Calibrating hydrologic models in flow-corrected time

Tyler Smith,[1] Lucy Marshall,[2] and Brian McGlynn[3]

Received 26 August 2013; revised 13 November 2013; accepted 13 December 2013; published 15 January 2014.

[1] Modeling streamflow hydrographs can be a highly complex problem, particularly due to difficulties caused by multiple dominant streamflow states, switching of dominant streamflow generation mechanisms temporally, and dynamic catchment responses to precipitation inputs based on antecedent conditions. Because of these complexities and the extreme heterogeneity that can exist within a single catchment, model calibration techniques are generally required to obtain reasonable estimates of the model parameters. Models are typically calibrated such that a best fit is determined over the entire period of simulation. In this way, each time step explicitly carries equal weight during the calibration process. Data transformations (e.g., logarithmic or square root) are a common way of modifying the calibration process by scaling the magnitude of the observations. Here we consider a data transformation that is focused on the time domain rather than the data domain. This approach, previously employed in transit time modeling literature, conceptually stretches time during high streamflows and compresses it during low streamflow periods, dynamically weighting streamflows in the time domain. The transformation, known as flow-corrected time, is designed to provide greater weight to time periods with larger hydrologic flux. Here the flow-corrected time transformation is compared to a baseline untransformed case and the commonly employed logarithmic transformation. Considering both visual and numerical (Nash-Sutcliffe efficiency) assessments, we demonstrate that over the time periods that dominate hydrologic flux the flow-corrected time transformation resulted in improved fits to the observed hydrograph.

Citation: Smith, T., L. Marshall, and B. McGlynn (2014), Calibrating hydrologic models in flow-corrected time, *Water Resour. Res.*, *50*, 748–753, doi:10.1002/2013WR014635.

1. Introduction

[2] The use of predictive models necessitates a careful examination of model assumptions, fit, and reliability. This is particularly important to hydrologic modeling, where the underlying streamflow generating mechanisms are not always well understood or measured at the site and/or scale of interest. As a consequence, the use of predictive streamflow models is almost exclusively predicated on generating a good estimate of model parameters via some type of calibration procedure [*Madsen et al.*, 2002].

[3] In general, all approaches to model calibration in hydrologic settings, whether they be global optimization techniques [e.g., *Duan et al.*, 1992; *Tolson and Shoemaker*, 2007] or uncertainty-based methods [e.g., *Beven and Binley*, 1992; *Kuczera and Parent*, 1998] are driven by the use of cost functions to evaluate the goodness-of-fit between the observed and model predicted streamflows [refer to *Gupta and Kling*, 2011]. For the vast majority of cases, such fit measures are evaluated as the correspondence between observed and modeled streamflow for each time step. In other words, each time step carries equal weight when estimating the fit. However, many traditional goodness-of-fit measures used in hydrologic applications, such as Nash-Sutcliffe Efficiency (NSE) and root mean squared error (RMSE), can result in a degree of implicit weighting of the model fit toward better agreement at observations with larger errors. This often occurs at high flows due to the typical distribution of model errors (i.e., larger flows tend to have larger error). Methods, such as multiobjective model calibration schemes [e.g., *Gupta et al.*, 1998; *Madsen*, 2003; *Vrugt et al.*, 2003; *Wagener et al.*, 2001], have been proposed and applied to address this natural weighting, though such approaches are less commonly utilized than single objective calibration. Regardless, the intent of typically employed objective functions for hydrologic model calibration is to assign equal importance to all observations across the duration of the model simulation.

[4] But, should this always be the goal in model optimization? What are the unintended consequences of attempting to fit all observations equally? Are all observations of equal importance? Indeed, there are numerous instances where the most important times are those responsible for generating the majority of the hydrologic flux. Examples are numerous

[1]Department of Civil and Environmental Engineering, Clarkson University, Potsdam, New York, USA.
[2]School of Civil and Environmental Engineering, University of New South Wales, Sydney, New South Wales, Australia.
[3]Division of Earth and Ocean Sciences, Duke University, Durham, North Carolina, USA.

Corresponding author: T. J. Smith, Department of Civil and Environmental Engineering, Clarkson University, PO Box 5710, 8 Clarkson Ave., Potsdam, NY 13699, USA. (tsmith@clarkson.edu)

0043-1397/14/10.1002/2013WR014635

and include any number of variables that vary nonlinearly with streamflow, such as nutrient, contaminant, and sediment transport. Water flux tracking or management (e.g., reservoirs) could also benefit from flux proportional observation weighting. With this in mind, we sought to investigate the value of an alternative transformation that is weighted proportional to hydrologic flux, and thus, emphasizing the model performance for those times of increased interest.

