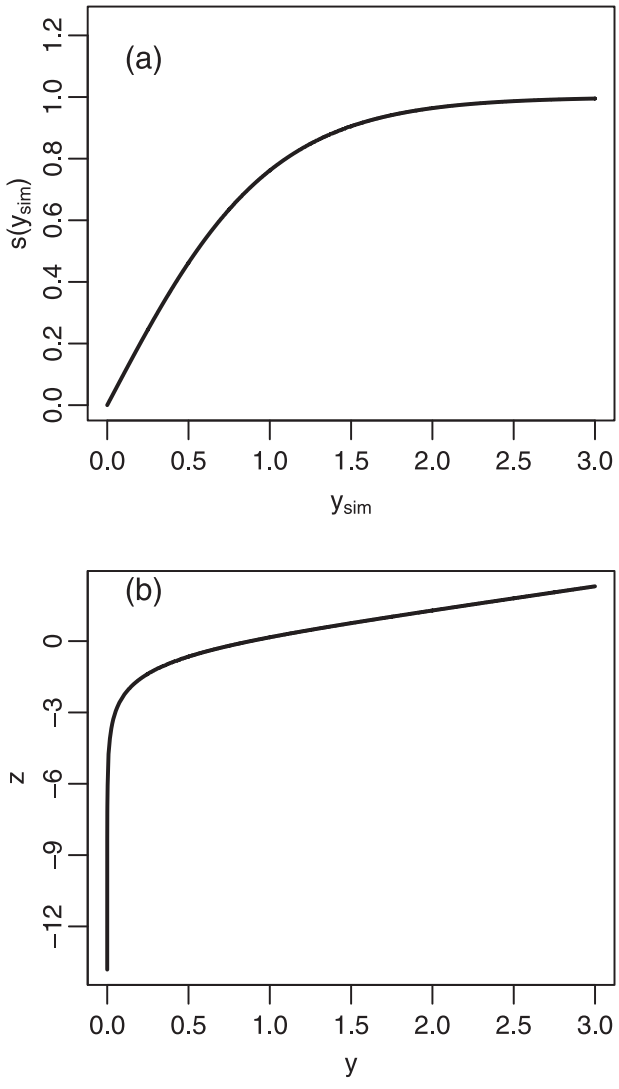

Figure 2. (a) Standard deviation as a function of simulated variable (for equation (4) with $s_0 = 1$, $a = 0$, and $b = 1$). (b) Transformed variable versus untransformed variable (for equation (6) with $a = 0$ and $b = 1$).

[9] A general variance stabilizing transformation may be deduced as ([*Huber et al.*, 2002]; a note at http://www.stat.ufl.edu/~winner/sta6207/transform.pdf)

$$z = \int^{y} \frac{1}{s(y_{\text{sim}})} dy_{\text{sim}}. \tag{3}$$

[10] In deriving the new transformation, we assume that the standard deviation follows

$$s(y_{\text{sim}}) = s_0 \tanh(a + b y_{\text{sim}}), \tag{4}$$

where s_0, a, and b are parameters. Equation (4) approximates the kind of spread pattern of errors shown in Figure 1. The standard deviation increases with y_{sim}, but the rate of increase tapers off and the standard deviation approaches s_0 as y_{sim} becomes larger (Figure 2a). In sections 3 and 4, we will demonstrate that equation (4) is a reasonable assumption.

[11] Substituting (4) into (3) and taking the integration yields

$$z = \frac{1}{s_0}\frac{1}{b}\log(\sinh[a + by]) + c. \tag{5}$$

[12] The constants s_0 and c are removed from (5) to give the transformation a cleaner form,

$$z = \frac{1}{b}\log(\sinh[a + by]). \tag{6}$$

[13] We call it the log-sinh transformation. The errors in the transformed space have a constant variance when (4) is satisfied.

