AGU PUBLICATIONS

Water Resources Research

RESEARCH ARTICLE

10.1002/2016WR019736

Key Points:
- Functional approach to watersheds' flow partitioning is used to assess models
- Proposed methodology improves hydrologic partitioning performed by models
- Partitioning-based model calibration improves parameter identifiability

Correspondence to:
M. Shafii,
mshafiih@uwaterloo.ca

Citation:
Shafii, M., N. Basu, J. R. Craig, S. L. Schiff, and P. Van Cappellen (2017), A diagnostic approach to constraining flow partitioning in hydrologic models using a multiobjective optimization framework, *Water Resour. Res.*, *53*, 3279–3301, doi:10.1002/2016WR019736.

Received 31 AUG 2016
Accepted 21 MAR 2017
Accepted article online 28 MAR 2017
Published online 21 APR 2017

A diagnostic approach to constraining flow partitioning in hydrologic models using a multiobjective optimization framework

Mahyar Shafii[1,2], Nandita Basu[1,2,3], James R. Craig[2,3], Sherry L. Schiff[1,2], and Philippe Van Cappellen[1,2]

[1]Department of Earth and Environmental Sciences, University of Waterloo, Waterloo, Ontario, Canada, [2]Water Institute, University of Waterloo, Waterloo, Ontario, Canada, [3]Department of Civil and Environmental Engineering, University of Waterloo, Waterloo, Ontario, Canada

Abstract Hydrologic models are often tasked with replicating historical hydrographs but may do so without accurately reproducing the internal hydrological functioning of the watershed, including the flow partitioning, which is critical for predicting solute movement through the catchment. Here we propose a novel partitioning-focused calibration technique that utilizes flow-partitioning coefficients developed based on the pioneering work of L'vovich (1979). Our hypothesis is that inclusion of the L'vovich partitioning relations in calibration increases model consistency and parameter identifiability and leads to superior model performance with respect to flow partitioning than using traditional hydrological signatures (e.g., flow duration curve indices) alone. The L'vovich approach partitions the annual precipitation into four components (quick flow, soil wetting, slow flow, and evapotranspiration) and has been shown to work across a range of climatic and landscape settings. A new diagnostic multicriteria model calibration methodology is proposed that first quantifies four calibration measures for watershed functions based on the L'vovich theory, and then utilizes them as calibration criteria. The proposed approach is compared with a traditional hydrologic signature-based calibration for two conceptual bucket models. Results reveal that the proposed approach not only improves flow partitioning in the model compared to signature-based calibration but is also capable of diagnosing flow-partitioning inaccuracy and suggesting relevant model improvements. Furthermore, the proposed partitioning-based calibration approach is shown to increase parameter identifiability. This model calibration approach can be readily applied to other models.

Plain Language Summary Hydrologic models are often tasked with replicating historical hydrographs but may do so without accurately reproducing the internal hydrological functioning of the watershed, including the flow partitioning between low and high flows, which is critical for predicting solute movement through the catchment. Here we propose a novel model calibration framework that utilizes an empirical understanding about flow partitioning developed by L'vovich (1979) to constrain the outcomes of watershed models. Our hypothesis is that this approach increases model consistency leads to superior model performance. This method is also capable of diagnosing model structural errors (in flow partitioning) and suggesting relevant model improvements. Overall, this work is a step toward getting the right answer from hydrologic model for the right reasons.

1. Introduction

1.1. Calibration of Hydrologic Models

The exchange of water and energy between land and atmosphere and the movement of water across the land surface are the fundamentals of land-surface hydrology and have led to the development of several hydrologic models. Traditionally, the calibration of hydrologic models is conducted by adjusting parameter values to minimize the discrepancy between the time series of observed and simulated streamflow. The discrepancy is quantified using a number of goodness-of-fit metrics [*Gupta et al.*, 2009; *Krause et al.*, 2005; *Legates and McCabe*, 1999], such as sum of squared errors (SSE) and Nash-Sutcliffe Efficiency (NSE), and minimization is achieved using various parameter estimation methods, e.g., stochastic procedures [*Clarke*, 1973;

Johnston and Pilgrim, 1976; *Kuczera*, 1983; *Sorooshian and Dracup*, 1980] and informal optimization-based techniques [e.g., *Duan et al*., 1992; *Tolson and Shoemaker*, 2007].

It has, however, been argued that measures such as NSE or SSE do not provide the modeler with enough ability to discriminate between different model structural forms, since different representations of storage and discharge in a model can lead to the same aggregate statistics [e.g., *Gupta et al*., 2009; *Martinez and Gupta*, 2010; *McCuen et al*., 2006; *Schaefli and Gupta*, 2007]. Recognizing this deficiency, diagnostic model evaluations have been advocated [*Gupta et al*., 2008] based on hydrological signatures mostly derived based on streamflow time series, e.g., the flow duration curve (FDC) [e.g., *Westerberg et al*., 2011; *Yilmaz et al*., 2008; *Yokoo and Sivapalan*, 2011], the spectral density of runoff [e.g., *Montanari and Toth*, 2007; *Vogel and Sankarasubramanian*, 2003; *Winsemius et al*., 2009], the rising limb density [*Shamir et al*., 2005b; *Yadav et al*., 2007], the base flow index [e.g., *Arnold and Allen*, 1999; *Sawicz et al*., 2011], or from other sources of data such as soil moisture [*McMillan et al*., 2014, 2011].

Hydrological signatures serve as model evaluation diagnostics for different purposes such as models comparison [*Coxon et al*., 2014; *Euser et al*., 2013; *Hrachowitz et al*., 2014; *Martinez and Gupta*, 2011; *Viglione et al*., 2013; *Vogel and Sankarasubramanian*, 2003], top-down model development [*Eder et al*., 2003; *Farmer et al*., 2003; *Jothityangkoon et al*., 2001; *Littlewood et al*., 2003], and constraining model outcomes against different aspects of the hydrograph [*Bulygina and Gupta*, 2009; *Pokhrel et al*., 2012; *Shafii and Tolson*, 2015; *Shamir et al*., 2005a; *Westerberg et al*., 2011; *Winsemius et al*., 2009; *Yilmaz et al*., 2008]. Among other strategies for improving the calibration of hydrologic models include stepwise calibration based on hydrograph partitioning [*He et al*., 2015], incorporating expert knowledge and process/parameter constraints in model development process [*Gharari et al*., 2014], etc. Even though implementing signature-based calibration approaches results in the selection of models that are more consistent with watershed's hydrological behavior and the reduction of predictive uncertainty [e.g., *Hrachowitz et al*., 2014; *Pokhrel et al*., 2012; *Yilmaz et al*., 2008], it fails to explicitly address the partitioning of water among the different flow paths, which become critically important when modeling solute transport.

1.2. Flow Partitioning and the Use of Tracers

Solute concentrations are a function of both the flow magnitude and the flow pathway, and thus, it has been advocated that tracer and isotope data should be routinely used for a proper identification of flow partitioning [e.g., *Didszun and Uhlenbrook*, 2008; *Klaus and McDonnell*, 2013; *Tetzlaff and Soulsby*, 2008]. Among other approaches to improving flow partitioning are using "soft" data related to partitioning between event and preevent water [*Vaché and McDonnell*, 2006; *Vaché et al*., 2004] or groundwater table fluctuations [*Seibert and McDonnell*, 2002], fraction of annual groundwater nitrate load [*Yen et al*., 2014], and estimates of monthly nitrogen oxide [*Ahmadi et al*., 2014]. Such approaches can take us one step closer to reducing equifinality, and getting the right answers for the right reasons [*Davies et al*., 2011, 2013; *Hrachowitz et al*., 2013; *Kirchner*, 2006; *McDonnell and Beven*, 2014].

Although the concurrent use of tracers and discharge for routine development of watershed models is desirable, such measurements are not available at most places. Thus, while we explicitly recognize that tracers (if available) have to be used for flow path identification, we pose the question: Is there greater information content in the rainfall and runoff time series that is not currently captured in traditional aggregate measures of goodness of fit and/or more recent hydrologic signatures, and that might be used to further constrain model partitioning skill and parameter identifiability? We propose the use of the empirical L'vovich framework for flow partitioning as a way to constrain hydrologic models. *L'vovich* [1979], as explained in details in section 2.2, explicitly considered the partitioning of precipitation into major catchment functions (e.g., infiltration, runoff, subsurface drainage, and evapotranspiration), and applied his approach to a large number of catchments located in a variety of ecoregions of the world. Mathematical models derived based on the L'vovich theory also proved to be robust for the characterization of water balance in a large number of catchments across the world [*Ponce and Shetty*, 1995a, 1995b; *Sivapalan et al*., 2011].

Given the simplicity and hydrological meaningfulness involved in the L'vovich approach, it has also been employed in multiple recent research studies for different purposes including the assessment of runoff elasticity to precipitation [*Harman et al*., 2011] and getting ecohydrological understanding by analyzing controls on the water balance of catchment ecosystems across the United States [*Thompson et al*., 2011]. More recently, *Wang et al*. [2015] defended the L'vovich flow-partitioning scheme through the use of an

optimality principle (Maximum Entropy Production). Overall, there is sufficient evidence in the literature that the L'vovich approach adequately characterizes watersheds flow partitioning at the annual time scale, and the forms of the L'vovich equations differ between catchments. However, at the finer time scales, proper simulation of watershed temporal dynamics requires the application of detailed physically based models. The question now is how we can constrain the responses of these models to get the flow partitioning correct. We contend that the L'vovich partitioning provides a unique and accessible representation of catchment internal functioning, and thus is a useful constraint upon flow partitioning in a model, particularly in cases where tracer or another ancillary data are not available.

1.3. Objectives of the Paper

The aim of this paper is to develop a novel partitioning-based multicriteria calibration technique that utilizes both L'vovich partitioning relations and hydrograph goodness-of-fit measures as calibration objectives to better constrain the flow partitioning in hydrologic models. Our hypothesis is that inclusion of the L'vovich partitioning increases model consistency and parameter identifiability and leads to superior model performance with respect to flow-partitioning than using hydrologic signatures only. To test this hypothesis, we first use the L'vovich approach to assess the hydrologic partitioning skill of models that were constrained using conventional signatures. Second, we improve the partitioning skill of models by including L'vovich partitioning in the calibration process. Third, we investigate the possibility of improving parameter identifiability by incorporating the partitioning-focused signatures in the model calibration process. Our proposed methodology is considered as a diagnostic approach to model selection and parameter identification that has been advocated in the literature [*Gupta et al*., 2008; *Hrachowitz et al*., 2014; *Kirchner*, 2006; *Martinez and Gupta*, 2011], and its novelty is in the simultaneous use of rainfall and streamflow time series in defining partitioning-focused calibration objectives.

The paper is organized as follows. Section 2 provides a background on functional approaches to water budget analysis (section 2.1), especially the L'vovich framework (section 2.2), and elaborates on our proposed methodology and the numerical experiments in which we compare our proposed calibration methodology with the traditional signature-based model calibration approach (section 2.3). Case studies and results are explained in sections 3 and 4, respectively. Finally, we provide the concluding remarks from our study in section 5.

2. Methodology

2.1. Functional Approaches to Hydrologic Partitioning

Functional approaches to hydrologic analysis are methods that focus on how the catchment *functions*, for example, the partitioning of annual precipitation into soil moisture storage, evapotranspiration [*Wagener et al*., 2007]. Even though process-based models have shown to be useful for modeling catchment functions mediated by climatic properties [e.g., *Eagleson*, 1978; *Milly*, 1994; *Reggiani et al*., 2000; *Salvucci and Entekhabi*, 1995], we focus here on empirical functional approaches that use only hydrologic data [e.g., *Budyko*, 1974; *Horton*, 1933; *L'vovich*, 1979; *Troch et al*., 2009]. The most famous example of this is likely the study by Budyko [*Budyko*, 1974] in which the conceptual limits for the ratio of actual evaporation to precipitation are identified based on the ratio of potential evapotranspiration to precipitation. Whereas the Budyko framework and its more recent developments [e.g., *Yang et al*., 2007; *Zhang et al*., 2001] describe only the long-term annual water balance, the temporal variability of the water balance components has been described by the Horton and the L'vovich models [*Horton*, 1933; *L'vovich*, 1979; *Ponce and Shetty*, 1995a, 1995b; *Troch et al*., 2009].

2.2. Hydrologic Partitioning: L'vovich and Ponce and Shetty Models

L'vovich [1979] developed an empirical theory of annual flow partitioning based on a simple two-stage precipitation (P) partitioning approach (see Figure 1). At the first stage, P is partitioned into a quick flow component, S, and an infiltrated component called the catchment wetting, W. At the second stage, W is further partitioned into a slow-flow component, U, and an energy-dependent vaporization component, V, which is composed of evaporation and transpiration. The terms S, W, U, and V are henceforth called L'vovich components. This hydrologic partitioning, neglecting the carryover of storage between consecutive years (an underlying assumption in the L'vovich approach), is mathematically represented using four simple equations,

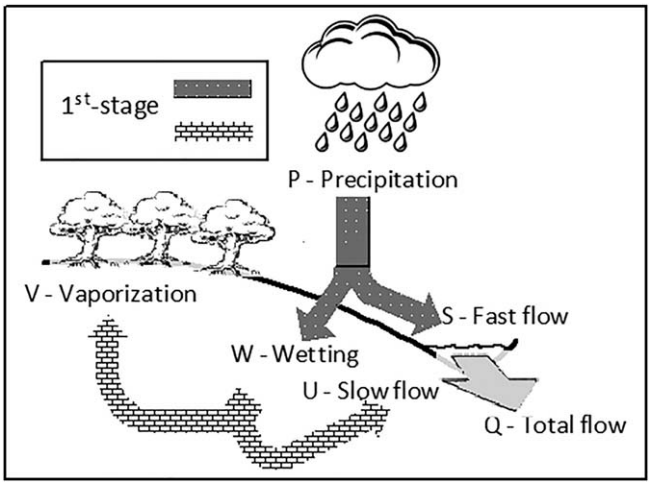

Figure 1. Schematics of the two-stage flow partitioning in the L'vovich approach.

