Front. Earth Sci. 2020, 14(1): 110–123
https://doi.org/10.1007/s11707-019-0755-y

RESEARCH ARTICLE

Predicting the temporal transferability of model parameters through a hydrological signature analysis

Dilhani Ishanka JAYATHILAKE[1], Tyler SMITH (✉)[1,2]

1 Institute for a Sustainable Environment, Clarkson University, Potsdam, NY 13699, USA
2 Department of Civil and Environmental Engineering, Clarkson University, Potsdam, NY 13699, USA

Abstract Attention has recently increased on the use of hydrological signatures as a potential tool for assessing the fidelity of model structures and providing insights into the transfer of model parameters. The utility of hydrological signatures as model performance/reliability indicators in a calibration-validation testing scenario (i.e., the temporal transfer of model parameters) is the focus of this study. The Probability Distributed Model, a flexible conceptual hydrological model, is used to test the approach across a number of catchments included in the MOPEX data set. We explore the change in model performance across calibration and validation time periods and contrast it to the corresponding change in several hydrological signatures to assess signature worth. Results are explored in finer detail by utilizing a moving window approach to calibration and validation time periods. The results of this study indicated that the most informative signature can vary, both spatially and temporally, based on physical and climatic characteristics and their interaction to the model parameterization. Thus, one signature could not adequately illustrate complex watershed behaviors nor predict model performance in new analysis periods.

Keywords streamflow, hydrological signature, validation testing, model calibration

1 Introduction

Data is insufficient for many catchments across the globe and thus, an understanding of the hydrological process is poor. The transfer of information from gauged catchments to ungauged catchments via relationships derived between model parameters and catchment characteristics is, therefore, an important area of research (Kay et al., 2007; Zhang et al., 2011; Patil and Stieglitz, 2015; Zhang et al. 2016). Recent studies have focused on the transfer of information by identifying hydrological similarities between gauged and ungauged catchments (Kay et al., 2007; Parajka et al., 2007; Yadav et al., 2007; Sawicz et al., 2011; Zhang et al. 2016). In general, hydrological similarities are characterized by means of physical proximity measures (e.g., topography, soil type, and land cover) or spatial proximity measures (e.g., distance) (Parajka et al., 2007; Samaniego et al., 2010; Samuel et al., 2011). However, the information transfer among catchments of interest can occur across a wide range of climatic and geographic regions and remains highly uncertain (Kay et al., 2007; Parajka et al., 2007).

Temporal transfer of parameters, the application of the calibrated model at varied time periods within the same catchment, is perhaps the most common and straightforward procedure used in hydrological modeling. An implicit assumption is made that the calibrated model parameters are stable beyond the calibration period. Hydrological studies often reveal that calibrated models suffer from significant performance loss outside the calibration period (Seibert, 2003; Vaze et al., 2010; Merz et al., 2011). Current hydrological studies focus on evaluating the range of conditions covered by the calibration data (Vaze et al., 2010) to improve the robustness of models and aid in their applications aimed at addressing major issues in hydrology.

At the same time, greater attention has been given to the use of hydrological signatures as a potential tool for improving the transfer of model parameters, most commonly under a "predictions in ungauged basins" scenario, where parameters are transferred spatially (Yadav et al., 2007; Sawicz et al., 2011; Patil and Stieglitz, 2012; Blöschl et al., 2013; Donnelly et al., 2016). Hydrological signatures represent indices of a time series of a response characteristic that reflects the functional behaviors of a catchment (Yadav et al., 2007; Sawicz et al.,

Received November 19, 2018; accepted April 8, 2019

E-mail: tsmith@clarkson.edu

2011; Casper et al., 2012) and can be considered comparable to "soft" data approaches (e.g., Seibert and McDonnell, 2002). Many typical hydrological signatures of a watershed can be derived using commonly available time series data including precipitation, evapotranspiration, temperature, and streamflow (in addition to other response variables) (Yadav et al., 2007; Sawicz et al., 2011). Several studies have used hydrological signatures in the context of model calibration and examined their usefulness in improving the task of model parameter estimation by establishing relationships between catchment attributes and runoff signatures (Hingray et al., 2010; Westerberg et al., 2011). Hingray et al. (2010) presented a hybrid parameter regionalization and signature-based calibration method, where they calibrated the model on hydrological time series data and signatures. In streamflow predictions in ungauged basins, hydrological signatures can be regionalized to watershed physical characteristics. Regression relationships can then be developed between the response characteristics and watershed physical characteristics at gauged watersheds (Yadav et al., 2007; Masih et al., 2010; Beck et al., 2016). Evaluation of model performance based on criteria such as the goodness-of-fit of hydrological signatures can provide even more information about the hydrological behavior of the modeled catchments (Hrachowitz et al., 2014; Montanari and Toth, 2007). Indeed, this type of approach (signatures, soft data, etc.) was found useful in reducing uncertainty in model responses by specifying additional criteria to judge model simulations in calibration and validation processes (e.g., Seibert and McDonnell, 2002; Hrachowitz et al., 2013; Westerberg et al., 2016). Hydrological signatures have also been recently utilized for catchment classification using methods such as Bayesian cluster analysis (Sawicz et al., 2011) and a pooling group approach for spatial generalization of parameters (Kay et al., 2007).

It is believed that utilizing hydrological signatures can serve to strengthen the link between hydrological models and the underlying hydrological processes while allowing for the transfer of model parameters (Yadav et al., 2007; Hingray et al., 2010). Several studies have focused on the transfer of information temporally to validate the performance of the model (Wagener et al., 2003; Merz et al., 2011). However, parameter transferability, both spatially and temporally, necessitates further exploration due to the unreliability encountered in model performance when calibrated parameters are used under different conditions. In our view, understanding further insights into the suitability of calibrated parameters beyond the calibration period is critical to reducing model output uncertainty. The overall goal of this study is to identify signatures that best reflect changes in model performance across calibration and validation periods (i.e., a signature whose change is indicative of parameter transferability potential). Model appropriateness is tested based on the ability of the model to reproduce several hydrological signatures. The ability of signatures to reflect changes in model performance are explored under direct and moving window split sample approaches to the calibration and validation periods.

2 Methods and materials

2.1 Study area and data

The Model Parameter Estimation Project (MOPEX) data set described by Duan et al. (2006) was used in this study. The data set provides mean areal precipitation, potential evapotranspiration, daily streamflow, and daily maximum and minimum air temperatures from 1948 to 2003 for 438 basins across the USA. The precipitation data are from the National Climate Data Center and the Natural Resources Conservation Service Snow Telemetry (SNOTEL) network. Daily streamflow data are derived from the United States Geological Survey (USGS) and the potential evapotranspiration (PET) is based on the National Oceanic and Atmospheric Administration (NOAA) Pan Evaporation Atlas.

Some of the sites included in the MOPEX data set contain periods of missing streamflow data. Therefore, in this study, the most recent seven years of data with no missing streamflow days were chosen in all sites. To remove the impacts of parameter/model component interactions due to snow, only the 18 sites with the lowest percentage of likely snow days from the 438 sites were used for this study (Fig. 1). Snow days were identified using a simple algorithm, where any precipitation occurring at an air temperature less than 3°C was considered snow. The rain-snow threshold temperature of 3°C was chosen as a conservative metric for identifying snow-free catchments (e.g., Auer, 1974; Dai, 2008).

2.2 Rainfall-runoff model

The Probability Distributed Model (PDM) (Moore, 1985 and 2007) was selected for this study. The PDM is well-established within hydrological modeling literature and has been widely applied to a variety of catchments worldwide, including the USA, UK, Belgium, and India (Moore, 2007). For example, the PDM was used across catchments in the UK to develop river flow and groundwater level data sets for climate change impact assessments (Prudhomme et al., 2013) and to develop a spatial generalization for model parameters based on site similarity (Kay et al., 2007).

