Khatami Sina (Orcid ID: 0000-0003-1149-5080)
Peel Murray Cameron (Orcid ID: 0000-0002-3255-3692)
Peterson Tim J. (Orcid ID: 0000-0002-1885-0826)
Western Andrew William (Orcid ID: 0000-0003-4982-146X)

Confidential manuscript submitted to Water Resources Research

Flux Mapping: a new approach to evaluating model process representation under uncertainty

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

[1] Department of Infrastructure Engineering, University of Melbourne, Parkville, Victoria, 3010, Australia

Corresponding author: Sina Khatami (sina.khatami@unimelb.edu.au)

Key Points:

- Characterised different facets of model-equifinality and discussed them within the context of conceptual hydrological modelling.
- Introduced the new model evaluation method of Flux Mapping to explore model behaviour, 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.

Abstract

Uncertainty analysis is an integral part of any scientific modelling, particularly within the domain of hydrological sciences given the various types and sources of uncertainty. At the centre of uncertainty rests the concept of *equifinality*, i.e. reaching a given endpoint (*finality*) through different pathways. The operational definition of equifinality in hydrological modelling 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, i.e. 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 behaviour that cannot be captured by conventional model evaluation methods. We discuss the advantages of flux space, as a sub-space 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 modelling of open complex systems under uncertainty.

1 Introduction

Understanding, modelling 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 modelling and prediction of real-world hydrological processes are the issues of: scale (Blöschl & Sivapalan, 1995) and commensurability (Beven, 2012b, p. 245) between observed and modelled 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); modelling hydrological responses to change (Schaefli et al., 2011) and resilience to disturbance (Peterson & Western, 2014; Peterson et al., 2014); "*numerical daemons*" of conceptual hydrological modelling (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 modelling. 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 modelling, for a given criterion of reproducibility (e.g. an objective function), a number of rainfall-runoff models (i.e. distributed or lumped, process-based or black-box, etc. such as TOPMODEL, HBV, Sacramento, etc.) 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 generalised sensitivity (Hornberger & Spear, 1981) and GLUE (Generalised Likelihood Uncertainty Estimation, (Beven & Binley, 1992)).

In this paper we argue that equifinality—like uncertainty—is a multi-faceted 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 modelling of hydrological systems under uncertainty (Khatami et al., under review). While facets of model-equifinality are not mutually exclusive and in fact are intertwined, each facet underscore a particular aspect (sub-space) 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 characterise model process-representation. Using a modelling experiment (section 3.4), we demonstrate how flux mapping—analysing model-equifinality through the lens of model internal fluxes instead of model parameters—provides new insights into model internal behaviour and process-representation, which cannot be (easily) captured/characterised using conventional model evaluation schemes (e.g. objective functions, dotty plots, and parameter distributions) (sections 4 and 5.1). In other words, reprojecting model behaviour (e.g. response surface) from parameter space to the flux space can give new insights into model internal behaviour 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 behaviour 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) towards a particular hypothesis for explaining a given phenomenon. Theoretically, to explain real-world processes, MWH is a never-ending process within which *hypotheses*, i.e. 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 towards an approximation of the actual reality. The value and significance of pursuing

MWH in hydrological modelling 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 modelling to any field of scientific modelling dealing with conceptual modelling 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 modelling 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, non-uniqueness, under-determination or indeterminacy, system convergence, etc (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 make the case that there is more to model-equifinality than parameter uncertainty (distribution) by characterising 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 behaviour (e.g. discharge hydrograph) could be seen as MWH; with each model structure *representing* catchment behaviour differently, and hence equifinality of model structures. Various hydrological modelling 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 utilised 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 judgements and preferences of modellers (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, i.e. which conceptual model fits better to our perceptual model. In fact, the choice of model structure, like other subjective decisions in modelling e.g. the choice of objective function (Crochemore et al., 2015) (discussed below), is often an act of *will* (i.e. modeller'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 modelling approaches. Different modelling approaches, such as a top-down vs bottom-up, process-based vs black box, distributed vs lumped, may lead to similar results in a given modelling 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., 2017; Kelleher et al., 2015; Kirchner, 2016; Tang & Zhuang, 2008; Teweldebrhan et al., 2018b; 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, i.e. 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 e.g. Bayesian approach where different degrees of belief are assigned to the sampled parameter sets, approaches where parameter sets below a certain threshold of acceptability or outside particular limits of acceptability are rejected (e.g. GLUE) (Beven, 2009; Vrugt & Beven, 2018). Regardless of the approach, parameter uncertainty is typically then expressed in terms of likelihood, i.e. a parameter distribution.

