Water Resources Research

RESEARCH ARTICLE
10.1029/2018WR023750

Key Points:
- We characterized different facets of model equifinality and discussed them within the context of conceptual hydrological modeling
- We introduced the new model evaluation method of Flux Mapping to explore model behavior, particularly process representation
- Even within a very narrow margin of model error/performance, different modes of model response (i.e., internal flux dynamics) can be equally active

Supporting Information:
- Supporting Information S1

Correspondence to:
S. Khatami,
sina.khatami@unimelb.edu.au

Citation:
Khatami, S., Peel, M. C., Peterson, T. J., & Western, A. W. (2019). Equifinality and flux mapping: A new approach to model evaluation and process representation under uncertainty. *Water Resources Research*, *55*. https://doi.org/10.1029/2018WR023750

Received 23 JUL 2018
Accepted 18 JUN 2019
Accepted article online 26 JUN 2019

Equifinality and Flux Mapping: A New Approach to Model Evaluation and Process Representation Under Uncertainty

Sina Khatami[1], Murray C. Peel[1], Tim J. Peterson[1], and Andrew W. Western[1]

[1]Department of Infrastructure Engineering, University of Melbourne, Parkville, Victoria, Australia

Abstract Uncertainty analysis is an integral part of any scientific modeling, particularly within the domain of hydrological sciences given the various types and sources of uncertainty. At the center of uncertainty rests the concept of *equifinality*, that is, reaching a given endpoint (*finality*) through different pathways. The operational definition of equifinality in hydrological modeling is that various model structures and/or parameter sets (i.e., *equal pathways*) are equally capable of reproducing a similar (not necessarily identical) hydrological outcome (i.e., finality). Here we argue that there is more to model equifinality than model structures/parameters, that is, other model components can give rise to model equifinality and/or could be used to explore equifinality within model space. We identified six facets of model equifinality, namely, model structure, parameters, performance metrics, initial and boundary conditions, inputs, and internal fluxes. Focusing on model internal fluxes, we developed a methodology called *flux mapping* that has fundamental implications in understanding and evaluating model process representation within the paradigm of multiple working hypotheses. To illustrate this, we examine the equifinality of runoff fluxes of a conceptual rainfall-runoff model for a number of different Australian catchments. We demonstrate how flux maps can give new insights into the model behavior that cannot be captured by conventional model evaluation methods. We discuss the advantages of flux space, as a subspace of the model space not usually examined, over parameter space. We further discuss the utility of flux mapping in hypothesis generation and testing, extendable to any field of scientific modeling of open complex systems under uncertainty.

1. Introduction

Understanding, modeling, and predicting hydrological systems—realistically and viably—is the Holy Grail of hydrological sciences. There are barriers in this quest, particularly in a world undergoing rapid and large-scale changes (Peel & Blöschl, 2011). Among numerous difficulties with modeling and prediction of real-world hydrological processes are the issues of scale (Blöschl & Sivapalan, 1995) and commensurability (Beven, 2012b, p. 245) between observed and modeled variables; dependency upon the quantity (Boughton, 2007) and quality (Beven & Westerberg, 2011; Yew Gan et al., 1997) of available data and their information content (Nearing & Gupta, 2015); model complexity (Perrin et al., 2001; Yew Gan et al., 1997); the chaotic nature of many hydrological processes (Khatami, 2013a, 2013b; Sivakumar, 2000; Sivakumar et al., 2001); modeling hydrological responses to change (Schaefli et al., 2011) and resilience to disturbance (Peterson et al., 2014; Peterson & Western, 2014); *numerical daemons* of conceptual hydrological modeling (Clark & Kavetski, 2010; Kavetski & Clark, 2011); and the ill conditionedness or ill posedness of environmental models (Beck, 1987; Yeh, 1986). The latter is also referred to as *equifinality* (Beck, 2002; Beven, 2006; Ebel & Loague, 2006; Kelleher et al., 2017).

Beven (1975), p. 14) first used the term equifinality in the domain of hydrological modeling. Later, Beven (1993) proposed a concept of equifinality for model evaluation and uncertainty analysis. Based on his suggested concept, the operational definition of equifinality is that different model structures and/or parameter sets (i.e., *equal pathways*) are equally capable of reproducing a similar (not necessarily identical) hydrological outcome (i.e., *finality*). For example, in the case of rainfall-runoff modeling, for a given criterion of reproducibility (e.g., an objective function), a number of rainfall-runoff models (i.e., distributed or lumped, and process based or black box such as TOPMODEL, HBV, and Sacramento) and/or various parameter sets might be able to equally reproduce a particular observed runoff. This operational definition of equifinality is closely related to structural and parameter uncertainty and is the cornerstone of sensitivity and uncertainty estimation frameworks such as generalized sensitivity (Hornberger & Spear, 1981) and generalized likelihood uncertainty estimation (Beven & Binley, 1992).

In this paper we argue that equifinality—like uncertainty—is a multifaceted concept, and various model components other than model structure and parameters could also give rise to model equifinality. We first, briefly outline various facets of model equifinality (section 2), namely, equifinality of model structures and/or parameters, objective functions (or model performance metrics), model initial/boundary conditions, model inputs, and model internal fluxes. We should mention that there are other facets of equifinality than model equifinality that we discussed under a comprehensive theoretical framework of scientific inquiry and modeling of hydrological systems under uncertainty. While facets of model equifinality are not mutually exclusive and in fact are intertwined, each facet underscores a particular aspect (subspace) of the overall model space. We develop a new model evaluation scheme, called *flux mapping* (section 3), to examine the degree of equifinality of model internal fluxes and to explore and characterize model process representation. Using a modeling experiment (section 3.4), we demonstrate how flux mapping—analyzing model equifinality through the lens of model internal fluxes instead of model parameters—provides new insights into model internal behavior and process representation, which cannot be (easily) captured/characterized using conventional model evaluation schemes (e.g., objective functions, dotty plots, and parameter distributions; sections 4 and 5.1). In other words, reprojecting model behavior (e.g., response surface) from parameter space to the flux space can give new insights into model internal behavior that are not inferable from parameter space. To this end, we showcase and discuss the results of flux mapping for a number of Australian catchments (section 4).

Flux mapping is an approach to generate and explore multiple working hypotheses (MWH) based on model internal behavior and process representation. Chamberlin (1890) argued for the paradigm of MWH in scientific inquiries as this paradigm is more robust to reduce bias (i.e., assure impartiality) toward a particular hypothesis for explaining a given phenomenon. Theoretically, to explain real-world processes, MWH is a never-ending process within which *hypotheses*, that is, a set of plausible explanations of real-world phenomenon, are generated, evaluated, revised/refined, and further evaluated with the hope that our refined hypotheses converge toward an approximation of the actual reality. The value and significance of pursuing MWH in hydrological modeling is discussed in the literature (Beven, 2012a; Beven et al., 2012; Buytaert & Beven, 2011; Clark et al., 2011, 2012). We further discuss the utility and exploratory power of flux mapping in hypothesis forming/testing and process understanding (section 5.2). Flux mapping is extendable beyond hydrological modeling to any field of scientific modeling dealing with conceptual modeling of open complex systems under uncertainty.

2. Facets of Model Equifinality: Theoretical Discussion

Throughout the hydrological literature when the term equifinality is used, it is predominantly referring to model equifinality; different model structures and/or parameter sets could produce a similar outcome given some available (uncertain) observations and a particular (incomplete) metric of acceptability (e.g., model performance above a subjective value of one/multiple objective functions). So model equifinality is conditional on the model configuration, performance metric(s), and the information content of the data used. Beven (1975), p. 14) was the first to use the term equifinality in hydrology, and later Beven (1993) discussed its implication in hydrological modeling in terms of multiple acceptable model structures and/or parameter sets as a preferred alternative to the notion of a single optimum parameter set. There also has been other studies that referred to model equifinality using other terms such as ambiguity, identifiability, empirical equivalence, nonuniqueness, underdetermination or indeterminacy, and system convergence (e.g., Beck, 1987; Bethke, 1992; Carrera & Neuman, 1986; Gupta & Sorooshian, 1983; Hornberger & Spear, 1981; Konikow & Bredehoeft, 1992; Oreskes et al., 1994; Quine, 1975; Sorooshian & Gupta, 1983; Yeh, 1986).

The question of model equifinality is often reduced to model parameter equifinality as parameter uncertainty expressed in probabilistic terms (i.e., parameter distribution), although there are other studies that attempted to take other sources of uncertainty into account such as model inputs (e.g., Blazkova & Beven, 2009; Haydon & Deletic, 2009; Kavetski et al., 2006; Liu et al., 2009; Vrugt et al., 2008) or structural uncertainty (Ajami et al., 2007; Bulygina & Gupta, 2009, 2010, 2011; Butts et al., 2004; Renard et al., 2010). In this section, we

10.1029/2018WR023750

make the case that there is more to model equifinality than parameter uncertainty (distribution) by characterizing six different, yet interconnected, facets of model equifinality.

2.1. Equifinality of Model Structures

Multiple model structures with different degrees of complexity (i.e., number of model parameters, fluxes, and/or other components) that are almost equally capable of reproducing a hydrologic behavior (e.g., discharge hydrograph) could be seen as MWH, with each model structure representing catchment behavior differently and hence equifinality of model structures. Various hydrological modeling frameworks have been developed arguably based on this very facet of model equifinality, whether or not this facet was explicitly acknowledged, including, but not limited to, SUMMA (Clark et al., 2015a; Clark et al., 2015b), FUSE (Clark et al., 2008), and SUPERFLEX (Fenicia et al., 2011; Kavetski & Fenicia, 2011). Evaluating the *realism* of model structures (process representation) is fundamentally difficult, regardless of the number of models utilized and their (dis)agreement, as not all catchment internal processes are known or observed even at the scale of interest. Developing/choosing the model structure is majorly dependent upon the personal judgments and preferences of modelers (Addor & Melsen, 2019; Holländer et al., 2009) and influenced by politics (Heymann & Dalmedico, 2019). Thus, there is a "problem of decidability" (Beven, 2006) between feasible representations of the real world, that is, which conceptual model fits better to our perceptual model. In fact, the choice of model structure, like other subjective decisions in modeling, for example, the choice of objective function (Crochemore et al., 2015; discussed below), is often an act of *will* (i.e., modeler's personal or institutional preference) rather than *rationality/objectivity* (i.e., model adequacy or fit for purpose). For instance, Addor and Melsen (2019) demonstrated that in most cases that they investigated, the affiliation of the first author was a clear predictor of model selection, while the role model adequacy given the research objectives was less clear.

It is worth mentioning that model-structure-equifinality could be seen as a special case of equifinality of modeling approaches. Different modeling approaches, such as a top-down versus bottom-up, process-based versus black box, and distributed versus lumped, may lead to similar results in a given modeling case.

2.2. Equifinality of Model Parameters

This is the most widely studied facet of equifinality in hydrology (Arsenault & Brissette, 2014; Beven & Binley, 2014; Kelleher et al., 2015; Kelleher et al., 2017; Kirchner, 2016; Tang & Zhuang, 2008; Teweldebrhan et al., 2018; Vrugt & Beven, 2018). The operational definition of equifinality in hydrological literature is in fact parameter equifinality. Equifinal parameters are uncertain. Within the hydrological literature parameter equifinality and uncertainty are treated similarly and interchangeably. In simple terms, parameter uncertainty means that there is no certain/true parameter set and it is conventionally represented probabilistically as parameter distributions (commonly presented as marginal distributions). There are multiple acceptable/working parameter sets, that is, equifinal parameters, within the larger set of all uncertain parameters. Parameter uncertainty is generally addressed by searching for multiple acceptable parameter sets (e.g., set theoretic approaches and Monte Carlo experiments) given single/multiple measures of model performance. There are different ways to address parameter uncertainty, for example, Bayesian approach where different degrees of belief are assigned to the sampled parameter sets and approaches where parameter sets below a certain threshold of acceptability or outside particular limits of acceptability are rejected (e.g., generalized likelihood uncertainty estimation; Beven, 2009; Vrugt & Beven, 2018). Regardless of the approach, parameter uncertainty is typically then expressed in terms of likelihood, that is, a parameter distribution.

In the above, parameter equifinality is determined based on the model performance—that is, the value of objective (or likelihood) function(s)—and the physical significance/plausibility of the so-called equifinal/behavioral parameter sets is often not examined, while, typically expressed probabilistically, parameter equifinality could be represented in other nonprobabilistic forms (see section 4 for more details). It should be further noted that it is difficult to draw a sharp dividing line between model structure and parameter equifinality, as the two are intertwined. Model structures (e.g., equations of model fluxes and storages) can be dependent upon the parameter values (driven by the input data), and parameter values change the function of a model component, sometimes drastically. For instance, for the SIMHYD model used in this study (see Figure 5a), parameter *INSC* (interception store capacity) can vary between 0 and 20. $INSC = 0$

means the interception storage of the model is nonexistent and hence a significant change in process description/representation within the model structure. Similarly, K (baseflow recession parameter) can vary between 0 and 1. $K = 0$ means a nonexistent flux otherwise a linear flux equation with slopes decreasing from 1 (i.e., $K = 1$) asymptotic to the x axis (horizontal line at 0) as $K \rightarrow 0$. That is, each of these parameter values leads to an effectively different model structure.