2. Flow-Corrected Time Transformation

[5] To address the issue of model fit during periods of increased hydrologic flux, we introduce an objective function based on a flow-corrected time transformation [refer to *Rodhe et al*., 1996]. The concept of flow-corrected time (FCT) was originally used in transit time modeling to address assumptions of time invariance and constant volume [*McGuire and McDonnell*, 2006; *Rodhe et al*., 1996]. Under the transform, "clock time" is mapped to "flow time" and is defined as

$$t_f = \int_{t_0}^{t} q(t)dt/\overline{q} \tag{1}$$

where q is observed streamflow, $\overline{q}$ is the mean of observed streamflow, and t is "clock time". For discrete time intervals, this simply means that for any time, t, the equivalent flow time, t_f, is simply the sum of the observed streamflows from t_0 to t divided by the mean observed streamflow.

[6] Applying this transform to a streamflow hydrograph functionally results in the stretching of time during wet periods and the compressing of time during dry periods of the year (Figure 1). This results in more weight being given to times where the streamflow exceeds the mean streamflow and ultimately favors simulations that accurately predict such conditions of increased importance. Though this will naturally impact the predictions of low flows (less than mean), high flows often represent a disproportionally large percentage of total streamflow volume relative to percentage of streamflow observations (Figure 1).

[7] Because this transformation is based on the time domain rather than the data domain, there are some secondary modeling consequences. Projecting the data in a flow-corrected time domain results in observations at partial (or decimal) flow times with unequal time steps (noting that the data is unchanged in magnitude). This can cause difficulty in applying modeling approaches that rely on equal time steps in the new flow-corrected time domain. Given this constraint, there are two approaches for comparing observed and simulated data in the flow time domain: (1) resample the streamflow data in flow time onto equal time steps or (2) weight the streamflow data according to its flow time (duration) equivalency and model the transformed data in clock time.

[8] Each approach can lead to improved simulations at times of increased importance by assigning greater weight to times of greater hydrologic flux. Resampling the data onto uniform flow times (1, 2, … n) is attractive in that it preserves the concept of a time domain transformation, however, it requires the use of an interpolation algorithm. Here we choose to focus on weighting the streamflow data according to its flow time duration. We believe this approach is more

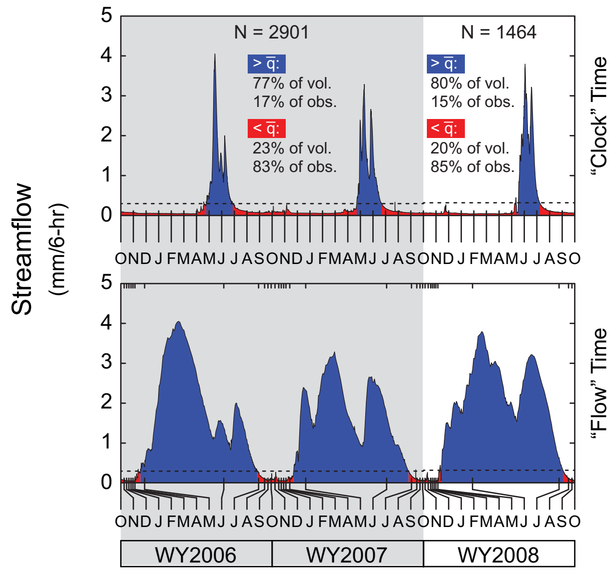

Figure 1. The Stringer Creek streamflow hydrograph visualized in both (top) clock time and (bottom) flow-corrected time. Water years 2006–2007 were used for calibration (shaded) and water year 2008 was used for validation.

straightforward and interpretable to those accustomed to the use of other data transformations common to hydrologic modeling. Using this direct weighting approach to modeling, the transformed streamflow data can be expressed as

$$q^{fct} = w_{fct} \cdot q \tag{2}$$

where q^{fct} are the flow-corrected time weighted data, q are the original data, and w_{fct} are the flow-corrected time weights (which sum to n, the number of observations of q). The flow-corrected time weights (w_{fct}) are simply calculated as

$$w_{fct} = q/\overline{q} \tag{3}$$

where all terms are as previously defined. The flow-corrected time data transformation can be used within any goodness-of-fit measure. The utility of this transformation in improving streamflow predictions will be investigated in the following section.