The PDM is a flexible rainfall-runoff model that is conceptualized as a simple bucket. The runoff production at a point is controlled by the absorption capacity of the soil. Water exceeding the soil absorption capacity is routed to a surface storage, while all other water drains into a subsurface storage. Outflows from two storage components result in the total runoff of the watershed. The spatial variability of the storage capacity is represented by a

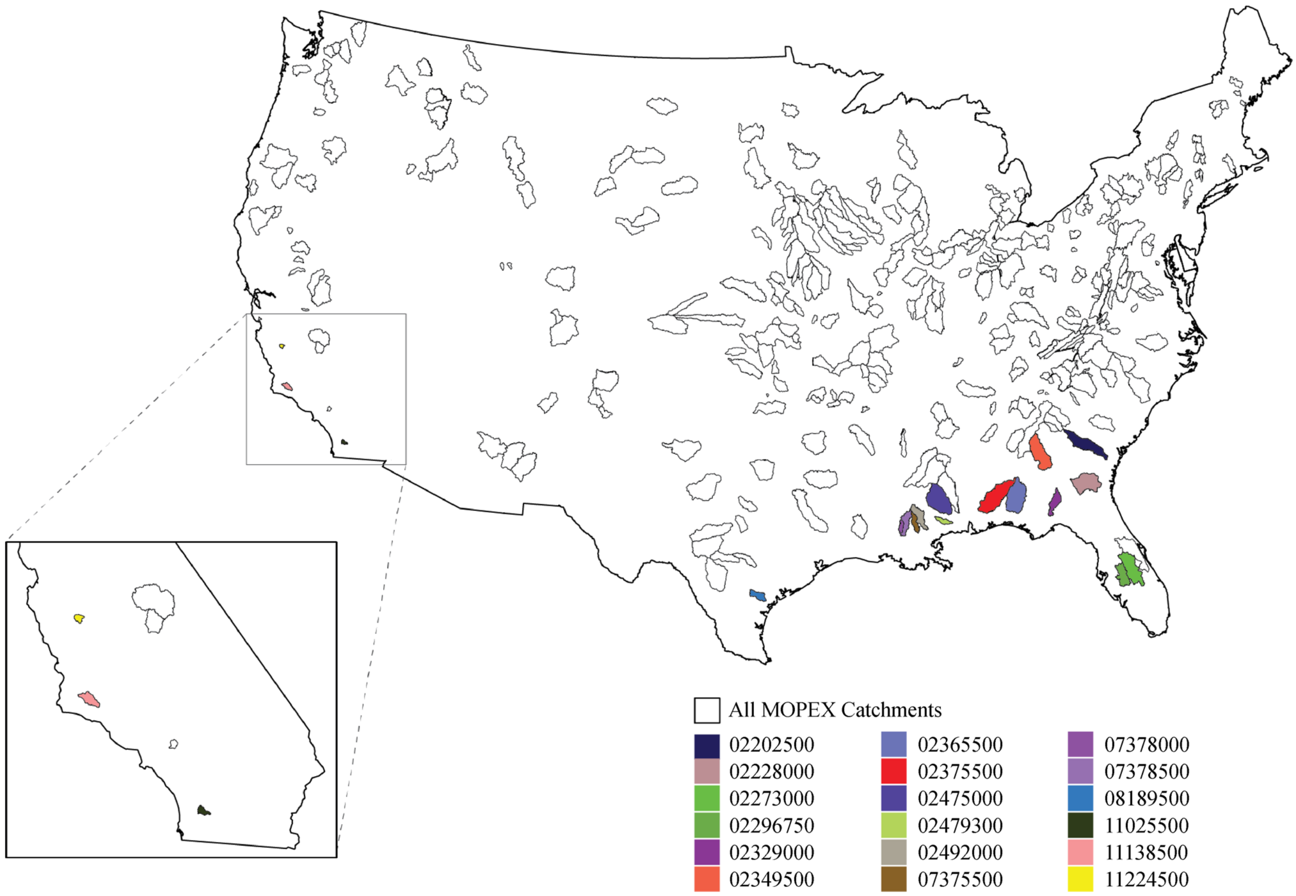

Fig. 1 A subset of 18 catchments from the MOPEX data set located across the southern USA.

probability distribution. Although the PDM is a simple conceptual model, in contrast to many other simple bucket models, it maintains the ability to describe spatial variability with a single parameterized bucket rather than using additional parameters to represent multiple buckets of varying capacities (Moore, 2007; Ewen, 2011). The model as applied here requires the estimation of the six parameters as shown in Table 1.

Table 1 Parameters of the probability distributed model

Parameter	Description
C_{max}	The maximum soil water content capacity/mm
B	The spatial variability within the watershed
Kb	The rate of drainage into subsurface storage/hr
$Tres1$	The fraction of subsurface storage released to outflow
$Tres2$	The fraction of surface storage released to outflow
CF	The soil storage threshold for subsurface inflow/mm

2.3 Model calibration

Due to the conceptual nature of the PDM, model calibration is required to determine appropriate parameter values. The Dynamically Dimensioned Search (DDS) algorithm developed by Tolson and Shoemaker, 2007 was employed for parameter estimation. The DDS algorithm is based on a simple stochastic neighborhood search algorithm that explores the neighborhood around the best solution and compares it with candidate solutions to determine if an update is required to the current "best" solution (Tolson and Shoemaker, 2007 and 2008). The DDS algorithm was used with the Nash-Sutcliffe Efficiency (*NSE*) (Nash and Sutcliffe, 1970) as the objective function in the model calibration process. The *NSE* is a widely used goodness-of-fit metric for the calibration and evaluation of hydrological models with observed data. The *NSE* quantifies model performance as the comparative ability of a chosen model with respect to a benchmark model (the mean of the observations, as typically defined).

2.4 Hydrological signatures

Hydrological signatures are meant to provide a summary metric of the hydrological behavior of a catchment (Wagener et al., 2007; Sawicz et al., 2011; Patil and Stieglitz, 2012) and are seen as useful indicators of model realism and, by extension, parameter transferability. Signatures have been used in numerous hydrological studies for the transfer of hydrological information (model parameter values), catchment classification, and generalization of hydrological process understanding and catchment responses (Wagener et al., 2007; Castiglioni et al. 2010; Sawicz et al., 2011). For this study, we examined a commonly employed suite of hydrological signatures that are derived from widely available time series data; the slope of the flow duration curve (*SFDC*), runoff ratio (*RR*), streamflow elasticity (*EQP*), and flashiness index (*FI*).

Runoff Ratio (*RR*)

The runoff ratio is defined as the ratio of average annual streamflow (Q) to average annual precipitation (P).

$$RR = \frac{Q}{P}. \tag{1}$$

It is a metric that represents the long-term water balance separation between water that exits the catchment as runoff and as evapotranspiration. A catchment with a high runoff ratio is considered to be streamflow dominated, whereas a catchment with a low runoff ratio is evapotranspiration dominated (Sawicz et al., 2011; Patil and Stieglitz; 2012).

Slope of the Flow Duration Curve (*SFDC*)

The flow duration curve (FDC) is a representation of the distribution of probabilities of streamflow being greater than or equal to a specific streamflow value. To quantify an index of flow variability, the slope of the flow duration curve (*SFDC*) is calculated between the 33rd (Q_{33}) and 66th (Q_{66}) streamflow percentiles as this represents a relatively linear part of the FDC on semi-log scale (Yadav et al., 2007; Sawicz et al., 2011).

$$SFDC = \frac{\ln(Q_{66}) - \ln(Q_{33})}{0.66 - 0.33}. \tag{2}$$

A high value of *SFDC* indicates that the catchment is subject to high flow variability, while a low *SFDC* value is typical of a catchment with a damped response behavior and stable streamflow. The FDC represents the ability of a catchment to produce discharge values of different magnitudes. Therefore, this metric is strongly sensitive to the vertical distribution of soil moisture storages within a basin (Sawicz et al., 2011; Patil and Stieglitz, 2012).

Streamflow Elasticity (*EQP*)

Streamflow elasticity is the ratio of the change in annual streamflow to the change in annual precipitation. *EQP* is defined as

$$EQP = \mathrm{median}\left(\frac{(Q_i - \overline{Q})}{(P_i - \overline{P})}\, \frac{\overline{P}}{\overline{Q}}\right), \tag{3}$$

where $\overline{P}$ and $\overline{Q}$ are mean annual precipitation and streamflow for the period of study; P_i and Q_i are annual precipitation and streamflow for the i^{th} year. Streamflow elasticity is an indicator of the sensitivity of the streamflow response of a catchment to changes in precipitation. An *EQP* value of 1 indicates a catchment with a linear relationship between precipitation change and streamflow change. A value greater than 1 indicates a catchment being elastic or more sensitive to the change in precipitation, while a value less than 1 indicates that the catchment is inelastic or insensitive to the change in precipitation (Sawicz et al., 2011; Patil and Stieglitz, 2012).