[14] A plot of z versus y for $a = 0$ and $b = 1$ is shown in Figure 2b. The transformation stretches out the z range far

19447973, 2012, 5, Downloaded from https://agupubs.onlinelibrary.wiley.com/doi/10.1029/2011WR010973 by Readcube (Labtiva Inc.), Wiley Online Library on [22/12/2022]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

more when y is small than when y is large. This has the effect of normalizing positively skewed variables such as flows, and consequently the variable errors in the transformed space are also more normally distributed than before transformation.

[15] In normalizing variables and stabilizing variances, the transformation brings about opportunities for successfully applying simple error models to quantify model prediction uncertainty.

3. Case Study 1

[16] Here we first demonstrate the application of the log-sinh transformation to the example shown earlier in Figure 1b. Monthly flow volume y was simulated as y_{sim} using the WAPABA model for the Lake Hume catchment (catchment area 12,185 km^2; annual rainfall 882 ± 228 mm; annual runoff 230 ± 137 mm; annual potential evapotranspiration 1281 ± 49 mm). The model was calibrated for 1950–1955 to achieve a good fit between the observed flows and simulated flows [*Wang et al.*, 2011]. It was then used to simulate flows for 1956–2008. The aim here is to construct an error model to describe the uncertainty of the model simulations.

[17] The observed flow and simulated flow are first log-sinh transformed to z and z_{sim}, respectively. The error in the transformed space is assumed to be normally distributed with a constant standard deviation σ, giving

$$p(z) = \mathrm{N}(z_{sim}, \sigma) \tag{7}$$

and

$$p(y) = J_{z\to y} p(z), \tag{8}$$

where $J_{z\to y}$ is the Jacobian determinant of the transformation from z to y,

$$J_{z\to y} = \frac{dz}{dy} = \coth(a + by). \tag{9}$$

[18] A Bayesian inference of the parameters a, b, and σ is made by assuming that the prior for the parameters is inversely proportional to σ [*Gelman et al.*, 2004; *Wang and Robertson*, 2011]. Zero flow events are treated as having censored data that are known to be equal or below zero but not known for their precise values [*Wang and Robertson*, 2011]. Point estimates of the parameters are obtained through a maximum a posteriori (MAP) solution, and parameter uncertainty is not further considered. Figure 3a shows that the errors after transformation are normally distributed. Figure 3b shows that the assumed theoretical relationship between the standard deviation and simulated flow, as represented by equation (4) (but evaluated here through Monte Carlo simulations), approximates well the standard error estimated from data. Figures 3c and 4a give the 0.05 and 0.95 quantiles of the modeled error distributions. The quantiles appear to be consistent with the data.

[19] For the purpose of comparison, we also apply Box-Cox transformations to the same data set. A two-parameter Box-Cox transformation has the form [*Box-Cox*, 1964]

$$z = \frac{(y+\varepsilon)^{\lambda} - 1}{\lambda}. \tag{10}$$

[20] The most widely used is the one-parameter Box-Cox transformation with ε set to zero. In this study, however, we apply a one-parameter Box-Cox transformation by fixing ε to 1.00×10^{-6} to cope with zero flow events. We also apply the more general two-parameter Box-Cox transformation with both ε and λ being determined from data.

[21] The results from using the one-parameter Box-Cox transformation are included in Figures 3 and 4. The results from using the two-parameter Box-Cox transformation are nearly identical to using the one-parameter transformation and therefore not included. The difference between the Box-Cox transformations and the log-sinh transformation is mainly in the high-simulated flow range, where the log-sinh transformation gives a smaller standard deviation (Figure 3b)