$P=S+W$, $W=U+V$, $P=V+Q$, and $Q=S+U$. This formulation requires the availability of continuous precipitation and streamflow data.

L'vovich [1979] derived the empirical relationships for S and W versus P, and U and V versus W in the form of nomographs and tables for a large number of watersheds in the world [see *Sivapalan et al.*, 2011, Figure 2, for example]. *Ponce and Shetty* [1995a, 1995b] developed simple mathematical models based on the first-stage and second-stage partitioning in the L'vovich theory [see *Ponce and Shetty*, 1995a, 1995b, for details], which were fully defined by four parameters. Recently, *Sivapalan et al.* [2011] applied the Ponce-Shetty model to 431 catchments in the United States and elaborated on how the four parameter of this model should be estimated (using nonlinear regression). Subsequently, the model estimates annual values of L'vovich components (such as wetting and quick flow) based on observed precipitation and streamflow data. Note that Ponce-Shetty mathematical relationships assume that there is no carryover of the soil storage between years. The violation of this assumption might introduce uncertainties in these relationships, e.g., considerable scatter in the relationships as shown in *Sivapalan et al.* [2011].

The L'vovich approach to hydrologic partitioning relies on base flow separation from the observed streamflow time series. Multiple base flow separation techniques exist in the literature [see for example, *Arnold and Allen*, 1999] and the choice of separation approach may change the estimated hydrologic partitioning. However, *Troch et al.* [2009] previously demonstrated that the estimation of annual water balance metrics was not highly sensitive to the method of base flow separation. As a result, in this study, we followed *Sivapalan et al.* [2011] and applied a simple one-parameter low-pass filter developed by *Lyne and Hollick* [1979] to partition the streamflow into slow flow and quick flow; the value of 0.925 was imposed for the parameter of the base flow separation equation.

2.3. Model Calibration Framework Based on Flow Partitioning

We hypothesize that the empirical relationships developed in *L'vovich* [1979] and *Ponce and Shetty* [1995a, 1995b] can be used to constrain the responses of any hydrologic model if there is a one-to-one correspondence between L'vovich components and two or more model components. For instance, consider a simple saturation-excess one-bucket model (Figure 2) for streamflow simulation in which precipitation (P) enters the bucket as infiltration (INF) and outflows via base flow (B) and runoff (Q); the bucket is also exposed to evaporation (E). Estimating the parameters of this bucket model based only on the fit between simulated and observed streamflow would not guarantee consistent hydrologic partitioning (the same streamflow value may be generated by different combinations of B and Q). However, matching different components of the model to the corresponding L'vovich components (W,S,U, and V that are also shown in Figure 2) will force the model to capture the flow partitioning, in addition to fitting the output hydrograph.

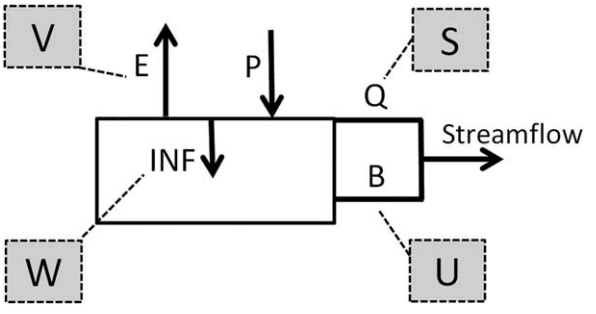

Figure 2. Schematic representation of single-bucket saturation-excess model with different processes including precipitation (P), evapotranspiration (E), infiltration (INF), surface runoff (Q), and base flow (B). Also shown are the corresponding L'vovich components W, S, U, and V.

To properly reproduce flow partitioning, simple hydrological signatures are traditionally

used to inform calibration, e.g., use of a base flow index to constrain the slow-flow component. However, streamflow at any point in time is composed of water going through different pathways, and standard signatures such as quick and slow flow volume of water do not necessarily manifest hydrologic partitioning and the relation between different processes. We propose that, to evaluate the correctness of the flow partitioning conducted by signature-based constrained models, one can utilize Ponce-Shetty model as the benchmark and generate acceptability intervals after fitting this model to the data (section 2.3.1). We also demonstrate that one can use a set of new partitioning-based calibration diagnostics formulated from the Ponce-Shetty model (described in section 2.3.1) and employ them to constrain flow partitioning according to the L'vovich theory (section 2.3.2).

2.3.1. Developing Scoring Functions Based on L'vovich Components

The identification of the Ponce-Shetty model is a curve fitting process; this model is not a perfect representation of reality and is prone to uncertainties. The confidence and prediction intervals associated with the curve fitting process characterize the uncertainty in the Ponce-Shetty model. To account for this uncertainty, we used the prediction intervals as acceptability bounds [*Beven*, 2006; *Liu et al*., 2009], instead of solely using the mean Ponce-Shetty values. In other words, we ensure that the modeled quick flow and infiltration fall within the prediction intervals of S and W, and the modeled slow flow and vaporization fall within the prediction intervals of U and V, respectively. This is illustrated by an example in Figure 3 where the blue circles represent the annual $P-S$ calculated using base flow separation of 40 years of observed streamflow data. Following *Sivapalan et al*. [2011], nonlinear regression is used to fit the Ponce-Shetty model to this annual data (solid line in Figure 3) by estimating the underlying parameters minimizing the deviation between observed S and the S generated by the Ponce-Shetty model. In such a standard nonlinear regression experiment, assuming an error series with zero mean and an unknown constant variance, the prediction intervals (for a given future value) can be extracted as well (dashed lines in Figure 3). We use MATLAB's statistical package to find the model estimate and the prediction intervals. Note that other uncertainty estimation method such as Bayesian inference, similar to *Harman et al*. [2011] can also be used to quantify the prediction intervals, which may result in slightly different intervals. Nonetheless, our proposed calibration approach is independent from the uncertainty estimation methodology and can be carried out as long as the acceptability intervals are built. Upon the identification of the model and its prediction intervals, they can be extrapolated toward both the origin (i.e., very small rainfall) and extreme precipitation events that have not been observed in the historical records.

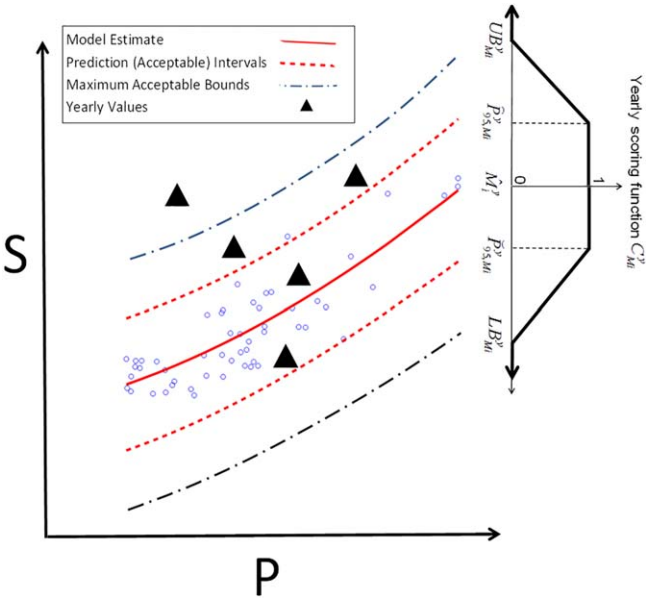

Figure 3. Schematic representation of precipitation-surface runoff partitioning based on L'vovich theory (circles representing annual values), the Ponce-Shetty model estimate (solid line), the prediction intervals associated with model estimate (dashed lines), maximum bounds for predictions intervals (dash-dotted lines; see equation (1)), and yearly values generated by a hypothetical hydrologic model (triangles). The vertical diagram in the right-hand side shows the scoring function for L'vovich component S.

Next, a given hydrologic model is used and the component S is derived. The black triangles in Figure 3 are the predictions made by the model during a 5 year simulation period. Note that the simulation period (5 years in this case) may be shorter than the entire observation period (40 years). Now we need a quantitative metric or scoring function to evaluate the partitioning consistency of the given hydrologic model. The proposed scoring function (described below) quantifies the deviation between the model results (black triangles) and the prediction intervals (dashed lines) in terms of a number between 0 and 1. Note that these scoring functions will be incorporated as the calibration criteria into a multicriteria optimization algorithm. This is similar to the study by *Shafii and Tolson* [2015] who

developed a signature-based multicriteria calibration methodology where any signatures that characterize some part of the hydrological behavior can be implemented as calibration criteria.

Given a parameter set $\boldsymbol{\theta}$ of a hydrologic model, the scoring functions for the L'vovich component M_i, $i \in \{S, W, U, V\}$, are calculated through five steps:

i. Calculate M_i^y for each year y in the calibration period, $y \in \{1, 2, \ldots, Y\}$ where Y is the total number of years in the period (e.g., M_S^y is the quick flow component for year y).
ii. Find the estimated value of that component by the Ponce-Shetty model $\left(\hat{M}_i^y\right)$ and the corresponding upper $\left(\hat{P}^{\,y}_{95,Mi}\right)$ and lower $\left(\breve{P}^{\,y}_{95,Mi}\right)$ 95% prediction intervals as the acceptability bounds (e.g., for component S, the x coordinate of a given year is the annual precipitation in that year, used to find the corresponding prediction intervals from the Ponce-Shetty model).
iii. Consider maximum limits for $\hat{P}^{\,y}_{95,Mi}$ and $\breve{P}^{\,y}_{95,Mi}$ using the following equations:

$$LB_{Mi}^y = \hat{M}_i^y - \alpha \times \breve{P}^{\,y}_{95,Mi} \tag{1a}$$

$$UB_{Mi}^y = \hat{M}_i^y + \alpha \times \hat{P}^{\,y}_{95,Mi} \tag{1b}$$

where $\alpha > 1$; we use a value of 2, which corresponds to approximately 4 times the standard deviation away (on both sides) from the model estimate.

iv. Calculate the scoring function for each year based on two possible cases, (1) if $M_i^y > \hat{M}_i^y$, then:

$$C_{Mi}^y = \begin{cases} 0, & M_i^y > UB_{Mi}^y \\ 1, & M_i^y \le \hat{P}^{\,y}_{95,Mi} \\ \dfrac{UB_{Mi}^y - M_i^y}{UB_{Mi}^y - \hat{P}^{\,y}_{95,Mi}}, & \text{Otherwise} \end{cases} \tag{2}$$

and (2) if $M_i^y < \hat{M}_i^y$, then:

$$C_{Mi}^y = \begin{cases} 0, & M_i^y < \overline{LB_{Mi}^y} \\ 1, & M_i^y \ge LB_{Mi}^y \\ \dfrac{\overline{LB_{Mi}^y} - M_i^y}{\overline{LB_{Mi}^y} - LB_{Mi}^y}, & \text{Otherwise} \end{cases} \tag{3}$$

In this way, C_{Mi}^y falls in the range [0,1], where 1 means that the component M_i lies within acceptability levels in year i, 0 means that it is beyond the maximum plausible level, and numbers in the range (0,1) indicate that M_i is between the prediction interval and the maximum plausible level—this yearly scoring function is also shown in Figure 3.

v. Calculate the overall scoring function ($C_{overall}$) for the component M_i in two steps: (a) sort yearly scoring function values in descending order, and (b) treat the n^*+1^{th} value in the sorted vector as $C_{overall}(M_i)$ where n^* is the number of individual scoring functions that are equal to one. This value is the closest value to the acceptability intervals, and according to *Shafii and Tolson* [2015], selecting this value facilitates locating all yearly components within acceptability bounds in a sequential manner.

These steps are consistent with the methodology developed in *Shafii and Tolson* [2015], with the difference that L'vovich components (four overall scoring functions for S, W, U, and V) are used as hydrological signatures, rather than conventional hydrological signatures. The preferred solution is the one that yields a value of 1 for all four components (i.e., full consistency with respect to the Ponce-Shetty model in the presence of uncertainties). Since this model realization may not be achievable due to the existing imperfections, the multicriteria calibration yields a set of acceptable Pareto-optimal solutions. We then calculate a quality measure (called the coverage quality metric) for each L'vovich component that demonstrates how the partitioning of the Pareto solutions fall within acceptability bounds. As such, we first calculate the total number of years modeled by the entire Pareto set (i.e., $N_1 = n_p \times Y$ where n_p is the number of Pareto solutions and Y is 5 in the case shown in Figure 3). Next we find the number of years N_2 $(\le N_1)$ in which the corresponding L'vovich component is within acceptable intervals. The coverage metric is $N_2/N_1 \in [0, 1]$, desired to be as close as possible to unity.

The proposed formulation and scoring functions form a model-to-model comparison framework and it may be argued that, since we already have the Ponce-Shetty conceptual hydrologic model, application of another model is redundant. However, although the Ponce-Shetty model constrains the annual values (of *S*, *W*, *U*, and *V*), it cannot describe the within year variations. Therefore, one still needs a hydrologic model to account for the temporal dynamics of hydrologic fluxes in a given year, and the proposed approach helps in constraining such a model to yield more realistic flow partitioning. Note that multiple papers have shown the validity of the L'vovich approach to properly represent flow partitioning in any given watershed (see section 1.2 for references), so it can be considered as a benchmark.

2.3.2. Signature-Based Model Calibration Approach

Automatic model calibration generally involves using an optimization algorithm to maximize goodness of fit between simulated and observed streamflow, as well as the resulting hydrological signatures. To improve flow-partitioning consistency with respect to Ponce-Shetty model, we develop a multicriteria calibration approach where the scoring functions (related to *S*, *W*, *U*, and *V*) are applied to guide model calibration in addition to optimizing streamflow goodness of fit. We compare this to an approach in which only goodness-of-fit metrics such as *NSE* and conventional signatures are used. Section 3.2 elaborates on the underlying criteria in these two approaches, i.e., (i) the basic approach using hydrologic signatures and *NSE* and (ii) the proposed approach using the L'vovich components scoring functions and *NSE*.

