Advances in Water Resources 112 (2018) 245–253

Contents lists available at ScienceDirect

Advances in Water Resources

journal homepage: www.elsevier.com/locate/advwatres

Typecasting catchments: Classification, directionality, and the pursuit of universality

Tyler Smith[a,*], Lucy Marshall[b], Brian McGlynn[c]

[a] Department of Civil and Environmental Engineering, Clarkson University, Potsdam, NY 13699, USA
[b] School of Civil and Environmental Engineering, University of New South Wales, Sydney, NSW 2052, Australia
[c] Division of Earth and Ocean Sciences, Duke University, Durham, NC 27708, USA

ARTICLE INFO

Keywords:
Catchment
Classification
Runoff
Model
Transferability
Regionalization

ABSTRACT

Catchment classification poses a significant challenge to hydrology and hydrologic modeling, restricting widespread transfer of knowledge from well-studied sites. The identification of important physical, climatological, or hydrologic attributes (to varying degrees depending on application/data availability) has traditionally been the focus for catchment classification. Classification approaches are regularly assessed with regard to their ability to provide suitable hydrologic predictions – commonly by transferring fitted hydrologic parameters at a data-rich catchment to a data-poor catchment deemed similar by the classification. While such approaches to hydrology's grand challenges are intuitive, they often ignore the most uncertain aspect of the process – the model itself. We explore catchment classification and parameter transferability and the concept of universal donor/acceptor catchments. We identify the implications of the assumption that the transfer of parameters between "similar" catchments is reciprocal (i.e., non-directional). These concepts are considered through three case studies situated across multiple gradients that include model complexity, process description, and site characteristics. Case study results highlight that some catchments are more successfully used as donor catchments and others are better suited as acceptor catchments. These results were observed for both black-box and process consistent hydrologic models, as well as for differing levels of catchment similarity. Therefore, we suggest that similarity does not adequately satisfy the underlying assumptions being made in parameter regionalization approaches regardless of model appropriateness. Furthermore, we suggest that the directionality of parameter transfer is an important factor in determining the success of parameter regionalization approaches.

1. Introduction

Catchment classification, the systematic arrangement of catchments into similar groups based on form and/or function, has the potential to be a powerful concept in our understanding and conceptualization of hydrologic processes and fundamental catchment characteristics across spatial and temporal scales (McDonnell and Woods, 2004; Wagener et al., 2007). As a tool, classification enables an objective framework within which hydrologic experiments, model conceptualization, and model testing are conducted and/or reconciled. A key anticipated outcome of such an approach is the ability to produce improved model predictions. This is of particular importance in catchments with no or limited data availability. In this case the information at a data-rich catchment can be shared with a similarly classified data-poor catchment to leverage available information.

The scope of the predictions in ungauged basins problem is vast and concerns many aspects of hydrologic understanding and prediction (Sivapalan et al., 2003). Here, we seek to focus on catchment classification and its frequent application in the context of predictions in ungauged basins (Hrachowitz et al., 2013). Under such scenarios, it is commonly assumed that the attribute similarity of catchments (physical, climatological, hydrological) leads to a parameter similarity (Bárdossy, 2007). Classification information is typically exploited through a regionalization approach to obtain knowledge about an ungauged location, such as hydrologic model parameters (e.g., Hundecha et al., 2008; Oudin et al., 2008; Di Prinzio et al., 2011), hydrologic function (e.g., Carrillo et al., 2011; Sawicz et al., 2011; Ye et al., 2012), or catchment form (Winter, 2001; e.g., Leibowitz et al., 2016).

Implicit in catchment classification and the associated regionalization studies is the concept of reciprocity – a mutual exchange of information, resources, services, etc. If the catchments are reciprocal, it should not matter which is acting as the donor (supplying information) and which is acting as the acceptor (receiving information) – reciprocal catchments have a mutual, non-directional exchange of information.

* Corresponding author.
E-mail address: tsmith@clarkson.edu (T. Smith).

https://doi.org/10.1016/j.advwatres.2017.12.020
Received 16 August 2017; Received in revised form 29 November 2017; Accepted 23 December 2017
Available online 30 December 2017
0309-1708/ © 2018 Elsevier Ltd. All rights reserved.

This means that information from one catchment may be successfully transferred to any similarly classified catchment with minimized impact to the resulting model application. Many regionalization approaches are predicated on this assumption (e.g., McIntyre et al., 2005; Oudin et al., 2008; Smith et al., 2014). However, it is probable that catchments differ in their ability to donate or accept model parameters within a classified group given the realities of model and observational errors as models are reconciled with data. In this way, individual catchments may be better considered as analogous to blood types (A, B, AB, O), where the donor 'type' controls the directionality of the transfer.

The inherent uncertainty associated with the hydrologic model itself (e.g., process representation, complexity, etc.) will impact assumptions made in developing robust regionalization approaches using catchment classification tools. Indeed, it is this disconnect between the mathematical model and catchment structure that has contributed to the lack of a reliable, universal classification framework in modeling applications. Addressing such issues requires developing a better understanding of assumptions, where they hold, where they fail, and how they propagate through the modeling process.

We sought to examine the underlying assumptions of catchment classification (i.e., reciprocity) through the lens of hydrologic model regionalization (using direct parameter transfer). We apply a cross validation approach, considering gradients across model structures (in terms of complexity, process fidelity) and diverse catchment settings. In particular, we were interested in addressing the questions:

- Does physical similarity, in a broad sense, guarantee parameter transfer reciprocity?
- Is there a relationship between physical and functional (parameter) similarity?
- How do model complexity and realism affect the transferability of model parameter sets?
- Are any relationships identified consistent across diverse sets of catchments?