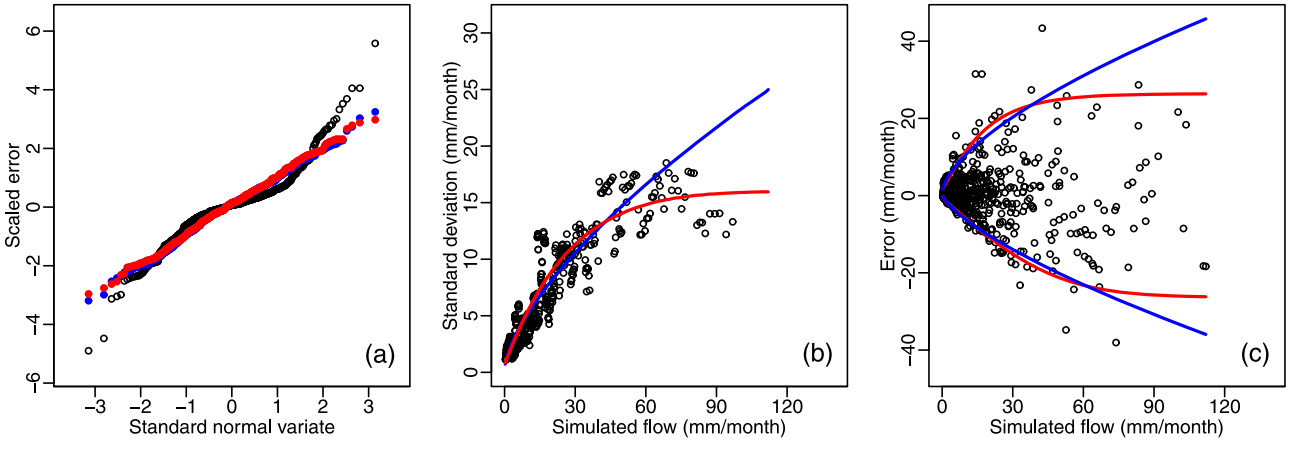

Figure 3. Diagnostic plots of the error models for WAPABA simulated monthly flows for the Lake Hume catchment. (a) Normal probability plot of scaled errors before (open circles) and after (solid circles) transformation of the simulated flows. Scaled errors are errors divided by all sample standard deviation. (b) Standard errors estimated from a moving window of 10 data points (circles), and theoretical relationship between standard deviation and simulated flow (line). (c) The 0.05 and 0.95 quantiles (lines) of modeled error distribution compared with errors of the simulated flows (circles). Colors: Red for the log-sinh transformation, and blue for the one-parameter Box-Cox transformation.

19447973, 2012, 5, Downloaded from https://agupubs.onlinelibrary.wiley.com/doi/10.1029/2011WR010973 by Readcube (Labtiva Inc.), Wiley Online Library on [22/12/2022]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