3. Case Study

3.1. Materials and Methods

[9] A simple application to highlight the potential benefits and limitations of the flow-corrected time transformation for hydrologic modeling studies is presented. We begin with the widely used Nash-Sutcliffe efficiency (NSE) [*Nash and Sutcliffe*, 1970] objective function

$$NSE = 1 - \frac{\sum (q_{obs} - q_{sim})^2}{\sum (q_{obs} - \overline{q}_{obs})^2} \tag{4}$$

where q_{obs} represents the observed streamflow, $\overline{q}_{obs}$ is the mean of the observed streamflow, and q_{sim} is the model

Table 1. Description of Parameters of the Probability Distributed Model

Parameter	Description
cmax	Maximum soil storage capacity (mm)
b	Spatial variability within the catchment
kb	Rate of drainage into subsurface storage
cf	Soil storage threshold for release of subsurface storage to streamflow (mm)
tr_1	Fraction of subsurface storage released to streamflow
tr_2	Fraction of surface storage released to streamflow

simulated streamflow. Applying the Nash-Sutcliffe objective function, we compared the results arising from using no data transformation, the FCT transformation (equation (1)), and the logarithmic case arising from the Box-Cox family of transformations [*Box and Cox*, 1964]

$$q^{bc} = \log(q+\lambda) \tag{5}$$

where q^{bc} is the Box-Cox transformed streamflow and λ is a transformation parameter used to prevent log(0). Here the value of λ was calibrated, though it could be fixed to a literature value.

[10] We selected the Probability Distributed Model (PDM) [*Moore*, 1985, 2007] for this application, based on previous studies where it was shown to perform well in our test catchment [*Smith and Marshall*, 2010]. The PDM is a flexible model structure that conceptualizes the catchment as a simple bucket whose depth is controlled by the shape of a soil storage probability distribution. Water exceeding the soil capacity is routed to a surface storage, while all other water drains into a subsurface storage. The model is represented with six parameters that are described in Table 1. Model parameters were estimated using the Dynamically Dimensioned Search (DDS) algorithm, which was developed and extensively tested on hydrologic model problems by *Tolson and Shoemaker* [2007].

[11] In our test application, we investigated the utility of the flow-corrected time transformation in the context of a snowmelt-dominated study catchment that features a relatively simple hydrograph with one main peak occurring during snowmelt. The Stringer Creek catchment (5.55 km^2) is situated within the Tenderfoot Creek Experimental Forest (TCEF) of Montana, USA (lat. 46°55′N, long. 110°52′W). For this study, 6 hourly precipitation, evapotranspiration, and streamflow data were obtained from October 2006 through September 2009 [*Nippgen et al.*, 2011]. The first 2 years were utilized to force and calibrate the PDM, while the third year was retained for model validation testing.

3.2. Modeling Results

[12] Following the model optimization for each of the data transformations applied within the NSE objective function (none, flow-corrected time, logarithmic), we investigated the differences and similarities that arose in terms of model specification and performance. The calibrated model parameters exhibited significant variability across the different transformations (Table 2). While this is not surprising considering the transformations effectively result in different, unique data sets, the model parameters (and therefore fit) were strongly affected by such transformations. In particular, the parameters controlling the soil storage conceptualization (size, shape; parameters *cmax*, *b*, *cf*) showed distinct differences. The model parameters optimized under flow-corrected time indicate larger soil storage capacity and an increased fraction of stored water being released as subsurface flow (parameter tr_1). This suggests that the increased antecedent soil moisture causes increased subsurface discharge, which results in larger peak streamflow and longer sustained recessions during and after saturated conditions. In this case, by modeling the streamflow in FCT there is greater importance placed on peak flow and the recession period that are responsible for the majority of the streamflow flux.