Flashiness Index (*FI*)

The FI measures oscillations in streamflow relative to the total streamflow of the catchment. Flashiness is equated with the rate of change of streamflow. *FI* is given by the ratio of the sum of absolute incremental differences of streamflow to total streamflow.

$$FI = \frac{\sum_{i=1}^{n} |Q_i - Q_{i-1}|}{\sum_{i=1}^{n} Q_i}. \tag{4}$$

A high value for *FI* indicates a rapid oscillating streamflow hydrograph relative to the total streamflow hydrograph, while a low value suggests a slow responding streamflow hydrograph (Baker et al., 2004).

2.5 Split sample testing approach

The predictive model performance was examined using a split-sample testing approach, where the available streamflow record is split into two segments such that the model is validated using a data period assumed independent of the period of model calibration. Two data sampling approaches were considered in this study: direct calibration-validation and moving window calibration-validation.

In the first approach, the seven years of streamflow data for each catchment were split, with the first four years used for calibration and the past three years for validation (Fig. 2). The second approach sought to enable a more detailed exploration of the results, where the model was calibrated over the first four years (as in the first approach) and validated over a four-year window, moving in three-month increments (Fig. 2).

Four hydrological signatures were computed using daily precipitation and streamflow data. In the direct calibration-validation approach, a single value was computed for both the calibration period (four years) and the validation period (three years) for each hydrological signature considered in this study. Similarly, in the moving window calibration-validation approach, a single value was computed for the four-year calibration period and for each of the 13 validation periods. Four-year data periods were gradually departed from the calibration period by three month increments. While data length is an important considera-

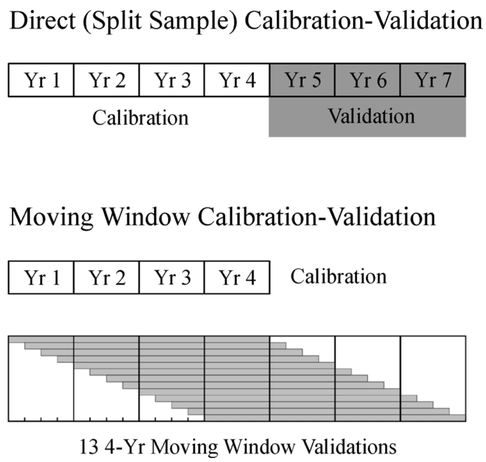

Fig. 2 Schematic representation of splitting data records in direct calibration-validation and moving window approaches.

tion in signature analyses due in part to limited data availability, this study seeks to understand the utility of signatures under typical data conditions (e.g., Euser et al., 2013; Westerberg et al., 2016).

3 Results

3.1 Hydrological signatures as predictors of model performance in validation mode

Climatic non-stationarity limits the ability of hydrological models to accurately predict future hydrological conditions. In light of this, a better understanding of the linkages between hydrological signatures, hydrological model parameters, and hydrological model performance was sought through a systematic implementation of the split-sample testing approach common to hydrological modeling. The change in both model performance and in hydrological signatures across the calibration and validation periods was evaluated under the direct and moving window validation approaches. Model performance was evaluated in terms of the *NSE*. During the calibration period, the *NSE* ranged from 0.39 to 0.86 with an average of 0.58. During the validation period, the *NSE* ranged from 0.30 to 0.80 with an average of 0.57. Reasonable model performance (e.g., $NSE = 0.55$) and calibration-validation consistency indicated that the PDM produced acceptable/accurate simulations of the basin hydrological processes. The hydrological variations were identified by four signatures, summarized in Table 2. The changes in *NSE* were contrasted to the changes in hydrological signatures (*SFDC*, *FI*, *RR*, and *EQP*; see Fig. 3).

Here, the percent differences of signatures and *NSE* across calibration and validation periods were plotted for each of the test catchments (Fig. 3). The percent differences in the signatures/*NSE* across calibration and validation periods were calculated relative to the calibration periods (i.e., signatures/*NSE* in calibration periods were used as the denominator of the ratio). In Fig. 3, each point represents the difference of *NSE*/signature value across calibration and validation periods. The correlation coefficient (*R*) between percent changes in signatures and *NSE* across calibration and validation periods was computed. A correlation coefficient of $+1$ indicates a perfect positive correlation, a correlation coefficient of -1 indicates a perfect negative correlation, and a correlation coefficient near 0 indicates no correlation.

Changes in model performance across sites and time periods (calibration-validation) showed a weak correlation to changes in hydrological signatures where $R_{SFDC} = 0.21$, $R_{EQP} = 0.17$, $R_{FI} = 0.14$, and $R_{RR} = 0.42$, excluding outliers. To more fully observe the variability in model performance and hydrological signatures across calibration and validation periods, results were also considered under a moving window validation approach. The validation periods (13 four-year windows moving by three-months each) gradually depart from the calibration period allowing model performance to be diagnosed under finer variation of the climatic conditions. Similar to the direct validation approach, the percent change in the signatures/*NSE* across calibration (i.e., the first moving window) and moving window validation periods were calculated relative to the calibration periods. The resulting values for the change in *NSE* and hydrological signatures over 13 four-year moving validation windows are shown in Fig. 4 for selected sites demonstrating the range of results. In Fig. 4, each point represents the difference in *NSE*/signature values between the calibration and each of the 13 moving validation periods. Due to space considerations, three representative sites are shown). The correlations of each signature in the moving window validation approach are given in Table 3 for all test catchments.

Table 2 Description of *NSE* and hydrological signatures values

Measure / Signature	High value	Low value
Nash-Sutcliffe Efficiency (*NSE*)	Good model performance	Poor model performance
Slope of Flow Duration Curve (*SFDC*)	Variable/responsive streamflow regime	Damped/unresponsive streamflow regime
Runoff ratio (*RR*)	Majority of water leaves catchment as streamflow	Majority of water leaves catchment as evapotranspiration
Streamflow Elasticity (*EQP*)	$EQP > 1$ – streamflow is sensitive to changes in precipitation	$EQP < 1$ – streamflow is not sensitive to changes in precipitation
Flashiness Index (*FI*)	Rapid short-term changes in streamflow response	Slow short-term changes in streamflow response