In the above, parameter equifinality is determined based on the model performance—i.e. the value of objective (or likelihood) function(s)—and the physical significance/plausibility of the so-called equifinal/behavioural parameter sets is often not examined. While, typically expressed probabilistically, parameter equifinality could be represented in other non-probabilistic 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 5-A), parameter INSC (Interception Store Capacity) can vary between 0 and 20. INSC = 0 means the interception storage of the model is non-existent, 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 non-existent 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 → 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 *modelling 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 modeller's decisions. Objective functions characterise the model performance as an aggregated measure of the matching between modelled 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), whether aggregated or distributed (Koch et al., 2017; Koch et al., 2016). Performance metrics reduce the complex behaviour of a system—often the integrated response of the catchment system, i.e. 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 behaviour, e.g. discharge hydrograph) could be the result of different objective functions with different mathematical structures. Although it is possible to improve metrics e.g. 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 & McMillan, 2015; Westerberg et al., 2016). 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—e.g. 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, i.e. 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, e.g. equifinality of model predictions from different stochastic realisations 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, 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, e.g. model runoff fluxes mimic catchment runoff generation mechanisms. Also, under a given model conditioning, i.e. 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 (sub-space) of the overall model space. That is, each facet (or their combinations) could be utilised 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 behaviour and to generate MWH, even if only model parameters are perturbed.

3 Flux Mapping: an approach to evaluate model internal dynamics

In this section we develop a method called *Flux Mapping* for evaluating model behaviour, based on the concept of model internal flux equifinality. We demonstrate that the new tool of *Flux Maps* can give new insights into model behaviour 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 potential evapotranspiration 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 AWAP (www.bom.gov.au/jsp/awap/) daily 5km grids (Jones et al., 2009), and also estimated areal potential evapotranspiration (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 modelling 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 optimisation algorithm of Shuffled Complex Evolution (SCE-UA) (Duan et al., 1992). For each case, the SCE-UA 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 $\leq$ NSE $\leq$ 0.88, were selected. It should be stressed that this pre-calibration step is not a part of the modelling 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 1-A 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 $\leq$ NSE $\leq$ 0.825) at a given catchment. It clearly shows the value of the objective function is not a reliable measure of model internal behaviour. For a large ensemble of model runs, equifinal fluxes can be summarised and visualised based on the percentage of their (volumetric) contribution to the total simulated Q on a plot we name a *Flux Map*. Figure 1-B 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 1-A.