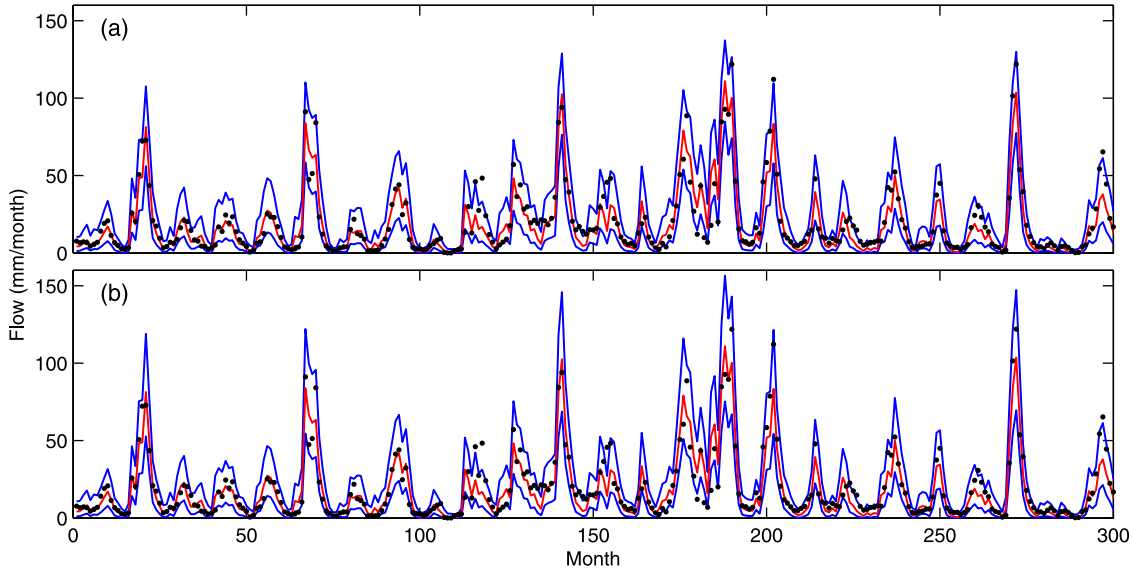

Figure 4. Sample time series plots of the error models for WAPABA simulated monthly flows for the Lake Hume catchment: (a) using the log-sinh transformation, and (b) using the one-parameter Box-Cox transformation. Legends: Black dots for observed flows, red line for modeled median, blue lines for 0.05 and 0.95 quantiles of modeled error distribution.

and therefore a narrower uncertainty band (Figure 3c). There are probably not sufficient data points in the high-simulated flow range for making any firm conclusions, but visually the log-sinh transformation appears to be more consistent with the data.

[22] Table 1 compares the performances in flow predictions of the three transformations in terms of the Akaike information criteria (AIC), continuous ranked probability score (CRPS) [e.g., *Matheson and Winkler*, 1976; *Wang et al.*, 2009], and percentage of data points falling into the 90% credible interval. The Box-Cox transformations are slightly better in AIC and percentage cover, while the log-sinh transformation is slightly better in CRPS.

[23] We conclude from this case study that the log-sinh transformation is able to model the error variance and normalize the data. Its performance in flow predictions is comparable to that of the Box-Cox transformations.

4. Case Study 2

[24] We now present a second example to contrast the log-sinh transformation with the one- and two-parameter Box-Cox transformations. The same after-transformation error model (equation (7)) is applied to GR4J simulated daily flows for the Allyn River catchment in New South Wales, Australia (catchment area 205 km^2; annual rainfall 1189 $\pm$ 255 mm; annual runoff 350 $\pm$ 193 mm; annual potential evapotranspiration 1271 $\pm$ 39 mm). This example is chosen because the daily flows there are highly skewed and require a strong transformation to be normalized. The GR4J model was calibrated for 1992–2006, and the parameters of the transformations and error model are inferred in the same way as for the Lake Hume example.

[25] The results are shown in Figures 5 and 6 and in Table 2. The errors after the transformations are approximately normally distributed for all three transformations (Figure 5a). The log-sinh transformation appears to model well the standard deviation (Figure 5b) and uncertainty band (Figure 5c). However, the Box-Cox transformations poorly model the standard deviation (Figure 5b) and uncertainty band (Figure 5c). For a moderate- to high-simulated flow range, the one-parameter Box-Cox transformation underestimate the standard deviation and uncertainty band, while the two-parameter Box-Cox transformation severely overestimate the standard deviation and uncertainty band. This is also reflected in the time series plots in Figure 6. Note that for the two-parameter Box-Cox transformation, the results presented in Figures 5 and 6 and in Table 2 have been calculated by applying a limit of flow not exceeding 10 times the highest historically observed flow. Without applying this limit, the standard deviation and uncertainty band would have rapidly approached infinity, because of the thick upper tail of the Box-Cox transformed normal distribution associated with a negative λ of -0.323 (Table 2).