2.3. Equifinality of Model Performance Metrics (or Objective Functions)

Objective functions—both their choice and function—are integral parts of the *modeling process*. For instance, the model output or response surface is the product of the interplay between model structure, parameters, objective function, data information content, and modeler's decisions. Objective functions characterize the model performance as an aggregated measure of the matching between modeled and observed; either as metrics of model residuals (Bennett et al., 2013; Davtalab et al., 2017; Fowler et al., 2018; Murphy, 1988) or as signatures of similarity (Addor et al., 2018; Fowler et al., 2016; Gupta et al., 2008; Kelleher et al., 2017; Pfannerstill et al., 2014; Sawicz et al., 2014; Schaefli, 2016; Yilmaz et al., 2008); whether a scalar metric/variable (single criterion) or a vector of metrics/variables (i.e., multiple criteria/multivariable; Efstratiadis & Koutsoyiannis, 2010; Gupta et al., 1998; Stisen et al., 2018); and whether aggregated or distributed (Koch et al., 2016; Koch et al., 2017). Performance metrics reduce the complex behavior of a system—often the integrated response of the catchment system, that is, discharge—from a higher dimension (e.g., a time series) to a single, or a few, point values; thus information loss is inevitable (Gong et al., 2013; Gupta & Nearing, 2014; Nearing & Gupta, 2015). Such aggregations, similar to the averaging process discussed by Savenije (2001), give rise to equifinality. That is, a similar model error (i.e., distance between the model output and observed behavior, e.g., discharge hydrograph) could be the result of different objective functions with different mathematical structures. Although it is possible to improve metrics, for example, by benchmarking (Schaefli & Gupta, 2007; Seibert, 2001) or reformulation (Chiew et al., 1993; Gupta et al., 2009; Legates & McCabe, 1999; Pool et al., 2018; Willmott et al., 2012), all metrics (whether single/vector or error-based/signature) have limitations and deficiencies (Pushpalatha et al., 2012; Santos et al., 2018; Westerberg et al., 2016; Westerberg & McMillan, 2015). The problem of metric equifinality will not be eliminated by developing more sophisticated metrics. There is no ultimate (set of) objective function(s), as all metrics are “underdetermined” (i.e., “do not describe unique error characteristics, even when many of them are used collectively” (Tian et al., 2016)). Tian et al. (2016) demonstrated how identical values of conventional metrics and their derivatives—for example, bias, correlation coefficient, and mean square error—can be achieved from vastly different time series.

2.4. Equifinality of Model Initial/Boundary Conditions

Given the unknowability of historical/future initial and boundary conditions in almost all cases (epistemic uncertainty), they are a source of model equifinality. That is, different initial/boundary conditions can lead to similar results. Ebel and Loague (2006) simulated five scenarios of different initial/boundary conditions (e.g., soil water content and permeability characteristic functions) for a distributed model of an experimental catchment, with NSE (Nash-Sutcliffe model efficiency (Nash & Sutcliffe, 1970) values between 0.66 to 0.82 for the discharge. That is, different initial/boundary conditions could lead to reasonably acceptable model performance for the discharge. To see through this “fog of equifinality” of models' discharge performances, they further compared the simulated and observed pressure head at three locations and found that the associated NSE values were all negative except for the scenario with discharge NSE of 0.76. They attempted to constrain the model equifinality and improve the model realism by looking at variables other than discharge, that is, introducing additional information. There can be many other realistic scenarios with different initial/boundary conditions (Pappenberger et al., 2006), and these scenarios may fail to simulate other catchment processes if new data are introduced.

2.5. Equifinality of Model Inputs

Different input variables with varying degrees of information content (i.e., different types, quantities, and qualities of inputs) could lead to similar model outcome, for example, equifinality of model predictions from different stochastic realizations of the input data (Zin, 2002) such as rainfall input (Ehlers et al., 2018). Newman et al. (2015) developed an ensemble of gridded observation-based daily precipitation and temperature for 1980–2012 for the contiguous United States, which could be used to account for uncertainty of gridded product and model forcing, as well as exploring the equifinality of model inputs. As another example, Oudin et al. (2005) investigated the use of different potential evapotranspiration (PET) inputs to four different rainfall-runoff models and found no systematic improvements in the calibrated model performance when using daily temporally varying PET instead of seasonal mean PET. This instance of model input equifinality may also be related to the insensitivity of the model (process representation) to input information content.

2.6. Equifinality of Model Internal Fluxes

Various combinations of model internal fluxes can lead to similar model output. Interception, evapotranspiration (ET), and runoff fluxes are examples of internal fluxes in the case of a conceptual rainfall-runoff model. These fluxes are essentially representations of real-world processes, for example, model runoff fluxes mimic catchment runoff generation mechanisms. Also, under a given model conditioning, that is, available data and model performance metrics, different routines for calculating internal fluxes (e.g., interception) can be equifinal. Grayson et al. (1992) demonstrated that a given observed hydrograph could be equally reproduced through Hortonian overland flow or saturated area runoff with very different distributed flow characteristics. Physical significance is a distinct characteristic of flux equifinality, over other facets of model equifinality like parameter equifinality, which is desirable for generating/testing hypotheses.

It should be noted that these six facets of model equifinality are not mutually exclusive, as different model components are intertwined as discussed above. That said, each facet accentuates a particular aspect (subspace) of the overall model space. That is, each facet (or their combinations) could be utilized as a way of generating model simulations or to investigate an ensemble of model runs in terms of MWH. In the next section, we demonstrate how replacing the emphasis from the equifinality of model parameters to model internal fluxes gives new insights into model behavior and to generate MWH, even if only model parameters are perturbed.

3. Flux Mapping: An Approach to Evaluate Model Behavior

In this section we develop a method called flux mapping for evaluating model behavior, based on the concept of model internal flux equifinality. We demonstrate that the new tool of *flux maps* can give new insights into model behavior that are not inferable from conventional model evaluation tools of dotty plots (i.e., projection of points on a model response or likelihood surface onto a single parameter axis (Beven, 2006)) or statistical parameter distributions.

3.1. Hydrological Model

For this study, we chose SIMHYD, a lumped conceptual daily rainfall-runoff model (Chiew et al., 2002; Peel et al., 2000). It has seven parameters, takes precipitation and areal PET (APET) as inputs, and generates streamflow as the output. SIMHYD incorporates runoff generating mechanisms, namely, infiltration excess (INFexc), interflow and saturation excess overland flows (INT & SATexc), and baseflow (BAS; Figure 5a). With three runoff generating mechanisms it is a suitable choice to examine the equifinality of internal (runoff) fluxes.

3.2. Study Area and Data Set

The Australian Network of Hydrologic Reference Stations (HRS) is a set of 222 catchments (http://www.bom.gov.au/water/hrs/) with minimal land use disturbances and water resource development and relatively high-quality data (Turner, 2012, p. 6) composed of daily time series of observed streamflow (Q). Fowler et al. (2016) calculated areal average precipitation (P) from Australian Water Availability Project (www.bom.gov.au/jsp/awap/) daily 5-km grids (Jones et al., 2009) and also estimated APET at the catchment centroid using Morton's Wet Environment method (Morton, 1983) using gridded estimates of Jeffrey et al., 2001; see Fowler et al., 2016, for further details of data set preparation).

The modeling experiment (flux mapping) is conducted on a subset of HRS catchments with a high level of SIMHYD performance, here defined as $NSE \geq 0.75$. To select them, SIMHYD was first calibrated to all HRS sites over their total streamflow record, using the global optimization algorithm of Shuffled Complex Evolution (Duan et al., 1992). For each case, the Shuffled Complex Evolution routine was repeated 20 times to ensure consistency in the calibration results. The highest NSE value from the 20 repeats for each catchment-model pair was selected as the upper bound of possible model performance (hereinafter SCE-NSE). A subset of the top 53 catchments, $0.75 \leq \text{NSE} \leq 0.88$, were selected. It should be stressed that this precalibration step is not a part of the modeling experiment and only serves to find the upper bound of model performance as a rough measure of sampling sufficiency (explained in the following section).

3.3. Flux Mapping

Figure 1a presents a conceptual example indicating that a model (here SIMHYD) can simulate an observed hydrograph through different combinations of model internal fluxes (here proportions of model runoff fluxes), leading to similar NSE values ($0.803 \leq \text{NSE} \leq 0.825$) at a given catchment. It clearly shows that the value of the objective function is not a reliable measure of model internal behavior. For a large ensemble of model runs, equifinal fluxes can be summarized and visualized based on the percentage of their (volumetric) contribution to the total simulated Q on a plot we name a flux map. Figure 1b is a conceptual example of flux map and is used to inspect the runoff flux space, a subset of the model flux space, of SIMHYD model by mapping the three runoff fluxes (i.e., INFexc, INT & SATexc, and BAS) for three equifinal model runs demonstrated on Figure 1a.

A flux map is a ternary plot where each dimension represents a model runoff flux, and each model run is projected as a single point based on the proportions of its equifinal runoff fluxes to the total simulated Q. The cloud pattern can vary from very constrained (Figures 3a, one flux is dominating the simulation of total Q), through intermediated cases (Figures 3b and 3c, two fluxes are codominating the simulation of total Q with possibly small contribution of the third flux), up to filling the entire plausible flux space (i.e., the entire triangle, Figures 3d and 4). Thus, the point cloud on the flux maps is an expression of the model flux equifinality; filling a larger space on the flux map indicates higher degrees of model flux equifinality. In the case of two fluxes, the plausible flux space would shrink to a line showing the interplay between the contribution of either flux. In case of more than three fluxes, the flux map can be simply presented as a series of two-by-two scatterplots.

3.4. Experiment Design

The flux space of each catchment model is explored using simulations of 10^6 parameter sets sampled from a uniform parameter (prior) distribution using Latin Hypercube Sampling (LHS). The sample size was determined by comparing the difference between SCE-NSE (the best NSE value from 20 repeats of precalibration) and the highest NSE value from the LHS ensemble (hereinafter Ensemble-NSE) at a few trial catchments. Selecting 10^6 samples generally led to the Ensemble-NSE being within 3% of the SCE-NSE, suggesting that 10^6 samples was sufficient to adequately explore the parameter space and the consequent flux space. It should be mentioned that for higher-dimensional parameter spaces, LHS or random sampling are very inefficient and often insufficient; efficient searching/sampling strategies should be used instead such as dynamically dimensioned search (Tolson & Shoemaker, 2008).

To track the emerging pattern of the point cloud of the flux maps, model runs are evaluated against a set of thresholds of equifinality/acceptability, defined as SCE-NSE $\times$ {0.99, 0.98, 0.97, 0.96, 0.95, 0.90, 0.85, 0.80}, that is, gradually relaxing the threshold—hereinafter referred to as *thresholding*. Equifinal model runs above a given threshold are considered as acceptable and the rest as unacceptable.

4. Results

The overall model performance in the example catchments is $0.75 \leq$ SCE-NSE ≤ 0.80 and for the extreme case is SCE-NSE = 0.82. Yet as shown in Figures 3 and 4 the flux maps are vastly different. We chose four classes of general flux map behavior (Classes I to IV), within the 53 catchments studied, for further discussion. For each class, we present flux maps of an example catchment for the *strict* (0.95 $\times$ SCE-NSE) and *relaxed* (0.85 $\times$ SCE-NSE) thresholds (Figures 3 and 4), color coded based on the corresponding Ensemble-NSE values. For Class IV, an additional example of an extreme case is also presented (Figure 4). Flux maps of other thresholds are presented in the supporting information (Table S1) to demonstrate the emerging pattern of the flux maps. Also, the corresponding two-by-two flux maps of Figures 3 and 4 are available (see Khatami et al., 2017). To the extent possible, the four examples have similar hydrological characteristics, namely, catchment area of 125–170 km^2, mean annual 950–1353 mm, mean annual Q = 162–335 mm, mean annual $APET$ = 1,222–1,532 mm, and annual runoff coefficient 0.17–0.25. The