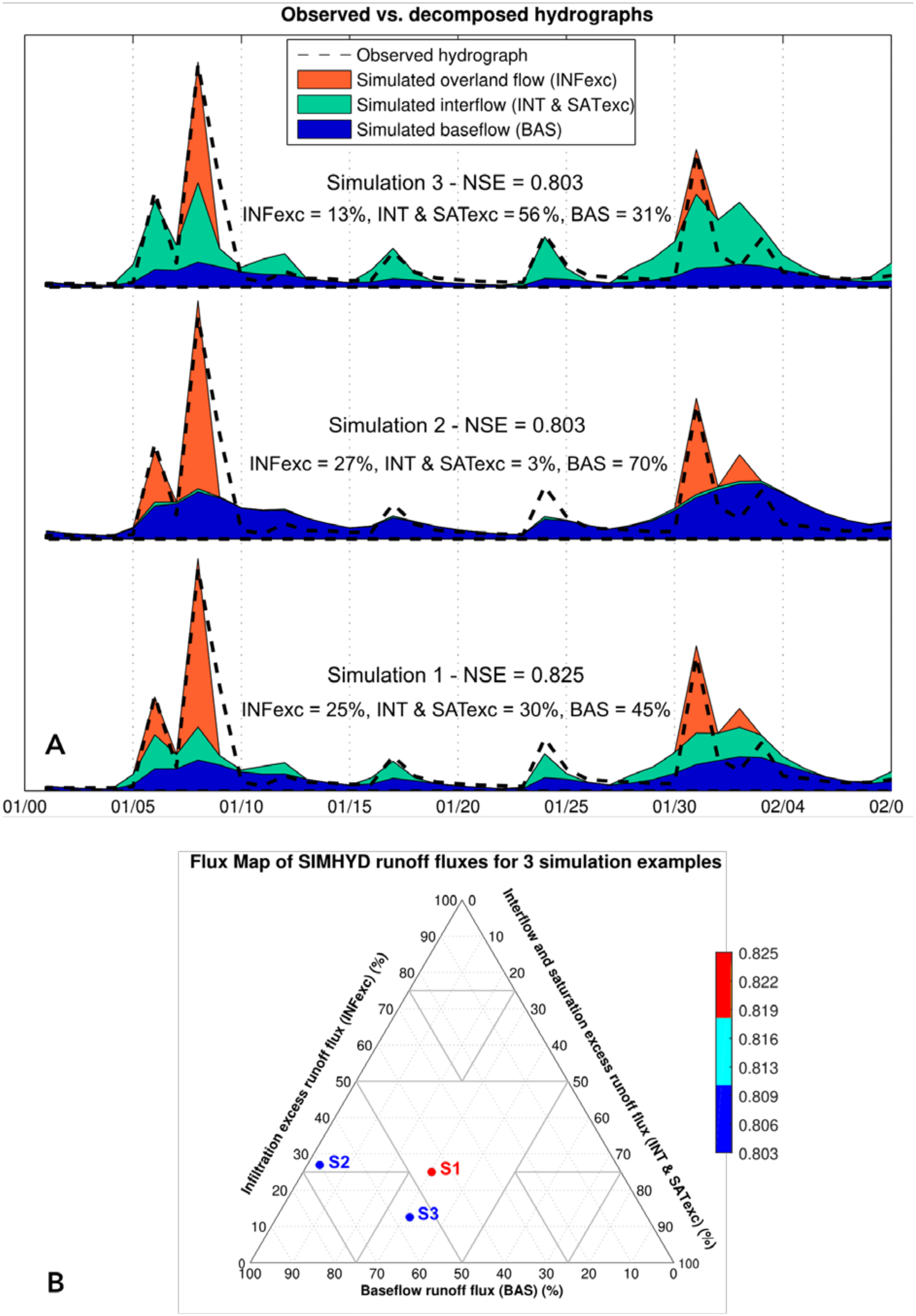

Observed vs. decomposed hydrographs
Observed hydrograph
Simulated overland flow (INFexc)
Simulated interflow (INT & SATexc)
Simulated baseflow (BAS)
Simulation 3 - NSE = 0.803
INFexc = 13%, INT & SATexc = 56%, BAS = 31%
Simulation 2 - NSE = 0.803
INFexc = 27%, INT & SATexc = 3%, BAS = 70%
Simulation 1 - NSE = 0.825
INFexc = 25%, INT & SATexc = 30%, BAS = 45%
A
01/00
01/05
01/10
01/15
01/20
01/25
01/30
02/04
02/0
Flux Map of SIMHYD runoff fluxes for 3 simulation examples
Infiltration excess runoff flux (INFexc) (%)
Interflow and saturation excess runoff flux (INT & SATexc) (%)
Baseflow runoff flux (BAS) (%)
S1
S2
S3
0.825
0.822
0.819
0.816
0.813
0.809
0.806
0.803
B

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) summarised on a flux map (B), colour-coded based on the simulation performance (NSE value). The triangle (B) represents the plausible flux space for a model with three runoff fluxes.