[13] Quantitative assessment of model fit using the Nash-Sutcliffe efficiency (Table 3) is complicated when considering multiple transformation domains. The model was fit within each domain, therefore, comparisons within any single domain will necessarily be biased against the data-model combinations fitted in different domains. As a consequence, here we compare NSE values determined in each transformation space (akin to comparing the NSE of a model fitted at multiple catchments). Such comparisons revealed that while the model parameterization was widely variable across all transformation cases, effective differences were minimal between the logarithmic and untransformed cases with NSEs of 0.884 and 0.881, respectively, during the calibration period. Model performance was improved under the FCT transformation with an NSE during the calibration period of 0.962. These results are logical when the characteristics of the hydrograph are considered (Figure 1). Though only 17% of the total number of observations during the calibration period exceeded the mean streamflow, they were responsible for 77% of the total volume of streamflow. In terms of weights, this means that for the FCT transformation 77% of the weight was assigned to these 500 observations (17% of total), whereas the untransformed and logarithmic cases assigned just 17% of the weight to the same 500 observations.

[14] To further investigate these results, a complete analysis of the NSE for each transformation was performed, including for observations above/below the mean

Table 2. Optimal Parameters of the Probability Distributed Model

	Parameter						
Transformation	*cmax* (mm)	*b*	*kb*	*cf* (mm)	tr_1	tr_2	λ (mm/6 h)
None	373.9	0.71	0.50	359.3	8.4×10^{-5}	4.4×10^{-2}	–
Box-Cox	292.8	0.36	0.14	291.9	1.6×10^{-4}	4.0×10^{-2}	0.99
FCT	452.5	0.95	0.59	383.4	2.9×10^{-3}	6.5×10^{-2}	–

Table 3. Comparison of Nash-Sutcliffe Model Efficiency Across Data Transformations[a]

	Calibration: WY 2006–2007				Validation: WY 2008			
Streamflow State	$>\bar{q}$	$<\bar{q}$	$>\bar{q}$	$<\bar{q}$	$>\bar{q}$	$<\bar{q}$	$>\bar{q}$	$<\bar{q}$
% Streamflow Observations	17	83	17	83	15	85	15	85
% Streamflow Volume	77	23	77	23	80	20	80	20
Transformation NSE	Transformation Space		Original Space		Transformation Space		Original Space	
None	0.8808		0.8808		0.9159		0.9159	
	0.90	0.72	0.90	0.72	0.91	0.93	0.91	0.93
Box-Cox	0.8842		0.8495		0.8817		0.7975	
	0.92	0.66	0.85	0.84	0.89	0.84	0.78	0.94
FCT	0.9621		0.8399		0.9433		0.8944	
	0.96	1.00	0.90	0.34	0.94	1.00	0.91	0.75

[a]The NSE has been reported for each transform in both the transformation (model) space and the original (untransformed) space. The NSE was further reported for streamflows greater than and less than the mean observed streamflow ($\bar{q}$).

streamflow and in transformed and original data space for both the calibration and validation periods (Table 3). As an example, consider the FCT-weighted transformation for the calibration period. In transformation space, the NSE is actually higher for low flows (1.0) than for high flows (0.96). This may seem counterintuitive, but is a result of the low flows receiving very little weight that results in very small errors. Contrasting this to the original data space results (NSE for low flows of 0.34), it is clear that the fit at low flows is linked to the weighting. Similarly, the inherent weighting that occurs as a consequence of nonconstant error variance can be seen in the untransformed case where the fit is greater for high flows than low flows.

[15] While the NSE values provide a summary of the model fit, much can also be gleaned from a visual inspection of the model simulations themselves (Figure 2).