As suggested by *Shafii and Tolson* [2015], this multicriteria optimization problem is solved using a Pareto-based multicriteria optimization algorithm, rather than aggregating all criteria into a single criterion and then using a single-criterion optimization algorithm to maximize the aggregated metric. Both the traditional calibration and the proposed calibration approach are performed using the same strategy suggested in *Shafii and Tolson* [2015]. The optimization objective functions in these two calibration approaches can be presented as follows:

$$\underset{\theta\in\Theta}{Max} \quad \mathbf{F}(\boldsymbol{\theta}) = [F_1(\boldsymbol{\theta}) \;\; F_2(\boldsymbol{\theta}) \;\; \ldots \;\; F_M(\boldsymbol{\theta})]^T \tag{4}$$

where $\mathbf{F}$ is the vector of objective functions or criteria composed of both goodness-of-fit and signature-based metrics, $\boldsymbol{\theta}$ is the model parameters vector located in the optimization search space Θ of which the limits are constrained by the prior range of the parameters of a given hydrologic model (see sections A1 and A2), and M is the number of objective functions, which may vary depending upon the calibration approach (see section 3.2 for detailed information on the calibration criteria). We use AMALGAM optimization algorithm [*Vrugt and Robinson*, 2007] to solve the problem defined by equation (4). The outcome is a number of Pareto-optimal solutions (not a single solution) that have shown to be acceptable with respect to at least one scoring function. Then, L'vovich components associated with the Pareto-optimal solutions are used to assess the partitioning consistency of the hydrologic model with respect to the Ponce-Shetty model.

3. Case Studies

3.1. Hydrologic Models

Our overall objective is to demonstrate that the proposed flow-partitioning-based model calibration framework is applicable to any rainfall-runoff model that is generally compatible with Ponce-Shetty model (i.e., one-to-one correspondence between a subset of the processes involved in the model and those considered in the Ponce-Shetty model). We consider two different conceptual (bucket) hydrologic models in this study (sections 3.1.1 and 3.1.2) whose structural differences include: (i) one does and the other one does not model snowmelt, (ii) vertical movement of water in soil is modeled differently, and (iii) one is lumped (HYMOD; section 3.1.1) [*Boyle*, 2000], while the other one is semidistributed developed in a flexible hydrological modeling platform (Raven; section 3.1.2) [*Craig*, 2015]; the smallest computational units are hydrological response units (HRUs). HYMOD is applied to a watershed in Florida, the United States, and the semidistributed model is applied to a watershed in Ontario, Canada.

3.1.1. Lumped HYMOD Model

HYMOD is a five-parameter rainfall-runoff model introduced by *Boyle* [2000], of which different versions have been used in numerous studies [e.g., *Kollat et al*., 2012; *Vrugt et al*., 2003; *Wagener et al*., 2001]. The simplest version of HYMOD (Figure 4 (top)) consists of a relatively simple rainfall excess model, described in

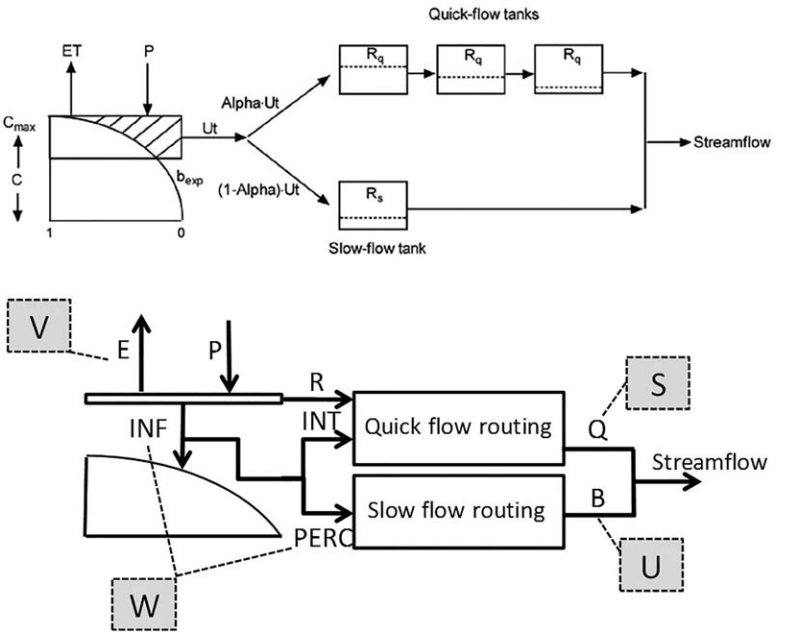

Figure 4. Schematic representation of (top) HYMOD after *Vrugt et al.* [2003] and (bottom) matching processes in HYMOD with L'vovich components.

detail by *Moore* [1985], connected with two series of linear reservoirs (three in series for quick flow and a single reservoir for the slow response). A closer look at this model structure (Figure 4 (bottom)) indicates that the processes involved in HYMOD include precipitation (*P*), evapotranspiration (*E*), infiltration (*INF*) to soil, surface runoff (*R*), interflow (*INT*) or lateral flow from soil to quick flow routing, percolation (*PERC*) or vertical drainage to slow flow routing, base flow (*B*) and quick flow (*Q*). The excess rainfall model determines the amount of *INF*, *INT*, and *PERC*. Whenever the soil is saturated, *R* occurs. Section A1 elaborates on the mathematical representation of these components. Putting HYMOD in the L'vovich context, *W* consists of the sum of *INF* and *PERC* (because *INF* and *PERC* are two separate parts of the water infiltrated to the system), *S* and *U* would be equal to *Q* and *B*, respectively, and *V* would be equal to *E*. The partitioning components of HYMOD are also demonstrated in the bottom part of Figure 4. These components are directly compared with the L'vovich partitioning components given by Ponce-Shetty model.

The above version of HYMOD requires the estimation of five parameters via calibration, the maximum storage capacity in the catchment, $C_{\max}\,(L)$, the degree of spatial variability of the soil moisture capacity within the catchment, $b_{\exp}$, the factor distributing the excess rainfall between interflow and percolation, α, and the residence time of the quick flow and base flow reservoirs, K_q and $K_s(T)$, respectively.

It is worth noting that there is no snowmelt module in the standard version of HYMOD and therefore we apply it to a watershed in Florida where there is no snow. This watershed is one of the MOPEX catchments [*Duan et al.*, 2006] with the USGS streamflow station ID of 02296750. The precipitation, potential evapotranspiration, and discharge data are used for model calibration over the time frame October 1995 to September 2000 (i.e., five water years). We also consider a 5 year validation period between October 1990 and September 1995 (five water years). To extract L'vovich components, 42 years (1948–1989) of precipitation and streamflow data are used.

3.1.2. Two-Bucket Semidistributed Hydrologic Model

The second model is developed using the modular modeling framework RAVEN [*Craig*, 2015] to simulate streamflow in a semidistributed mode. RAVEN is an object-oriented hydrological model that is compatible with a wide range of hydrologic process representations and model structures. In RAVEN, a watershed-scale model is composed of a number of subwatersheds, each of which contains a stream/river channel and subdivided into one or more HRUs. The vertical water and energy balance is applied to each HRU producing streamflow, which is then routed downstream (on the land surface) and laterally between subwatersheds. A two-bucket rainfall-runoff model (shown in Figure 5) was developed within RAVEN. The formulations

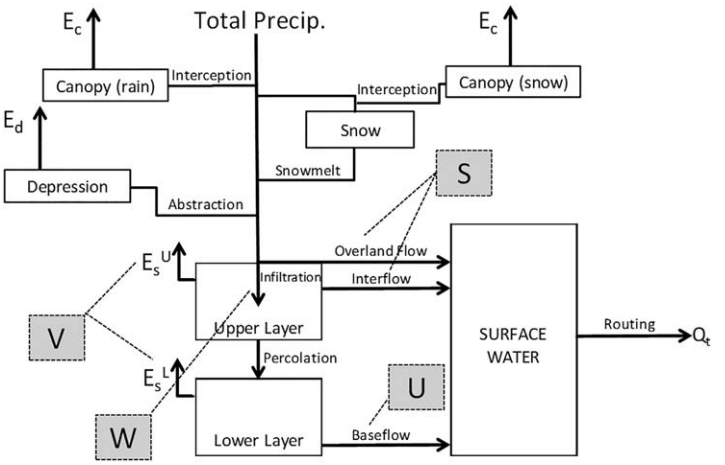

Figure 5. Two-bucket hydrologic model developed with RAVEN framework with L'vovich components shown in gray boxes.

associated with different processes in the model are described in Appendix A, but a brief description of the model structure is provided in the rest of this section.

Precipitation is first partitioned into rain and snow where each term is partly intercepted by rain and snow canopy, and then subject to the evaporation process. Snow storage is transformed to snowmelt using a simple degree-day approach. Excess rainfall and snowmelt are then combined and a part of this combination is stored in depression areas, where it is also partly evaporated. The remainder of excess water is then partitioned into infiltration to the upper soil layer and overland runoff; the latter is added to the surface water storage. Next, the evapotranspiration process is applied to the upper layer soil moisture, and if the remaining storage exceeds the storage at the field capacity, water drainage starts both vertically (i.e., percolation to the lower soil layer) and horizontally (i.e., interflow to surface water storage). Evapotranspiration is also applied to the lower layer soil moisture and then base flow occurs to the surface water storage. Finally, surface water (composed of overland runoff, interflow and base flow) is routed to the subwatersheds outlet (via a process called in-catchment routing) and then through the stream toward the subwatersheds outlet (via in-channel routing). Note that evapotranspiration from the two soil layers can exhaust soil moisture down to saturation at the wilting point and not more.

In terms of L'vovich components in this model (also shown in Figure 5), *W* corresponds to the infiltration, *S* consists of the sum of overland runoff and interflow, base flow is equivalent to *U*, and *V* is composed of both evapotranspiration terms to the atmosphere. Note that the percolation does not contribute to *W* in this model because it is deducted from the water infiltrated into the upper soil.

The developed hydrologic model is applied to one of the headwater subwatersheds in the Grand River Watershed (GRW) located in Southwestern Ontario in Canada. This subwatershed is associated with Water Survey of Canada (WSC) gauge 02GA039 on the Conestogo River with a drainage area of 274 km^2. Upon spatial delineation and overlay of digital elevation model, soil data, and land use information, the subwatershed is divided into 5 subbasins represented by 16 HRUs. Nine years of climatic and hydrometry data are used, where the time frames October 2004 to September 2009 (five water years) and October 2009 to September 2013 (five water years) are used as the calibration and validation period, respectively. To extract L'vovich components, 31 years of precipitation and streamflow data (1973–2003) are employed.

3.2. Experimental Setup

Two calibration scenarios are considered in this paper. The first scenario (called BASIC) involves automatic calibration using the AMALGAM optimization algorithm considering *NSE* calculated based on both daily and monthly streamflow (two fitness criteria) as well as the bias in four hydrological signatures including overall runoff ratio, high-flow volume in monthly and daily FDC, midsegment slope in daily FDC, and base flow index (i.e., seven optimization criteria in total). Table 1 provides detailed information on the criteria used in the BASIC calibration scenario, which eventually results in a number of Pareto-optimal solutions after applying AMALGAM to calibrate models.

In the second scenario (called LVOVICH), we calibrate models with six calibration criteria including monthly and daily *NSEs* plus four L'vovich scoring functions calculated according to the formulation described in section 2.3.1 (i.e., six optimization criteria in total, also shown in Table 1). Similar to the previous scenario, the LVOVICH scenario also yields a number of Pareto-optimal solutions, all of which are used in the comparison between the results of BASIC and LVOVICH scenarios.

Table 1. Criteria and Signatures (B1–B5: Traditional Hydrological Signatures; L1–L4: L'ovich-Based Signatures) Used in the BASIC and LVOVICH Calibration Scenarios; Acceptability Intervals are Also Shown in Brackets[a]

		Scenario		
Criteria	Description	BASIC	LVOVICH	Reference
1 [N/A]	Nash-Sutcliffe Efficiency (NSE); daily flow	X	X	*Nash and Sutcliffe* [1970]
2 [N/A]	NSE; monthly flow	X	X	*Nash and Sutcliffe* [1970]
Signature				
B1 [±5%][b]	Overall runoff ratio	X		*Shafii and Tolson* [2015]
B2 [±5%][b]	Monthly FDC high-flow volume	X		*Shafii and Tolson* [2015]
B3 [±10%][b]	Daily FDC midsegment slope	X		*Shafii and Tolson* [2015]
B4 [±10%][b]	Monthly FDC high-flow volume	X		*Shafii and Tolson* [2015]
B5 [±20%][b]	Base flow index	X		*Sawicz et al.* [2011]
L1 [95%][c]	Surface runoff		X	This paper
L2 [95%][c]	Infiltration		X	This paper
L3 [95%][c]	Subsurface flow		X	This paper
L4 [95%][c]	Evapotranspiration		X	This paper

[a]B1–B5 are traditional hydrological signatures, calculated using the time series of observed or simulated flows. L1–L4 are partitioning-based signatures developed in this study according to the L'ovich theory and the Ponce-Shetty model (section 2.3.1).
[b]Percent deviation between simulated and observed values.
[c]Prediction intervals associated with Ponce-Shetty model estimate obtained from nonlinear regression.

Note that the calibration problem in both BASIC and LVOVICH scenario is addressed utilizing the methodology developed in *Shafii and Tolson* [2015] with identical calibration settings except for the scoring functions employed for model constraining (BASIC uses traditional signatures, while LVOVICH uses the proposed signatures focused on flow partitioning—see Table 1). To consider the impact of initialization, AMALGAM is run in 10 different trials of both BASIC and LVOVICH scenarios, where each trial starts from an initial population of 100 solutions and runs up to 1000 simulations, i.e., 10,000 simulations in total for each calibration case. Afterward, the Pareto front solutions of all trials are pooled and sorted to find the resulting Pareto solution set. The number of simulations per trial in this paper (1000) was selected based on computational considerations as well as some preliminary experiments to make sure the calibrated models perform satisfactorily with respect to *NSE*, i.e., *NSEs* greater than 0.7 are obtained.