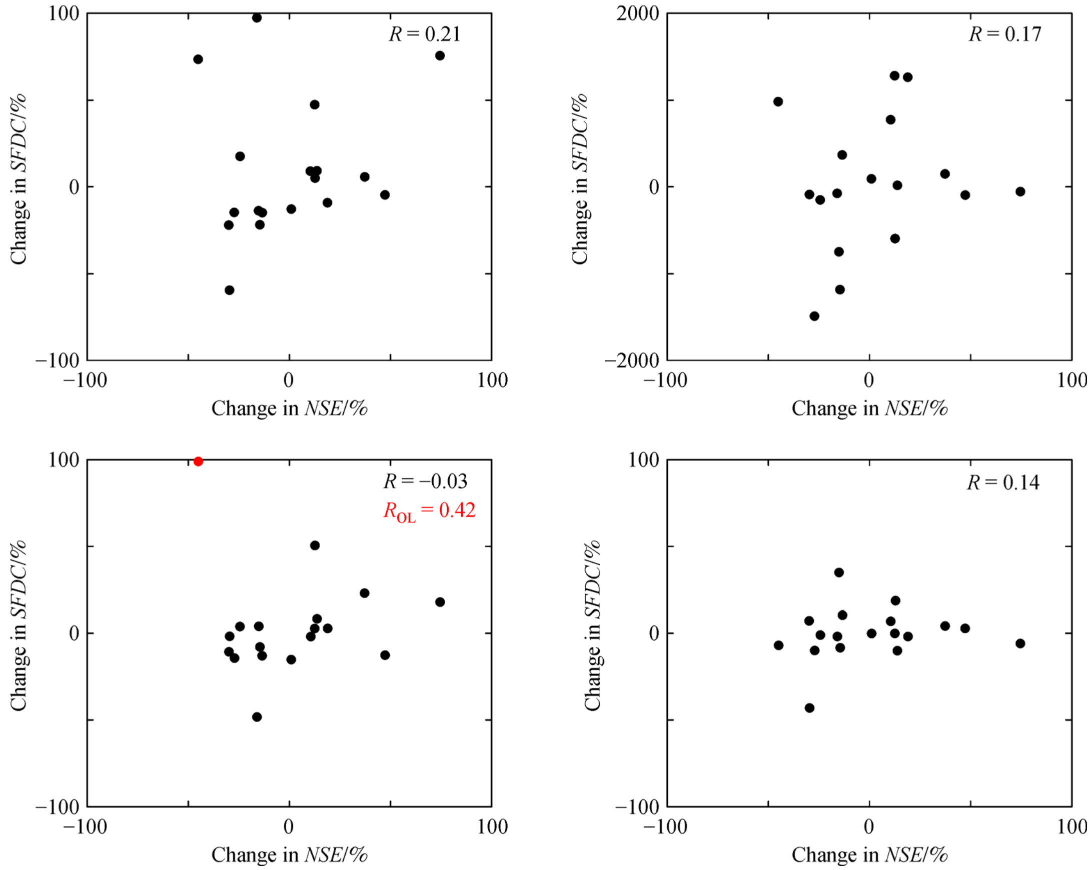

Fig. 3 Change in model performance (*NSE*) and hydrological signatures across 18 MOPEX sites under direct calibration-validation testing. Correlation coefficients (*R*) between change in hydrological signature and NSE are given for each plot. Note: R_{OL} represents the correlation coefficient when outliers were not considered.

In general, the magnitude of correlation values was relatively higher in the moving window validation approach than the direct validation approach (Fig. 4 and Table 3). While *EQP* showed a weak correlation (i.e., $R < 0.39$) in the majority of the test sites (14 sites), *SFDC*, *FI*, and *RR* showed stronger correlations with *NSE*. However, the correlations (Table 3) varied from negative to positive across sites in all signatures indicating an inconsistency in the relationship (e.g., R_{SFDC} = 0.60 in site 02479300 and R_{SFDC} = –0.15 in site 11224500). Additionally, the signature with the highest correlation to *NSE* varied across sites indicating a spatial variability of the most influential hydrological signature (Fig. 5). For example, in sites 02273000, 02479300, 07375500, 07378000, and 08189500, *SFDC* showed the strongest correlation to model performance, while in sites 02296750, 02492000, 11025500, and 11224500, *RR* was the most strongly correlated (Table 3).

3.2 Assessment of model realism

Due to the critical importance of accurate process representation within a hydrological modeling application, the realism of the PDM structure was evaluated for each test catchment and analysis period (calibration and validation) utilizing hydrological signatures. The model appropriateness test applied here is defined by the ability of the model to recreate hydrological signatures in the calibration and validation periods (Euser et al., 2013). The observed and modeled signature values and the model performance in the calibration and validation periods were evaluated under direct and moving window approaches. Model performance was evaluated in terms of *NSE*. In the direct calibration-validation approach, changes in *NSE* were contrasted to changes in the (scaled) hydrological signatures (Fig. 6).

Hydrological signatures calculated using observed streamflow data (observed signature values) indicated significant climatic variability across the test catchments in both calibration and validation periods, most notably in terms of the runoff ratio (*RR*) and the flashiness index (*FI*). The PDM reproduced *RR* and streamflow elasticity (*EQP*) very closely for most catchments over both the calibration and validation periods (i.e., observed and predicted values fall along the 1:1 reference line; Fig. 6). The PDM model showed a lesser ability to simulate *SFDC* and *FI*. *SFDC* was simulated by the PDM fairly consistently at the low end of the scale over both the calibration and validation

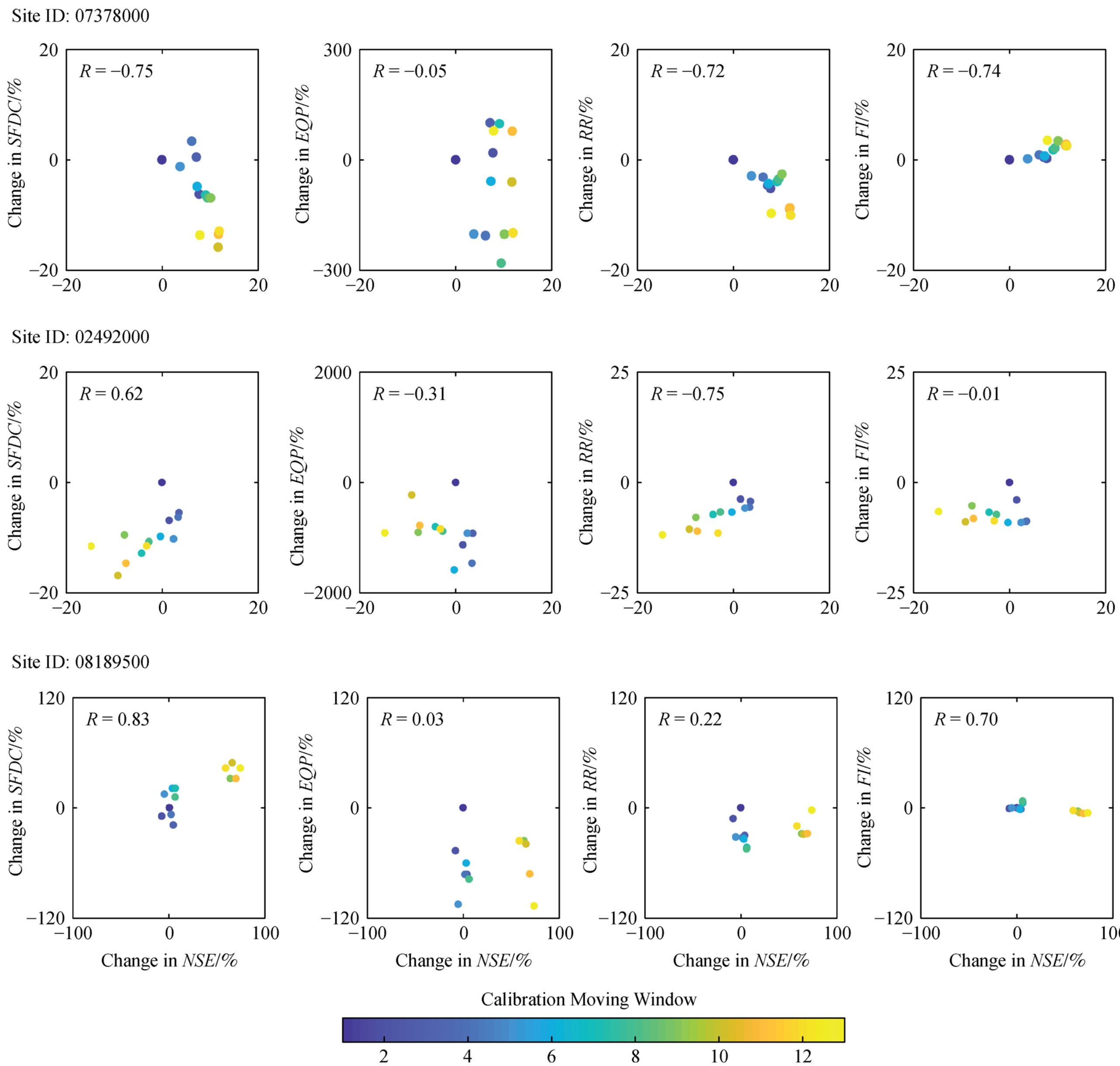

Fig. 4 *NSE* and hydrological signatures (at 3 representative sites) for 13 moving windows. Correlation coefficients (*R*) between change in signature and NSE are given on each plot.

periods, but had substantial variation in reproducibility at the high end of the scale. There was good reproduction of *FI* for most sites in the calibration period, but showed significant degradation during the validation period (Fig. 6).

NSE varied to a greater extent in both the calibration and validation periods across the test catchments. Although the model was able to replicate hydrological signatures in most cases, the fit between observed and predicted streamflow hydrograph was not always desirable. For example, while the model reproduced *RR* in the calibration period for most test catchments, the *NSE* performance was not equally consistent, i.e., *NSE* ranged from 0.40 to 0.90 across sites (Fig. 6, where *NSE* is represented by color scale). Additionally, although observed and modeled *SFDC* was not matched closely in site 11224500 (yellow point in lower right corner), the *NSE* value was the highest for all sites, including those where the observed and modeled *SFDC* matched closely (see Fig. 6, *SFDC* calibration period).