2. Analysis scenarios

Three scenarios were investigated to explore cases across gradients of catchment attributes and model realism. Scenarios 1 and 2 feature several well-studied and highly monitored catchments located within the Tenderfoot Creek Experimental Forest in Montana. Combined, these scenarios explore the importance of model realism on the directionality of model parameter transferability. Scenario 3 features a widely-used collection of catchments with basic hydrometric and catchment attribute data located in southeastern Australia. In conjunction with Scenario 2, the importance of catchment attribute similarity on the direction of model parameter transferability is explored. The directionality of model parameter transferability was considered within each scenario and across scenarios to leverage the combined results. In each of the subsequent scenarios, a group of catchments was identified that could be classified as being similar based on climate, topography, proximity, vegetation, or the parametric similarity of an optimized hydrologic model. Then, a hydrologic model was selected to represent the perceived dominant hydrologic processes (with different levels of realism based on our *a priori* knowledge). The hydrologic model parameters were calibrated to observed streamflow at each individual catchment to be considered. Following this, a leave-one-out transferability analysis was performed, where each catchment was used in-turn as a donor to the other catchments. Changes in model performance were tracked not only for the ability of a catchment to serve as a donor to other catchments, but also for its ability to serve as a recipient from other catchments. Donor suitability was assessed based on an average reduction in NSE (where each catchment was benchmarked to its own site-specific calibration). Such an approach allows for the validity of the assumption that catchment similarity is an effective surrogate for model parameter similarity to be considered. If this assumption is valid, it is expected that catchment performance should be similar when transferring model parameters to another, "similar" catchment in *both* directions (i.e., as donor and as recipient).

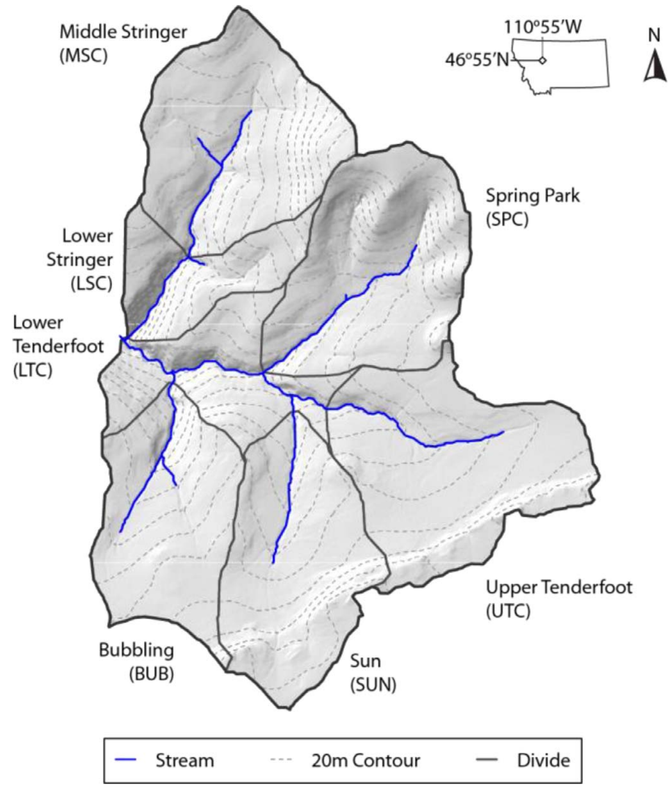

Fig. 1. The Tenderfoot Creek Experimental Forest, with each of the seven catchments analyzed in this study, following Smith et al. (2016).

2.1. Scenario 1: process-consistent model applied to tenderfoot creek experimental forest

The Tenderfoot Creek Experimental Forest (TCEF), located in Montana, USA, encompasses an area of 22.8 km^2 and is drained by Tenderfoot Creek, a tributary to the Smith River. TCEF (lat. 46°55'N, long. 1110°52'W; Fig. 1) is a highly-monitored research site, where many previous scientific investigations have been conducted (e.g., Jencso et al., 2009; Jencso et al., 2010; Jencso and McGlynn, 2011; Nippgen et al., 2011; Payn et al., 2009, 2012; Smith et al., 2014; Smith and Marshall, 2008, 2010). TCEF is a well-suited research site to consider catchment classification due to the expansive, detailed knowledge gained through collocated field- and model-based investigations. In this scenario, six sub-catchments nested within the Lower Tenderfoot Creek (LTC) catchment are considered – Bubbling Creek (BUB), Lower Stringer Creek (LSC), Middle Stringer Creek (MSC), Spring Park Creek (SPC), Sun Creek (SUN), and Upper Tenderfoot Creek (UTC). TCEF has a historical mean annual temperature of 0 °C and an average annual precipitation of 880 mm (approximately 75% as snow). For additional descriptions of the study area, refer to Jencso et al. (2009, 2010), Jencso and McGlynn (2011), and Nippgen et al. (2011).

The Catchment Connectivity Model (CCM) (Smith et al., 2013, 2016), which was developed for use at TCEF, was applied to each of the seven TCEF catchments. The CCM simulates shallow groundwater connectivity by exploiting the empirical relationship between connectivity and hillslope size (as upslope accumulated area) observed at TCEF (Jencso et al., 2009). The storage within each hillslope is computed by mass balance and its connectivity to the stream is achieved when the storage exceeds a threshold value based on hillslope size (as upslope accumulated area). Water is released from connected hillslopes

Table 1
Names and description of CCM parameters[a].

Parameter	Description	Limits	Units
q^*	the volumetric discharge from each hillslope per unit time	0 200	m³/day
κ	the Pareto parameter describing the shape of the UAA-connectivity relationship	0 1	--
τ	the residence time parameter of the exponential transfer function	0 2	day

[a]Each vertical line under "Limits" represent the optimal value from a Scenario 1 catchment.

as shallow groundwater at a rate equal to q^* (a calibrated parameter). Only connected hillslopes contribute water to total discharge at the catchment outlet. The CCM structure is a spatially explicit, process verifiable, three-parameter hydrologic model (Table 1), is underpinned by the concept that the frequency of hydrologic connections of the hillslopes to the stream drives streamflow rather than the magnitude of any single hydrologic connection (Jencso et al., 2009).

The CCM model was implemented on a 6-hourly time step with model forcing data, including SNOTEL precipitation (cf. Nippgen et al., 2011) and eddy-covariance derived evapotranspiration (Emanuel et al., 2010). Model calibration to streamflow data (Nippgen et al., 2011) for the period of October 2005 through September 2009 at each catchment. Topographic data (10 m DEMs derived from 1 m LiDAR) were used to determine catchment upslope accumulated areas (UAA) – a model input. Shallow groundwater connectivity data (Jencso et al., 2009) was utilized for independent model process verification.

CCM parameters were estimated using a Monte Carlo sampling approach (e.g., Kuczera and Parent, 1998), where 100,000 parameter sets were randomly sampled and the top 20% (based on Nash-Sutcliffe efficiency) retained for additional model analysis. Initial parameter ranges were selected considering field-based observations of parameter constraints following the findings presented in Smith et al. (2016), where such an approach was found to yield more physically consistent model simulations. The average Nash-Sutcliffe Efficiency for the best performing CCM parameter set at each of the seven TCEF catchments was 0.82 (max. = 0.88 at LTC; min. = 0.69 at UTC).