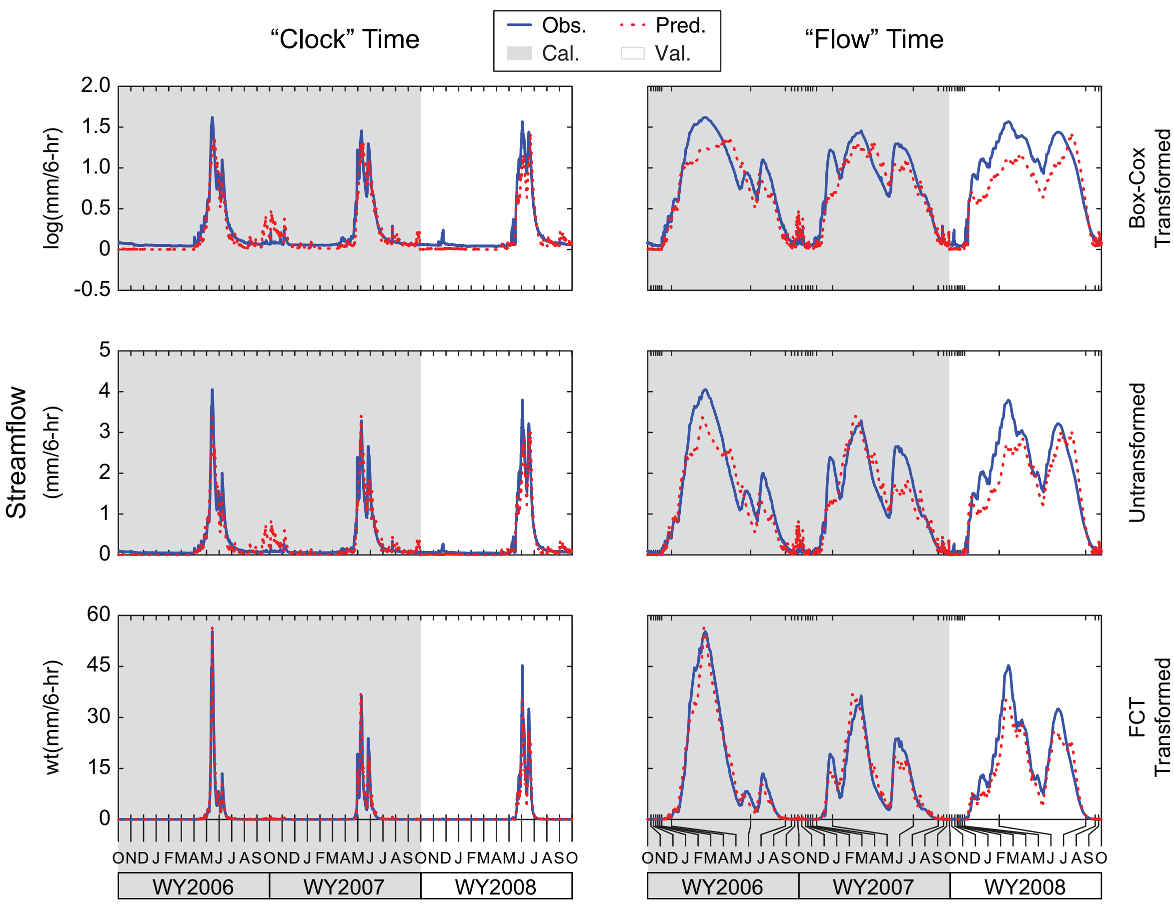

Figure 2. The Box-Cox transformed, untransformed, and FCT (weighted) transformed streamflow hydrograph for the Stringer Creek catchment application. Results are shown for each transform in both "clock" time and "flow" time. The calibration period (2 years) is shown in the shaded region, followed by the validation period (1 year).

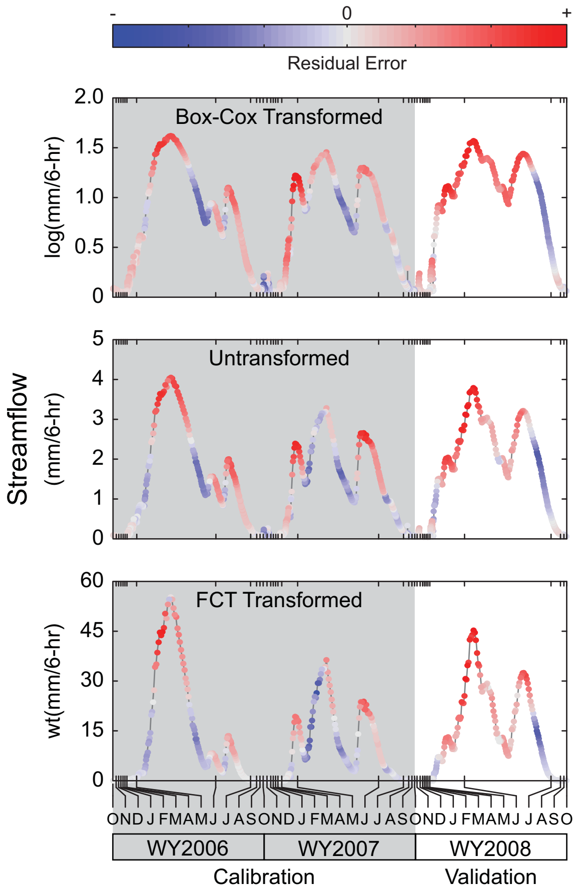

Figure 3. Observed streamflow hydrographs (in respective transformation spaces) displayed in the flow-corrected time domain, with residual error represented by the color map. Note positive (+) residual error is indicative of under prediction by the model and negative (−) residual error is representative of over prediction; the "wt" label on units for the FCT transformed streamflow indicates that these are the weighted streamflows.