[26] Interestingly, the two-parameter Box-Cox transformation results in the best AIC value and overall percentage cover (Table 2). There may be two reasons for this. First,

Table 1. Parameter Values and Performances of Error Models for WAPABA Simulated Monthly Flows for the Lake Hume Catchment

Transformation	Parameter Values	Akaike Information Criterion (AIC)	Continuous Ranked Probability Score (CRPS)	Data Points Falling Into 90% Credible Interval
Log-sinh	$a = 0.0359$; $b = 0.0286$; $\sigma = 16.0$	3574	3.554	88.2%
One-parameter Box-Cox	($\varepsilon = 1.00 \times 10^{-6}$), $\lambda = 0.335$; $\sigma = 1.07$	3553	3.563	88.5%
Two-parameter Box-Cox	$\varepsilon = 2.69 \times 10^{-9}$, $\lambda = 0.335$; $\sigma = 1.07$	3555	3.563	88.5%

19447973, 2012, 5, Downloaded from https://agupubs.onlinelibrary.wiley.com/doi/10.1029/2011WR010973 by Readcube (Labtiva Inc.), Wiley Online Library on [22/12/2022]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

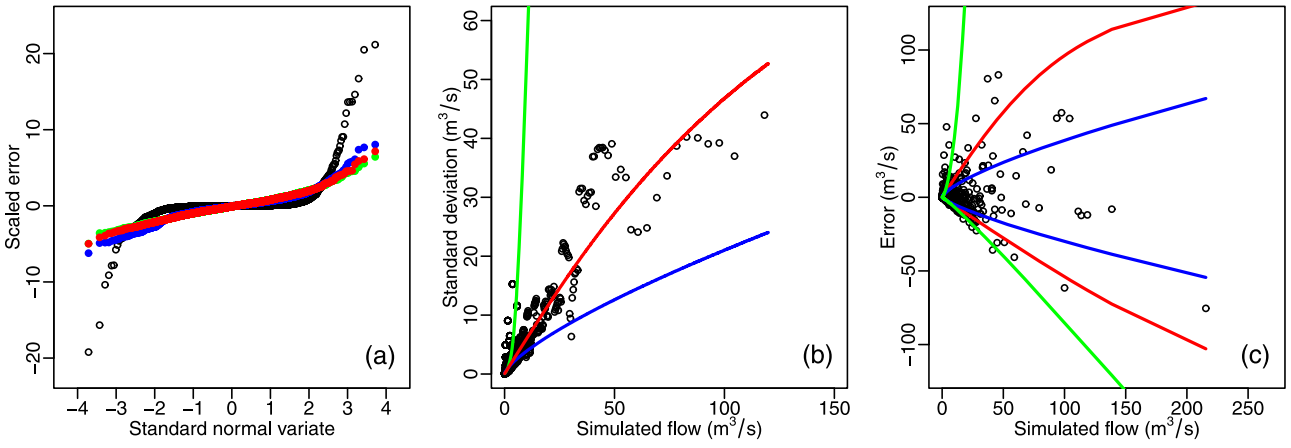

Figure 5. Diagnostic plots of the error models for GR4J simulated daily flows for the Allyn River catchment. (a) Normal probability plot of scaled errors before (open circles) and after (solid circles) transformation of the simulated flows. Scaled errors are errors divided by all sample standard deviation. (b) Standard errors estimated from a moving window of 10 data points (circles), and theoretical relationship between standard deviation and simulated flow (line). (c) The 0.05 and 0.95 quantiles (lines) of modeled error distribution compared with errors of the simulated flows (circles). Colors: Red for the log-sinh transformation, blue for the one-parameter Box-Cox transformation, and green for the two-parameter Box-Cox transformation.