Using the site-specific calibration performance as a benchmark for parameter transferability, the average reduction in model performance was computed (as a percent change: $\otimes NSE = 1 - [Regional\ NSE\ /\ Site\ NSE]$). This metric was produced for each catchment as both a parameter donor and a parameter acceptor (Fig. 2). When the catchment is used as a donor, this metric is calculated as the average change in fit across the other six TCEF catchments (each benchmarked to their own site-specific NSE). When the catchment is used as an acceptor, this metric is calculated as the average change in fit when utilizing the optimum parameter set from each of the other six TCEF catchments (again, benchmarked to site-specific NSE). To facilitate comparison, an average reduction in model performance threshold of 10% was included to represent "acceptable" model performance degradation due to model transfer.

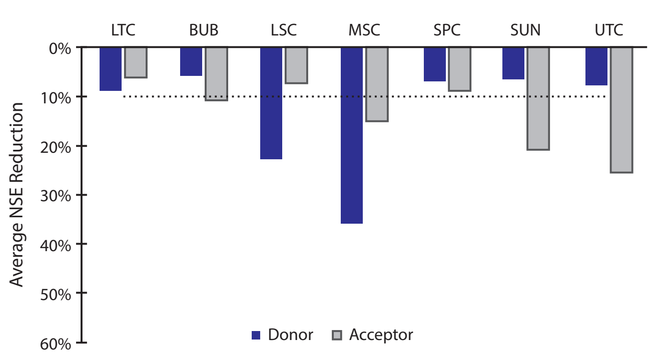

Fig. 2. Average reduction in model performance (as percentage of site-specific NSE) at each TCEF catchment when used as a donor or an acceptor with the Catchment Connectivity Model. The dashed line represents a threshold reduction value of 10%.

The CCM parameter transferability results (Fig. 2) clearly refute the implicit assumption that model transferability is reciprocal (i.e., if Catchment A is a good donor for Catchment B, then Catchment B will be a good donor for Catchment A). Although variation is to be expected, some catchments (e.g., MSC and LSC) show a propensity for being acceptors and not donors, while other catchments (e.g., SUN and UTC) show a propensity for being donors and not acceptors. The remaining catchments (e.g., LTC, BUB, and SPC) exhibit a more equitable correspondence between donor and acceptor performance.

To understand the causes of these relationships, we explored the parameter space for each of the site-specific calibrations (Fig. 3). The regions of high model performance (hot colors) are clearly variable from catchment to catchment, both in terms of the location within the parameter space and the spread across the parameter space. Strong variabilities in terms of the range and location of suitable values are particularly noticeable across the catchments in terms of the parameter controlling the discharge of stored water from hydrologically connected hillslopes (q^*). Catchments where model performance (as NSE) was less sensitive to q^* (e.g., LSC and MSC) were found to be well suited as parameter acceptors but poorly suited as parameter donors (Fig. 2). This is a result of such catchments having a large range of q^* values resulting in "good" NSE values, but having optimal q^* values at non-coincident locations in the parameter space, relative to the other catchments. The opposite was true for catchments where model performance was more sensitive to q^* (e.g., SUN and UTC). In these catchments, the optimal parameter set to be donated to the other catchments correspond to areas of high model performance in the accepting catchments. Variability in the sensitivity of model performance to the stored water release parameter (q^*) has been linked to catchment structure, specifically the distribution of catchment upslope accumulated area (Smith et al., 2016).

2.2. Scenario 2: conceptual model applied to tenderfoot creek experimental forest

The analysis presented in Scenario 1 was repeated using the GR4J model (Perrin et al., 2003) to understand the degree to which the model structure itself controls the transferability between catchments. The GR4J model is a simple conceptual model (4 calibrated parameters; Table 2) based on soil moisture accounting and unit hydrograph theory

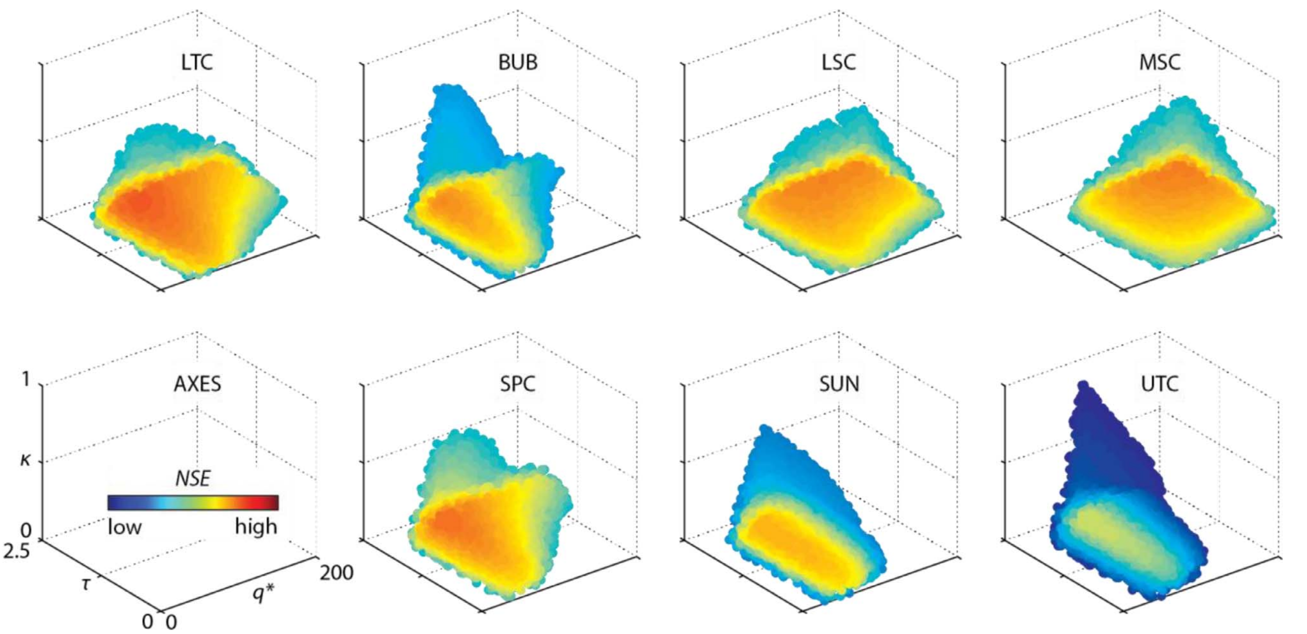

Fig. 3. Four-dimensional surface plots of the CCM parameter space and model performance (as NSE) for each TCEF catchment. Note that all subplots utilize common axes and color scale. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

that has been applied extensively and shown to perform well across a range of catchment conditions. The GR4J model structure was selected to provide a distinct alternative to the CCM structure used in Scenario 1, where CCM represents a specific runoff generating mechanism that is intrinsically linked to catchment form and condition and GR4J represents a structure aimed at flexibly representing a range of catchment conditions and hydrologic processes (Perrin et al., 2003). The same catchments (Fig. 1), input data, and temporal resolution were maintained in this scenario.