4. Results and Discussions

4.1. HYMOD Case Study

The top row in Figure 6 shows *NSE* values, the deviations for the five signatures associated with the Pareto solutions, the regime curve, and monthly and daily FDC for scenario BASIC. These results indicate high *NSE* values and perfect match between the set of simulated signatures and observed signatures. Also, no model structural failure is diagnosable by the signatures used in this case study. The bottom row in Figure 6, however, demonstrates that flow partitioning in the model is distinctly different from the Ponce-Shetty prediction intervals (model components fall outside the acceptability levels). Indeed, HYMOD produces too much surface runoff, insufficient wetting and too small base flow than would be typically expected in sandy Floridian soils. Note that this difference appears despite the fact that HYMOD captures FDC and the other signatures consistently well. Overall, although the traditional signature-based calibration (scenario BASIC) yields hydrologically consistent results according to the commonly used signatures, it does not necessarily guarantee correct flow partitioning.

The HYMOD-predicted L'ovich components associated with the Pareto front solutions for scenario LVOVICH are shown in the bottom row in Figure 7. After postprocessing the streamflow time series obtained in this scenario, we calculated the same signatures that were used in scenario BASIC, which are shown in the top row in Figure 7. These signatures similarly indicate that the resulting simulations are hydrologically consistent. Moreover, it is observed that further optimizing to the L'ovich components leads to a more realistic representation of the flow partitioning. The flow partitioning was still not adequate and to address this we doubled the computational budget considered and ran the model; however, L'ovich components were still not contained within the acceptability limits (results now shown here). This finding indicated a critical structural inadequacy of the model to capture the hydrologic partitioning.

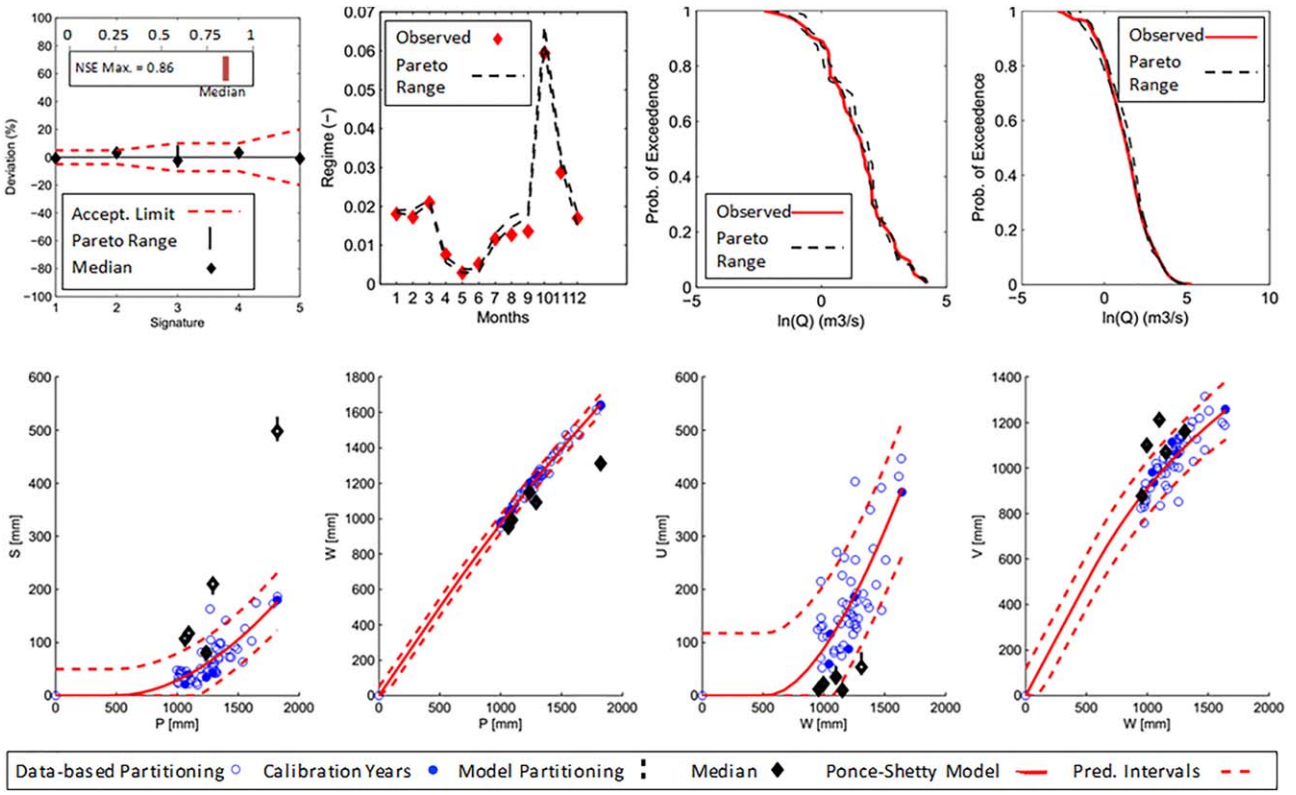

Figure 6. HYMOD results for BASIC scenario in calibration period. (top row from left to right) (i) Deviations of the five signatures described in Table 1 and the range, median, and maximum NSE, (ii) regime curve (ratio between monthly flow to annual precipitation) starting from January, (iii, iv) monthly and daily FDC. (bottom row from left to right) (i) P-S, (ii) P-W, (iii) W-U, and (iv) W-V relationships; blue circles are L'vovich partitioning-based on observed streamflow data. Vertical lines are L'vovich components of Pareto solutions (black diamonds show median); solid line is the estimated Ponce-Shetty model; dashed lines are prediction intervals (used as acceptability bounds) associated with the Ponce-Shetty model.

To address the above issue, we decided to improve flow partitioning by applying some changes to the model structure. Since the model overestimated the surface runoff, we increased the number of quick flow reservoirs (n_q), increased the slow-flow component by changing the base flow algorithm from linear to nonlinear with the degree of nonlinearity (n), and added a correction factor for evapotranspiration (e_c); n_q, n, and e_c are additional calibration parameters, thus the modified HYMOD has eight parameters estimated using both BASIC and LVOVICH scenarios (called MOD-BASIC and MOD-LVOVICH hereafter). Figure 8 illus- illustrates the modified HYMOD-generated flow partitioning when it is calibrated using scenario BASIC (top row) and LVOVICH (bottom row). The conventional signatures obtained in both cases (not shown here) are as satisfactory as those obtained in the previous cases. However, Figure 8 indicates that, in the MOD-BASIC case, flow partitioning is only slightly improved, whereas the flow-partitioning components generated by the L'vovich-guided calibration approach (MOD-LVOVICH) fall within the acceptability limits. Also, the resulting *NSE* values are similarly high in both cases. Hence, the proposed calibration approach proves to outperform the traditional signature-based calibration approach with respect to model performance as well as the flow partitioning according to the L'vovich theory.

The calibration results and parameter sets are carried over to the 5 year validation period. Figure 9 shows the results obtained for the modified HYMOD case study during the validation period. Top and bottom rows are related to the MOD-BASIC and MOD-LVOVICH scenario, respectively. We note that the simulations generated by the modified model of which the calibration is guided by flow partitioning (i.e., the MOD-LVOVICH scenario) are consistent with the Ponce-Shetty model during the validation period as catchment functions are contained within the acceptability bounds for all years. Moreover, the resulting *NSE* ranges for both scenarios are very close to one another.

The coverage quality metric values obtained in all previously explained calibration cases are shown in Figure 10 (note that ideally all coverage values should be equal to 1). It is observed in Figure 10a that even though the LVOVICH scenario results in higher coverage values than BASIC, these values are still not equal

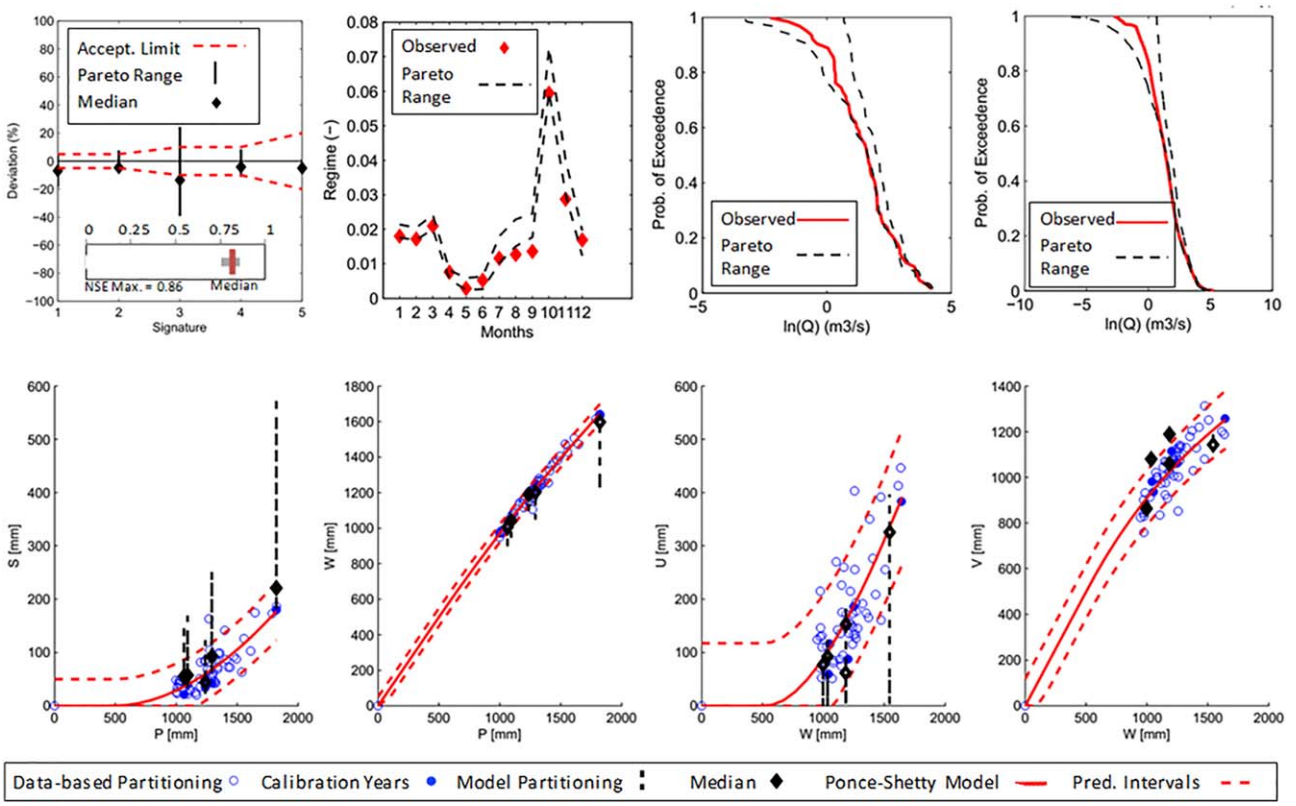

Figure 7. HYMOD results for LVOVICH scenario in calibration period. Details on these plots are provided in the caption of Figure 6.

to 1 in any of the components. By modifying the model structure most of these values become equal to 1 for the MOD-LVOVICH scenario, which indicates a higher level of consistency in flow partitioning. Regarding component *V*, the fact that the quality metric cannot be equal to 1 indicates that, even after modifying the model structure and introducing a correction factor for *PET*, the method of simulating evapotranspiration is not wholly satisfactory in this model. This finding shows that the proposed method is also capable of diagnosing model structural inadequacies.

Overall, the HYMOD results imply that the proposed partitioning-based model calibration represents a promising means of constraining the model outcome with respect to the functioning and hydrologic partitioning within the watershed in both calibration and validation period. Unlike the proposed approach, the traditional signature-based model calibration does not necessarily improve flow partitioning, and it may not diagnose the deficiencies in the model structure properly.

4.2. Two-Bucket Semidistributed Hydrologic Model Case Study

Similar to Figure 8 for the HYMOD case study, Figure 11 illustrates the results obtained in the second modeling case study, where BASIC and LVOVICH scenarios results are provided in the top and bottom row, respectively. Unlike HYMOD, there is no modified version of the model in this case. It is observed that the model results of the LVOVICH scenario are more tightly contained within the acceptability bounds compared to scenario BASIC. Moreover, the *NSE* plots in Figure 11 also show that it is possible to get similar daily *NSE* in the second scenario, but with more consistent flow partitioning. It is worth mentioning that the values we obtained for Ponce-Shetty model parameters in this case study are completely in agreement with those reported in previous papers, *Sivapalan et al*. [2011] and *Ponce and Shetty* [1995a].