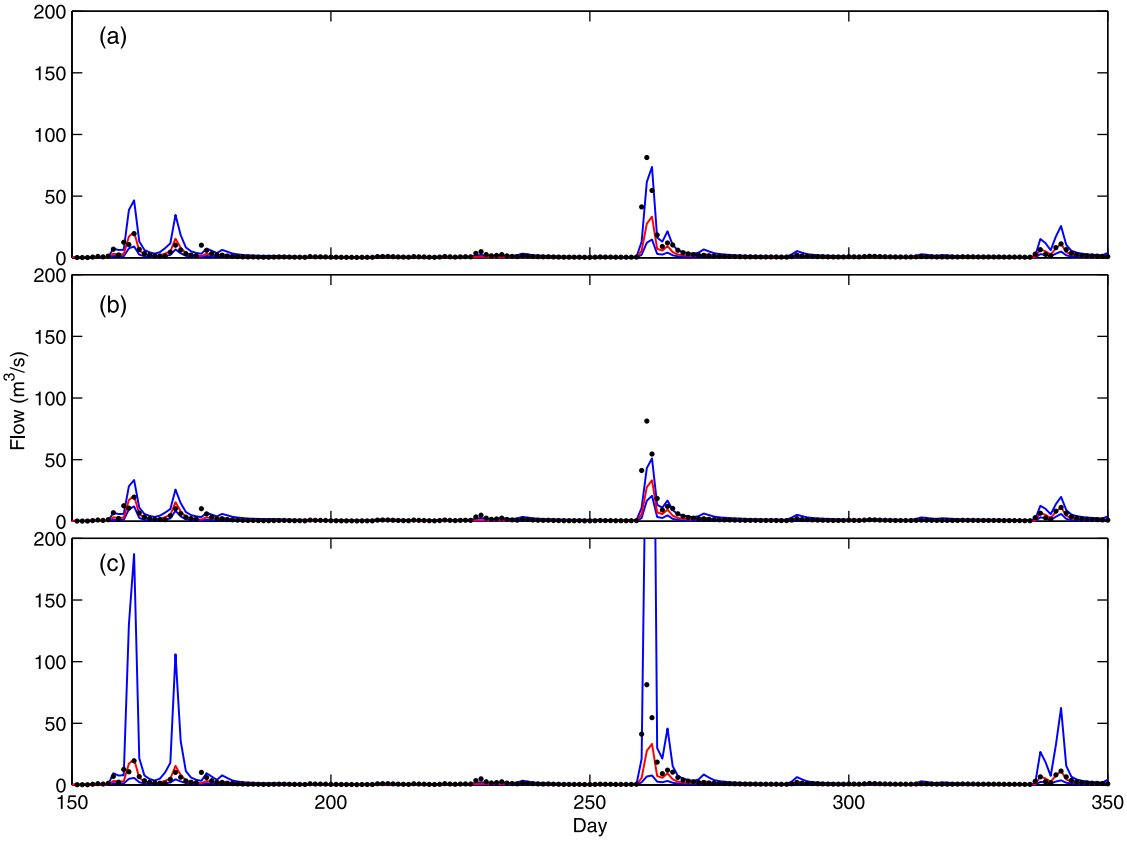

Figure 6. Sample time series plots of the error models for GR4J simulated daily flows for the Allyn River catchment: (a) using the log-sinh transformation, (b) using the one-parameter Box-Cox transformation, and (c) using the two-parameter Box-Cox transformation. Legends: Black dots for observed flows, red lines for modeled median, blue lines for 0.05 and 0.95 quantiles of modeled error distribution.

19447973, 2012, 5, Downloaded from https://agupubs.onlinelibrary.wiley.com/doi/10.1029/2011WR010973 by Readcube (Labtiva Inc.), Wiley Online Library on [22/12/2022]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Table 2. Parameter Values and Performances of Error Models for GR4J Simulated Daily Flows for the Allyn River Catchment