The variability in model performance and modeled and observed hydrological signatures were also considered under a three-month moving window approach to evaluate the results over a finer temporal resolution (Fig. 7). The PDM simulations were able to reproduce each of the four signatures (*FI*, *RR*, *SFDC*, and *EQP*) in most of the

Table 3 Correlation coefficients between the change (Δ) in hydrologic signatures and model performance for the moving window calibration-validation approach[a]

USGS ID	Correlation Coefficients (*R*)			
	Δ*NSE*−Δ*SFDC*	Δ*NSE*−Δ*EQP*	Δ*NSE*−Δ*RR*	Δ*NSE*−Δ*FI*
02202500	−0.17	−0.61	0.11	0.41
02228000	−0.15	0.63	0.15	0.26
02273000	−0.63	0.03	0.55	−0.38
02296750	0.27	−0.17	−0.82	−0.69
02329000	0.32	0.25	0.35	−0.86
02349500	−0.50	−0.29	−0.50	−0.24
02365500	0.05	0.28	0.06	0.15
02375500	0.47	0.08	0.69	0.83
02475000	0.26	0.00	0.07	0.40
02479300	0.60	0.09	0.20	0.35
02492000	0.62	−0.31	0.75	−0.01
07375500	0.60	0.11	0.05	−0.10
07378000	−0.75	−0.05	−0.72	0.74
07378500	0.50	−0.02	0.30	−0.65
08189500	0.83	0.03	0.22	−0.70
11025500	0.49	−0.59	−0.88	0.78
11138500	−0.53	−0.57	−0.59	0.82
11224500	−0.15	−0.33	−0.62	0.33

[a]Where (yellow) represents $|R| < 0.4$, (orange) $0.4 \leqslant |R| \leqslant 0.6$, and (green) $|R| > 0.6$.

catchments across all 13 moving windows, i.e., regardless of how far removed the validation period was from the calibration period. However, again, the model fit (as *NSE*) was not equally as good at several catchments (e.g., sites 02329000 and 022967509), despite the model's ability to replicate signatures in those cases. Some catchments with a close fit between observed and modeled signatures (e.g., site 02329000) still resulted in low *NSE* values relative to catchments which were unable to closely replicate signatures (e.g., site 110255006).

4 Discussion

The reliability of spatial and temporal model transfer remains a major challenge encompassing numerous uncertainties. Given this, a better understanding of the link between hydrological attributes, model performance, and ultimately, model structure, is vital in the search for robust model predictions. In this study, the temporal transfer of model parameters using signatures as indicators of model realism and transferability potential was explored. The relationship between model performance and hydrological signatures, using both a traditional split-sample approach and a moving window approach to model validation, was investigated.

4.1 Can simple hydrological models replicate signatures?

Hydrological signatures have been used to evaluate the suitability of model structures (Euser et al., 2013), where the suitability or realism of the model is assessed based on the ability of the model structure to adequately reproduce hydrological signatures. In this application, the PDM structure was able to reproduce hydrological signatures consistently both temporally (across calibration and validation periods) and spatially (across a majority of test catchments). Nevertheless, the model performance (in terms of *NSE*) was not equally consistent from calibration to validation periods (Fig. 6). High model performance (i.e., *NSE*) and its degree of consistency when transferred temporally can provide useful information about the suitability of a certain model structure for a certain catchment. In addition, poor performance and poor consistency of a certain model structure can be an indicator of missing processes in the model structure or the calibration period used for the catchment. A higher degree of process realism does not necessarily denote better performance for a specific objective function; however, it generally yields a more comparable result for the performance across calibration and validation periods, and therefore, smaller predictive uncertainty.

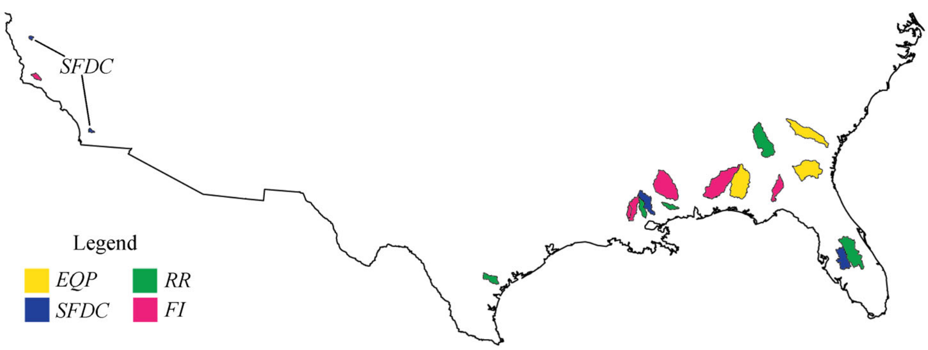

Fig. 5 Spatial distribution of hydrological signature most highly correlated to *NSE* under the moving window calibration-validation approach.

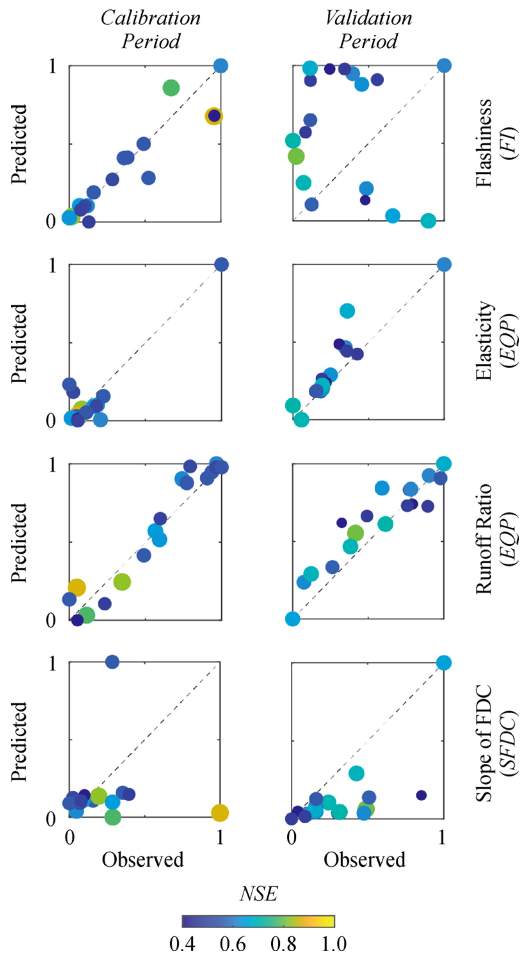

Fig. 6 Scatterplots of predicted vs. observed hydrological signatures across calibration and validation periods for all 18 MOPEX sites used in this study. Note, each point's color (and size) represents model performance (as *NSE*).

The results of this study demonstrate that the reproducibility of several hydrological signatures does not itself guarantee hydrological model transferability. This result also calls into question the utility of hydrological signatures to predict the realism of model structures as is commonly assumed in current literature (e.g., Euser et al., 2013). Signatures are summary measures for specific hydrograph characteristics. The same value of a summary measure can result from two data sets with differing temporal dynamics. For example, two data sets might yield similar mean annual streamflow values but have extremely different time series. Therefore, assessing the model structural realism solely based on the ability of a model to reproduce one or more hydrological signatures should be more carefully considered.

4.2 What do signatures indicate about temporal parameter transferability?

In the direct split-sample approach, the variation of model performance across calibration and validation periods showed a weak relationship (i.e., $R < 0.4$) with the variation of *SFDC*, *EQP*, *FI*, and *RR* (e.g., Fig. 3). In this approach, we represented the relationship between a hydrological signature and model performance by providing one value of correlation for all different catchments. Therefore, this weak correlation is a function of the many implicit assumptions that are made about the hydrological model in terms of process representation (realism). Additionally, this can be attributed to the spatial variability of dominant catchment attributes (i.e., the most important hydrological signatures) across catchments resulting in a weak correlation between signatures and model performance.

When the variations in model performance were contrasted to the corresponding variations in signatures across a finer scale of temporal variation (i.e., moving window validation), *SFDC*, *FI*, and *RR* showed moderate (i.e., $0.4 \leqslant R \leqslant 0.6$) to strong correlations (i.e., $R > 0.6$) in most catchments, while correlations with *EQP* were weak (Table 3). The moving window approach can aid in identifying relationships between hydrological signatures and model performance for each catchment. Therefore, a strong correlation between catchment and signature combinations could result when the PDM model is physically realistic to the catchments and the dominant attributes match the signature. However, the sign of the correlation for both *SFDC* and *RR* varied from negative to positive (i.e., positive and negative relationships) indicating relationship inconsistencies.