The top row in Figure 12 shows the signature ranges associated with the Pareto solutions obtained in scenario BASIC and LVOVICH (note that these signatures are directly incorporated in the BASIC calibration, whereas they are postcalculated in the scenario LVOVICH). The signature results indicate that there is not much difference between the consistency of model simulations. In the validation period (results not shown here), the resulting flow partitioning from BASIC and LVOVICH scenarios indicates that the proposed

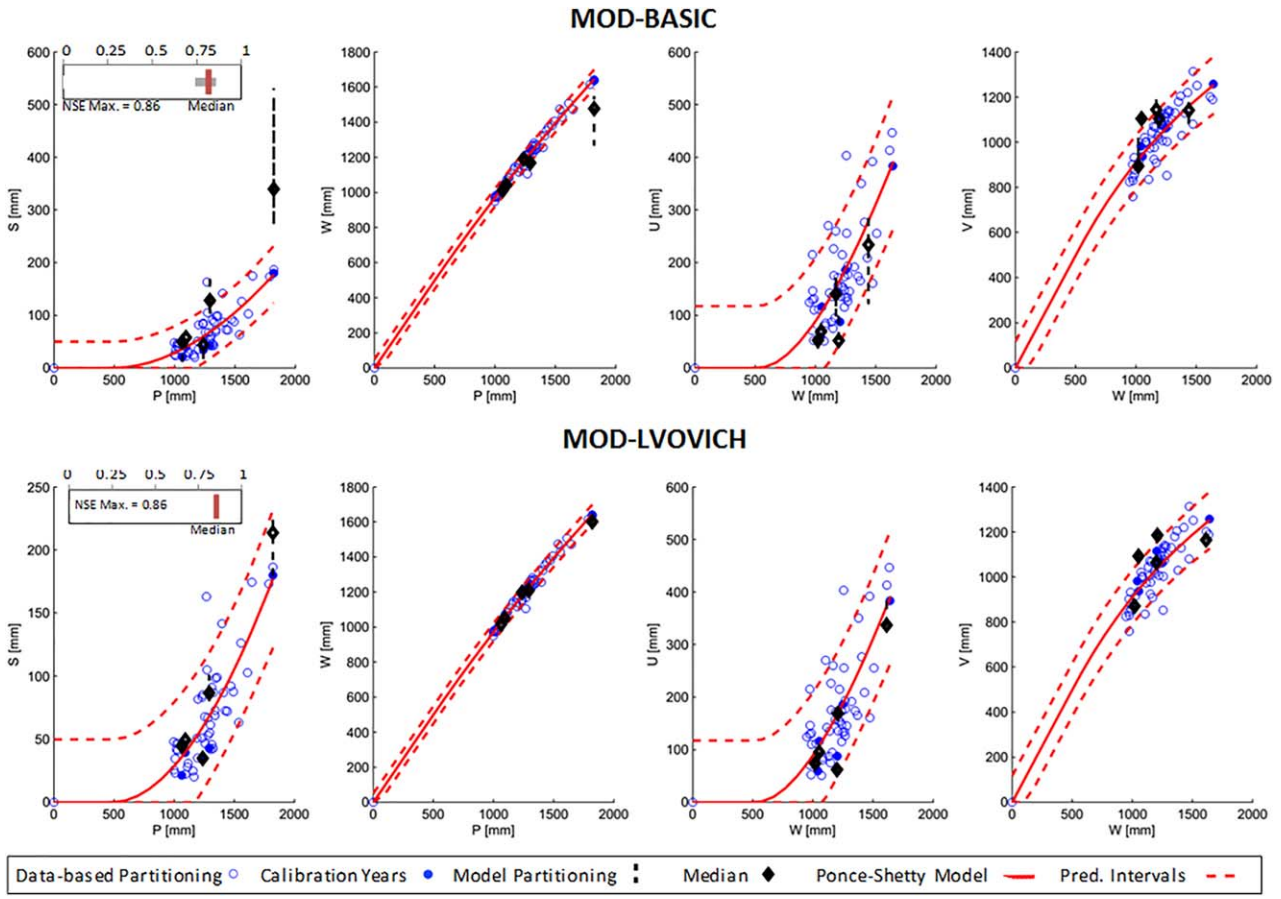

Figure 8. Flow partitioning based on the modified HYMOD model's results for MOD-BASIC and MOD-LVOVICH scenarios in calibration period. Details on these plots are provided in the caption of Figure 6.

approach reproduces the L'vovich components more properly in comparison with the traditional signature-based calibration. *NSE* values for these two scenarios are also fairly similar in the validation period scenarios (maximum *NSE* of 0.63 and 0.61 in BASIC and LVOVICH scenarios, respectively). The bottom row in Figure 12 shows the coverage quality metric for BASIC and LVOVICH scenarios, which indicate that coverage improves in the latter compared to the former both in calibration and validation period (i.e., higher coverage values for the LVOVICH scenario). Thus, the proposed model calibration strategy produces comparably appropriate goodness of fit while it improves the consistency of flow partitioning with the Ponce-Shetty theory.

The findings again highlight the usefulness of the proposed model calibration framework to constrain model outcomes with respect to flow partitioning. It should be noted that this framework can be used either for evaluating the consistency of an existing model or for guiding model calibration toward the regions in the search space that yield consistent simulations with observed flow partitioning. In the former case, the modeler can apply changes to model structure and/or parameterization to improve the consistency of the modified model. The latter case facilitates the improvement of consistency through process-based model calibration.

4.3. Parameter Identifiability

To investigate the possibility of improving parameters identifiability using the proposed partitioning-based calibration, we focus on the range of parameter values contained in the set of Pareto-optimal solution in each calibration scenario. Figure 13 illustrates the range between the minimum and maximum parameter values in the Pareto set (vertical lines) as well as the range between the first and third quartiles, i.e., 25 and 75 percentile (vertical bars), obtained in the modified HYMOD case using BASIC (red color) and LVOVICH (black color) scenario. It is observed that the parameter ranges obtained in scenario LVOVICH are smaller than those of scenario BASIC for all but one parameter. RAVEN results (not shown here) also demonstrate

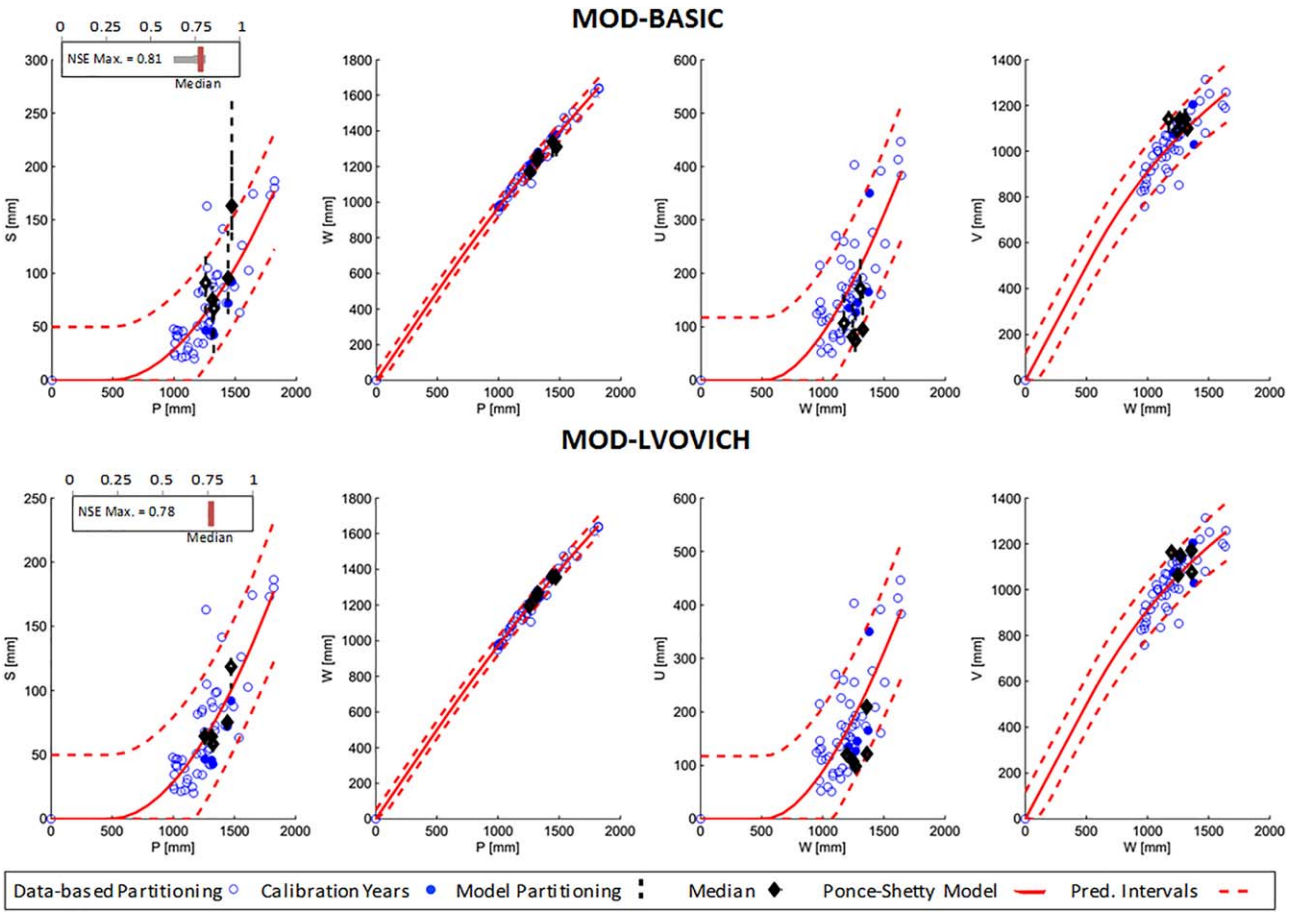

Figure 9. Modified HYMOD model's results for scenario (top row) MOD-BASIC, and (bottom row) MOD-LVOVICH during the validation period. Details on these plots are provided in the caption of Figure 6.

that the Pareto range of 16 (out of 21) model parameters is reduced in scenario LVOVICH. Such range reduction in both models indicates the improvement in parameters' identifiability. The reduction in parameter identifiability also results in smaller ranges for the partitioning components and Figures (8 and 9), and 11 clearly demonstrate this fact.

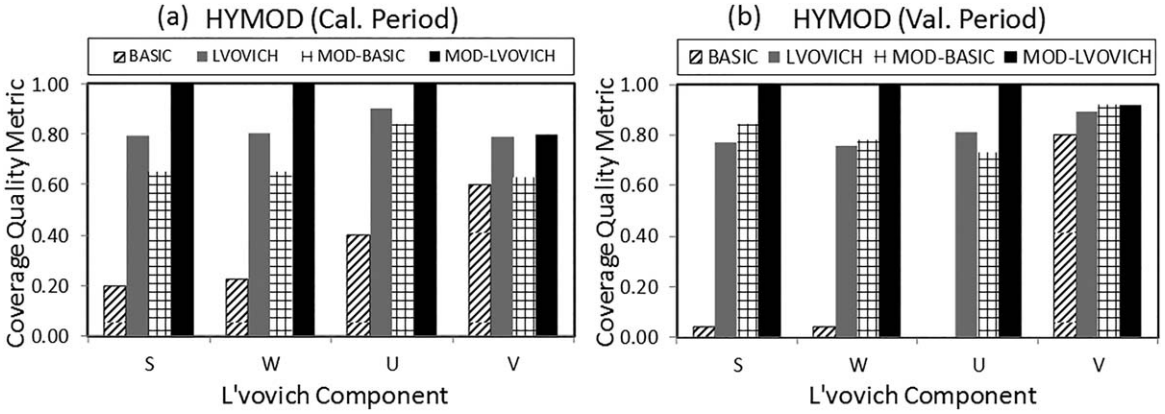

Figure 10. Coverage quality metric for L'vovich components S, W, U, and V in HYMOD case study calculated based on the Pareto solutions of different calibration experiments. Results are for (left) calibration and (right) validation periods.

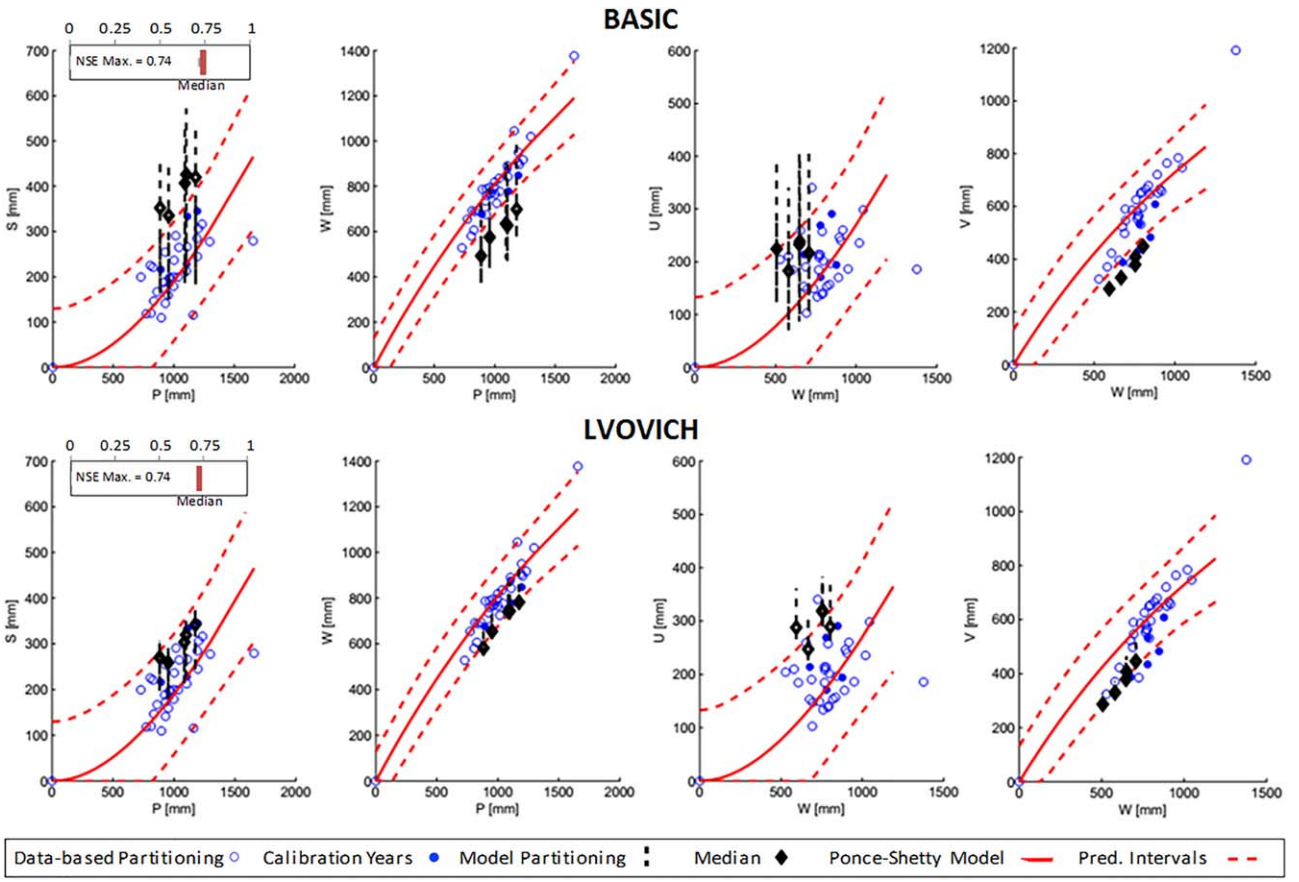

Figure 11. Two-bucket model results for different scenarios in calibration period: (top row) BASIC and (bottom row) LVOVICH. Details on these plots are provided in the caption of Figure 6.