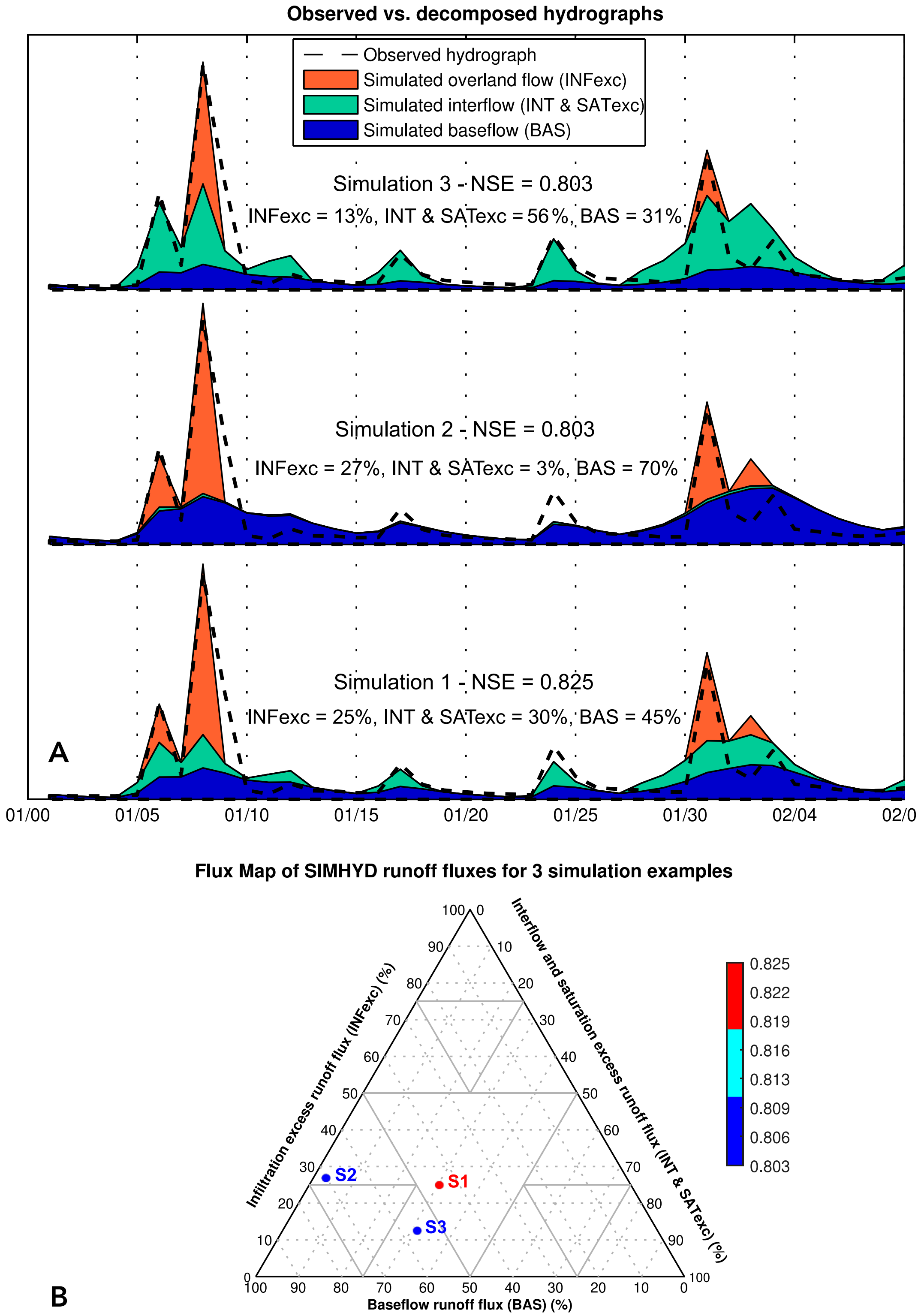

Figure 1. (a) Three simulations (S1–S3) of an observed hydrograph through different combinations of runoff generating mechanisms (volumetric contribution to the total simulated runoff) summarized on a flux map (b), color coded based on the simulation performance (NSE value). The triangle (*B*) represents the plausible flux space for a model with three runoff fluxes. NSE = Nash-Sutcliffe model efficiency.

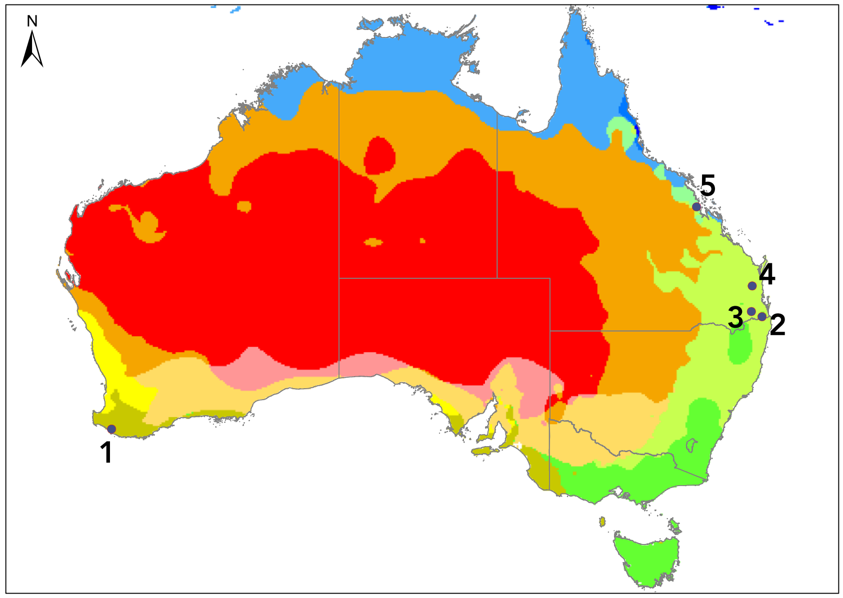

Figure 2. Map of the Australian example catchments for the modeling experiment of flux mapping. The color scheme is based on the Köppen-Geiger climate classification by Peel et al. (2007).

extreme case of Class IV flux maps is quite similar to the other four examples except for a smaller catchment area, 83 km^2, and a higher runoff coefficient of 0.44. Figure 2 presents the location of example catchments based on the Köppen-Geiger climate classification (Peel et al., 2007), and catchment summaries are presented in Table 1. Specific simulation characteristics of each example are mentioned in the corresponding figure caption.

4.1. Class I

The point cloud on the flux map in Figures 3a1 and 3a2 is very constrained indicating a very low degree of flux equifinality, a baseflow-dominant runoff simulation, having zero infiltration excess runoff and a low saturation excess runoff in the simulations. For the strict threshold (Figure 3a1) there are 252 equifinal model runs, which all exhibit a very narrow range of possible flux contributions to total simulated Q. As the equifinality threshold lowers to the relaxed threshold (Figure 3a2) the number of equifinal runs increases, yet the general pattern of the flux map remains, i.e., predominantly baseflow oriented, with the possibility of up to 20% contribution from INT & SATexc.

Table 1
Catchments Summaries of the Examples Presented in Modeling Experiments

		Catchment						
Catchment no.	Corresponding figures (class)	Name	Location	Area (km^2)	Mean annual total P (mm)	Mean annual total Q (mm)	Mean annual $APET$ (mm)	Annual runoff ratio
1	Figures 3a1 & 3a2 (I)	Dombakup Brook at Malimup Track	Western Australia (115.98°E 34.58°S)	125.09	1,130.99	232.63	1,222.41	0.21
2	Figures 3b1 & 3b2 (II)	Albert River at Lumeah	Queensland (153.05°E 28.05°S)	167.39	1,353.61	335.14	1,451.17	0.25
3	Figures 3c1 & 3c2 (III)	Bremer River at Adams Bridge	Queensland (152.51°E 27.83°S)	126.09	951.07	162.19	1,506.55	0.17
4	Figures 3d1 & 3d2 (IV)	Kandanga Creek at Hygait	Queensland (152.65°E 26.39°S)	170.78	1,135.18	277.98	1,532.49	0.24
5	Figures 4a, 4b, & 4c (IV)	Carmila Creek at Carmila	Queensland (149.399° E 21.915°S)	83.8	1,275	538	1,736	0.42

Model runs with similar flux contributions but distinct flux dynamics (e.g., magnitude, shape, and sequencing of events) would be mapped as identical points on the flux map. Therefore, even a very constrained flux map might be unfolded to a number of model runs with distinct dynamics yet similar volumetric contribution of the fluxes. In other words, flux maps are in fact underestimating the flux equifinality.

4.2. Class II

The flux map for this class (Figures 3b1 and 3b2) is mainly constrained around a line that corresponds to the complementary possible contributions of the BAS and INT & SATexc runoff fluxes. The range of the line varies depending on the thresholding. For the strict threshold (Figure 3b1), the baseflow contribution to total runoff volume is around 45–90%, and as the threshold relaxes (Figure 3b2), baseflow contributes 15–100% to the total simulated flow. SIMHYD exhibits less than 5% contribution from INFexc at the relaxed threshold.

It should be noted that in producing flux maps, model runs with different performance but similar flux contributions are plotted on top of each other, with higher performing points (toward red) plotted on top. So, the marginal distributions of flux maps are also provided (supporting information Table S1). Regardless of the threshold, even the highest performing points (i.e., red ones in the cloud) are moderately spread across the two contributing fluxes. Color coding and marginal distributions of flux maps are not suggested as a way for identifying *optimal* or *high likelihood* regions on the flux maps. One cannot infer that model runs with higher performance values (red points on the flux maps) are more realistic. Model performance—particularly assessed in a scalar sense—is a weak, unreliable, and unrealistic measure for model evaluation, as model process representation cannot be measured with a single (or few) value(s) of performance metrics (for further details also see the discussion on hydrological signatures (Addor et al., 2018; Euser et al., 2013; Gupta et al., 2008; Kelleher et al., 2017; Schaefli, 2016; Yilmaz et al., 2008)). As demonstrated by Ebel and Loague (2006) and Seibert and McDonnell (2002) in detail, strict interpretation of objective functions is misleading, as model runs with slightly lower values of NSE might have better process representation than the higher ones. Model performance does not imply realism and may be a numerical artifact given various sources of uncertainty. Insufficient knowledge of catchment processes (e.g., runoff-generating mechanisms and the details of the sequencing of different storm and runoff events) makes it difficult to assign likelihood to model runs. Marginal distributions (Table S1) and color coding of the flux maps only serve a demonstrative purpose (the spread of equifinal fluxes on the flux map) and not a prescriptive one (necessarily indicating the realism of model runs). All the cloud points are equifinal by definition, that is, possible flux contributions given the modeling setup and equifinality threshold, unless/until additional data are used to reject or constrain them, which argues against a probability- or likelihood-based interpretation of the flux maps.

4.3. Class III

This class (Figures 3c1 and c2) is an extension of Class II to three active fluxes, that is, all three runoff fluxes can make noticeable contributions to simulating the observed flow, albeit with INFexc being smaller than the other two. For the strict case of acceptability (Figure 3c1), $0.74 \leq$ Ensemble-NSE ≤ 0.78, there exists an extreme model run with almost 95% contribution form INT & SATexc and another with almost 80% contribution from BAS. INFexc contributes up to around 15%. For the relaxed threshold (Figure 3c2), the range of all fluxes increases, most notably INFexc, which is the least constrained (being <15% at the strict threshold). For both cases of thresholding, particularly for the relaxed one, the flux contribution of high-performing model runs (red points) are widely spread on the flux map, indicating a high degree of flux equifinality even for the highest performing model runs.

4.4. Class IV

This class (Figures 3d) is an extension of Class III, that is, all three fluxes can vary widely. Even for the strict threshold, there are 1,270 acceptable model runs with a moderately wide range (≥50% variation) of possible flux contributions as shown in Figure 3d1 (compare this case with Figure 3b1). A key difference to the other classes is that INFexc is as variable as the other fluxes. In the case of the relaxed threshold, the cloud fills almost the entire space of the flux map (lower triangle), signifying a very high degree of flux equifinality. All three runoff fluxes vary around at least 90% of flux contributions. In other words, almost any combination of runoff flux (volumetric) contribution could lead to a similar model performance—all flux combinations are plausible/feasible.