4.3 What signatures/characteristics were most important?

In the direct, split-sample approach, the correlation between model performance and *SFDC*, *FI*, *RR*, and *EQP* was weak. In the moving window approach, *EQP* showed a weak correlation in most of the sites, while other signatures, *SFDC*, *FI*, and *RR* showed moderate to strong correlations at many sites (Table 3). While hydrological attributes, such as flow variability (e.g., *SFDC*), variation in water balance partitioning between evapotranspiration and streamflow (e.g., *RR*), and frequency and rapidity of short-term changes in streamflow (e.g., *FI*), influenced model performance (and by extension, model parameterization), other attributes, such as sensitivity of streamflow to changes in rainfall (*EQP*) were less influential. Additionally, the correlation between the signature and model performance varied across sites indicating the dependency of the signature's predictive skills on the characteristics of the site.

Identifying the most influential catchment characteristics and hydrological processes is important to develop a better understanding of the spatial and temporal variability/transferability of model parameters. The model calibration process is inherently a function of the current catchment conditions, which represent a discrete sample from all

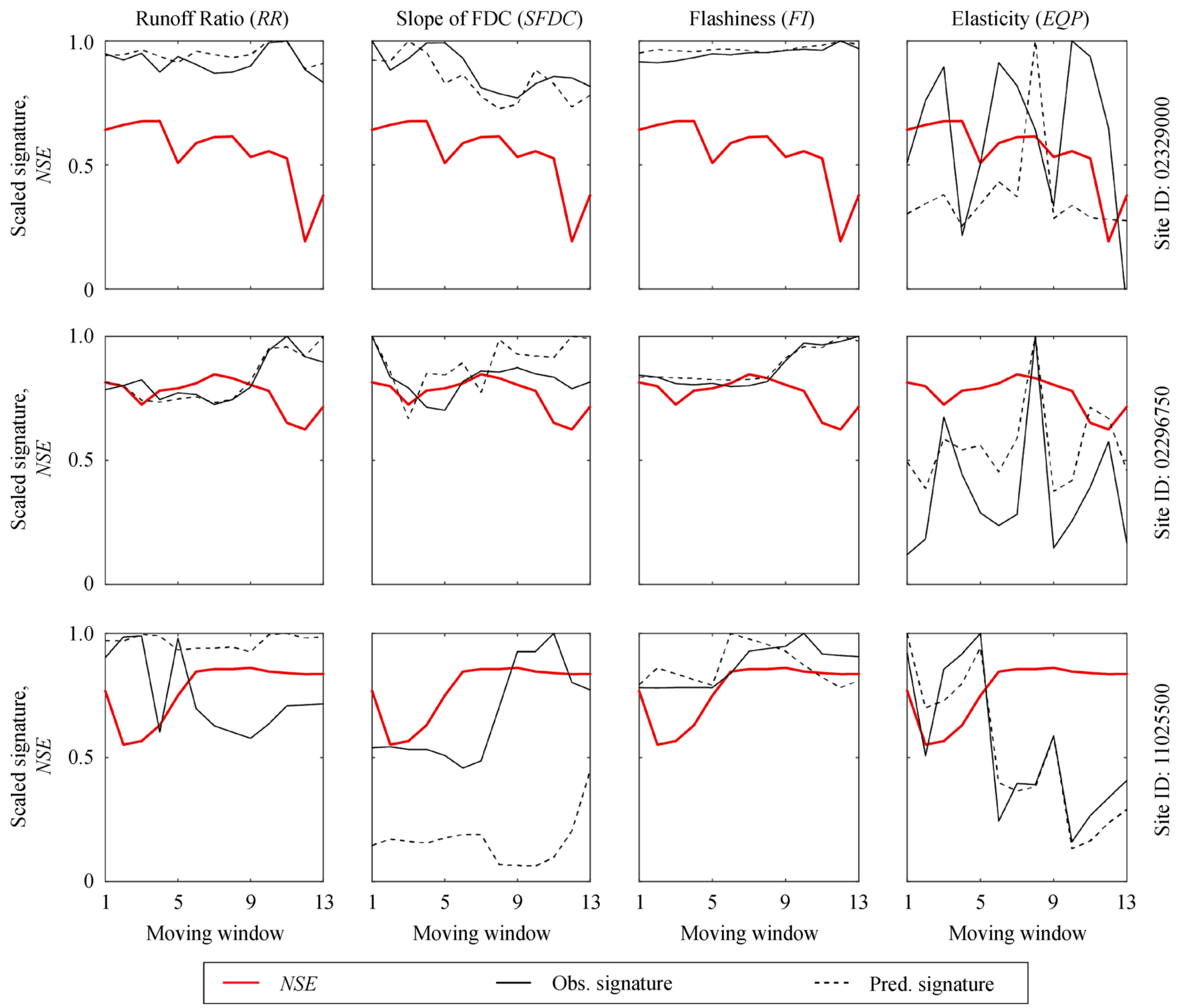

Fig. 7 Variations in model performance and predicted and observed hydrological signatures (for 3 representative sites) under the moving window approach.

possible hydrologic/climatic conditions. The optimum parameter set obtained through search algorithms only provides numerical optima for the conditions represented in the calibration period. As such, model parameters are inevitably tied to the space-time scales of their calibration (Koren et al., 1999). Moreover, the catchment is never observed over the complete range of possible climatic situations, given the limitations of the available data. As a consequence, some of the hydrological processes included in the model structure may not have been significantly activated during the calibration period, leading to a poor characterization of potentially important hydrological processes. Thus, the hydrological signatures/processes most influential on validation period model predictions were dependent on the site and model structure.

4.4 What are the barriers to a better relationship?

A model is a simplified representation of a real system and as such, is composed of aggregated physical processes. The heterogeneity of a watershed is a difficult concept to incorporate into a simple hydrological model (Grayson and Blöschl, 2001). The PDM structure, while physically realistic for many catchments, does not guarantee that the included watershed processes will behave as intended due to a lack of structural process rigidity. For example, a catchment might be dominated by baseflow, but the model might be parameterized as surface flow dominated. Although this difference is significant in terms of process realism, the resulting streamflow hydrograph may yield the same total runoff. Hence, the variability of model appropriateness across multiple test catchments is a barrier for a better understanding of the relationship between signatures and hydrological model predictions. An inconsistent relationship between model performance and hydrological signatures spatially across the test catchments reaffirms the role that unique site characteristics play in determining the functional relationship between model structure (parameterization) and model performance. Hydrological signatures may often be indicative of a certain hydrological change in the data from one period to the next, while still characterizing the important processes and how they manifest in the model parameterization. At the same time, the most informative signature can vary, both spatially and temporally, according to physical and climatic characteristics and their interaction. The degree of hydrological signature transferability (i.e., consistent explanation of dominant patterns) can vary between stream types (e.g., snowmelt-dominated, groundwater-dominated, or intermittent streams) (Olden and Poff, 2003). Thus, one signature is likely inadequate to illustrate complex watershed behaviors and unable to predict the model performance in new analysis (validation) periods.

4.5 Is the information content within selected signatures sufficient to characterize hydrological functions in PDM?

To understand the relationship between the selected hydrological signatures and the PDM parameters, we analyzed the sensitivity of each signature to changes in the PDM parameters. The general sensitivity analysis framework called Variogram Analysis of Response Surfaces (VARS) (Razavi and Gupta, 2016a) was used for the analysis, where we quantified the change in a signature value to the change in parameter value. The VARS framework provides a comprehensive set of “global” sensitivity metrics with minimal computational cost. Razavi and Gupta (2016b) have demonstrated the effectiveness, efficiency, and reliability (robustness) of the VARS framework on real-data case studies.