GR4J parameters were estimated using a Monte Carlo sampling approach (e.g. Kuczera and Parent, 1998), where 100,000 parameter sets were randomly sampled and the top 20% (based on Nash-Sutcliffe efficiency) retained for additional model analysis. Initial parameter ranges (Table 2) were selected based on the values suggested by Perrin et al. (2003). The average Nash-Sutcliffe Efficiency for the best performing GR4J parameter set at each of the seven TCEF catchments was 0.83 (max. = 0.87 at BUB; min. = 0.78 at SUN). Although average GR4J model performance was comparable to CCM performance (average NSE = 0.82; see Scenario 1), it should be noted that the degree to which the GR4J reproduces the actual hydrological processes is unknown due to a lack of corroboration to other hydrologic variables. Conversely, the fidelity of simulated hydrological processes at these catchments has been previously verified using auxiliary groundwater data utilizing the CCM applied in Scenario 1 (Smith et al., 2016).

Again, the site-specific calibration performance is utilized as a benchmark for GR4J parameter transferability, and the average reduction in model performance (as a percent change: $\Delta NSE = 1 - [Regional\ NSE\ /\ Site\ NSE]$) was computed for each catchment as both a parameter donor and a parameter acceptor (Fig. 4). The average reduction threshold of 10% was maintained to allow for direct comparison across the scenarios. The GR4J parameter transferability results (Fig. 4) demonstrate the same patterns as were observed in Scenario 1, with some catchments (e.g., MSC and LSC) acting as suitable acceptors and not donors, other catchments (e.g., SUN and UTC) acting as suitable donors and not acceptors, and the remaining catchments (e.g., LTC, BUB, and SPC) acting as (mostly) equitable donors and acceptors.

The parameter space (dotty plots) for each of the site-specific calibrations (Fig. 5) again highlighted a consistent variability in regions of high model performance from catchment to catchment. In Scenario 1 this variability was linked to the parameter controlling the discharge of

Table 2
Names and description of GR4J parameters[a].

Parameter	*Description*	*Limits*	*Units*
x1	maximum production storage	0 1000	mm
x2	groundwater exchange	-5 5	mm/day
x3	maximum routing storage	0 300	mm
x4	unit hydrograph time delay	0.5 5	day

[a]Each vertical line under "Limits" represent the optimal value from a Scenario 2 catchment.

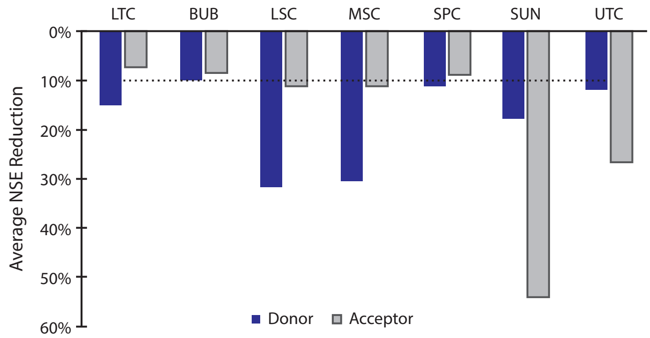

Fig. 4. Average reduction in model performance (as percentage of site-specific NSE) at each TCEF catchment when used as a donor or an acceptor with the GR4J model. The dashed line represents a threshold reduction value of 10%.

stored water from hydrologically connected hillslopes (q^*; Table 1). In this scenario, the variability can be linked to the parameter controlling the maximum capacity of the soil moisture storage (x_1; Table 2). Catchments where model performance (as NSE) was less sensitive to x_1 (e.g., LSC and MSC) were found to be well suited as parameter acceptors but poorly suited as parameter donors (Fig. 4). The opposite was true for catchments where model performance was more sensitive to x_1 (e.g., SUN and UTC). In these catchments, the optimal parameter set to be donated to the other catchments correspond to areas of high model performance in the accepting catchments. Variability in the sensitivity of model performance to the soil moisture storage capacity (x_1) and subsequent release (as percolation in the model structure, a function of x_1) suggests catchment structure plays an important role in model parameterization, even for models that lack an explicit connection to such information.

2.3. Scenario 3: conceptual model applied to collection of australian catchments

In order to place the analysis from Scenarios 1 and 2 into a broader context, previous work by Smith et al. (2014) that explored the role of model parameter similarity on model transferability and uncertainty, was extended to analyze transferability across a vastly different collection of catchments than was considered in Scenarios 1 and 2. In their study, Smith et al. (2014) applied the Probability Distributed Model (PDM; (Moore, 1985) to 118 catchments located across southeastern Australia. The PDM is a six-parameter soil moisture accounting model (Table 3) capable of representing a wide range hydrologic conditions. Runoff production is controlled by the spatial variability of soil capacities across the catchment via a flexible probability distributed soil storage capacity, and water is routed through two storages representing surface (quick) and subsurface (slow) response mechanisms to produce total streamflow.

Smith et al. (2014) sought to understand the value of donor catchments as a function of their similarity to the catchment of interest. As a result, they considered three donor catchment classifications – random, proximity, and cluster – that represented varying levels of classification information (Fig. 6). Random donors were simply selected at random from the collection of catchments, proximity donors were selected based on distance from the target catchment, and cluster donors were selected based on similarity of calibrated, optimal PDM parameters (where each parameter was given equal weight in the clustering). In this scenario, we repeated the transferability analysis utilized in the previous two scenarios across each of the three donor classifications from the study of Smith et al. to assess the degree to which parameter similarity influences reciprocal transferability between catchments and provide comparison points to the catchments studied in Scenarios 1 and 2.