We present model simulations in both clock and flow time. Notably, the predicted hydrograph for the untransformed Nash-Sutcliffe objective function and the logarithmic transformed NSE were flashier in their response to precipitation. At the same time, the FCT transformed NSE was less sensitive to the inputs (particularly at low streamflows), as expected given that flows less than the mean flow were assigned less weight. In terms of peak streamflow simulations, the FCT transformed NSE resulted in better simulations for each of the three primary peaks during maximum streamflow in terms of both timing and magnitude. Interestingly, when considering the difference between the observed and modeled hydrograph (Figure 3; i.e., the model residuals), the temporal patterns of model fit for each data transformation (none, logarithmic, FCT) resulted in similar patterns of prediction. For example, the peaks of the hydrograph were underestimated for each case, while the rising and falling limbs tended to be overestimated. The degree (magnitude) and extent (temporal persistence) to which these differences occurred are highlighted in the model fits (Table 3).

4. Discussion and Conclusions

[16] As the application of hydrologic models for water resource management increases, the development of optimization/calibration approaches that appropriately reflect the hydrologic characteristic of primary importance is becoming ever more necessary. Despite the significant number of water resource applications focused on times of increased hydrologic flux, to date model calibration research approaches have yet to focus on addressing this problem directly and unambiguously. Here we introduce and apply a new transformation, within the context of hydrologic model calibration, that utilizes the flow-corrected time transform, first introduced in transit time modeling [*Rodhe et al.*, 1996]. The nonuniform, partial time steps that are a result of the flow-corrected time calculation (equation (1)) pose a functional constraint of the method. To address this, we utilized the flow time durations as weights to transform the data. However, an alternative approach is to resample the flow-corrected streamflow data onto uniform time intervals.

[17] The results of our case study, featuring a snowmelt dominated annual streamflow regime, highlighted the utility of the FCT transformation for improved model fits during periods of increased hydrologic flux (Table 3 and Figures 2 and 3). While these results are promising, the use of any data transform comes with caveats. In this case, the method is focused on fitting the values of streamflow that exceed the mean streamflow. The use of the flow-corrected time transform could be most valuable where high streamflow events are of interest (e.g., floods), but could be counterproductive (in its current formulation) under circumstances concerned with baseflow events (e.g., droughts). However, with a simple modification to the FCT transform presented here, priority could be placed on low flows rather than peak flows. The FCT transform is referenced to mean streamflow, and as a result has increasing effect as the hydrograph becomes less symmetric about the mean.

[18] Beyond the limitations in terms of the "streamflow state" of interest, there are additional limitations in terms of real-time forecasting. Because the methodology downweights streamflows less than the mean observed streamflow, such streamflows are less influential on the model calibration. As a result, the reliability of model predictions for streamflows less than the mean streamflow diminishes with departure from the mean, noting that this approach is not suitable in ephemeral catchments, where no flow conditions would receive zero weight. Further complications may arise when auxiliary modeled variables of interest depend on "clock time" (e.g., groundwater storage). Under this scenario, internal model variables that are simulated as the result of threshold exceedances may be negatively impacted at low streamflows because a wider range of model parameters will produce acceptable fits (due to the decreased weight such observations receive).

[19] Despite these potential limitations, the flow-corrected time transform was shown to result in a superior fit to the observed streamflow record (as measured by the NSE in flow transformed space) than the alternatives investigated (as measured in their respective transformed spaces) in both the calibration and validation periods of study (Table 3 and Figures 2 and 3). Difficulties exist in comparing model

efficiencies across different data sets (in this case, due to different data transformations), but the results indicate that the FCT provides a robust approach to modeling periods of high hydrologic flux. While implicit weighting toward flows with large errors (often associated with large flows) occurs within squared error-type cost functions, such weighting is a function of errors in the forcing data, calibration data, and/or model structure. The FCT transformation provides an explicit and unambiguous, data-based weighting toward periods of increased hydrologic flux. Changes in fit came about due to both shifts in optimal model parameters (Table 2) in response to the differences in data (due to transformation) and the dynamic weighting of observations relative to their importance to total hydrologic flux (Figure 1). Future studies considering the concept of flow time as it relates to the propagation of predictive uncertainty and its utility within formal and informal uncertainty frameworks are planned.