The sensitivity of each signature to PDM model parameters was found to be consistent across all catchments considered in this study. Sensitivity results for a representative catchment (Site ID: 02228000) are shown in Fig. 8. *NSE* between modeled discharge and the observed data was most sensitive to PDM routing parameters *Tres*1 and *Tres*2 and the subsurface storage drainage rate parameter *Kb* (Fig. 8). While, *FI*, *EQP*, and *SFDC* were most sensitive to the fraction of surface storage released to outflow *Tres*2, they were also found to be sensitive to the fraction of subsurface storage released to outflow *Tres*1. The runoff ratio (*RR*) was more sensitive to the soil storage threshold for subsurface inflow *CF* (Fig. 8). The *RR* represents the long-term water balance separation between water that exits the catchment as runoff and evapotranspiration. Thus, *RR* was more sensitive to water balance parameters of the model, such as *CF* and $C_{\max}$ (maximum soil moisture capacity. The other signatures represent different aspects of streamflow variability in a catchment. For example, *SFDC* represents the relationship between the frequency and magnitude of flows, *FI* indicates the frequency and rapidity of short-term changes in the streamflow, and *EQP* indicates the climate elasticity of streamflow. Thus, they are more sensitive to the flow routing parameters that determine the timing of surface and subsurface flows.

Although the use of a single hydrological signature oversimplifies the problem and lacks adequate information to fully characterize streamflow dynamics, combining several signatures might be successful in producing better predictive performance. However, the results found in this study did not bear this out. Many hydrological signatures are correlated, given their mathematical formulation. The signatures chosen for this study are not necessarily

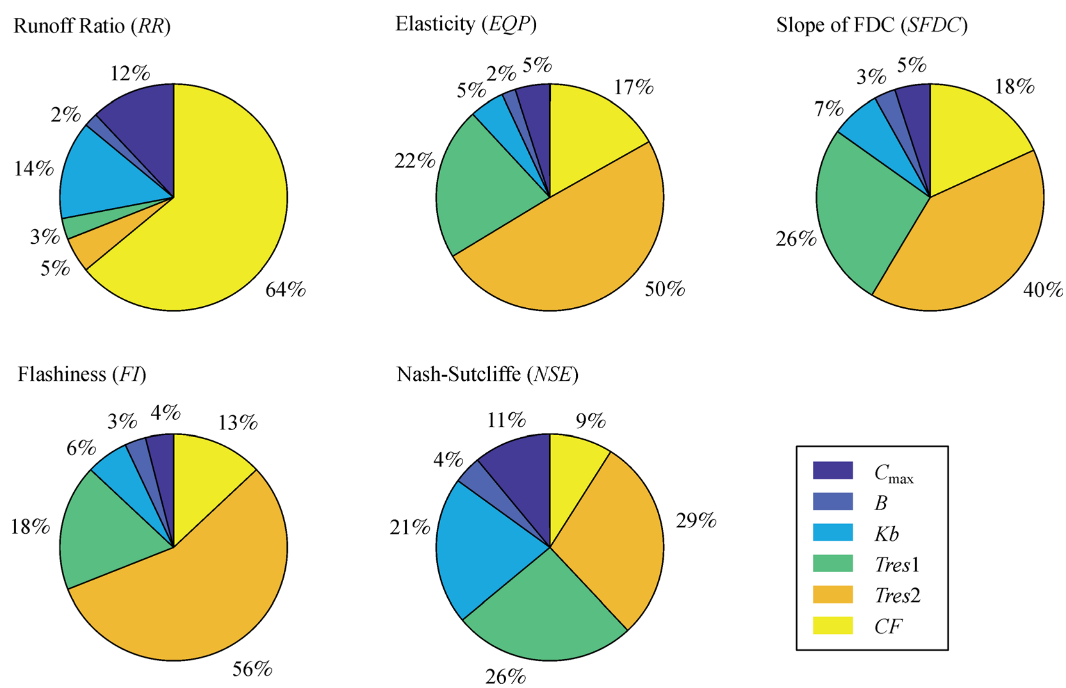

Fig. 8 Sensitivity of hydrological signatures and *NSE* to changes in PDM parameters for a representative site (Site ID: 02228000).

independent; the examined signatures have some degree of overlap in terms of the information they provide (Fig. 8; *FI*, *EQP*, *SFDC*). Including additional targeted, yet uncorrelated signatures based on the hydroclimatic region that explain different aspects of the hydrological response of a catchment could fortify the predictive skill of hydrological signatures. However, choosing a parsimonious subset of hydrological indices that represent critical streamflow characteristics in a non-redundant manner is not a trivial task.

5 Conclusions

Hydrological models typically require the transfer of calibrated parameters either in time (for streamflow forecasting) or space (for prediction at ungauged catchments) or both in an applied or real world setting. Consistent temporal transfer of hydrological model parameters without sacrificing model performance (i.e., fit) is an enduring challenge in hydrology. In this study, the identification of hydrological signatures capable of indicating model realism and potential for validation success by understanding the link between hydrological signatures and model performance (in the context of parameter transferability) was sought. Model appropriateness was examined based on the ability of a model parameterization to simulate observed hydrological signatures.

While the results indicated that the PDM was able to effectively reproduce signatures over both calibration and validation periods (at most of the test catchments considered), model performance (in terms *NSE*) was not equally robust. Although high model performance (in terms of streamflow) and internal variable consistency (i.e., reproduction of hydrological variables not used for parameter estimation) are typically assumed to be indicative of model appropriateness (or realism) for a certain catchment, the use of hydrological signatures as indicators of model realism was found to be of limited value. Hydrological signatures are summary measures of the hydrological variability in time series data and can be similar for two periods of vastly different temporal dynamics. Given this, the reproducibility of signatures should not be viewed as a verification of model appropriateness. As model appropriateness is often cited as a surrogate for parameter transferability (a long-held assumption in hydrological modeling), the results of this study provide direct evidence in contrast to this generalization.

When the variations in hydrological signatures were compared with variations in model performance in the traditional split-sampling approach, a weak relationship was observed. In contrast, in the moving window approach, the correlation between hydrological signatures and model performance ranged from moderate to strong for many sites, but was not consistent across all catchments or signatures tested. The most informative signature can vary, both spatially and temporally, according to physical and climatic characteristics and their interaction to the model parameterization.

Identification of predictive signatures for temporal parameter transferability is influenced by the realism of the model structure, spatially and temporally. The variability of the most informative signature is dependent on the site and the model structure. At the same time, hydrological signatures are only indicative of catchment patterns rather than specific processes. Thus, one signature was not adequate to illustrate complex watershed behaviors nor able to predict model performance in new analysis periods. Accommodating more hydrological signatures that are strongly correlated to the model parameterization process will benefit the search for effective signatures that are indicative of temporal parameter transfer success. The result of this study highlights the ability of hydrological models to provide long-term flow signatures at both annual and seasonal scales; however, caution should be taken when considering shorter temporal scales.

Although the methodology presented in this study utilizes a specific model, the framework is general and could easily be extended to other hydrological models. This study selected hydrological signatures derived by using readily available data, such as precipitation, temperature, and streamflow. However, hydrological signatures should be selected based on their correlation to the model structure/parameterization to understand the relationship between hydrological signature and model performance. Using more physically realistic models for many catchments may reduce the uncertainty in understanding the predictive power of hydrological signatures.

Acknowledgements We acknowledge the National Oceanic and Atmospheric Administration for making available the Model Parameter Estimation Experiment (MOPEX) data that was used for this study and S. Razavi for providing free access to the VARS toolbox. We thank the two anonymous reviewers for their contributions toward the improvement of this manuscript. Funding for this research was provided through a graduate scholarship from Clarkson University awarded to D. Jayathilake.