The optimal PDM parameters were estimated using the Dynamically Dimensioned Search (DDS) algorithm (Tolson and Shoemaker, 2007). The average Nash-Sutcliffe Efficiency for each of the classification levels were: 0.73 for cluster donors (max. = 0.88; min. = 0.59), 0.76 for proximity donors (max. = 0.86; min. = 0.60), and 0.71 for random donors (max. = 0.86; min. = 0.54). Note that each group consists of 10 unique member catchments and 1 common catchment, based on the conditions of Smith et al. (2014). As with the GR4J model performance from Scenario 2, the degree to which hydrological processes were accurately simulated with the PDM at these catchments is uncertain despite reasonable performance across classification groups.

The average reduction in model performance (as percent change in NSE) was computed for each classification group and each catchment as both a parameter donor and a parameter acceptor (Fig. 7). Catchment similarity (from the random to proximity class) directly impacts the magnitude of performance reduction (Fig. 7a and b; note individual subplot axes), a reflection of the role that catchment similarity plays in parameter transferability (in general) and reciprocal transferability (in particular). An additional improvement in transferability is observed

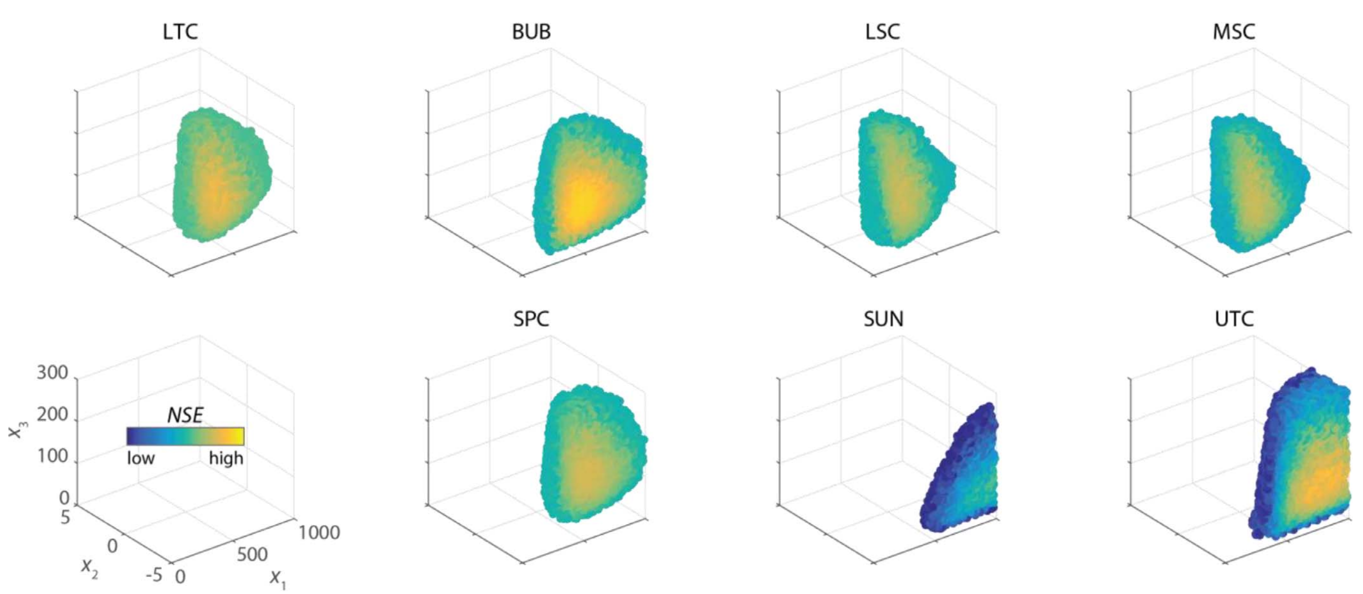

Fig. 5. Four-dimensional surface plots of the GR4J parameter space and model performance (as NSE) for each TCEF catchment. Note that all subplots utilize common axes and color scale. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Table 3
Names and description of PDM parameters[a,b].

Parameter	*Description*	*Limits*	*Units*
cmax	maximum soil storage capacity	0 400	mm
b	spatial variability within the catchment	0 0.4	--
kb	rate of drainage into subsurface storage	0 0.2	--
tr1	fraction of subsurface storage released to outflow	0 0.2	--
tr2	fraction of surface storage released to outflow	0 1	--
cf	soil storage threshold for release to subsurface storage	0 400	mm

[a]Each vertical line under "Limits" represent the optimal value from a Scenario 3 catchment.
[b]Green lines (top) represent randomly selected catchment, red lines (middle) represent proximity selected catchment, and blue lines (bottom) represent cluster selected catchment.

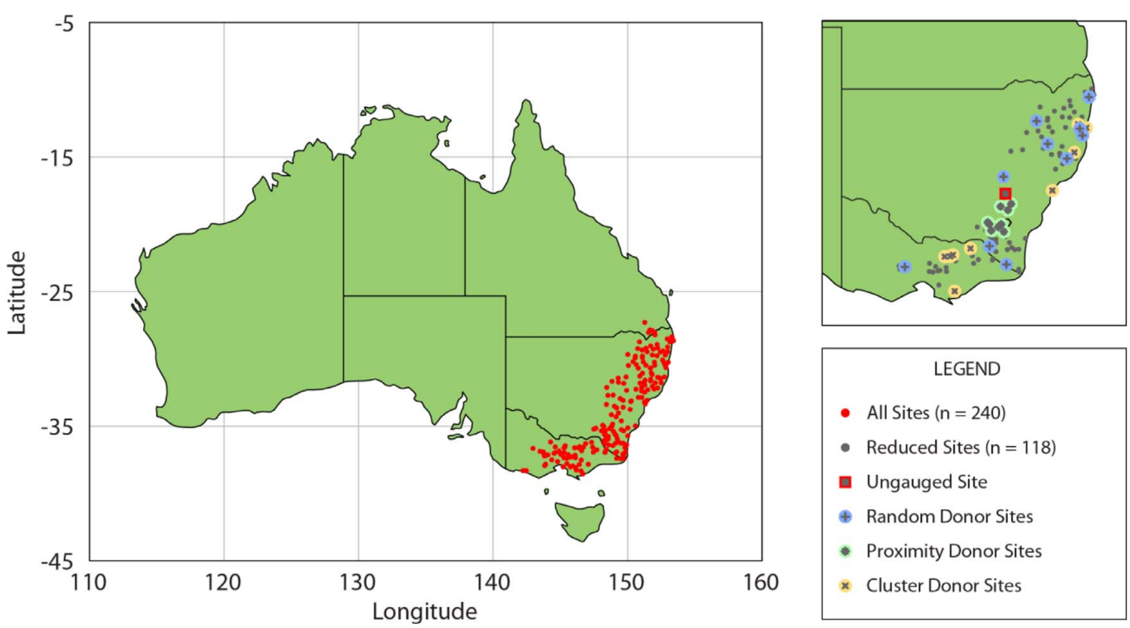

Fig. 6. Location of the catchments used in Smith et al. (2014). (left) Each of the 240 catchments in the data set is shown. (right) The reduced set of 118 catchments are shown (non-ephemeral with at least 10 years of continuous data), along with the sites used as donors for one of the ungauged target catchments based on random, spatial proximity, and hierarchical cluster analysis selection.
Reprinted from Smith et al. (2014).