10.1029/2018WR023750

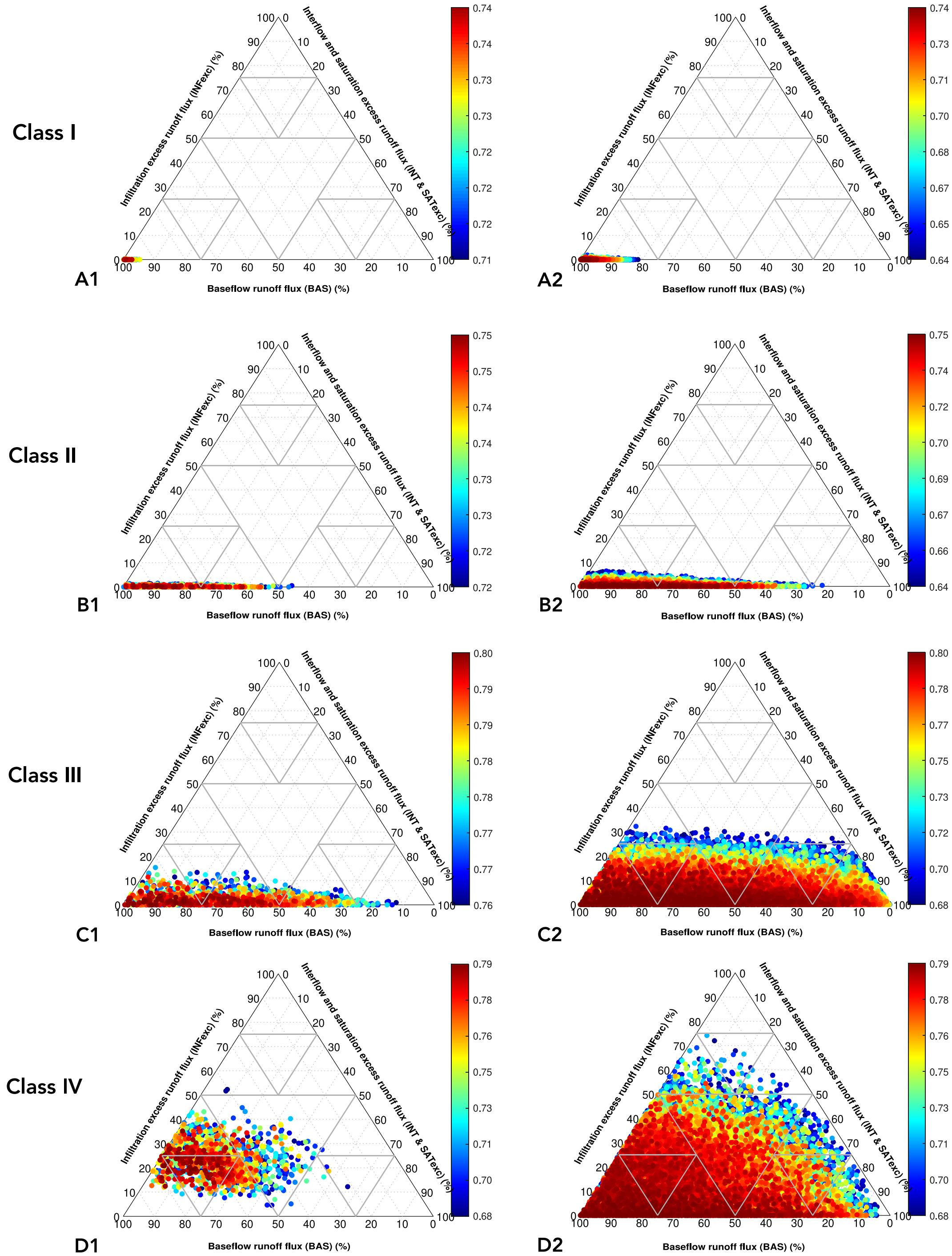

Figure 3. This figure represents how different model internal behavior, represented on flux maps, can emerge within a certain range of model performance. Flux maps for four different Australian catchments (catchments 1–4 on Table 1) for two different thresholding; each row presents a different catchment with strict threshold on the first column and relaxed threshold on the second column. Color bar represents the model performance in terms of Nash-Sutcliffe model efficiency value. (a) Class I flux maps for strict threshold ($0.95 \times$ SCE-NSE $= 0.71 \leq$ Ensemble-NSE $= 0.75 \leq$ SCE-NSE $= 0.75$) with 252 equifinal simulations (a1) and relaxed threshold ($0.85 \times$ SCE-NSE $= 0.64 \leq$ Ensemble-NSE $\leq$ SCE-NSE) with 3,036 equifinal simulations (a2). (b) Class II flux maps for strict threshold ($0.95 \times$ SCE-NSE $= 0.71 \leq$ Ensemble-NSE $= 0.75 \leq$ SCE-NSE $= 0.75$) with 592 equifinal simulations (b1) and relaxed threshold ($0.85 \times$ SCE-NSE $= 0.64 \leq$ Ensemble-NSE $\leq$ SCE-NSE) with 8,476 equifinal simulations (b2). (c) Class III flux maps for strict threshold ($0.95 \times$ SCE-NSE $= 0.74 \leq$ Ensemble-NSE $= 0.78 \leq$ SCE-NSE $= 0.78$) with 3,644 equifinal simulations (c1) and relaxed threshold ($0.85 \times$ SCE-NSE $= 0.66 \leq$ Ensemble-NSE $\leq$ SCE-NSE) with 47,702 equifinal simulations (c2). (d) Class IV flux maps for strict threshold ($0.95 \times$ SCE-NSE $= 0.71 \leq$ Ensemble-NSE $= 0.75 \leq$ SCE-NSE $= 0.75$) with 1,270 equifinal simulations (d1) and relaxed threshold ($0.85 \times$ SCE-NSE $= 0.64 \leq$ Ensemble-NSE $\leq$ SCE-NSE) with 65,126 equifinal simulations (d2). Catchment summaries are presented in Table 1.

10.1029/2018WR023750

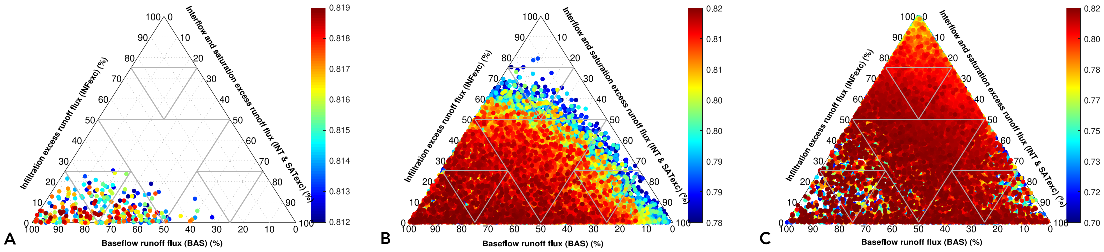

Figure 4. Class IV flux maps for very strict threshold (0.99 × SCE-NSE = 0.80 ≤ Ensemble-NSE = 0.80 ≤ SCE-NSE) with 287 equifinal simulations (a), strict threshold (0.95 × SCE-NSE = 0.77 ≤ Ensemble-NSE ≤ SCE-NSE) with 69,090 equifinal simulations (b), and relaxed threshold (0.85 × SCE-NSE = 0.69 ≤ Ensemble-NSE ≤ SCE-NSE) with 49,591 equifinal simulations only from 10^5 simulations (c). Catchment summaries are presented in Table 1.

To further illustrate Class IV, Figure 4 presents an extreme case. Even for a threshold as strict as 0.99 × SCE-NSE and as few as 287 equifinal runs (which is close to the number of equifinal runs in the case of strict thresholding in Class I with a constrained flux map, Figure 3a1), acceptable fluxes are remarkably scattered—occupying about 25% of the flux map plane, shown in Figure 4a. Even within 1% of SCE-NSE, significant differences in the dynamics of model runoff fluxes emerge. The flux map is almost space filled even for the strict threshold, with more than 69,000 equifinal model runs (Figure 4b). This shows an extraordinary degree of flux equifinality compared with previous cases. For such a degree of flux equifinality, even 10^5 thousand model runs were enough to achieve a space-filling flux map at the relaxed threshold (Figure 4c). Given that the SCE-NSE for this catchment is higher than the previous cases (0.81), one cannot simply associate the flux equifinality with the model performance.

5. Discussion

5.1. Flux Mapping and Model Evaluation: Flux Space Versus Parameter Space

A striking dimension of the results is the wide range of model internal dynamics that emerges from a similar level of model performance for catchments with a reasonably constrained range of physiographic characteristics, regardless of the thresholding and number of equifinal model runs. The different patterns/degrees of flux equifinality range from very constrained to almost space-filling flux maps. Such new insights into model behavior that flux mapping provides cannot (easily) be characterized by common model evaluation tools, like dotty plots and parameter distributions. The dotty plot is a tool to visualize model parameter equifinality and (in)sensitivity, and the statistical distribution of parameter values (commonly presented as marginal distributions) is a probability-based tool for expressing model parameter uncertainty. While dotty plots and parameter distributions are useful tools, they do not sufficiently characterize the model behavior, particularly its internal dynamics and process representation; hence, the flux map is a valuable complement to these tools. For instance, the differences in model behaviors for the aforementioned cases I–IV are not discernible from the marginal distribution of model parameters (Figure 5) nor the dotty plots (supporting information Table S1).

Figure 5 presents the marginal distributions of model parameters corresponding to the example catchments in Figure 3. Comparing Figures 3 and 5 shows that despite vastly different flux maps, the marginal distributions of the influencing parameters can be similar, and therefore, it would be difficult to translate the changes in distribution into flux dynamics. For instance, parameter distributions of both Classes III and IV (Figure 5b) are almost identical except for parameter SUB, which controls the INT & SATexc flux, while Figure 3 clearly shows that there are profound differences between the flux maps of these classes.

To explore more complex interactions of model parameters and their implication on model fluxes, we discuss further a few more cases. While given the model structure, some parameters are not influencing the partitioning of runoff fluxes (e.g., *K* only determines the timing of the baseflow reservoir); other parameters are strongly related to flux partitioning. For instance, both INSC and COEFF influence infiltration excess runoff, but INSC is almost identical across all cases hence not informative, and COEFF is very similar between Classes I, III, and IV. When parameter SUB, as a control of INT & SATexc flux, is highly constrained (Class I) the BAS flux is also constrained, and as the SUB marginal distribution becomes more uniform (toward Class IV) the BAS flux becomes less constrained. That said, while BAS flux is (almost)

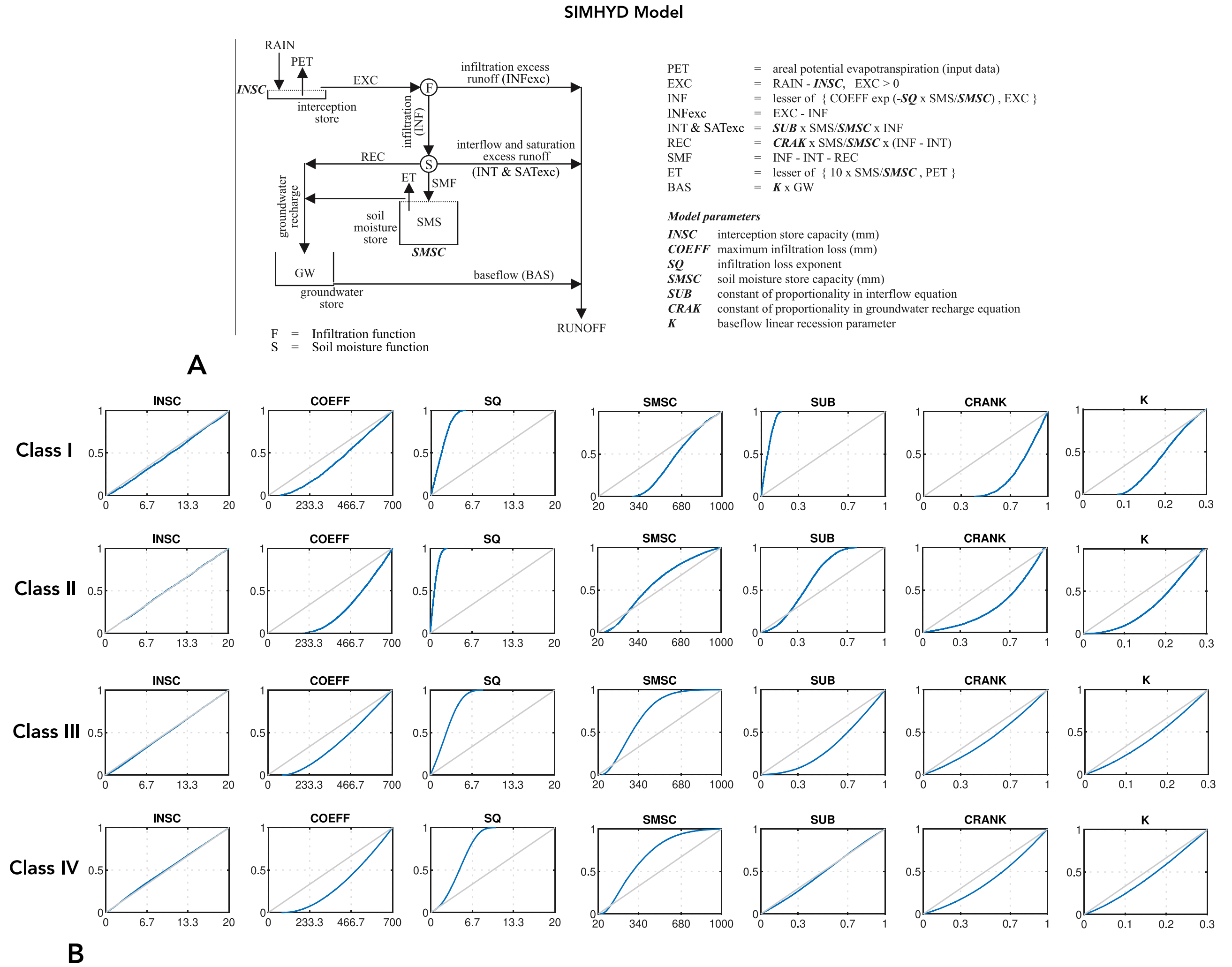

Figure 5. (a) Schematic illustration of the SIMHYD model structure together with the description of the model's internal fluxes and parameters (derived from Peel et al., 2000). (b) Marginal (cumulative) distribution of model parameters (gray is the prior, and blue is the posterior distributions) corresponding to the Classes I–IV flux maps (excluding the extreme case of Class IV) in section 4, for the relaxed threshold (0.85 × SCE-NSE).

equally variable for both Classes III and IV (varying between 0 and 100%), the corresponding marginal distributions of SUB parameter are different. This is due to the strong interaction of influencing parameters, as model fluxes are usually controlled by more than one parameter. The bottom line is, given the often complex and nonlinear interaction between model parameters, even in the case of influencing parameters, it is still very difficult to evaluate the impact of parameter distributions or their changes from one case (e.g., catchment and model structure) to another on flux dynamics (partitioning/contributions) in particular and model process representation in general. Hence, it is difficult—if not impossible—to infer model internal dynamics from parameter marginal distributions and/or their changes. It is instead much easier to map the model behavior into the flux space, that is, flux mapping. In technical terms, flux mapping is a nonlinear transformation of the model response surface from the parameter space into the flux space. Within parameter uncertainty estimation methods, *modeling uncertainties*—the interplay between data information content, modeling framework and model structure, performance metrics, and modeler's decisions/understanding—are lumped into and evaluated within the parameter space. Flux mapping can illustrate the impact of modeling uncertainties, regardless of the source and nature, on model internal behavior and process representation.