References

Auer A H Jr (1974). The rain versus snow threshold temperatures. Weatherwise, 27(2): 67

Baker D B, Richards R P, Loftus T T, Kramer J W (2004). A new flashiness index: characteristics and applications to midwestern rivers and streams. JAWRA J Am Water Res, 40(2): 503–522

Beck H E, van Dijk A I, de Roo A, Miralles D G, McVicar T R, Schellekens J, Bruijnzeel L A (2016). Global—scale regionalization of hydrologic model parameters. Water Resour Res, 52(5): 3599–3622

Beven K (2002). Towards a coherent philosophy for modelling the environment. P Royal Soc A-Math Phy, 458(2026): 2465–2484

Blöschl G, Sivapalan M, Savenije H, Wagener T, Viglione A, eds. (2013). Runoff Prediction in Ungauged Basins: Synthesis Across Processes, Places and Scales. Cambridge University Press

Casper M C, Grigoryan G, Gronz O, Gutjahr O, Heinemann G, Ley R, Rock A (2012). Analysis of projected hydrological behavior of catchments based on signature indices. Hydrol Earth Syst Sci, 16(2): 409–421

Castiglioni S, Lombardi L, Toth E, Castellarin A, Montanari A (2010). Calibration of rainfall-runoff models in ungauged basins: a regional maximum likelihood approach. Adv Water Resour, 33(10): 1235–1242

Dai A (2008). Temperature and pressure dependence of the rain-snow phase transition over land and ocean. Geophys Res Lett, 35(12)

Donnelly C, Andersson J C, Arheimer B (2016). Using flow signatures and catchment similarities to evaluate the E-HYPE multi-basin model across Europe. Hydrol Sci J, 61(2): 255–273

Duan Q, Schaake J, Andréassian V, Franks S, Goteti G, Gupta H V, Gusev Y M, Habets F, Hall A, Hay L, Hogue T, Huang M, Leavesley G, Liang X, Nasonova O N, Noilhan J, Oudin L, Sorooshian S, Wagener T, Wood E F (2006). Model parameter estimation experiment (MOPEX): an overview of science strategy and major results from the second and third workshops. J Hydrol (Amst), 320 (1–2): 3–17

Euser T, Winsemius H C, Hrachowitz M, Fenicia F, Uhlenbrook S, Savenije H H G (2013). A framework to assess the realism of model structures using hydrological signatures. Hydrol Earth Syst Sci, 17 (5): 1893–1912.

Ewen J (2011). Hydrograph matching method for measuring model performance. J Hydrol (Amst), 408(1–2): 178–187

Grayson R, Blöschl G (2001). Spatial patterns in catchment hydrology: observations and modelling. CUP Archive

Hingray B, Schaefli B, Mezghani A, Hamdi Y (2010). Signature-based model calibration for hydrological prediction in mesoscale Alpine catchments. Hydrolog Sci J, 55(6): 1002–1016

Hrachowitz M, Fovet O, Ruiz L, Euser T, Gharari S, Nijzink R, Freer J, Savenije H H G, Gascuel-Odoux C (2014). Process consistency in models: The importance of system signatures, expert knowledge, and process complexity. Water Resour Res, 50(9): 7445–7469

Hrachowitz M, Savenije H H G, Blöschl G, McDonnell J J, Sivapalan M, Pomeroy J W, Arheimer B, Blume T, Clark M P, Ehret U, Fenicia F, Freer J E, Gelfan A, Gupta H V, Hughes D A, Hut R W, Montanari A, Pande S, Tetzlaff D, Troch P A, Uhlenbrook S, Wagener T, Winsemius H C, Woods R A, Zehe E, Cudennec C (2013). A decade of predictions in ungauged basins (PUB)—a review. Hydrol Sci J, 58 (6): 1198–1255

Kay A L, Jones D A, Crooks S M, Kjeldsen T R, Fung C F (2007). An investigation of site-similarity approaches to generalisation of a rainfall-runoff model. Hydrol Earth Syst Sci Discuss, 11(1): 500–515

Koren V I, Finnerty B D, Schaake J C, Smith M B, Seo D J, Duan Q Y (1999). Scale dependencies of hydrologic models to spatial variability of precipitation. J Hydrol (Amst), 217(3–4): 285–302

Masih I, Uhlenbrook S, Maskey S, Ahmad M D (2010). Regionalization of a conceptual rainfall–runoff model based on similarity of the flow duration curve: a case study from the semi-arid Karkheh basin, Iran. J Hydrol (Amst), 391(1–2): 188–201

Merz R, Parajka J, Blöschl G (2011). Time stability of catchment model parameters: implications for climate impact analyses. Water Resour Res, 47(2): W02531

Montanari A, Toth E (2007). Calibration of hydrological models in the spectral domain: an opportunity for scarcely gauged basins? Water Resour Res, 43(5): W05434

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 Discuss, 11(1): 483–499

Nash J, Sutcliffe J V (1970). River flow forecasting through conceptual models part I—a discussion of principles. J Hydrol (Amst), 10(3): 282–290

Olden J D, Poff N L (2003). Redundancy and the choice of hydrologic indices for characterizing streamflow regimes. River Res Appl, 19(2): 101–121

Parajka J, Blöschl G, Merz R (2007). Regional calibration of catchment models: potential for ungauged catchments. Water Resour Res, 43(6): W06406

Patil S, Stieglitz M (2012). Controls on hydrologic similarity: role of nearby gauged catchments for prediction at an ungauged catchment. Hydrol Earth Syst Sci, 16(2): 551–562

Patil S D, Stieglitz M (2015). Comparing spatial and temporal transferability of hydrological model parameters. J Hydrol (Amst), 525: 409–417

Prudhomme C, Haxton T, Crooks S, Jackson C, Barkwith A, Williamson J, Kelvin J, Mackay J, Wang L, Young A, Watts G (2013). Future flows hydrology: an ensemble of daily river flow and monthly groundwater levels for use for climate change impact assessment across Great Britain. Earth Syst Sci Data, 5(1): 101–107

Razavi S, Gupta H V (2016a). A new framework for comprehensive, robust, and efficient global sensitivity analysis: 1. theory. Water Resour Res, 52(1): 423–439

Razavi S, Gupta H V (2016b). A new framework for comprehensive, robust, and efficient global sensitivity analysis: 2. application. Water Resour Res, 52(1): 440–455

Samaniego L, Bárdossy A, Kumar R (2010). Streamflow prediction in ungauged catchments using copula-based dissimilarity measures. Water Resour Res, 46(2): W02506

Samuel J, Coulibaly P, Metcalfe R A (2011). Estimation of continuous streamflow in Ontario ungauged basins: comparison of regionalization methods. J Hydrol Eng, 16(5): 447–459

Sawicz K, Wagener T, Sivapalan M, Troch P A, Carrillo G (2011). Catchment classification: empirical analysis of hydrologic similarity based on catchment function in the eastern USA. Hydrol Earth Syst Sci, 15(9): 2895–2911

Seibert J (2003). Reliability of model predictions outside calibration conditions. Nord Hydrol, 34(5): 477–42

Seibert J, McDonnell J J (2002). On the dialog between experimentalist and modeler in catchment hydrology: use of soft data for multicriteria model calibration. Water Resour Res, 38(11): 23–1

Tolson B A, Shoemaker C A (2007). Dynamically dimensioned search algorithm for computationally efficient watershed model calibration. Water Resour Res, 43(1): W01413

Tolson B A, Shoemaker C A (2008). Efficient prediction uncertainty approximation in the calibration of environmental simulation models. Water Resour Res, 44(4): W04411

Vaze J, Post D A, Chiew F H S, Perraud J M, Viney N R, Teng J (2010). Climate non-stationarity-validity of calibrated rainfall-runoff models for use in climate change studies. J Hydrol (Amst), 394(3–4): 447–457

Wagener T, McIntyre N, Lees M J, Wheater H S, Gupta H V (2003). Towards reduced uncertainty in conceptual rainfall-runoff modelling: dynamic identifiability analysis. Hydrol Processes, 17(2): 455–476

Wagener T, Sivapalan M, Troch P, Woods R (2007). Catchment classification and hydrologic similarity. Geogr Compass, 1(4): 901–931

Westerberg I K, Guerrero J L, Younger P M, Beven K J, Seibert J, Halldin S, Freer J E, Xu C Y (2011). Calibration of hydrological models using flow-duration curves. Hydrol Earth Syst Sci, 15(7): 2205–2227

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 Resour Res, 52(3): 1847–1865

Yadav M, Wagener T, Gupta H (2007). Regionalization of constraints on expected watershed response behavior for improved predictions in ungauged basins. Adv Water Resour, 30(8): 1756–1774

Zhang Y Q, Viney N R, Chiew F H S, Van Dijk A I J M, Liu Y Y (2011). Improving hydrological and vegetation modelling using regional model calibration schemes together with remote sensing data. In: Proceedings of the 19th International Congress on Modelling and Simulation (MODSIM’11): 3448–3454

Zhang Y, Zheng H, Chiew F H, Arancibia J P, Zhou X (2016). Evaluating regional and global hydrological models against streamflow and evapotranspiration measurements. J Hydrometeorol, 17(3): 995–1010