Transformation	Parameter Values	Akaike Information Criterion (AIC)	Continuous Ranked Probability Score (CRPS)	Data Points Falling Into 90% Credible Interval
Log-sinh	$a = 4.72 \times 10^{-4}$; $b = 5.59 \times 10^{-3}$; $\sigma = 88.6$	4662	0.665	91.0%
One-parameter Box-Cox	($\varepsilon = 1.00 \times 10^{-6}$), $\lambda = 0.254$; $\sigma = 0.668$	6749	0.693	92.5%
Two-parameter Box-Cox	$\varepsilon = 0.264$, $\lambda = -0.323$; $\sigma = 0.357$	4359	0.940	90.3%

the two-parameter Box-Cox transformation models well the low flows, which constitute most of the data points. Second, for moderate- and high-simulated flows, the observed flows fall into the middle range of the uncertainty distribution. Thus, the poorly defined distribution upper tail has little effect on the AIC evaluation. However, the CRPS as a measure of the performance of the full predictive distribution, is affected by the distribution upper tail. Indeed, the two-parameter Box-Cox transformation gives the poorest CRPS value, which would have been much worse if the upper flow limit were not applied. In comparison with the one-parameter Box-Cox transformation, the log-sinh transformation has much better AIC value as well as better CRPS and percentage cover.

[27] We conclude that for this second case study, the log-sinh transformation produces a reasonable uncertainty band, while the one-parameter Box-Cox transformation appears to result in an uncertainty band that is too narrow, and the two-parameter Box-Cox transformation too wide and unstable.

[28] The daily flow data in this case study are highly skewed and thus contain many more data points in the low flow range than high flow range. For this reason, the Bayesian MAP estimate of the error model parameters is expected to be heavily influenced by the data in the low flow range. This may partially explain why the fitted Box-Cox transformations deviate from the data in the high flow range. However, it is also likely that the intrinsic relationships of error variance and simulated flow, as implied by the Box-Cox transformations, are inappropriate, making it difficult to fit the full range of data without applying further constraints to the parameters. In contrast, the log-sinh transformation, which has been developed based on an empirical relationship of error variance and simulated flow, is able to appropriately track the data well for both low flows and high flows. Our analyses of flow and modeling data from a range of catchments generally support the use of the log-sinh transformation in preference to the Box-Cox transformations.

5. Concluding Discussions

[29] For quantifying model prediction uncertainty, it is in general mathematically convenient to model errors that are normally distributed with a constant variance. Transformation is often applied to normalize data and stabilize variance. In this paper, a log-sinh transformation is derived. The transformation is particularly suited to applications where prediction variables are positively skewed and the spread of errors is seen to first increase rapidly, then slowly, and eventually approach a constant as the prediction variables become greater.

[30] In our first case study, we show that the log-sinh transformation performs similarly to the Box-Cox transformations. In our second case study, however, we show that the log-sinh transformation is clearly superior to the Box-Cox transformations. For moderate- and high-simulated flow range, the one-parameter Box-Cox transformation produces an uncertainty band that is too narrow, the two-parameter Box-Cox transformation one too wide and unstable. In contrast, the log-sinh transformation is able to produce an uncertainty band that appears to be consistent with the data.

[31] To keep this study focused, the rainfall-runoff models are calibrated first and the error models fitted next. Furthermore, autocorrelations in model errors (lag-1 autocorrelation coefficient 0.34 in case study 1 and 0.10 in case study 2) are ignored. In practice, however, it makes more sense to jointly infer the parameters of rainfall-runoff and error models (including autocorrelations) to allow for parameter interactions. This is indeed the approach we take in our routine applications.

[32] We also suggest that point estimation or probabilistic inference of the log-sinh transformation parameters are performed on $log(a)$ and $log(b)$ instead of a and b. This is because the transformation is highly sensitive to a and b when they are of small values and becomes less so as they get larger. The sensitivity is more uniform on the $log(a)$ and $log(b)$ parameter space. For this reason, parameter estimation or inference on the $log(a)$ and $log(b)$ parameter space tends to be more straightforward. In our experience, a starting range of $[-15, 0]$ may be reasonable for both $log(a)$ and $log(b)$ for many applications. One is cautioned against allowing high values of $log(a)$ and $log(b)$, as the log-sinh transformation may simply approach a linear shift, which has no effect on data normalization and variance stabilization.