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 co-dominating 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 2-by-2 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 pre-calibration) 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 × {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 behaviour (Class 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×SCE-NSE) and *relaxed* (0.85×SCE-NSE) thresholds (Figures 3 and 4), colour-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 2-by-2 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 P = 950-1353 mm, mean annual Q = 162-335 mm, mean annual APET = 1222-1532 mm, and annual runoff coefficient 0.17-0.25. The 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.

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

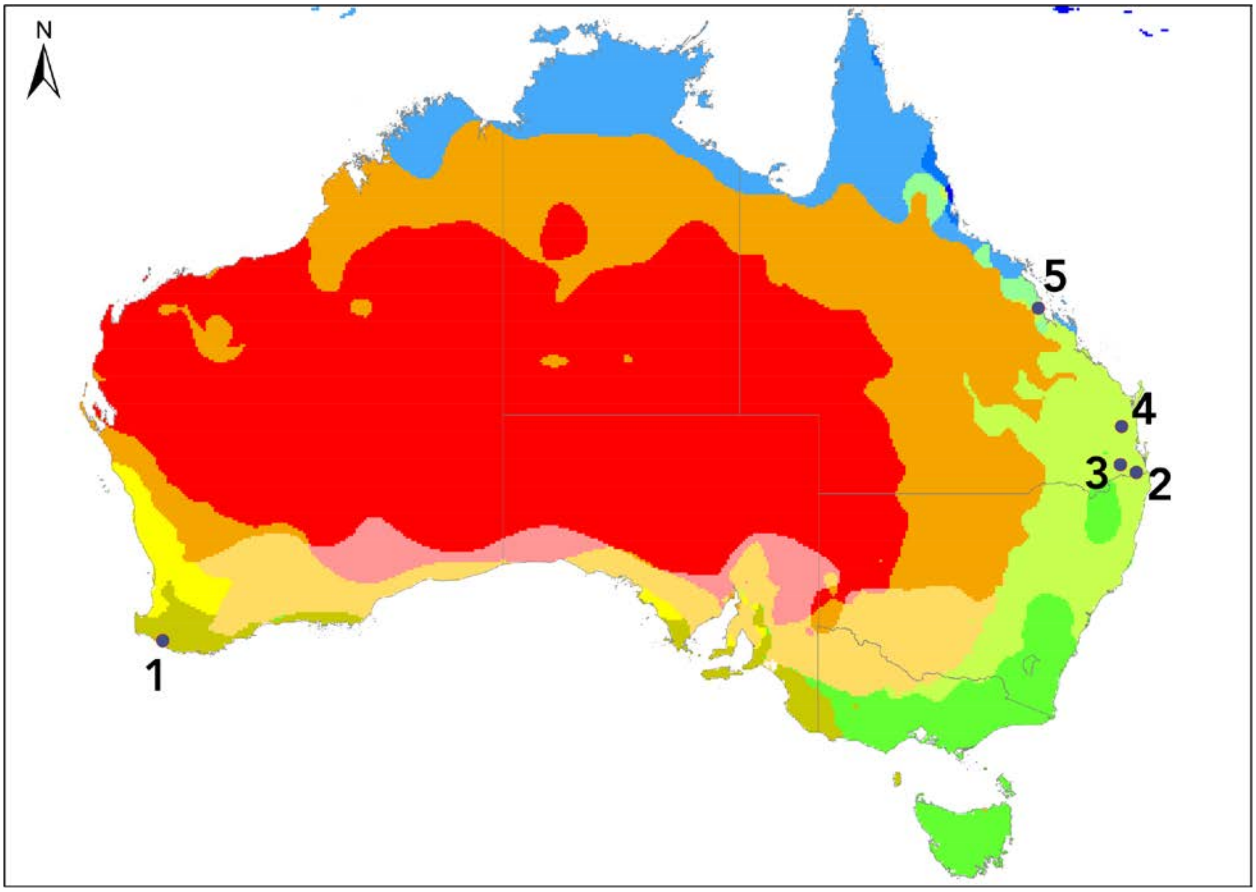


Table 1. Catchments summaries of the examples presented in modelling experiments

Catchment No.	Corresponding figures (Class)	Catchment						
		Name	Location	Area (km^2)	Mean annual total P (mm)	Mean annual total Q (mm)	Mean annual APET (mm)	Annual runoff ratio
1	**Figs 3-A1 & 3-A2 (Class I)**	Dombakup Brook at Malimup Track	Western Australia (115.98° E 34.58° S)	125.09	1130.99	232.63	1222.41	0.21

2	**Figs 3-B1 & 3-B2 (Class II)**	Albert River at Lumeah	Queensland (153.05° E 28.05° S)	167.39	1353.61	335.14	1451.17	0.25
3	**Figs 3-C1 & 3-C2 (Class III)**	Bremer River at Adams Bridge	Queensland (152.51° E 27.83° S)	126.09	951.07	162.19	1506.55	0.17
4	**Fig 3-D1 & 3-D2 (Class IV)**	Kandanga Creek at Hygait	Queensland (152.65° E 26.39° S)	170.78	1135.18	277.98	1532.49	0.24
5	**Fig 4-A, 4-B, & 4-C (Class IV)**	Carmila Creek at Carmila	Queensland (149.399° E 21.915° S)	83.8	1275	538	1736	0.42

4.1 Class I

The point cloud on the flux map in Figures 3A 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 3-A1) 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 3-A2) 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.