In this study, flux maps could be interpreted as a new visualization of parameter equifinality within the model flux space. Although in this study we only perturbed model parameters to explore the dynamics of internal fluxes, other model components could be taken into account (e.g., using an ensemble of model inputs or various scenarios of boundary conditions, as discussed in section 2) to generate an ensemble of model runs. No matter how the ensemble of model runs is produced, we could use flux maps (i.e., the particular facet of model internal fluxes of model equifinality) to explore, summarize, and visualize model internal behavior. It should also be mentioned that the overall pattern of the point cloud of flux maps is independent of the thresholding, the number of acceptable model runs, and/or how constrained the parameter distributions are. Flux maps of each modeling example for different thresholdings are presented in Table S1. The bottom line is that equifinal model fluxes, compared to equifinal model parameters, provide a more insightful basis to generate and explore MWH based on model process representation.

Furthermore, runoff flux space has lower dimensions than parameter space. For instance, in the case of SIMHYD, visualizing and investigating a 3-D runoff flux space instead of a 7-D parameter space is more convenient. Moreover, flux mapping is extendable to any combination of model fluxes. It is particularly of interest for fluxes of physical significance such as actual evaporation. Although in this work model realizations are the product of perturbing model parameters, other components of the model could also be used for generating flux maps, for example, model inputs and/or input/boundary conditions.

5.2. Flux Mapping, Process Representation, and Multiple Working Hypotheses

Flux space also has the advantage over parameter space of being (more) physically meaningful/relatable. Modeling uncertainties in general, and parameter uncertainty in particular, are the result of various types/sources of uncertainties. However, placing the emphasis only on parameter distributions, that is, uncertainty estimation as if all uncertainties are aleatory (due to variability and randomness; Beven, 2016), neglects the crucial role of epistemic uncertainties. But, within the flux mapping approach, we deal with and can pose questions of an epistemic nature. For instance, the flux map of Class I (Figures 3a1 and 3a2) indicates a baseflow-dominant system, a hypothesis to be further tested across multiple model structures and/or performance metrics (error/efficiency metrics and hydrological signatures). To see whether different modeling setups (given modeler's judgments/decisions) are in (dis)agreement with each other and to what extent and eventually pose an essential question of conceptual modeling, that is, how well model output (and internal behavior) corresponds to catchment response (and its internal processes). In other words, changing the emphasis of model equifinality from model parameters to fluxes enables us to develop MWH that are process-based and testable both across different model-based hypotheses as well as our understanding of real-world catchment processes. It should be noted that defining the flux space—that is, selecting a subset of the model's entire flux space for flux mapping—is closely related to the model structure and nature of the hypotheses of interest, that is, what processes are represented in the model or meant to be examined.

The importance of independent estimates of *catchment* internal fluxes/storages and their exploitation as diagnostic tools has been discussed in the literature (for further details see Clark et al., 2011). However, evaluating model behavior in terms of their internal fluxes—model flux equifinality—has received little attention compared with parameter equifinality, although it can be a valuable source of insight. For instance, Guo et al. (2017) presented an example of evaluating the relative realism of ET process representation within three conceptual rainfall-runoff models by comparing the simulated actual ET (AET) with measurements. They observed some unrealistic behavior in the simulated AET. Given that ET process representation can have significant impacts on the sensitivity of runoff projections under climatic changes, assessing the realism of model AET flux is essential. Li et al. (2015) also demonstrated that while a conceptual model can generally simulate the total observed streamflow well for various catchments with different characteristics, it may fail to match the observed baseflow and quickflow fluxes. Such unrealistic internal model behavior/dynamics are not easily discernible from model performance solely or parameter distributions (if at all) but are crucial for rejecting unfit model runs, improving the model realism and reducing model structural and prediction uncertainties.

Furthermore, if additional measurements are available, such as flow path information (baseflow measurements/estimates), the conflicting modeling scenarios (as MWH) could then be evaluated in terms

10.1029/2018WR023750

of their process representation. That is, which of the so-called equifinal fluxes should be rejected as not physically plausible/significant (less realistic) and which of them could be seen as plausible working hypotheses. Therefore, the curious follow-up question would be to investigate the relationship between catchment characteristics and flux map patterns: which catchment characteristics are controlling/determining flux maps? For instance, other hydrologic variables such as filed data on surface runoff measurement and soil moisture (Western & Grayson, 1998), water quality, for example, salt load (Nathan & Mudgway, 1997), and isotopes (Beria et al., 2018; Kendall & McDonnell, 1998) may provide diagnostic information about the sources of dominant flow processes that could be used to further evaluate/constrain model fluxes. Also, integrating hard and soft data (Winsemius et al., 2009) and diagnostic approaches (e.g., hydrological signatures) could provide valuable sources of information to understand the interplay between flux maps and catchment dynamics particularly dominant processes. While knowledge about dominant processes within the catchment system is a major control for developing/improving models (Seibert & McDonnell, 2002), the modeler's personal judgment and experience is crucial in deciding the dominant processes (Holländer et al., 2009). So combining hydrologic signatures with expert knowledge of mechanisms and processes of real-world catchment systems, that is, expert elicitation, we can improve model realism and process representation by imposing relative model parameter/flux constraints (Euser et al., 2013; Fenicia et al., 2014; Gharari et al., 2014; Hrachowitz et al., 2014).

We can also further assess how modifying different model components (e.g., interception, evaporation, and soil water routines) can influence and hopefully improve the flux map pattern. Therefore, flux mapping has an *exploratory* power to evaluate the impact of modeling uncertainties on model behavior and to explore model *capacity* for process representation, at least partly and in a lumped way. This can pave the way to go beyond only evaluating model performance (i.e., model capability for yielding a high value of some error metrics), possibly together with some estimation of parameter uncertainties, based on only the model output. We emphasize that flux mapping alone cannot improve model realism, and additional information is required for refining models. Also, flux mapping provides *explanatory* opportunities to postulate hypotheses of possible explanations of real-world processes (e.g., catchment processes and their changes) based on the (internal) behavior of their corresponding conceptual models under all modeling uncertainties. The relationship between different modes of model process representation (i.e., model equifinality) and catchment internal processes is a crucial question and remains an open one, which can only be addressed by bridging the gap between modeling and experimental hydrology (Seibert & McDonnell, 2002). That said, various catchment internal processes are unknown or unknowable (knowledge uncertainty). There are myriad (ever evolving and dynamic) flow paths in any real catchment that we cannot expect conceptual models to represent them. Such epistemic uncertainties and hence modeling limitations may be a contributing factor to model equifinality. It is unclear to what extent a projected flux equifinality resembles the plausible hypotheses of internal processes of a given catchment and to what extent it is a modeling artifact due to modeling uncertainties, for instance, to what extent a constrained or space-filling flux map is due to catchment characteristics or due to modeling setup. What occurred in reality in a particular period would be mapped to a single unknown *true point* on the flux map. Model equifinality, for example, flux equifinality, results in a point cloud that necessarily does not encompass the true point (i.e., reality). What we can hope for is to refine our hypotheses to approximate reality by improving model realism—by improving model structures, evaluation schemes, and data quality and accounting for their uncertainties—and hence reducing knowledge uncertainty.

5.3. Equifinality, Hydrological Systems, and Beyond

The theoretical framework presented, and its implications for model evaluation (going beyond model output and accounting for model behavior under modeling uncertainties) and hypothesis generating, could be further extended to other domains (Blöschl et al., 2019). It could also be used to improve understanding/modeling in data-poor catchments/regions (Davtalab et al., 2017) and in regional generalization (also known as regionalization or prediction in ungauged basins; Peel et al., 2000; Reichl et al., 2006). Moreover, although the proposed theoretical framework and flux mapping method are mainly discussed within the context of hydrological systems, it can be further extended to any field of scientific modeling concerned with understanding/modeling open complex systems in the face of uncertainties. Real-world processes and their perceptual and conceptual models are all open systems and hence give rise to equifinality.

Therefore, it is important to go beyond evaluating only the (equifinality of) model output and account for model fluxes (or other facets of model equifinality) as well, particularly physically meaningful fluxes of conceptual models that may be used in, for example, data assimilation (Alvarez-Garreton et al., 2014; Teweldebrhan et al., 2018) and flood forecasting (Alvarez-Garreton et al., 2015), catchment classification (Kelleher et al., 2015; Knoben et al., 2018; Sawicz et al., 2014), sociohydrological systems (Di Baldassarre et al., 2015; Khazaei et al., 2019; van Emmerik et al., 2014; Westerberg et al., 2017), environmental (Chowdhury et al., 2016) and ecological systems (Luo et al., 2009), natural hazard and risk assessment (Beven et al., 2017), decision making under uncertainty (Madani & Lund, 2011; Maier et al., 2016), agent-based modeling (Billari et al., 2006; Madani et al., 2014), sustainability transitions (de Haan et al., 2016; Moallemi & Köhler, 2019) and exploratory modeling (Kwakkel & Pruyt, 2013; Moallemi et al., 2017; Moallemi & Malekpour, 2017) under deep uncertainty (Haasnoot et al., 2013; Maier et al., 2016; Moallemi et al., 2018), demand modeling/forecasting in energy (van Ruijven et al., 2010), and traffic (Flyvbjerg et al., 2006; Saberi et al., 2018) and network design (Chen et al., 2011).

6. Conclusion

We outlined different facets of model equifinality in the context of hydrological modeling. We developed a novel model evaluation scheme, flux mapping, based on a particular facet of model equifinality, namely, model internal fluxes. We demonstrated how flux mapping can give new insights into model behavior that cannot be inferred from conventional evaluation methods. That is, even within a very narrow margin of model error/performance, different modes of model response, that is, internal runoff generating fluxes of the model, can be equally active. In other words, different dynamics of model internal fluxes can equally reproduce a given observed hydrograph within a narrow margin of model error. Flux mapping can be extended to any field of scientific modeling dealing with conceptual modeling of open complex systems under uncertainty. We argued that equifinality plays a central role in scientific modeling, particularly within the paradigm of MWH.

Acknowledgments
The authors gratefully acknowledge the support of the University of Melbourne and Australian Government in carrying out this research. Sina Khatami is supported by Melbourne International Research and Fee Remission Scholarships (MIRS and MIFRS), Murray Peel is the recipient of an Australian Research Council Future Fellowship (FT120100130), and Tim Peterson jointly funded by Australian Research Council Linkage Project LP130100958, Bureau of Meteorology (Australia), Department of Environment, Land, Water and Planning (Vic., Australia), Department of Economic Development, Jobs, Transport and Resources (Vic., Australia), and Power and Water Corporation (N.T., Australia). The authors would also like to acknowledge the rigorous yet delightful review process that helped to significantly improve the manuscript by sincerely thanking Andrew Binley for suggesting the use of ternary plots reviewing the first draft of the manuscript, as well as the encouraging and constructive reviews of Hoshin Gupta, Grey Nearing, and two anonymous reviewers. Sina Khatami is extremely grateful to Keith Beven for selflessly giving his time to teaching the excellent course of "Uncertainty in Environmental Modelling," as well as Sven Halldin for administrating it at Department of Earth Sciences at Uppsala University. He is also sincerely thankful to Keirnan Fowler for preparing and sharing the entire rainfall-runoff data set of this study and his occasional feedback throughout the evolution of this work as well as the manuscript. Streamflow data used in this project are from the Australian Bureau of Meteorology's (BOM) Hydrologic Reference Station project website (www.bom.gov.au/hrs). Rainfall data are from the Australian Water Availability Project (AWAP) project (www.bom.gov.au/jsp/awap/). Potential evapotranspiration data are from the SILO project (www.longpaddock.qld.gov.au/silo/).