[33] **Acknowledgments.** We would like to thank our colleagues Tom Pagano and Prasantha Hapuarachchi for making available the GR4J calibration data for the Hurdle Creek and Allyn River catchments. Dr. Pagano also made useful comments on a draft of the paper. We are grateful to editor Hoshin Gupta and three anonymous reviewers for their valuable suggestions.

References

Bates, B. C., and E. P. Campbell (2001), A Markov chain Monte Carlo scheme for parameter estimation and inference in conceptual rainfall-runoff modeling, *Water Resour. Res.*, *37*(4), 937–947, doi:10.1029/2000WR900363.

Box, G. E. P., and D. R. Cox (1964), An analysis of transformations, *J. R. Stat. Soc., Ser. B*, *26*, 296–311.

Engeland, K., B. Renard, I. Steinsland, and S. Kolberg (2010), Evaluation of statistical models for forecast errors from the HBV model, *J. Hydrol.*, *384*(1–2), 142–155.

Gelman, A., J. B. Carlin, H. S. Stern, and D. B. Rubin (2004), *Bayesian Data Analysis*, 668 pp., Chapman and Hall, London, U. K.

Huber, W., A. von Heydebreck, H. Sültmann, A. Poustka, and M. Vingron (2002), Variance stabilization applied to microarray data calibration and to the quantification of differential expression, *J. Bioinformatics*, *18*(Suppl. 1), S96–S104 doi:10.1093/bioinformatics/18.suppl_1.S96.

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

19447973, 2012, 5, Downloaded from https://agupubs.onlinelibrary.wiley.com/doi/10.1029/2011WR010973 by Readcube (Labtiva Inc.), Wiley Online Library on [22/12/2022]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Matheson, J. E., and R. L. Winkler (1976), Scoring rules for continuous probability distributions, *Manage. Sci.*, *22*, 1087–1095, doi:10.1287/mnsc.22.10.1087.

Montanari, A., and A. Brath (2004), A stochastic approach for assessing the uncertainty of rainfall-runoff simulations, *Water Resour. Res.*, *40*, W01106, doi:10.1029/2003WR002540.

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

Sakia, R. M. (1992), The Box-Cox transformation technique: A review, *J. R. Stat. Soc., Ser. D*, *41*(2), 169–178.

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

Thyer, M., G. Kuczera, and Q. J. Wang (2002), Quantifying parameter uncertainty in stochastic models using the Box-Cox transformation, *J. Hydrol.*, *265*(1–4), 246–257, doi:10.1016/S0022-1694(02)00113-0.

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

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

Wang, Q. J., D. E. Robertson, and F. H. S. Chiew (2009), A Bayesian joint probability modeling approach for seasonal forecasting of streamflows at multiple sites, *Water Resour. Res.*, *45*, W05407, doi:10.1029/2008WR007355.

Wang, Q. J., T. C. Pagano, S. L. Zhou, H. A. P. Hapuarachchi, L. Zhang, and D. E. Robertson (2011), Monthly versus daily water balance models in simulating monthly runoff, *J. Hydrol.*, *404*, 166–175, doi:10.1016/j.jhydrol.2011.04.027.

Yang, J., P. Reichert, K. C. Abbaspour, and H. Yang (2007), Hydrological modelling of the Chaohe basin in China: Statistical model formulation and Bayesian inference, *J. Hydrol.*, *340*(3–4), 167–182.

P. Pokhrel, D. E. Robertson, D. L. Shrestha, and Q. J. Wang, CSIRO Land and Water, PO Box 56, Highett, Victoria, VIC 3190, Australia. (QJ.Wang@csiro.au)

19447973, 2012, 5, Downloaded from https://agupubs.onlinelibrary.wiley.com/doi/10.1029/2011WR010973 by Readcube (Labtiva Inc.), Wiley Online Library on [22/12/2022]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License