The reduction in the ranges of the model parameters and the corresponding partitioning components also has implications with respect to model's sensitivity to parameter values. Smaller parameter ranges might mean that model is more sensitive to that parameter, but this hypothesis is not in the scope of this work, and will be explored in future. Another research direction is to investigate the sensitivity of hydrologic model parameters to the criteria used in the proposed diagnostic model evaluation, especially in comparison to the traditional signature-based metrics. For example, one can evaluate model's sensitivity to different parameters taking into account a particular process (e.g., infiltration) rather than the overall performance of model quantified in terms of aggregate goodness-of-fit metrics.

The two-stage partitioning relationship by Ponce and Shetty is derived based on the proportionality hypothesis, which is also the basis of the curve number approach at the event scale, i.e., Soil Conservation Service or SCS method [*Mockus*, 1972]. According to this hypothesis, runoff (and vaporization in the second partitioning stage) does not start until the corresponding initial abstraction has been met. HYMOD is a saturation-excess runoff model and does not take into account initial abstractions and thus has to be calibrated to the Ponce and Shetty relationship to capture the flow partitioning. But the value of our methodology is model independent. This is apparent in the observation that the second model used (RAVEN) does take into account initial abstractions, but still fails to reproduce the Ponce and Shetty relationships, and has to be constrained to satisfy it. Understanding which model process descriptions would lead to a better match with the Ponce and Shetty relationship without additional calibration can be explored in future papers.

4.4. Limitations

The biggest limitation of the proposed methodology is the assumption that the L'vovich flow partitioning is "correct," and thus constraining the model to match it is a valid approach. The L'vovich framework and

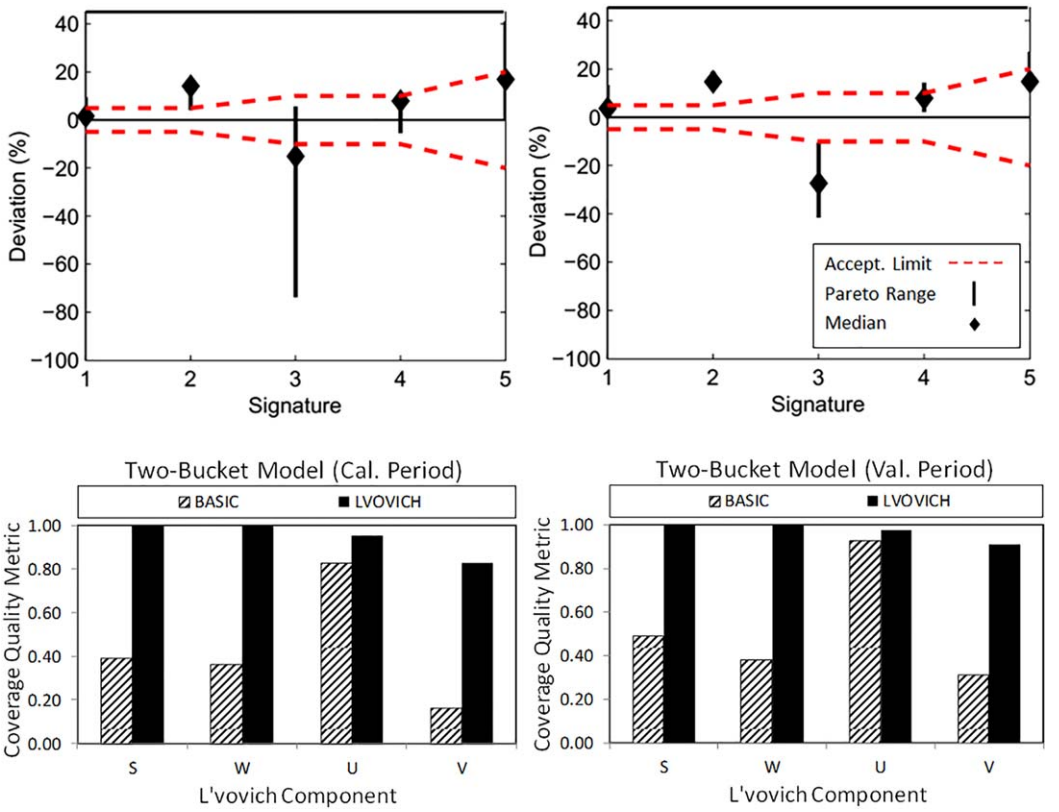

Figure 12. Hydrological signatures associated to the Pareto-optimal solutions obtained in (top row) scenario BASIC and LVOVICH in calibration period, and coverage quality metric for L'vovich components S, W, U, and V calculated based on the Pareto solutions in both scenarios during (left) calibration and (right) validation period.

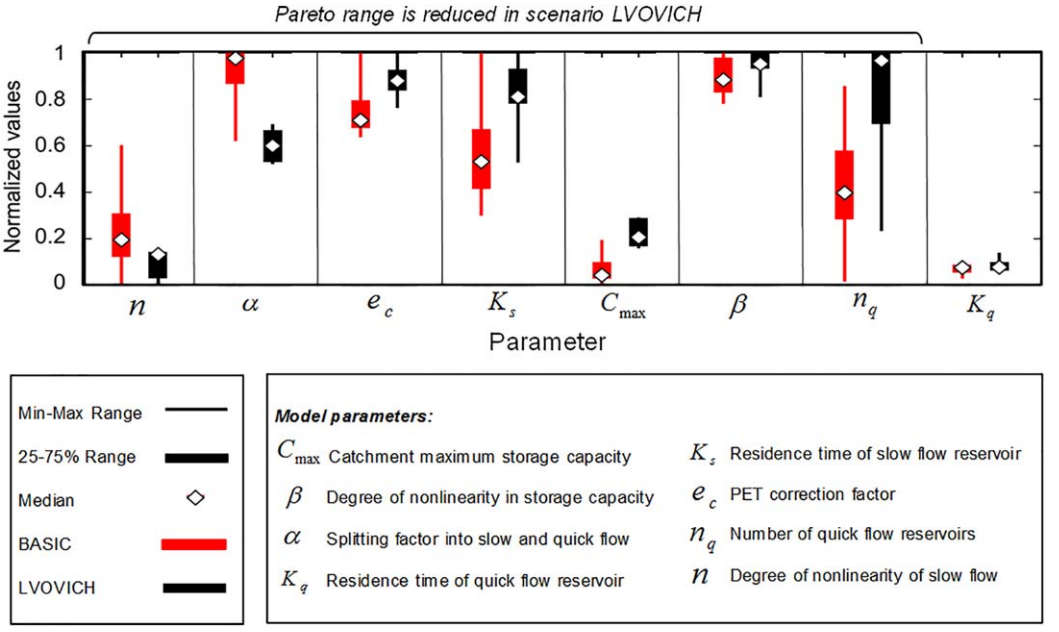

Figure 13. Parameter ranges (minimum to maximum: line; interquartile: bar) and median values associated with the Pareto set obtained for BASIC (red) and LVOVICH (black) scenarios in the modified HYMOD case study. Those parameters of which the Pareto range is smaller in LVOVICH compared to BASIC are separated from other parameters. Parameters description is also provided in the table embedded in this figure.

subsequent studies by Ponce and Shetty have indeed been able to provide a reasonable means of describing the flow partitioning in several catchments. Model parameters have also proven to be consistent with the climatic and physiographic settings. For example, *Ponce and Shetty* [1995a] demonstrate that their model results in zero initial abstraction, low to average wetting potential, low base flow, and low vaporization potential in cold region catchments, which is corroborated by the frozen ground and low temperature in these areas. A comprehensive study by *Sivapalan et al.* [2011] across the continental U.S. also shows that the spatial patterns of the L'vovich parameters are consistent with spatial variations of climatic and landscape properties. Furthermore, in our Florida case study, the flow constraining based on L'vovich did lead to more realistic surface runoff and base flow magnitudes that is expected in Floridian soils. It is worth noting that, despite the suitability of this theory, we have not used it as a crisp benchmark and considered the uncertainties involved in the estimation of Ponce-Shetty model to derive acceptability bounds.

It is also worth noting that the L'vovich theory and Ponce-Shetty model assume that there is no storage carryover over years. If the proposed calibration approach is utilized, simulated storage carryover has to be verified to meet this assumption. The interannual water storage carryover was negligible for both our models (less than 0.5 mm for HYMOD and less than 2.5 mm for RAVEN case study) and no systematic patterns were found in them over time. Recently, *Chen and Wang* [2015] developed a new partitioning formulation at seasonal scale that accounts for storage carryover, which should be taken into account in cases where there exists such carryover.

Future work would involve further exploring the ability of the L'vovich approach to describe flow partitioning using tracers and/or stable isotopes. Moreover, as base flow separation is a key element in this methodology, this has to be conducted carefully and ultimately validated using independent tracer studies. It is expected that constraining hydrologic partitioning-based on the L'vovich theory may be more appropriate for watersheds that are not regulated by large dams, because reservoir operation may change the natural behavior of the watershed significantly. It is also worth noting that using this approach requires long records of precipitation and streamflow data. Acknowledging this fact, the method cannot be applied directly to ungauged basins. However, one possible strategy to benefit from this approach in ungauged basins may be to regionalize the Ponce-Shetty model parameters with respect to watershed's physical properties.

In this study, we used an optimization-based calibration approach utilizing NSE (as an informal residual-based performance metric) as well as L'vovich scoring functions to find optimal solutions that are consistent with respect to the partitioning as well. An alternative approach to the identification of such parameter sets is to conduct the search by minimizing robust and statistically correct functions of model residuals (i.e., frequentist or Bayesian methods), and subsequently, select those solutions that are consistent with the L'vovich understanding of watershed response. Such a stepwise parameter estimation (using other metrics of catchment functions) has been previously reported in the literature [e.g., *Ahmadi et al.*, 2014; *Yen et al.*, 2014] and incorporating L'vovich scoring functions in it is considered as a future research avenue.

5. Conclusions

This research extends the classical theoretical approach to hydrologic partitioning introduced by *L'vovich* [1979] and the corresponding mathematical representations explored in subsequent studies [*Ponce and Shetty*, 1995a, 1995b; *Sivapalan et al.*, 2011] to constrain the outcome of hydrologic models based on the interactions among major catchment functions: annual quick and slow lateral flow, infiltration, and evapotranspiration. The main objective is to investigate whether hydrologic models can be guided to better reproduce the flow partitioning in a watershed given these additional constraints. As such, we developed a model calibration framework that employs four partitioning-based signatures quantified based on the L'vovich approach and employed it to constrain the simulations generated with two hydrologic models (one lumped and one semidistributed model). To do so, we guided the calibration process with the proposed partitioning-based quality metrics and compared the results with those obtained from traditional signature-based model calibration.

Our results demonstrate that the L'vovich partitioning-based calibration provided more realistic representation of the surface and subsurface flow components, compared to the traditional signature-based calibration approach. Indeed in the first case study, the traditional signature-based calibration approach produced

acceptable goodness-of-fit metrics but led to much higher surface runoff and lower base flow than can be typically expected in Floridian soils. The proposed methodology led to more reasonable flow-partitioning behavior, while producing hydrological fits that were comparable to the traditional signature-based methodology. Furthermore, as shown in one of the case studies in this research, the proposed model evaluation framework can also help to improve the model structure based on the observed inconsistencies. Finally, partitioning-based model calibration was observed to increase parameter identifiability and decrease model uncertainty.

Our approach is designed to complement, and not replace, the novel methods that have been proposed by others. For example, if additional data sources like isotopes are available, the use of such information would be preferred over our approach. It would be interesting to compare our approach to isotope-based partitioning when such data sources are available. Our approach is similar, but distinct, from that of *He et al*. [2015] in which the hydrograph is partitioned based on the dominant runoff generation mechanisms (such as snowmelt, glacial melt, and groundwater discharge), and the model is calibrated separately to each mechanism via a stepwise approach. We not only partition the hydrograph but explicitly consider the interactions between the various flow pathways, using L'vovich theory, and calibrate the model to each of these distinct interactions. The expert knowledge-based constraining approach proposed in *Gharari et al*. [2014] demonstrates that how conditioning a model based on experts' knowledge can improve its predictive power. It would be interesting to not only compare our approach to these novel methods but also evaluate if using them in conjunction leads to more consistent hydrologic predictions.

Overall, the proposed framework is deemed to be a suitable diagnostic approach to increasing the chance of "getting the right answer for the right reasons" in hydrological modeling [*Kirchner*, 2006]. This is crucially important in modeling exercises where flow pathways dictate biogeochemical sources and transformation processes. Therefore, the proposed method is a step toward more reliable water quantity and quality predictions. There is a significant body of research on the diagnostic model evaluation and the use of signatures (e.g., references in section 1.1); however, as demonstrated in this study, the existing approaches do not adequately capture surface and subsurface flow partitioning. This is the contribution of our study, and we believe that our strategy synthesizes process-based modeling with empirical data-based water balance relationships and presents opportunities for progress in hydrologic modeling, as suggested by *Wang and Tang* [2014]. The proposed partitioning-focused model calibration should also be explored in the context of coupled hydrological-biogeochemical modeling, and research in this area is ongoing by authors.

Appendix A

A1. HYMOD Mathematical Representation

Table A1 shows the parameters and variables involved in HYMOD along with parameters' prior ranges considered in model calibration.