References

Addor, N., & Melsen, L. A. (2019). Legacy, rather than adequacy, drives the selection of hydrological models. *Water Resources Research*, *55*, 378–390. https://doi.org/10.1029/2018WR022958
Addor, N., Nearing, G., Prieto, C., Newman, A. J., Le Vine, N., & Clark, M. P. (2018). A ranking of hydrological signatures based on their predictability in space. *Water Resources Research*, *54*, 8792–8812. https://doi.org/10.1029/2018WR022606
Ajami, N. K., Duan, Q., & Sorooshian, S. (2007). An integrated hydrologic Bayesian multimodel combination framework: Confronting input, parameter, and model structural uncertainty in hydrologic prediction. *Water Resources Research*, *43*, W01403. https://doi.org/10.1029/2005WR004745
Alvarez-Garreton, C., Ryu, D., Western, A. W., Crow, W. T., & Robertson, D. E. (2014). The impacts of assimilating satellite soil moisture into a rainfall–runoff model in a semi-arid catchment. *Journal of Hydrology*, *519*, 2763–2774. https://doi.org/10.1016/j.jhydrol.2014.07.041
Alvarez-Garreton, C., Ryu, D., Western, A. W., Su, C. H., Crow, W. T., Robertson, D. E., & Leahy, C. (2015). Improving operational flood ensemble prediction by the assimilation of satellite soil moisture: comparison between lumped and semi-distributed schemes. *Hydrology and Earth System Sciences*, *19*(4), 1659–1676. https://doi.org/10.5194/hess-19-1659-2015
Arsenault, R., & Brissette, F. P. (2014). Continuous streamflow prediction in ungauged basins: The effects of equifinality and parameter set selection on uncertainty in regionalization approaches. *Water Resources Research*, *50*, 6135–6153. https://doi.org/10.1002/2013WR014898
di Baldassarre, G., Viglione, A., Carr, G., Kuil, L., Yan, K., Brandimarte, L., & Blöschl, G. (2015). Debates—Perspectives on socio-hydrology: Capturing feedbacks between physical and social processes. *Water Resources Research*, *51*, 4770–4781. https://doi.org/10.1002/2014WR016416
Beck, M. B. (1987). Water quality modeling: A review of the analysis of uncertainty. *Water resources research*, *23*(8), 1393–1442. https://doi.org/10.1029/WR023i008p01393
Beck, M. B. (2002). *Environmental foresight and models: A manifesto*. Oxford: Elsevier.
Bennett, N. D., Croke, B. F. W., Guariso, G., Guillaume, J. H. A., Hamilton, S. H., Jakeman, A. J., et al. (2013). Characterising performance of environmental models. *Environmental Modelling & Software*, *40*, 1–20. https://doi.org/10.1016/j.envsoft.2012.09.011
Beria, H., Larsen, J. R., Ceperley, N. C., Michelon, A., Vennemann, T., & Schaefli, B. (2018). Understanding snow hydrological processes through the lens of stable water isotopes. *Wiley Interdisciplinary Reviews: Water*, *5*(6), e1311. https://doi.org/10.1002/wat2.1311
Bethke, C. M. (1992). The question of uniqueness in geochemical modeling. *Geochimica et Cosmochimica Acta*, *56*(12), 4315–4320. https://doi.org/10.1016/0016-7037(92)90274-M
Beven, K. (1975). A deterministic, spatially distributed model of catchment hydrology. University of East Anglia,
Beven, K. (1993). Prophecy, reality and uncertainty in distributed hydrological modelling. *Advances in Water Resources*, *16*(1), 41–51. https://doi.org/10.1016/0309-1708(93)90028-E
Beven, K. (2006). A manifesto for the equifinality thesis. *Journal of Hydrology*, *320*(1–2), 18–36. https://doi.org/10.1016/j.jhydrol.2005.07.007

Beven, K. (2009). *Environmental modelling: An uncertain future?* CRC Press.
Beven, K. (2012a). Causal models as multiple working hypotheses about environmental processes. *Comptes rendus geoscience*, *344*(2), 77–88. https://doi.org/10.1016/j.crte.2012.01.005
Beven, K. (2012b). *Rainfall-runoff modelling: The primer*, (2nd ed.). Chicheseter, UK: John Wiley & Sons, Ltd. https://doi.org/10.1002/9781119951001
Beven, K. (2016). Facets of uncertainty: Epistemic uncertainty, non-stationarity, likelihood, hypothesis testing, and communication. *Hydrological Sciences Journal*, *61*(9), 1652–1665. https://doi.org/10.1080/02626667.2015.1031761
Beven, K., & Binley, A. (1992). The future of distributed models: Model calibration and uncertainty prediction. *Hydrological Processes*, *6*(3), 279–298. https://doi.org/10.1002/hyp.3360060305
Beven, K., & Binley, A. (2014). GLUE: 20 years on. *Hydrological Processes*, *28*(24), 5897–5918. https://doi.org/10.1002/hyp.10082
Beven, K., Smith, P., Westerberg, I., & Freer, J. (2012). Comment on “Pursuing the method of multiple working hypotheses for hydrological modeling” by P. Clark et al. *Water Resources Research*, *48*, W11801. https://doi.org/10.1029/2012WR012282
Beven, K., & Westerberg, I. (2011). On red herrings and real herrings: Disinformation and information in hydrological inference. *Hydrological Processes*, *25*(10), 1676–1680. https://doi.org/10.1002/hyp.7963
Beven, K. J., Aspinall, W. P., Bates, P. D., Borgomeo, E., Goda, K., Hall, J. W., et al. (2017). Epistemic uncertainties and natural hazard risk assessment. 2. What should constitute good practice? *Natural Hazards and Earth System Sciences*, *2017*, 1–25. https://doi.org/10.5194/nhess-2017-251
Billari, F. C., Fent, T., Prskawetz, A., & Scheffran, J. (2006). *Agent-based computational modelling: Applications in demography, social, economic and environmental sciences*. Physica-Verlag Heidelberg.
Blazkova, S., & Beven, K. (2009). A limits of acceptability approach to model evaluation and uncertainty estimation in flood frequency estimation by continuous simulation: Skalka catchment, Czech Republic. *Water Resources Research*, *45*, W00B16. https://doi.org/10.1029/2007WR006726
Blöschl, G., Bierkens, M. F. P., Chambel, A., Cudennec, C., Destouni, G., Fiori, A., et al. (2019). Twenty-three Unsolved Problems in Hydrology (UPH)—A community perspective. *Hydrological Sciences Journal*. https://doi.org/10.1080/02626667.2019.1620507
Blöschl, G., & Sivapalan, M. (1995). Scale issues in hydrological modelling: A review. *Hydrological Processes*, *9*(3-4), 251–290. https://doi.org/10.1002/hyp.3360090305
Boughton, W. C. (2007). Effect of data length on rainfall–runoff modelling. *Environmental Modelling & Software*, *22*(3), 406–413. https://doi.org/10.1016/j.envsoft.2006.01.001
Bulygina, N., & Gupta, H. (2009). Estimating the uncertain mathematical structure of a water balance model via Bayesian data assimilation. *Water Resources Research*, *45*, W00B13. https://doi.org/10.1029/2007WR006749
Bulygina, N., & Gupta, H. (2010). How Bayesian data assimilation can be used to estimate the mathematical structure of a model. *Stochastic Environmental Research and Risk Assessment*, *24*(6), 925–937. https://doi.org/10.1007/s00477-010-0387-y
Bulygina, N., & Gupta, H. (2011). Correcting the mathematical structure of a hydrological model via Bayesian data assimilation. *Water Resources Research*, *47*, W05514. https://doi.org/10.1029/2010WR009614
Butts, M. B., Payne, J. T., Kristensen, M., & Madsen, H. (2004). An evaluation of the impact of model structure on hydrological modelling uncertainty for streamflow simulation. *Journal of Hydrology*, *298*(1-4), 242–266. https://doi.org/10.1016/j.jhydrol.2004.03.042
Buytaert, W., & Beven, K. (2011). Models as multiple working hypotheses: Hydrological simulation of tropical alpine wetlands. *Hydrological Processes*, *25*(11), 1784–1799. https://doi.org/10.1002/hyp.7936
Carrera, J., & Neuman, S. P. (1986). Estimation of aquifer parameters under transient and steady state conditions: 1. Maximum likelihood method incorporating prior information. *Water Resources Research*, *22*(2), 199–210. https://doi.org/10.1029/WR022i002p00199
Chamberlin, T. C. (1890). The method of multiple working hypotheses. *Science*, *15*, 92–96.
Chen, A., Zhou, Z., Chootinan, P., Ryu, S., Yang, C., & Wong, S. C. (2011). Transport network design problem under uncertainty: A review and new developments. *Transport Reviews*, *31*(6), 743–768. https://doi.org/10.1080/01441647.2011.589539
Chiew, F., Peel, M., & Western, A. (2002). Application and testing of the simple rainfall-runoff model SIMHYD. In V. P. Singh & D. Frevert (Eds.), *Mathematical models of small watershed hydrology and applications* (pp. 335–367). Water Resources Pub.
Chiew, F. H. S., Stewardson, M. J., & McMahon, T. A. (1993). Comparison of six rainfall-runoff modelling approaches. *Journal of Hydrology*, *147*(1-4), 1–36. https://doi.org/10.1016/0022-1694(93)90073-I
Chowdhury, R. B., Moore, G. A., Weatherley, A. J., & Arora, M. (2016). A novel substance flow analysis model for analysing multi-year phosphorus flow at the regional scale. *Science of The Total Environment*, *572*, 1269–1280. https://doi.org/10.1016/j.scitotenv.2015.10.055
Clark, M. P., & Kavetski, D. (2010). Ancient numerical daemons of conceptual hydrological modeling: 1. Fidelity and efficiency of time stepping schemes. *Water Resources Research*, *46*, W10510. https://doi.org/10.1029/2009WR008894
Clark, M. P., Kavetski, D., & Fenicia, F. (2011). Pursuing the method of multiple working hypotheses for hydrological modeling. *Water Resources Research*, *47*, W09301. https://doi.org/10.1029/2010WR009827
Clark, M. P., Kavetski, D., & Fenicia, F. (2012). Reply to comment by K. Beven et al. on “Pursuing the method of multiple working hypotheses for hydrological modeling”. *Water Resources Research*, *48*, W11802. https://doi.org/10.1029/2012WR012547
Clark, M. P., Nijssen, B., Lundquist, J. D., Kavetski, D., Rupp, D. E., Woods, R. A., et al. (2015a). A unified approach for process-based hydrologic modeling: 1. Modeling concept. *Water Resources Research*, *51*, 2498–2514. https://doi.org/10.1002/2015WR017198
Clark, M. P., Nijssen, B., Lundquist, J. D., Kavetski, D., Rupp, D. E., Woods, R. A., et al. (2015b). A unified approach for process-based hydrologic modeling: 2. Model implementation and case studies. *Water resources research*, *51*, 2515–2542. https://doi.org/10.1002/2015wr017200
Clark, M. P., Slater, A. G., Rupp, D. E., Woods, R. A., Vrugt, J. A., Gupta, H. V., et al. (2008). Framework for Understanding Structural Errors (FUSE): A modular framework to diagnose differences between hydrological models. *Water Resources Research*, *44*, W00B02. https://doi.org/10.1029/2007WR006735
Crochemore, L., Perrin, C., Andréassian, V., Ehret, U., Seibert, S. P., Grimaldi, S., et al. (2015). Comparing expert judgement and numerical criteria for hydrograph evaluation. *Hydrological Sciences Journal*, *60*(3), 402–423. https://doi.org/10.1080/02626667.2014.903331
Davtalab, R., Mirchi, A., Khatami, S., Gyawali, R., Massah, A., Farajzadeh, M., & Madani, K. (2017). Improving continuous hydrologic modeling of data-poor river basins using hydrologic engineering center's hydrologic modeling system: Case study of Karkheh River basin. *Journal of Hydrologic Engineering*, *22*(8). https://doi.org/10.1061/(ASCE)HE.1943-5584.0001525
Duan, Q., Sorooshian, S., & Gupta, V. (1992). Effective and efficient global optimization for conceptual rainfall-runoff models. *Water Resources Research*, *28*(4), 1015–1031. https://doi.org/10.1029/91WR02985
Ebel, B. A., & Loague, K. (2006). Physics-based hydrologic-response simulation: Seeing through the fog of equifinality. *Hydrological Processes*, *20*(13), 2887–2900. https://doi.org/10.1002/hyp.6388

Efstratiadis, A., & Koutsoyiannis, D. (2010). One decade of multi-objective calibration approaches in hydrological modelling: A review. *Hydrological Sciences Journal*, *55*(1), 58–78. https://doi.org/10.1080/02626660903526292

Ehlers, L. B., Sonnenborg, T. O., & Refsgaard, J. C. (2018). Observational and predictive uncertainties for multiple variables in a spatially distributed hydrological model. *Hydrological Processes*. https://doi.org/10.1002/hyp.13367

van Emmerik, T. H. M., Li, Z., Sivapalan, M., Pande, S., Kandasamy, J., Savenije, H. H. G., et al. (2014). Socio-hydrologic modeling to understand and mediate the competition for water between agriculture development and environmental health: Murrumbidgee River basin, Australia. *Hydrology and Earth System Sciences*, *18*(10), 4239–4259. https://doi.org/10.5194/hess-18-4239-2014

Euser, T., Winsemius, H. C., Hrachowitz, M., Fenicia, F., Uhlenbrook, S., & Savenije, H. H. G. (2013). A framework to assess the realism of model structures using hydrological signatures. *Hydrology and Earth System Sciences*, *17*(5), 1893–1912. https://doi.org/10.5194/hess-17-1893-2013

Fenicia, F., Kavetski, D., & Savenije, H. H. G. (2011). Elements of a flexible approach for conceptual hydrological modeling: 1. Motivation and theoretical development. *Water Resources Research*, *47*, W11510. https://doi.org/10.1029/2010wr010174

Fenicia, F., Kavetski, D., Savenije, H. H. G., Clark, M. P., Schoups, G., Pfister, L., & Freer, J. (2014). Catchment properties, function, and conceptual model representation: Is there a correspondence? *Hydrological Processes*, *28*(4), 2451–2467. https://doi.org/10.1002/hyp.9726