between the proximity and cluster based catchments (Fig. 7b and c). However, improvements due to catchment similarity (physical / climatological characteristics; Fig. 7a and b) or model parameter similarity (Fig. 7b and c) do not provide a strict control on the reciprocity of such transfers. In each grouping, some catchments were more suitable as donors, while others were more suitable as acceptors. The degree to which that is true is a function of the sensitivity of the model performance to model parameter variability.

3. Discussion

Catchment classification provides a means within which the complexity of hydrologic systems can be organized to facilitate more fruitful applications and discovery. For this ideal to be met, however, the assumptions of the classification must be satisfied; i.e., the group members within the classification are functionally similar. Donor reciprocity is a simple concept inherent to most catchment classification approaches in the context of regionalization. It simply represents the idea that once classified as similar, the transfer of information between the catchments can occur in either direction with equivalent value. In this

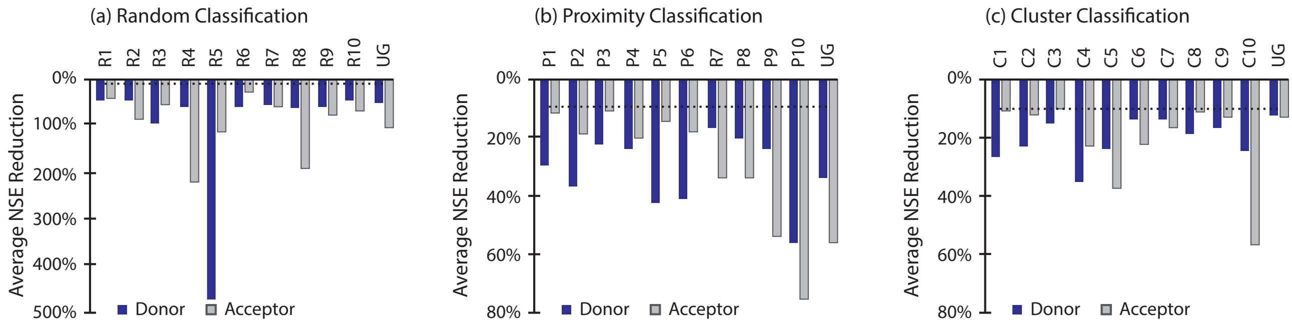

Fig. 7. Average reduction in model performance (as percentage of site-specific NSE) for each catchment within a classification group (a–c) when used as a donor or an acceptor with the probability distributed model. The dashed line represents a threshold reduction value of 10%.

study, we have demonstrated through three scenarios, across gradients of model realism and catchment attributes, that parameters are not transferred bi-directionally with equivalent results, and that the directionality can be attributed to model sensitivities and uncertainties. If the model and the classification were perfect, the transferability of model parameters should be bi-directional, however, model uncertainty (as defined by parameter response surfaces) has a confounding effect on the results. What do these results suggest for the usefulness of the catchment classification from a modeling perspective?

3.1. Do similar catchments act as universal donors / acceptors?

The catchments of the Tenderfoot Creek Experimental Forest receive the same climatological inputs and are similar in terms of most typical similarity metrics – each has similar soils, parent geology, and vegetation, as well as other physical characteristics (e.g., relief, elevation, average slope, area, distance from creek, gradient to creek, etc.). Given the apparent similarity of these catchments, we sought to illustrate the directionality of information transfer from a model parameterization perspective. Extending the study of Smith et al. (2016) on parameter transferability of the process-verifiable Catchment Connectivity Model applied to the Tenderfoot Creek Experimental Forest, we plotted 4D response surfaces (Fig. 3) that clearly demonstrate the variability in the optimal parameter space across the "similar", neighboring catchments. This analysis was repeated for the conceptual GR4J model (Fig. 5) to better understand the results with respect to differences in model structures. Similar results were observed for both models tested, strengthening the universality of the results. From a classification/regionalization perspective (using a simple direct transfer of the "optimal" hydrologic model parameter set), this parameter sensitivity/variability will play a determining role in model performance at the target (accepting) catchment. At Tenderfoot Creek Experimental Forest, a transfer of the optimal parameter set from SUN to LSC will result in marginal performance degradation, while a transfer from LSC to SUN will result in significant performance degradation (Figs. 2 and 5). This directionality occurs due to the "optimal" parameter spaces not being coincident.

3.2. Is there a relationship between physical and functional (parameter) similarity?

The sensitivity of model parameters is directly linked to model structure (and its errors and uncertainties). Unquestionably, the use of conceptual rainfall-runoff models presents a challenge with regard to regionalization approaches (via catchment classification based on physical characteristics) due to a lack of (verifiable) process consistency, though this issue may persist despite model fidelity (see Scenario 1). The PUB decadal initiative bore this out, with no classification approach consistently outperforming another (across diverse catchment datasets) for regionalization (Hrachowitz et al., 2013). This should necessarily reflect *both* the model structure and the assumptions of the regionalization approach as sources of great uncertainty.

Smith et al. (2014) present an alternative approach to classification, where the typical regionalization problem was inverted – rather than classifying catchments by physical characteristics they were instead classified by optimal parameter values. Such an approach, where the analysis begins with the desired answer (i.e., catchments with similar hydrologic performance/fit) and works backwards to the question (i.e., does catchment classification by physical, climatological, or hydrological attributes work), can be particularly instructive in better understanding the assumptions and limitations of the underlying methodology (i.e., model regionalization). The results suggest that catchment similarity can be predicted from a modeling perspective, but that significant model error can result in sub-optimal performance.