Soil moisture storage capacity, c (mm), is assumed to vary across the catchment and the spatial variability of soil moisture capacity is described by the following distribution function:

$$F(c) = \frac{1}{1+\beta}\left[\left(1 - c(t)/C_{\max}\right)^{1+\beta}\right] \tag{A1}$$

Table A1. Variables and Calibration Parameters Associated With HYMOD

Variable/Parameter	Description	Unit	Prior Range
Variable			
PET	Potential evapotranspiration rate	mm/d	
P_{ex}	Excess rainfall	mm/d	
Parameter			
$C_{\max}$	Catchment maximum storage capacity	mm	1–500
β	Degree of spatial variability in storage capacity		0.1–2.0
α	Factor splitting excess rainfall into interflow and percolation		0.1–0.5
K_q	Residence time of quick flow routing reservoirs	Day	0.05–0.5
K_s	Residence time of slow flow routing reservoir	Day	0.001–0.5

Table A2. Variables and Calibration Parameters Associated With the Two-Bucket Hydrologic Model

Variable/Parameter	Description	Unit	Prior Range
Variable			
PET	Potential evapotranspiration rate	mm/d	
F_c	Forest coverage proportion		
φ_{soil}	Soil layer water content	mm	
$\varphi_{\max}=Hn$	Maximum soil storage	mm	
H	Soil layer thickness	mm	
n	Porosity		
$\varphi_{tens}=\varphi_{FC}-\varphi_{WP}$	Tension storage	mm	
φ_{FC}	Storage at field capacity	mm	
φ_{WP}	Storage at wilting point	mm	
ξ	Root fraction		
Parameter			
F_{sparse}	Vegetation sparseness		0.1–0.99
F_{ow}	Fraction of PET to apply to open water evaporation		0.1–0.5
$M_{p,\max}$	Maximum percolation rate	mm/d	0.1–30
$M_{i,\max}$	Maximum interflow rate	mm/d	0.01–1.0
k	Base flow coefficient	1/d	0.001–0.5
β	Infiltration parameter		1.5–2.5
S_{rain}	Rain canopy storage	mm	1–10
S_{snow}	Snow canopy storage	mm	1–10
S_{dep}	Depression storage	mm	1–50
C_{PET}	PET correction factor		0.5–1.5
ΔT	Rain-snow mixture temperature range	°C	0.0–3.0
M_a	DD factor	mm/d/°C	2.0–5.0

Given the storage at the beginning of time step t as $c_1(t)$, when rainfall $P(t)$ happens during the time step, all stores of capacity less than $P(t)$ (identified using equation (A1)) are filled with water and the storage at the end of time step or $c_2(t)$ is calculated, i.e., infiltration to the soil is $INF(t)=c_2(t)-c_1(t)$. Evapotranspiration is calculated as $E(t)=PET(t).\min\left(INF(t)/C_{\max},1\right)$, interflow is $INT(t)=\alpha.P_{ex}(t)$ where $P_{ex}(t)=P(t)-INF(t)$, and if $P_{ex}(t)>C_{\max}$, surface runoff occurs $R(t)=P_{ex}(t)-C_{\max}$. Percolation to slow flow routing is calculated as $PERC(t)=(1-\alpha).P_{ex}(t)$. Quick flow is routed using three linear reservoirs, where in each reservoir the outflow is $OUT_q(t)=K_q^{-1}.S_{q,i}(t)$; $S_{q,i}$ is the depth of water in reservoir i at the end of time step t. Slow flow routing works similarly using one linear reservoir, i.e., $OUT_s(t)=K_s^{-1}.S_s(t)$. Streamflow at the end of time step t is composed of the outflow from the third quick flow routing reservoir and the slow flow routing reservoir. In the modified version of HYMOD, corrected evapotranspiration is calculated as $E'(t)=e_cE(t)$ where e_c is the evapotranspiration correction factor, the number of quick flow is estimated by calibration, and the modified base flow is calculated as $OUT_s'(t)=K_s^{-1}.S_s(t)^n$ where n is a calibration parameter.

A2. Two-Bucket Semidistributed Model Formulation

This appendix provides the equations implemented in the model described in section 3.1.2 to quantify different hydrological processes. Table A2 elaborates on different variables used in the model formulation, as well as the global parameters (and their prior range) to be adjusted via calibration—see Table A2.

Potential evapotranspiration (PET): Calculated based on the formulations provided in *Hargreaves and Samani* [1985]. The Hargreaves equation is an empirical approach based solely on temperature and incoming solar radiation. The latter is approximated using the air temperature (see RAVEN's user manual [*Craig*, 2015] for more information). The calculated *PET* is then corrected using a correction factor (a calibration parameter).

Canopy: Upon the rainfall-snow partitioning, 5% of the rainfall and snow is intercepted in the rain and snow canopy storage, respectively.

Canopy evaporation rate (E_c): Calculated as $E_c=PET.F_c$ and applied to both rain and snow canopy storages.

Depression and open water evaporation rate (E_d): A part of excess rainfall and snowmelt is stored in the depression storage, and *PET* is then applied to it to evaporate open water with the rate $E_d=PET.F_{ow}$.

Snow modeling: The first step in the estimation of snowmelt rate is to partition total precipitation into rain and snow. Rather than using a simple threshold temperature for partitioning, a linear transition between all snow and all rain is determined based on the average daily temperature, or $\alpha=0.5-T_a/\Delta T$ where T_a is the

average air temperature (°C), and ΔT is a temperature range (°C). In this way, alpha (i.e., the proportion of snow) varies from 1 to 0 as temperature changes from $-\Delta T/2$ to $\Delta T/2$. Upon rain and snow partitioning, the potential melt rate is calculated using the equation $M_{melt,TI}=M_a.\max(T_a,0)$ where $M_{melt,TI}$ is the potential melt rate (mm/d), T_a is air temperature (°C), and M_a is the DD factor (mm/d/°C).

Infiltration rate ($M_{\inf}$) and overland flow ($M_{overlan}$): Calculated as $M_{\inf}=R(1-(\varphi_{soil}/\varphi_{\max})^{\beta})$ where R is the combination of excess rainfall and snowmelt (mm/d), and β is a coefficient. The overland flow is then calculated as $M_{overland}=R-M_{\inf}$, which is stored in the surface water storage in the model.

Soil evaporation (E_s): This process is linearly proportional to the soul saturation but distributed by root fraction between upper and lower soil layers. If soil moisture exceeds the tension storage φ_{tens}, soil evaporation is calculated as $E_s^U=PET.\xi_U.\min\left(\varphi_{soil}^U/\varphi_{tens}^U,1\right)$ and $E_s^L=PET.\xi_L.\min\left(\varphi_{soil}^L/\varphi_{tens}^L,1\right)$ where U and L refer to the upper and lower soil layers, respectively. ξ_U and ξ_L are set to 0.7 and 0.3, respectively. Soil evaporation can exhaust the storage down to the wilting point storage and not more than that.

Percolation rate: Percolation becomes active when soil moisture in the upper layer exceeds the storage at field capacity, and is calculated as $M_{perc}=M_{p,\max}((\varphi_{soil}-\varphi_{FC})/(\varphi_{\max}-\varphi_{FC}))$.

Interflow rate (M_{int}): Similar to the percolation, this process also becomes active when there is more water in the upper soil layer than saturation at field capacity, and is calculated as $M_{\text{int}}=M_{i,\max}((\varphi_{soil}-\varphi_{FC})/(\varphi_{\max}-\varphi_{FC}))$. The percolation and interflow can be generally considered as the drainage of water out of the upper later in the vertical and horizontal directions, respectively, with different potentials expressed as $M_{p,\max}$ and $M_{i,\max}$.

Base flow rate (M_{base}). Calculated as $M_{base}=k\varphi_{soil}$.

Routing: During each time step, the overland flow, interflow, and base flow components are stored in the surface water, which is then routed toward the watershed outlet. This routing is called in-catchment routing. Furthermore, as each watershed is composed of smaller subwatersheds, the water accumulated in the outlet of upstream subwatersheds is routed in stream toward the outlet of downstream subwatersheds. This routing component is called in-channel routing. The in-catchment routing is a function of the time of concentration (T_c) in each small subwatersheds, i.e., if T_c is greater than one day, a part of the streamflow generated during a time step will get to the outlet in the next time step. However, the T_c values we calculated using empirical formulations in *Fang et al*. [2008] were all less than a day, which means that water generated in each time step can be transferred to the outlet on the same day, i.e., no in-catchment routing.

The in-channel routing is a function of recent time history of upstream and (upbasin) inflows to the channel as well as a vector of channel parameters (e.g., channel rating curves, and channel and bank roughness). Given the channel properties, we use applied the iterative Newton's root-finding algorithm to the following discretization of the storage relationship:

$$\frac{V\left(Q_{out}^{n+1}\right)-V\left(Q_{out}^{n}\right)}{\Delta t}=\frac{1}{2}\left(Q_{in}^{n}+Q_{in}^{n+1}\right)-\frac{1}{2}\left(Q_{out}^{n}+Q_{out}^{n+1}\right) \tag{A2}$$

where $V(Q)$ is the channel volume as a function of discharge Q, Q_{in}^t, and Q_{out}^t are the inflow and outflow in time step t, respectively. This method is a root-finding problem for Q_{out}^{n+1}, which is performed in each time step.

Acknowledgments
This research was supported with funding from NSERC Strategic Partnership grant (STPGP-447692-2013), and also with funding from Canada Excellence Research Chair in Ecohydrology in the Department of Earth and Environmental Sciences at University of Waterloo. The data used in this research can be requested from the corresponding author by e-mail.

References

Ahmadi, M., M. Arabi, J. C. Ascough Ii, D. G. Fontane, and B. A. Engel (2014), Toward improved calibration of watershed models: Multisite multiobjective measures of information, *Environ. Modell. Software*, *59*, 135–145.

Arnold, J. G., and P. M. Allen (1999), Automated methods for estimating baseflow and ground water recharge from streamflow records, *J. Am. Water Resour. Assoc.*, *35*(2), 411–424.

Beven, K. (2006), A manifesto for the equifinality thesis, *J. Hydrol.*, *320*(1–2), 18–36.

Boyle, D. P. (2000), Multicriteria calibration of hydrological models, PhD dissertation thesis, Univ. of Ariz., Tucson, AZ.

Budyko, M. (1974), *Climate and Life*, Elsevier, New York.

Bulygina, N., and H. Gupta (2009), Estimating the uncertain mathematical structure of a water balance model via Bayesian data assimilation, *Water Resour. Res.*, *45*, W00B13, doi:10.1029/2007WR006749.

Chen, X., and D. Wang (2015), Modeling seasonal surface runoff and base flow based on the generalized proportionality hypothesis, *J. Hydrol.*, *527*, 367–379.

Clarke, R. T. (1973), A review of some mathematical models used in hydrology, with observations on their calibration and use, *J. Hydrol.*, *19*(1), 1–20.