Flyvbjerg, B., Skamris Holm, M. K., & Buhl, S. L. (2006). Inaccuracy in traffic forecasts. *Transport Reviews*, *26*(1), 1–24. https://doi.org/10.1080/01441640500124779

Fowler, K., Peel, M., Western, A., & Zhang, L. (2018). Improved rainfall-runoff calibration for drying climate: Choice of objective function. *Water Resources Research*, *54*, 3392–3408. https://doi.org/10.1029/2017WR022466

Fowler, K. J. A., Peel, M. C., Western, A. W., Zhang, L., & Peterson, T. J. (2016). Simulating runoff under changing climatic conditions: Revisiting an apparent deficiency of conceptual rainfall-runoff models. *Water Resources Research*, *52*, 1820–1846. https://doi.org/10.1002/2015WR018068

Gharari, S., Hrachowitz, M., Fenicia, F., Gao, H., & Savenije, H. H. G. (2014). Using expert knowledge to increase realism in environmental system models can dramatically reduce the need for calibration. *Hydrology and Earth System Sciences*, *18*(12), 4839–4859. https://doi.org/10.5194/hess-18-4839-2014

Gong, W., Gupta, H. V., Yang, D., Sricharan, K., & Hero, A. O. (2013). Estimating epistemic and aleatory uncertainties during hydrologic modeling: An information theoretic approach. *Water Resources Research*, *49*, 2253–2273. https://doi.org/10.1002/wrcr.20161

Grayson, R. B., Moore, I. D., & McMahon, T. A. (1992). Physically based hydrologic modeling: 1. A terrain-based model for investigative purposes. *Water Resources Research*, *28*(10), 2639–2658. https://doi.org/10.1029/92WR01258

Guo, D., Westra, S., & Maier, H. R. (2017). Impact of evapotranspiration process representation on runoff projections from conceptual rainfall-runoff models. *Water Resources Research*, *53*, 435–454. https://doi.org/10.1002/2016WR019627

Gupta, H. V., Kling, H., Yilmaz, K. K., & Martinez, G. F. (2009). Decomposition of the mean squared error and NSE performance criteria: Implications for improving hydrological modelling. *Journal of Hydrology*, *377*(1–2), 80–91. https://doi.org/10.1016/j.jhydrol.2009.08.003

Gupta, H. V., & Nearing, G. S. (2014). Debates—The future of hydrological sciences: A (common) path forward? Using models and data to learn: A systems theoretic perspective on the future of hydrological science. *Water Resources Research*, *50*, 5351–5359. https://doi.org/10.1002/2013WR015096

Gupta, H. V., Sorooshian, S., & Yapo, P. O. (1998). Toward improved calibration of hydrologic models: Multiple and noncommensurable measures of information. *Water Resources Research*, *34*(4), 751–763. https://doi.org/10.1029/97WR03495

Gupta, H. V., Wagener, T., & Liu, Y. (2008). Reconciling theory with observations: Elements of a diagnostic approach to model evaluation. *Hydrological Processes*, *22*(18), 3802–3813. https://doi.org/10.1002/hyp.6989

Gupta, V. K., & Sorooshian, S. (1983). Uniqueness and observability of conceptual rainfall-runoff model parameters: The percolation process examined. *Water resources research*, *19*(1), 269–276. https://doi.org/10.1029/WR019i001p00269

de Haan, F. J., Rogers, B. C., Brown, R. R., & Deletic, A. (2016). Many roads to Rome: The emergence of pathways from patterns of change through exploratory modelling of sustainability transitions. *Environmental Modelling & Software*, *85*, 279–292. https://doi.org/10.1016/j.envsoft.2016.05.019

Haasnoot, M., Kwakkel, J. H., Walker, W. E., & ter Maat, J. (2013). Dynamic adaptive policy pathways: A method for crafting robust decisions for a deeply uncertain world. *Global Environmental Change*, *23*(2), 485–498. https://doi.org/10.1016/j.gloenvcha.2012.12.006

Haydon, S., & Deletic, A. (2009). Model output uncertainty of a coupled pathogen indicator–hydrologic catchment model due to input data uncertainty. *Environmental Modelling & Software*, *24*(3), 322–328. https://doi.org/10.1016/j.envsoft.2008.09.004

Heymann, M., & Dalmedico, A. D. (2019). Epistemology and politics in Earth system modeling: Historical perspectives. *Journal of Advances in Modeling Earth Systems*. https://doi.org/10.1029/2018ms001526

Holländer, H. M., Blume, T., Bormann, H., Buytaert, W., Chirico, G. B., Exbrayat, J. F., et al. (2009). Comparative predictions of discharge from an artificial catchment (Chicken Creek) using sparse data. *Hydrology and Earth System Sciences*, *13*(11), 2069–2094. https://doi.org/10.5194/hess-13-2069-2009

Hornberger, G. M., & Spear, R. C. (1981). An approach to the preliminary analysis of environmental systems. *Journal of Environmental Management*, *12*, 7–18.

Hrachowitz, M., Fovet, O., Ruiz, L., Euser, T., Gharari, S., Nijzink, R., et al. (2014). Process consistency in models: The importance of system signatures, expert knowledge, and process complexity. *Water Resources Research*, *50*, 7445–7469. https://doi.org/10.1002/2014WR015484

Jeffrey, S. J., Carter, J. O., Moodie, K. B., & Beswick, A. R. (2001). Using spatial interpolation to construct a comprehensive archive of Australian climate data. *Environmental Modelling & Software*, *16*(4), 309–330. https://doi.org/10.1016/S1364-8152(01)00008-1

Jones, D. A., Wang, W., & Fawcett, R. (2009). High-quality spatial climate data-sets for Australia. *Australian Meteorological and Oceanographic Journal*, *58*(04), 233–248. https://doi.org/10.22499/2.5804.003

Kavetski, D., & Clark, M. P. (2011). Numerical troubles in conceptual hydrology: Approximations, absurdities and impact on hypothesis testing. *Hydrological Processes*, *25*(4), 661–670. https://doi.org/10.1002/hyp.7899

Kavetski, D., & Fenicia, F. (2011). Elements of a flexible approach for conceptual hydrological modeling: 2. Application and experimental insights. *Water Resources Research*, *47*, W11511. https://doi.org/10.1029/2011wr010748

Kavetski, D., Kuczera, G., & Franks, S. W. (2006). Bayesian analysis of input uncertainty in hydrological modeling: 1. Theory. *Water Resources Research*, *42*, W03407. https://doi.org/10.1029/2005WR004368

Kelleher, C., McGlynn, B., & Wagener, T. (2017). Characterizing and reducing equifinality by constraining a distributed catchment model with regional signatures, local observations, and process understanding. *Hydrology and Earth System Sciences*, *21*(7), 3325–3352. https://doi.org/10.5194/hess-21-3325-2017

Kelleher, C., Wagener, T., & McGlynn, B. (2015). Model-based analysis of the influence of catchment properties on hydrologic partitioning across five mountain headwater subcatchments. *Water Resources Research*, *51*, 4109–4136. https://doi.org/10.1002/2014WR016147

Kendall, C., & McDonnell, J. J. (1998). *Isotope Tracers In Catchment Hydrology*. Amsterdam: Elsevier.

Khatami, S. (2013a). Evidence of low-dimensional determinism in short time series of solute transport. Retrieved from Lund, Sweden: https://lup.lub.lu.se/search/publication/4139566

Khatami, S. (2013b). Nonlinear chaotic and trend analyses of water level at Urmia Lake, Iran. M.Sc. Thesis report: TVVR 13/5012, ISSN:1101-9824, Lund: Lund University, Retrieved from https://lup.lub.lu.se/search/publication/4253926

Khatami, S., Peel, M., Peterson, T., & Western, A. (2017). Equifinality and process-based modelling. Paper presented at the AGU Fall Meeting, New Orleans, LA.

Khazaei, B., Khatami, S., Alemohammad, S. H., Rashidi, L., Wu, C., Madani, K., et al. (2019). Climatic or regionally induced by humans? Tracing hydro-climatic and land-use changes to better understand the Lake Urmia tragedy. *Journal of Hydrology*, *569*, 203–217. https://doi.org/10.1016/j.jhydrol.2018.12.004

Kirchner, J. W. (2016). Aggregation in environmental systems—Part 2: Catchment mean transit times and young water fractions under hydrologic nonstationarity. *Hydrology and Earth System Sciences*, *20*(1), 299–328. https://doi.org/10.5194/hess-20-299-2016

Knoben, W. J. M., Woods, R. A., & Freer, J. E. (2018). A quantitative hydrological climate classification evaluated with independent streamflow data. *Water Resources Research*, *54*, 5088–5109. https://doi.org/10.1029/2018WR022913

Koch, J., Mendiguren, G., Mariethoz, G., & Stisen, S. (2017). Spatial sensitivity analysis of simulated land surface patterns in a catchment model using a set of innovative spatial performance metrics. *Journal of Hydrometeorology*, *18*(4), 1121–1142. https://doi.org/10.1175/jhm-d-16-0148.1

Koch, J., Siemann, A., Stisen, S., & Sheffield, J. (2016). Spatial validation of large-scale land surface models against monthly land surface temperature patterns using innovative performance metrics. *Journal of Geophysical Research: Atmospheres*, *121*, 5430–5452. https://doi.org/10.1002/2015JD024482

Konikow, L. F., & Bredehoeft, J. D. (1992). Ground-water models cannot be validated. *Advances in Water Resources*, *15*(1), 75–83. https://doi.org/10.1016/0309-1708(92)90033-X

Kwakkel, J. H., & Pruyt, E. (2013). Exploratory modeling and analysis, an approach for model-based foresight under deep uncertainty. *Technological Forecasting and Social Change*, *80*(3), 419–431. https://doi.org/10.1016/j.techfore.2012.10.005

Legates, D. R., & McCabe, G. J. (1999). Evaluating the use of "goodness-of-fit" measures in hydrologic and hydroclimatic model validation. *Water Resources Research*, *35*(1), 233–241. https://doi.org/10.1029/1998WR900018

Li, L., Lambert, M. F., Maier, H. R., Partington, D., & Simmons, C. T. (2015). Assessment of the internal dynamics of the Australian Water Balance Model under different calibration regimes. *Environmental Modelling & Software*, *66*, 57–68. https://doi.org/10.1016/j.envsoft.2014.12.015

Liu, Y., Freer, J., Beven, K., & Matgen, P. (2009). Towards a limits of acceptability approach to the calibration of hydrological models: Extending observation error. *Journal of Hydrology*, *367*(1-2), 93–103. https://doi.org/10.1016/j.jhydrol.2009.01.016

Luo, Y., Weng, E., Wu, X., Gao, C., Zhou, X., & Zhang, L. (2009). Parameter identifiability, constraint, and equifinality in data assimilation with ecosystem models. *Ecological Applications*, *19*(3), 571–574. https://doi.org/10.1890/08-0561.1

Madani, K., Hooshyar, M., Khatami, S., Alaeipour, A., & Moeini, A. (2014, 5-8 Oct. 2014). Nash-reinforcement learning (N-RL) for developing coordination strategies in non-transferable utility games. Paper presented at the 2014 IEEE International Conference on Systems, Man, and Cybernetics (SMC). https://doi.org/10.1109/SMC.2014.6974336

Madani, K., & Lund, J. R. (2011). A Monte-Carlo game theoretic approach for multi-criteria decision making under uncertainty. *Advances in water resources*, *34*(5), 607–616. https://doi.org/10.1016/j.advwatres.2011.02.009

Maier, H. R., Guillaume, J. H. A., van Delden, H., Riddell, G. A., Haasnoot, M., & Kwakkel, J. H. (2016). An uncertain future, deep uncertainty, scenarios, robustness and adaptation: How do they fit together? *Environmental Modelling & Software*, *81*, 154–164. https://doi.org/10.1016/j.envsoft.2016.03.014

Moallemi, E. A., de Haan, F., Kwakkel, J., & Aye, L. (2017). Narrative-informed exploratory analysis of energy transition pathways: A case study of India's electricity sector. *Energy Policy*, *110*, 271–287. https://doi.org/10.1016/j.enpol.2017.08.019

Moallemi, E. A., Elsawah, S., & Ryan, M. J. (2018). Model-based multi-objective decision making under deep uncertainty from a multi-method design lens. *Simulation Modelling Practice and Theory*, *84*, 232–250. https://doi.org/10.1016/j.simpat.2018.02.009

Moallemi, E. A., & Köhler, J. (2019). Coping with uncertainties of sustainability transitions using exploratory modelling: The case of the MATISSE model and the UK's mobility sector. *Environmental Innovation and Societal Transitions.*. https://doi.org/10.1016/j.eist.2019.03.005

Moallemi, E. A., & Malekpour, S. (2017). A participatory exploratory modelling approach for long-term planning in energy transitions. *Energy Research & Social Science.*, *35*, 205–216. https://doi.org/10.1016/j.erss.2017.10.022

Morton, F. I. (1983). Operational estimates of areal evapotranspiration and their significance to the science and practice of hydrology. *Journal of Hydrology*, *66*(1), 1–76. https://doi.org/10.1016/0022-1694(83)90177-4

Murphy, A. H. (1988). Skill scores based on the mean square error and their relationships to the correlation coefficient. *Monthly Weather Review*, *116*(12), 2417–2424. https://doi.org/10.1175/1520-0493(1988)116<2417:ssbotm>2.0.co;2

Nash, J. E., & Sutcliffe, J. V. (1970). River flow forecasting through conceptual models part I—A discussion of principles. *Journal of Hydrology*, *10*(3), 282–290. https://doi.org/10.1016/0022-1694(70)90255-6

Nathan, R. J., & Mudgway, L. B. (1997). Estimating salt loads in high water table areas. II: Regional salt loads. *Journal of Irrigation and Drainage Engineering*, *123*(2), 91–99. https://doi.org/10.1061/(ASCE)0733-9437(1997)123:2(91)

Nearing, G. S., & Gupta, H. V. (2015). The quantity and quality of information in hydrologic models. *Water Resources Research*, *51*, 524–538. https://doi.org/10.1002/2014WR015895

Newman, A. J., Clark, M. P., Craig, J., Nijssen, B., Wood, A., Gutmann, E., et al. (2015). Gridded ensemble precipitation and temperature estimates for the contiguous United States. *Journal of Hydrometeorology*, *16*(6), 2481–2500. https://doi.org/10.1175/jhm-d-15-0026.1

Oreskes, N., Shrader-Frechette, K., & Belitz, K. (1994). Verification, validation, and confirmation of numerical models in the earth sciences. *Science*, *263*(5147), 641–646.