[20] **Acknowledgments.** This study was funded by the National Science Foundation (EAR-0943640) awarded to L. Marshall and B. McGlynn. Additional support was awarded to T. Smith as a Graduate Fellow of the Montana Institute on Ecosystems (NSF EPSCoR Track-1, EPS-1101342). We thank Harald Kling and two anonymous reviewers for their contributions toward the improvement of this manuscript.

References

Beven, K., and A. Binley (1992), The future of distributed models: Model calibration and uncertainty prediction, *Hydrol. Processes*, *6*, 279–298, doi:10.1002/hyp.3360060305.

Box, G. E. P., and D. R. Cox (1964), An analysis of transformations, *J. R. Stat. Soc., Ser. B*, *26*(2), 211–252.

Duan, Q., S. Sorooshian, and V. Gupta (1992), Effective and efficient global optimization for conceptual rainfall-runoff models, *Water Resour. Res.*, *28*, 1015–1031, doi:10.1029/91WR02985.

Gupta, H. V., and H. Kling (2011), On typical range, sensitivity, and normalization of Mean Squared Error and Nash-Sutcliffe Efficiency type metrics, *Water Resour. Res.*, *47*, W10601, doi:10.1029/2011WR010962.

Gupta, H. V., S. Sorooshian, and P. O. Yapo (1998), Toward improved calibration of hydrologic models: Multiple and noncommensurable measures of information, *Water Resour. Res.*, *34*, 751–763, doi:10.1029/97WR03495.

Kuczera, G., and E. Parent (1998), Monte Carlo assessment of parameter uncertainty in conceptual catchment models: The Metropolis algorithm, *J. Hydrol.*, *211*(1–4), 69–85.

Madsen, H. (2003), Parameter estimation in distributed hydrological catchment modelling using automatic calibration with multiple objectives, *Adv. Water Resour.*, *26*(2), 205–216.

Madsen, H., G. Wilson, and H. C. Ammentorp (2002), Comparison of different automated strategies for calibration of rainfall-runoff models, *J. Hydrol.*, *261*(1–4), 48–59.

McGuire, K. J., and J. J. McDonnell (2006), A review and evaluation of catchment transit time modeling, *J. Hydrol.*, *330*(3–4), 543–563.

Moore, R. J. (1985), The probability-distributed principle and runoff production at point and basin scales, *Hydrol. Sci. J.*, *30*(2), 273–297.

Moore, R. J. (2007), The PDM rainfall-runoff model, *Hydrol. Earth Syst. Sci.*, *11*(1), 483–499.

Nash, J. E., and J. V. Sutcliffe (1970), River flow forecasting through conceptual models: Part I—A discussion of principles, *J. Hydrol.*, *10*(3), 282–290.

Nippgen, F., B. L. McGlynn, L. A. Marshall, and R. E. Emanuel (2011), Landscape structure and climate influences on hydrologic response, *Water Resour. Res.*, *47*, W12528, doi:10.1029/2011WR011161.

Rodhe, A., L. Nyberg, and K. Bishop (1996), Transit times for water in a small till catchment from a step shift in the oxygen 18 content of the water input, *Water Resour. Res.*, *32*, 3497–3511, doi:10.1029/95WR01806.

Smith, T. J., and L. A. Marshall (2010), Exploring uncertainty and model predictive performance concepts via a modular snowmelt-runoff modeling framework, *Environ. Modell. Software*, *25*(6), 691–701.

Tolson, B. A., and C. A. Shoemaker (2007), Dynamically dimensioned search algorithm for computationally efficient watershed model calibration, *Water Resour. Res.*, *43*, W01413, doi:10.1029/2005WR004723.

Vrugt, J. A., H. V. Gupta, L. A. Bastidas, and W. Bouten (2003), Effective and efficient algorithm for multiobjective optimization of hydrologic models, *Water Resour. Res.*, *39*(8), 1214, doi:10.1029/2002WR001746.

Wagener, T., D. P. Boyle, M. J. Lees, H. S. Wheater, H. V. Gupta, and S. Sorooshian (2001), A framework for development and application of hydrological models, *Hydrol. Earth Syst. Sci.*, *5*(1), 13–26.