Coxon, G., J. Freer, T. Wagener, N. A. Odoni, and M. Clark (2014), Diagnostic evaluation of multiple hypotheses of hydrological behaviour in a limits-of-acceptability framework for 24 UK catchments, *Hydrol. Processes*, *28*(25), 6135–6150.
Craig, J. (2015), *RAVEN's User Manual*, Dept. of Civ. and Environ. Eng., Uni. of Waterloo, Waterloo, Ont., Canada.
Davies, J., K. Beven, L. Nyberg, and A. Rodhe (2011), A discrete particle representation of hillslope hydrology: Hypothesis testing in reproducing a tracer experiment at Gårdsjön, Sweden, *Hydrol. Processes*, *25*(23), 3602–3612.
Davies, J., K. Beven, A. Rodhe, L. Nyberg, and K. Bishop (2013), Integrated modeling of flow and residence times at the catchment scale with multiple interacting pathways, *Water Resour. Res.*, *49*, 4738–4750, doi:10.1002/wrcr.20377.
Didszun, J., and S. Uhlenbrook (2008), Scaling of dominant runoff generation processes: Nested catchments approach using multiple tracers, *Water Resour. Res.*, *44*, W02410, doi:10.1029/2006WR005242.
Duan, Q., S. Sorooshian, and V. K. Gupta (1992), Effective and efficient global optimization for conceptual rainfall-runoff models, *Water Resour. Res.*, *28*(4), 1015–1031.
Duan, Q., et al. (2006), Model Parameter Estimation Experiment (MOPEX): An overview of science strategy and major results from the second and third workshops, *J. Hydrol.*, *320*(1–2), 3–17.
Eagleson, P. S. (1978), Climate, soil, and vegetation: 3. A simplified model of soil moisture movement in the liquid phase, *Water Resour. Res.*, *14*(5), 722–730.
Eder, G., M. Sivapalan, and H. P. Nachtnebel (2003), Modelling water balances in an Alpine catchment through exploitation of emergent properties over changing time scales, *Hydrol. Processes*, *17*(11), 2125–2149.
Euser, T., H. C. Winsemius, M. Hrachowitz, F. Fenicia, S. Uhlenbrook, and H. H. G. Savenije (2013), A framework to assess the realism of model structures using hydrological signatures, *Hydrol. Earth Syst. Sci.*, *17*, 1893–1912.
Fang, X., D. B. Thompson, T. G. Cleveland, P. Pradhan, and R. Malla (2008), Time of concentration estimated using watershed parameters determined by automated and manual methods, *J. Irrig. Drain. Eng.*, *134*(2), 202–211.
Farmer, D., M. Sivapalan, and C. Jothityangkoon (2003), Climate, soil, and vegetation controls upon the variability of water balance in temperate and semiarid landscapes: Downward approach to water balance analysis, *Water Resour. Res.*, *39*(2), 1035, doi:10.1029/2001WR000328.
Gharari, S., M. Shafiei, M. Hrachowitz, R. Kumar, F. Fenicia, H. V. Gupta, and H. H. G. Savenije (2014), A constraint-based search algorithm for parameter identification of environmental models, *Hydrol. Earth Syst. Sci.*, *18*(12), 4861–4870.
Gupta, H. V., T. Wagener, and Y. Liu (2008), Reconciling theory with observations: Elements of a diagnostic approach to model evaluation, *Hydrol. Processes*, *22*(18), 3802–3813.
Gupta, H. V., H. Kling, K. K. Yilmaz, and G. F. Martinez (2009), Decomposition of the mean squared error and NSE performance criteria: Implications for improving hydrological modelling, *J. Hydrol.*, *377*(1–2), 80–91.
Hargreaves, G. H., and Z. A. Samani (1985), Reference crop evapotranspiration from ambient air temperature, *Rep. Pap. 85-2517*, Am. Soc. of Agric. Eng., Chicago, Ill.
Harman, C. J., P. A. Troch, and M. Sivapalan (2011), Functional model of water balance variability at the catchment scale: 2. Elasticity of fast and slow runoff components to precipitation change in the continental United States, *Water Resour. Res.*, *47*, W02523, doi:10.1029/2010WR009656.
He, Z. H., F. Q. Tian, H. V. Gupta, H. C. Hu, and H. P. Hu (2015), Diagnostic calibration of a hydrological model in a mountain area by hydrograph partitioning, *Hydrol. Earth Syst. Sci.*, *19*(4), 1807–1826.
Horton, R. E. (1933), The role of infiltration in the hydrologic cycle, *Eos Trans. AGU*, *14*, 446–460.
Hrachowitz, M., H. Savenije, T. A. Bogaard, D. Tetzlaff, and C. Soulsby (2013), What can flux tracking teach us about water age distribution patterns and their temporal dynamics?, *Hydrol. Earth Syst. Sci.*, *17*(2), 533–564.
Hrachowitz, M., O. Fovet, L. Ruiz, T. Euser, S. Gharari, R. Nijzink, J. Freer, H. H. G. Savenije, and C. Gascuel-Odoux (2014), Process consistency in models: The importance of system signatures, expert knowledge, and process complexity, *Water Resour. Res.*, *50*, 7445–7469, doi:10.1002/2014WR015484.
Johnston, P. R., and D. H. Pilgrim (1976), Parameter optimization for watershed models, *Water Resour. Res.*, *12*(3), 477–486.
Jothityangkoon, C., M. Sivapalan, and D. L. Farmer (2001), Process controls of water balance variability in a large semi-arid catchment: Downward approach to hydrological model development, *J. Hydrol.*, *254*(1–4), 174–198.
Kirchner, J. W. (2006), Getting the right answers for the right reasons: Linking measurements, analyses, and models to advance the science of hydrology, *Water Resour. Res.*, *42*, W03S04, doi:10.1029/2005WR004362.
Klaus, J., and J. J. McDonnell (2013), Hydrograph separation using stable isotopes: Review and evaluation, *J. Hydrol.*, *505*, 47–64.
Kollat, J. B., P. M. Reed, and T. Wagener (2012), When are multiobjective calibration trade-offs in hydrologic models meaningful?, *Water Resour. Res.*, *48*, W03520, doi:10.1029/2011WR011534.
Krause, P., D. P. Boyle, and F. Bäse (2005), Comparison of different efficiency criteria for hydrological model assessment, *Adv. Geosci.*, *5*, 89–97.
Kuczera, G. (1983), Improved parameter inference in catchment models: 1. Evaluating parameter uncertainty, *Water Resour. Res.*, *19*(5), 1151–1162.
Legates, D. R., and G. J. McCabe (1999), Evaluating the use of "goodness-of-fit" measures in hydrologic and hydro-climatic model evaluation, *Water Resour. Res.*, *35*(1), 233–241.
Littlewood, I. G., B. F. W. Croke, A. J. Jakeman, and M. Sivapalan (2003), The role of 'top-down' modelling for Prediction in Ungauged Basins (PUB), *Hydrol. Processes*, *17*(8), 1673–1679.
Liu, Y., J. Freer, K. Beven, and P. Matgen (2009), Towards a limits of acceptability approach to the calibration of hydrological models: Extending observation error, *J. Hydrol.*, *367*, 93–103.
L'vovich, M. I. (1979), *World Water Resources and Their Future, translated from Russian by R. L. Nace*, 415 pp., AGU, Washington, D. C.
Lyne, V., and M. Hollick (1979), Stochastic time variable rainfall-runoff modelling, paper presented at Institute of Engineers Australia National Conference, Institute of Engineers Australia, Canberra.
Martinez, G. F., and H. V. Gupta (2010), Toward improved identification of hydrological models: A diagnostic evaluation of the "abcd" monthly water balance model for the conterminous United States, *Water Resour. Res.*, *46*, W08507, doi:10.1029/2009WR008294.
Martinez, G. F., and H. V. Gupta (2011), Hydrologic consistency as a basis for assessing complexity of monthly water balance models for the continental United States, *Water Resour. Res.*, *47*, W12540, doi:10.1029/2011WR011229.
McCuen, R. H., K. Zachary, and A. G. Cutter (2006), Evaluation of the Nash–Sutcliffe Efficiency index, *J. Hydrol. Eng.*, *11*(6), 597–602.
McDonnell, J. J., and K. Beven (2014), Debates—The future of hydrological sciences: A (common) path forward? A call to action aimed at understanding velocities, celerities and residence time distributions of the headwater hydrograph, *Water Resour. Res.*, *50*, 5342–5350, doi:10.1002/2013WR015141.

McMillan, H., M. Gueguen, E. Grimon, R. Woods, M. Clark, and D. E. Rupp (2014), Spatial variability of hydrological processes and model structure diagnostics in a 50 km^2 catchment, *Hydrol. Processes*, *28*(18), 4896–4913.
McMillan, H. K., M. P. Clark, W. B. Bowden, M. Duncan, and R. A. Woods (2011), Hydrological field data from a modeller's perspective: Part 1. Diagnostic tests for model structure, *Hydrol. Processes*, *25*, 511–522.
Milly, P. C. D. (1994), Climate, soil water storage, and the average annual water balance, *Water Resour. Res.*, *30*(7), 2143–2156.
Mockus, V. (1972), Chapter 10. Estimation of direct runoff from storm rainfall, in *SCS National Engineering Handbook, Section 4*, U.S. Dep. of Agric., SCS, Washington, D. C.
Montanari, A., and E. Toth (2007), Calibration of hydrological models in the spectral domain: An opportunity for scarcely gauged basins?, *Water Resour. Res.*, *43*, W05434, doi:10.1029/2006WR005184.
Moore, R. J. (1985), The probability-distributed principle and runoff production at point and basin scales, *Hydrol. Sci.*, *30*(2), 273–297.
Nash, J. E., and J. V. Sutcliffe (1970), River flow forecasting through conceptual models: Part 1—A discussion of principles, *J. Hydrol.*, *10*(3), 282–290.
Pokhrel, P., K. K. Yilmaz, and H. V. Gupta (2012), Multiple-criteria calibration of a distributed watershed model using spatial regularization and response signatures, *J. Hydrol.*, *418*, 49–60.
Ponce, V. M., and A. V. Shetty (1995a), A conceptual model of catchment water balance: 1. Formulation and calibration, *J. Hydrol.*, *173*(1–4), 27–40.
Ponce, V. M., and A. V. Shetty (1995b), A conceptual model of catchment water balance: 2. Application to runoff and baseflow modeling, *J. Hydrol.*, *173*(1–4), 41–50.
Reggiani, P., M. Sivapalan, and S. M. Hassanizadeh (2000), Conservation equations governing hillslope responses: Exploring the physical basis of water balance, *Water Resour. Res.*, *36*(7), 1845–1863.
Salvucci, G. D., and D. Entekhabi (1995), Hillslope and climatic controls on hydrologic fluxes, *Water Resour. Res.*, *31*(7), 1725–1739.
Sawicz, K., T. Wagener, M. Sivapalan, P. A. Troch, and G. Carrillo (2011), Catchment classification: Empirical analysis of hydrologic similarity based on catchment function in the eastern USA, *Hydrol. Earth Syst. Sci.*, *15*(9), 2895–2911.
Schaefli, B., and H. V. Gupta (2007), Do Nash values have value?, *Hydrol. Processes*, *21*(15), 2075–2080.
Seibert, J., and J. J. McDonnell (2002), On the dialog between experimentalist and modeler in catchment hydrology: Use of soft data for multicriteria model calibration, *Water Resour. Res.*, *38*(11), 1241, doi:10.1029/2001WR000978.
Shafii, M., and B. A. Tolson (2015), Optimizing hydrological consistency by incorporating hydrological signatures into model calibration objectives, *Water Resour. Res.*, *51*, 3796–3814, doi:10.1002/2014WR016520.
Shamir, E., B. Imam, H. V. Gupta, and S. Sorooshian (2005a), Application of temporal streamflow descriptors in hydrologic model parameter estimation, *Water Resour. Res.*, *41*, W06021, doi:10.1029/2004WR003409.
Shamir, E., B. Imam, E. Morin, H. V. Gupta, and S. Sorooshian (2005b), The role of hydrograph indices in parameter estimation of rainfall–runoff models, *Hydrol. Processes*, *19*(11), 2187–2207.
Sivapalan, M., M. A. Yaeger, C. J. Harman, X. Xu, and P. A. Troch (2011), Functional model of water balance variability at the catchment scale: 1. Evidence of hydrologic similarity and space-time symmetry, *Water Resour. Res.*, *47*, W02522, doi:10.1029/2010WR009568.
Sorooshian, S., and J. A. Dracup (1980), Stochastic parameter estimation procedures for hydrologie rainfall-runoff models: Correlated and heteroscedastic error cases, *Water Resour. Res.*, *16*(2), 430–442.
Tetzlaff, D., and C. Soulsby (2008), Sources of baseflow in larger catchments—Using tracers to develop a holistic understanding of runoff generation, *J. Hydrol.*, *359*(3–4), 287–302.
Thompson, S. E., C. J. Harman, P. A. Troch, P. D. Brooks, and M. Sivapalan (2011), Spatial scale dependence of ecohydrologically mediated water balance partitioning: A synthesis framework for catchment ecohydrology, *Water Resour. Res.*, *47*, W00J03, doi:10.1029/2010WR009998.
Tolson, B. A., and C. A. Shoemaker (2007), Dynamically dimensioned search algorithm for computationally efficient watershed model calibration, *Water Resour. Res.*, *43*, W01413, doi:10.1029/2005WR004723.
Troch, P. A., G. F. Martinez, V. R. N. Pauwels, M. Durcik, M. Sivapalan, C. Harman, P. D. Brooks, H. Gupta, and T. Huxman (2009), Climate and vegetation water use efficiency at catchment scales, *Hydrol. Processes*, *23*(16), 2409–2414.
Vaché, K. B., and J. J. McDonnell (2006), A process-based rejectionist framework for evaluating catchment runoff model structure, *Water Resour. Res.*, *42*, W02409, doi:10.1029/2005WR004247.
Vaché, K. B., J. J. McDonnell, and J. Bolte (2004), On the use of multiple criteria for a posteriori model rejection: Soft data to characterize model performance, *Geophys. Res. Lett.*, *31*, L21504, doi:10.1029/2004GL021577.
Viglione, A., J. Parajka, M. Rogger, J. L. Salinas, G. Laaha, M. Sivapalan, and G. Bloschl (2013), Comparative assessment of predictions in ungauged basins—Part 3: Runoff signatures in Austria, *Hydrol. Earth Syst. Sci.*, *17*(6), 2263–2279.
Vogel, R. M., and A. Sankarasubramanian (2003), Validation of a watershed model without calibration, *Water Resour. Res.*, *39*(10), 1292, doi:10.1029/2002WR001940.
Vrugt, J., and B. Robinson (2007), Improved evolutionary optimization from genetically adaptive multimethod search, *Proc. Natl. Acad. Sci. U. S. A.*, *104*(3), 708–711.
Vrugt, J. A., H. V. Gupta, W. Bouten, and S. Sorooshian (2003), A Shuffled Complex Evolution Metropolis algorithm for optimization and uncertainty assessment of hydrologic model parameters, *Water Resour. Res.*, *39*(8), 1201, doi:10.1029/2002WR001642.
Wagener, T., D. P. Boyle, M. J. Lees, H. S. Wheater, H. V. Gupta, and S. Sorooshian (2001), A framework for development and application of hydrological models, *Hydrol. Earth Syst. Sci.*, *5*(1), 13–26.
Wagener, T., M. Sivapalan, P. Troch, and R. Woods (2007), Catchment classification and hydrologic similarity, *Geogr Compass.*, *1*(4), 901–931.
Wang, D., and Y. Tang (2014), A one-parameter Budyko model for water balance captures emergent behavior in darwinian hydrologic models, *Geophys. Res. Lett.*, *41*, 4569–4577, doi:10.1002/2014GL060509.
Wang, D., J. Zhao, Y. Tang, and M. Sivapalan (2015), A thermodynamic interpretation of Budyko and L'vovich formulations of annual water balance: Proportionality Hypothesis and maximum entropy production, *Water Resour. Res.*, *51*, 3007–3016, doi:10.1002/2014WR016857.
Westerberg, I. K., J.-L. Guerrero, P. M. Younger, K. J. Beven, J. Seibert, S. Halldin, E. Freer, and C.-Y. Xu (2011), Calibration of hydrological models using flow-duration curves, *Hydrol. Earth Syst. Sci.*, *15*, 2205–2227.
Winsemius, H. C., B. Schaefli, A. Montanari, and H. H. G. Savenije (2009), On the calibration of hydrological models in ungauged basins: A framework for integrating hard and soft hydrological information, *Water Resour. Res.*, *45*, W12422, doi:10.1029/2009WR007706.
Yadav, M., T. Wagener, and H. Gupta (2007), Regionalization of constraints on expected watershed response behavior for improved predictions in ungauged basins, *Adv. Water Resour.*, *30*(8), 1756–1774.

Yang, D., F. Sun, Z. Liu, Z. Cong, G. Ni, and Z. Lei (2007), Analyzing spatial and temporal variability of annual water-energy balance in nonhumid regions of China using the Budyko hypothesis, *Water Resour. Res.*, *43*, W04426, doi:10.1029/2006WR005224.

Yen, H., R. T. Bailey, M. Arabi, M. Ahmadi, M. J. White, and J. G. Arnold (2014), The role of interior watershed processes in improving parameter estimation and performance of watershed models, *J. Environ. Qual.*, *43*(5), 1601–1613.

Yilmaz, K. K., H. V. Gupta, and T. Wagener (2008), A process-based diagnostic approach to model evaluation: Application to the NWS distributed hydrologic model, *Water Resour. Res.*, *44*, W09417, doi:10.1029/2007WR006716.

Yokoo, Y., and M. Sivapalan (2011), Towards reconstruction of the flow duration curve: Development of a conceptual framework with a physical basis, *Hydrol. Earth Syst. Sci.*, *15*(9), 2805–2819.

Zhang, L., W. R. Dawes, and G. R. Walker (2001), Response of mean annual evapotranspiration to vegetation changes at catchment scale, *Water Resour. Res.*, *37*(3), 701–708.