Exploring this concept further, we performed an additional analysis on the model simulations performed by Smith et al. (2014), who utilized a catchment classification approach that compared classes consisting of random members, spatially proximal members, and functionally proximal members (using cluster analysis to group catchments with similar optimal parameterizations). Contrasting the conditions examined in Scenarios 1 and 2, the catchments utilized here spanned a wide range of physical and climatological conditions (particularly in the random and cluster selected catchments). The lack of spatial proximity of the cluster selected catchments (Fig. 6), in and of itself, suggests that the link between model parameterization and catchment similarity is weak. Results from this analysis again demonstrated the directionality inherent to hydrologic model transferability (Fig. 7), regardless of the similarity of the catchments (or their optimal parameterization). This indicates the weakness of single parameter set transfers in a regionalization context, and suggests that including parameter variability / uncertainty is a necessary component to achieve successful, reliable transfer.

3.3. How do model complexity and realism affect the transferability of model parameter sets?

Three conceptual hydrologic models (CCM, GR4J, and PDM) were considered in this study. Although each had similar numbers of model parameters (3, 4, and 6, respectively), the models make unique assumptions that impact their complexity and realism. Holding the catchments constant, Scenarios 1 and 2 explored the impact of changing the model structure from the process verifiable CCM (e.g., Smith et al., 2013, 2016) to the black box implementation of the unit hydrograph-based GR4J. Despite the differences in verifiable process fidelity, both models resulted in sensitivities to parameters controlling the soil moisture accounting of the watersheds. For the CCM, the affected parameter (q^*) can be tied directly to the catchment structure; the distribution of hillslope sizes (i.e., the elements of the watershed upon which the model carries out computations). In effect, this allows for *a priori* identification of parameter sensitivity and, therefore, parameter transferability, and this is the primary advantage process consistent models hold over more black box approaches, where this information can only be obtained *a posteriori* based on the model parameterization (as was found with Scenario 2).

3.4. Are any relationships identified consistent across diverse sets of catchments?

When classifying catchments as similar using physical (area, topography, shape, etc.), climatological, or hydrological (signatures) metrics, there is an implicit assumption made. Namely, that the relationship is reciprocal; i.e., if Catchment A and Catchment B share the determining metric(s), Catchment A can be used as a donor to Catchment B *and* Catchment B can be used as a donor to Catchment A. Our findings indicate that this assumption is largely inappropriate. We found that across multiple gradients (hydrologic model complexity/fidelity, dominant runoff processes, climate, size, degree of physical similarity, degree of optimal parameter similarity; Figs. 2, 4, and 7) most catchments were ill-suited as both donor and acceptor. Rather, a subset of catchments was suitable only as donors, while another subset was suitable only as acceptors. Ultimately, parameter sensitivity controls the success and directionality of direct parameter transferability. Understanding the link between physical / climatological characteristics and optimal hydrologic parameters remains a challenge. This study better defines the domain of the problem, and how such characteristics propagate into parameter sensitivity.

4. Conclusion

Catchment classification has long been viewed as a Rosetta stone for hydrologic modeling, representing the promise of unlocking the ability to transfer model information across space and time. A key component in this thinking is the assumption that classification based on catchment properties yields similar model properties (e.g., parameter values). And while this is perhaps logical on the surface, it necessarily ignores the uncertainties of the model (and data), as well as the potential for non-reciprocal relationships to exist. The results of this study highlight the potential pitfalls of classifying physical characteristics to inform the regionalization/transfer of hydrologic model information. The three test cases, utilizing diverse data sets and hydrologic model selection approaches (catchment-specific v. flexible), provided examples across hydrologic conditions, scale, and detail. We found that (1) belonging to a group of physically similar (classified) catchments is not sufficient to ensure the transfer of model information (e.g., parameterization) is reciprocal (i.e., non-directional); (2) there is not a clear relationship between physical and functional (i.e., parameter) similarity due to model/data uncertainty; (3) the fidelity of the model structure to observed hydrologic functioning does little to affect the ability of catchments to serve as donors/acceptors; and (4) the (lack of) relationship between catchment similarity and model similarity is not data- or site-specific. Given these observations, it is clear that while catchment classification has the potential to improve water resources management, significant work remains in understanding the relationships between catchment form and hydrologic model functioning.

Acknowledgments

Data for Tenderfoot Creek Experimental Forest can be obtained, at request, from the U.S. Forest Service (http://www.fs.fed.us/rm/tenderfoot-creek/data/) for streamflow, from the National Resources Conservation Service (http://www.wcc.nrcs.usda.gov/snow/snotel-data.html) for meteorological variables, and from the National Center for Airborne Laser Mapping (http://dx.doi.org/10.5069/G92F7KCN) for topography. Data for the Australian catchments can be obtained, at request, from the Australian Bureau of Meteorology (http://www.bom.gov.au/water/hrs).

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.advwatres.2017.12.020.

References

Bárdossy, A., 2007. Calibration of hydrological model parameters for ungauged catchments. Hydrol. Earth Syst. Sci. 11 (2), 703–712. http://dx.doi.org/10.5194/hess-11-703-2007.

Carrillo, G., Troch, P.A., Sivapalan, M., Wagener, T., Harman, C., Sawicz, K., 2011. Catchment classification: hydrological analysis of catchment behavior through process-based modeling along a climate gradient. Hydrol. Earth Syst. Sci. 15 (11), 3411–3430. http://dx.doi.org/10.5194/hess-15-3411-2011.

Emanuel, R.E., Epstein, H.E., McGlynn, B.L., Welsch, D.L., Muth, D.J., D'Odorico, P., 2010. Spatial and temporal controls on watershed ecohydrology in the northern Rocky Mountains. Water Resour. Res. 46 (11), W11553. http://dx.doi.org/10.1029/2009WR008890.

Hrachowitz, M., et al., 2013. A decade of predictions in ungauged basins (PUB)—a review. Hydrol. Sci. J. 58 (6), 1198–1255. http://dx.doi.org/10.1080/02626667.2013.803183.

Hundecha, Y., Ouarda, T.B.M.J., Bárdossy, A., 2008. Regional estimation of parameters of a rainfall-runoff model at ungauged watersheds using the “spatial” structures of the parameters within a canonical physiographic-climatic space. Water Resour. Res. 44, W01427. http://dx.doi.org/10.1029/2006WR005439.

Jencso, K.G., McGlynn, B.L., Gooseff, M.N., Wondzell, S.M., Bencala, K.E., Marshall, L.A., 2009. Hydrologic connectivity between landscapes and streams: transferring reach- and plot-scale understanding to the catchment scale. Water Resour. Res. 45, W04428. http://dx.doi.org/10.1029/2008WR007225.