Oudin, L., Michel, C., & Anctil, F. (2005). Which potential evapotranspiration input for a lumped rainfall-runoff model?: Part 1—Can rainfall-runoff models effectively handle detailed potential evapotranspiration inputs? *Journal of Hydrology*, *303*(1–4), 275–289. https://doi.org/10.1016/j.jhydrol.2004.08.025

Pappenberger, F., Matgen, P., Beven, K. J., Henry, J.-B., Pfister, L., & Fraipont, P. (2006). Influence of uncertain boundary conditions and model structure on flood inundation predictions. *Advances in Water Resources*, *29*(10), 1430–1449. https://doi.org/10.1016/j.advwatres.2005.11.012

Peel, M. C., & Blöschl, G. (2011). Hydrological modelling in a changing world. *Progress in Physical Geography*, *35*(2), 249–261. https://doi.org/10.1177/0309133311402550

Peel, M. C., Chiew, F. H., Western, A. W., & McMahon, T. A. (2000). Extension of unimpaired monthly streamflow data and regionalisation of parameter values to estimate streamflow in ungauged catchments. Retrieved from Report prepared for the National Land and Water Resources Audit, In Australian Natural Resources Atlas, Pages 37.: http://people.eng.unimelb.edu.au/mpeel/NLWRA.pdf

Peel, M. C., Finlayson, B. L., & McMahon, T. A. (2007). Updated world map of the Köppen-Geiger climate classification. *Hydrology and Earth System Sciences*, *11*(5), 1633–1644. https://doi.org/10.5194/hess-11-1633-2007

Perrin, C., Michel, C., & Andréassian, V. (2001). Does a large number of parameters enhance model performance? Comparative assessment of common catchment model structures on 429 catchments. *Journal of Hydrology*, *242*(3–4), 275–301. https://doi.org/10.1016/S0022-1694(00)00393-0

Peterson, T. J., & Western, A. W. (2014). Multiple hydrological attractors under stochastic daily forcing: 1. Can multiple attractors exist? *Water Resources Research*, *50*, 2993–3009. https://doi.org/10.1002/2012WR013003

Peterson, T. J., Western, A. W., & Argent, R. M. (2014). Multiple hydrological attractors under stochastic daily forcing: 2. Can multiple attractors emerge? *Water Resources Research*, *50*, 3010–3029. https://doi.org/10.1002/2012WR013004

Pfannerstill, M., Guse, B., & Fohrer, N. (2014). Smart low flow signature metrics for an improved overall performance evaluation of hydrological models. *Journal of Hydrology*, *510*, 447–458. https://doi.org/10.1016/j.jhydrol.2013.12.044

Pool, S., Vis, M., & Seibert, J. (2018). Evaluating model performance: Towards a non-parametric variant of the Kling-Gupta efficiency. *Hydrological Sciences Journal*, *63*(13-14), 1941–1953. https://doi.org/10.1080/02626667.2018.1552002

Pushpalatha, R., Perrin, C., Moine, N. L., & Andréassian, V. (2012). A review of efficiency criteria suitable for evaluating low-flow simulations. *Journal of Hydrology*, *420-421*, 171–182. https://doi.org/10.1016/j.jhydrol.2011.11.055

Quine, W. V. (1975). On empirically equivalent systems of the world. *Erkenntnis*, *9*(3), 313–328. https://doi.org/10.1007/BF00178004

Reichl, J., Chiew, F. H., & Western, A. (2006). Model averaging, equifinality and uncertainty estimation in the modelling of ungauged catchments. Paper presented at the 3rd International Congress on Environmental Modelling and Software (iEMSs), BURLINGTON, VERMONT.

Renard, B., Kavetski, D., Kuczera, G., Thyer, M., & Franks, S. W. (2010). Understanding predictive uncertainty in hydrologic modeling: The challenge of identifying input and structural errors. *Water Resources Research*, *46*, W05521. https://doi.org/10.1029/2009WR008328

van Ruijven, B., de Vries, B., van Vuuren, D. P., & van der Sluijs, J. P. (2010). A global model for residential energy use: Uncertainty in calibration to regional data. *Energy*, *35*(1), 269–282. https://doi.org/10.1016/j.energy.2009.09.019

Saberi, M., Rashidi, T. H., Ghasri, M., & Ewe, K. (2018). A complex network methodology for travel demand model evaluation and validation. *Networks and Spatial Economics.*, *18*(4), 1051–1073. https://doi.org/10.1007/s11067-018-9397-y

Santos, L., Thirel, G., & Perrin, C. (2018). Technical note: Pitfalls in using log-transformed flows within the KGE criterion. *Hydrology and Earth System Sciences*, *22*(8), 4583–4591. https://doi.org/10.5194/hess-22-4583-2018

Savenije, H. H. G. (2001). Equifinality, a blessing in disguise? *Hydrological Processes*, *15*(14), 2835–2838. https://doi.org/10.1002/hyp.494

Sawicz, K. A., Kelleher, C., Wagener, T., Troch, P., Sivapalan, M., & Carrillo, G. (2014). Characterizing hydrologic change through catchment classification. *Hydrology and Earth System Sciences*, *18*(1), 273–285. https://doi.org/10.5194/hess-18-273-2014

Schaefli, B. (2016). Snow hydrology signatures for model identification within a limits-of-acceptability approach. *Hydrological Processes*, *30*(22), 4019–4035. https://doi.org/10.1002/hyp.10972

Schaefli, B., & Gupta, H. V. (2007). Do Nash values have value? *Hydrological Processes*, *21*(15), 2075–2080. https://doi.org/10.1002/hyp.6825

Schaefli, B., Harman, C. J., Sivapalan, M., & Schymanski, S. J. (2011). HESS Opinions: Hydrologic predictions in a changing environment: Behavioral modeling. *Hydrology and Earth System Sciences*, *15*(2), 635–646. https://doi.org/10.5194/hess-15-635-2011

Seibert, J. (2001). On the need for benchmarks in hydrological modelling. *Hydrological Processes*, *15*(6), 1063–1064. https://doi.org/10.1002/hyp.446

Seibert, J., & McDonnell, J. J. (2002). On the dialog between experimentalist and modeler in catchment hydrology: Use of soft data for multicriteria model calibration. *Water Resources Research*, *38*(11), 1241. https://doi.org/10.1029/2001WR000978

Sivakumar, B. (2000). Chaos theory in hydrology: Important issues and interpretations. *Journal of Hydrology*, *227*(1–4), 1–20. https://doi.org/10.1016/S0022-1694(99)00186-9

Sivakumar, B., Berndtsson, R., Olsson, J., & Jinno, K. (2001). Evidence of chaos in the rainfall-runoff process. *Hydrological Sciences Journal*, *46*(1), 131–145. https://doi.org/10.1080/02626660109492805

Sorooshian, S., & Gupta, V. K. (1983). Automatic calibration of conceptual rainfall-runoff models: The question of parameter observability and uniqueness. *Water resources research*, *19*(1), 260–268. https://doi.org/10.1029/WR019i001p00260

Stisen, S., Koch, J., Sonnenborg, T. O., Refsgaard, J. C., Bircher, S., Ringgaard, R., & Jensen, K. H. (2018). Moving beyond run-off calibration—Multivariable optimization of a surface–subsurface–atmosphere model. *Hydrological Processes*, *32*(17), 2654–2668. https://doi.org/10.1002/hyp.13177

Tang, J., & Zhuang, Q. (2008). Equifinality in parameterization of process-based biogeochemistry models: A significant uncertainty source to the estimation of regional carbon dynamics. *Journal of Geophysical Research*, *113*, G04010. https://doi.org/10.1029/2008JG000757

Teweldebrhan, A., Burkhart, J., Schuler, T., & Xu, C.-Y. (2018). Improving the informational value of MODIS fractional snow cover area using fuzzy logic based ensemble smoother data assimilation frameworks. *Remote Sensing*, *11*(1), 28. https://doi.org/10.3390/rs11010028

Teweldebrhan, A. T., Burkhart, J. F., & Schuler, T. V. (2018). Parameter uncertainty analysis for an operational hydrological model using residual-based and limits of acceptability approaches. *Hydrology and Earth System Sciences*, *22*(9), 5021–5039. https://doi.org/10.5194/hess-22-5021-2018

Tian, Y., Nearing, G. S., Peters-Lidard, C. D., Harrison, K. W., & Tang, L. (2016). Performance metrics, error modeling, and uncertainty quantification. *Monthly Weather Review*, *144*(2), 607–613. https://doi.org/10.1175/mwr-d-15-0087.1

Tolson, B. A., & Shoemaker, C. A. (2008). Efficient prediction uncertainty approximation in the calibration of environmental simulation models. *Water Resources Research*, *44*, W04411. https://doi.org/10.1029/2007wr005869

Turner, M. (2012). *Hydrologic reference station selection guidelines*. Australia: Retrieved from Melbourne. http://www.bom.gov.au/water/hrs/media/static/papers/Selection_Guidelines.pdf

Vrugt, J. A., & Beven, K. J. (2018). Embracing equifinality with efficiency: Limits of Acceptability sampling using the DREAM (LOA) algorithm. *Journal of Hydrology*, *559*, 954–971. https://doi.org/10.1016/j.jhydrol.2018.02.026

Vrugt, J. A., ter Braak, C. J. F., Clark, M. P., Hyman, J. M., & Robinson, B. A. (2008). Treatment of input uncertainty in hydrologic modeling: Doing hydrology backward with Markov chain Monte Carlo simulation. *Water Resources Research*, *44*, W00B09. https://doi.org/10.1029/2007WR006720

Westerberg, I. K., Di Baldassarre, G., Beven, K. J., Coxon, G., & Krueger, T. (2017). Perceptual models of uncertainty for socio-hydrological systems: A flood risk change example. *Hydrological Sciences Journal*, *62*(11), 1705–1713. https://doi.org/10.1080/02626667.2017.1356926

Westerberg, I. K., & McMillan, H. K. (2015). Uncertainty in hydrological signatures. *Hydrology and Earth System Sciences*, *19*(9), 3951–3968. https://doi.org/10.5194/hess-19-3951-2015

Westerberg, I. K., Wagener, T., Coxon, G., McMillan, H. K., Castellarin, A., Montanari, A., & Freer, J. (2016). Uncertainty in hydrological signatures for gauged and ungauged catchments. *Water Resources Research*, *52*, 1847–1865. https://doi.org/10.1002/2015WR017635

Western, A. W., & Grayson, R. B. (1998). The Tarrawarra data set: Soil moisture patterns, soil characteristics, and hydrological flux measurements. *Water Resources Research*, *34*(10), 2765–2768. https://doi.org/10.1029/98WR01833

Willmott, C. J., Robeson, S. M., & Matsuura, K. (2012). A refined index of model performance. *International Journal of Climatology*, *32*(13), 2088–2094. https://doi.org/10.1002/joc.2419

Winsemius, H. C., Schaefli, B., Montanari, A., & Savenije, H. H. G. (2009). On the calibration of hydrological models in ungauged basins: A framework for integrating hard and soft hydrological information. *Water Resources Research*, *45*, W12422. https://doi.org/10.1029/2009WR007706

Yeh, W. W. G. (1986). Review of parameter identification procedures in groundwater hydrology: The inverse problem. *Water resources research*, *22*(2), 95–108. https://doi.org/10.1029/WR022i002p00095

Yew Gan, T., Dlamini, E. M., & Biftu, G. F. (1997). Effects of model complexity and structure, data quality, and objective functions on hydrologic modeling. *Journal of Hydrology*, *192*(1–4), 81–103. https://doi.org/10.1016/S0022-1694(96)03114-9

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

Zin, I. (2002). Incertitudes et ambiguïté dans la modélisation hydrologique. (Thèse de Doctorat), Grenoble, France.