Jencso, K.G., McGlynn, B.L., Gooseff, M.N., Bencala, K.E., Wondzell, S.M., 2010. Hillslope hydrologic connectivity controls riparian groundwater turnover: Implications of catchment structure for riparian buffering and stream water sources. Water Resour. Res. 46 (10), W10524. http://dx.doi.org/10.1029/2009WR008818.

Jencso, K.G., McGlynn, B.L., 2011. Hierarchical controls on runoff generation: Topographically driven hydrologic connectivity, geology, and vegetation. Water Resour. Res. 47 (11), W11527. http://dx.doi.org/10.1029/2011wr010666.

Kuczera, G., Parent, E., 1998. Monte Carlo assessment of parameter uncertainty in conceptual catchment models: the Metropolis algorithm. J. Hydrol. 211 (1–4), 69–85. http://dx.doi.org/10.1016/S0022-1694(98)00198-X.

Leibowitz, S.G., Comeleo, R.L., Wigington, P.J., Weber, M.H., Sproles, E.A., Sawicz, K.A., 2016. Hydrologic landscape characterization for the pacific Northwest, USA. JAWRA J. Am. Water Resour. Assoc. 52 (2), 473–493. http://dx.doi.org/10.1111/1752-1688.12402.

McDonnell, J.J., Woods, R., 2004. On the need for catchment classification. J. Hydrol. 299 (1–2), 2–3. http://dx.doi.org/10.1016/j.jhydrol.2004.09.003.

McIntyre, N., Lee, H., Wheater, H., Young, A., Wagener, T., 2005. Ensemble predictions of runoff in ungauged catchments. Water Resour. Res. 41, W12434. http://dx.doi.org/10.1029/2005WR004289.

Moore, R.J., 1985. The probability-distributed principle and runoff production at point and basin scales. Hydrol. Sci. J. 30 (2), 273–297. http://dx.doi.org/10.1080/02626668509490989.

Nippgen, F., McGlynn, B.L., Marshall, L.A., Emanuel, R.E., 2011. Landscape structure and climate influences on hydrologic response. Water Resour. Res. 47 (12), W12528. http://dx.doi.org/10.1029/2011wr011161.

Oudin, L., Andréassian, V., Perrin, C., Michel, C., Le Moine, N., 2008. Spatial proximity, physical similarity, regression and ungaged catchments: a comparison of regionalization approaches based on 913 French catchments. Water Resour. Res. 44, W03413. http://dx.doi.org/10.1029/2007WR006240.

Payn, R.A., Gooseff, M.N., McGlynn, B.L., Bencala, K.E., Wondzell, S.M., 2009. Channel water balance and exchange with subsurface flow along a mountain headwater stream in Montana, United States. Water Resour. Res. 45, W11427. http://dx.doi.org/10.1029/2008WR007644.

Payn, R.A., Gooseff, M.N., McGlynn, B.L., Bencala, K.E., Wondzell, S.M., 2009. Exploring changes in the spatial distribution of stream baseflow generation during a seasonal recession. Water Resour. Res. 48 (4), W04519. http://dx.doi.org/10.1029/2011wr011552.

Perrin, C., Michel, C., Andréassian, V., 2003. Improvement of a parsimonious model for streamflow simulation. J. Hydrol. 279 (1), 275–289. http://dx.doi.org/10.1016/S0022-1694(03)00225-7.

Di Prinzio, M., Castellarin, A., Toth, E., 2011. Data-driven catchment classification: application to the pub problem. Hydrol. Earth Syst. Sci. 15 (6), 1921–1935. http://dx.doi.org/10.5194/hess-15-1921-2011.

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. http://dx.doi.org/10.5194/hess-15-2895-2011.

Sivapalan, M., et al., 2003. IAHS decade on predictions in ungauged basins (PUB), 2003–2012: shaping an exciting future for the hydrological sciences. Hydrol. Sci. J. 48 (6), 857–880. http://dx.doi.org/10.1623/hysj.48.6.857.51421.

Smith, T.J., Marshall, L.A., 2008. Bayesian methods in hydrology: A study of recent advancements in Markov chain Monte Carlo techniques. Water Resour. Res. 44 (12), W00B05. http://dx.doi.org/10.1029/2007WR006705.

Smith, T.J., Marshall, L.A., 2010. Exploring uncertainty and model predictive performance concepts via a modular snowmelt-runoff modeling framework. Environ.

Modell. Software 25 (6), 691–701. http://dx.doi.org/10.1016/j.envsoft.2009.11.010.
Smith, T., Marshall, L., McGlynn, B., Jencso, K., 2013. Using field data to inform and evaluate a new model of catchment hydrologic connectivity. Water Resour. Res. 49 (10), 6834–6846. http://dx.doi.org/10.1002/wrcr.20546.
Smith, T., Marshall, L., Sharma, A., 2014. Predicting hydrologic response through a hierarchical catchment knowledgebase: a Bayes empirical Bayes approach. Water Resour. Res. 50 (2), 1189–1204. http://dx.doi.org/10.1002/2013WR015079.
Smith, T., Hayes, K., Marshall, L., McGlynn, B., Jencso, K., 2016. Diagnostic calibration and cross-catchment transferability of a simple process-consistent hydrologic model. Hydrol. Process. 30 (26), 5027–5038. http://dx.doi.org/10.1002/hyp.10955.
Tolson, B.A., Shoemaker, C.A., 2007. Dynamically dimensioned search algorithm for computationally efficient watershed model calibration. Water Resour. Res. 43, W01413. http://dx.doi.org/10.1029/2005WR004723.
Wagener, T., Sivapalan, M., Troch, P., Woods, R., 2007. Catchment classification and hydrologic similarity. Geogr. Compass 1 (4), 901–931. http://dx.doi.org/10.1111/j.1749-8198.2007.00039.x.
Winter, T.C., 2001. The concept of hydrologic landscapes. J. Am. Water Resour. Assoc. 37 (2), 335–349. http://dx.doi.org/10.1111/j.1752-1688.2001.tb00973.x.
Ye, S., Yaeger, M., Coopersmith, E., Cheng, L., Sivapalan, M., 2012. Exploring the physical controls of regional patterns of flow duration curves – part 2: role of seasonality, the regime curve, and associated process controls. Hydrol. Earth Syst. Sci 16, 4447–4465. http://dx.doi.org/10.5194/hess-16-4447-2012.