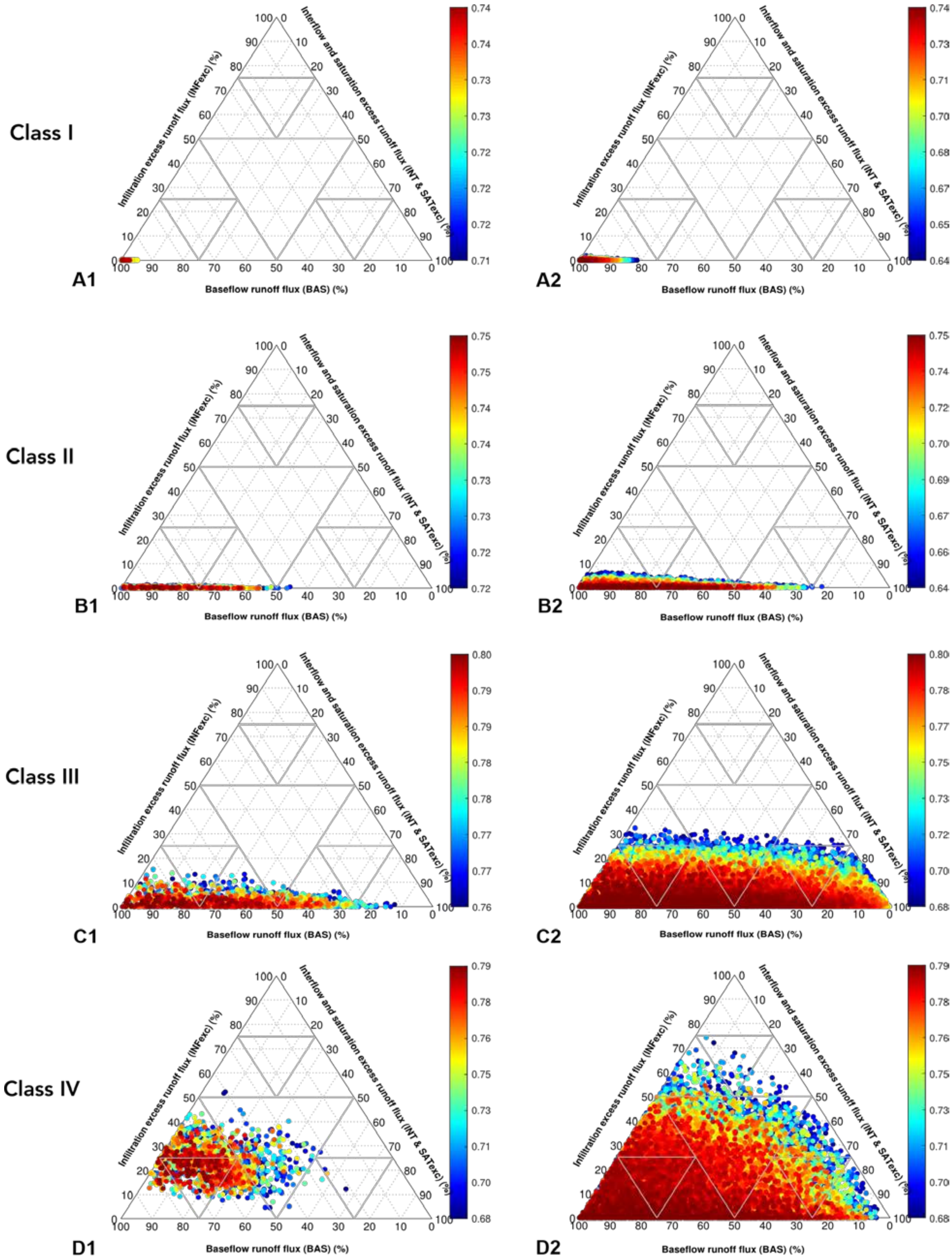

Class I
Class II
Class III
Class IV
A1
A2
B1
B2
C1
C2
D1
D2
Infiltration excess runoff flux (INFexc) (%)
Interflow and saturation excess runoff flux (INT & SATexc) (%)
Baseflow runoff flux (BAS) (%)

Figure 3. This figure represents how different model internal behaviour, represented on flux maps, can emerge within a certain range of model performance. Flux maps for 4 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. Colour bar represent the model performance in terms of NSE value. (A) Class I flux maps for strict threshold (0.95×SCE-NSE=0.71 ≤ Ensemble-NSE=0.75 ≤ SCE-NSE=0.75) with 252 equifinal simulations (A1), and relaxed threshold (0.85×SCE-NSE=0.64 ≤ Ensemble-NSE ≤ SCE-NSE) with 3036 equifinal simulations (A2). (B) Class II flux maps for strict threshold (0.95×SCE-NSE=0.71 ≤ Ensemble-NSE=0.75 ≤ SCE-NSE=0.75) with 592 equifinal simulations (B1), and relaxed threshold (0.85×SCE-NSE=0.64 ≤ Ensemble-NSE ≤ SCE-NSE) with 8476 equifinal simulations (B2). (C) Class III flux maps for strict threshold (0.95×SCE-NSE=0.74 ≤ Ensemble-NSE=0.78 ≤ SCE-NSE=0.78) with 3644 equifinal simulations (C1), and relaxed threshold (0.85×SCE-NSE=0.66 ≤ Ensemble-NSE ≤ SCE-NSE) with 47702 equifinal simulations (C2). (D) Class IV flux maps for strict threshold (0.95×SCE-NSE=0.71 ≤ Ensemble-NSE=0.75 ≤ SCE-NSE=0.75) with 1270 equifinal simulations (D1), and relaxed threshold (0.85×SCE-NSE=0.64 ≤ Ensemble-NSE ≤ SCE-NSE) with 65126 equifinal simulations (D2). Catchment summaries are presented in Table 1.

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 under-estimating the flux equifinality.

4.2 Class II

The flux map for this class (Figures 3B) is mainly constrained around a line which 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 3-B1), the baseflow contribution to total runoff volume is around 45-90%; and as the threshold relaxes (Figure 3-B2), 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 (towards 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. Colour-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 artefact 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) make it difficult to assign likelihood to model runs. Marginal distributions (Table S1) and colour-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; i.e. *possible* flux contributions given the modelling 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 3C) is an extension of Class II to three active fluxes, i.e. 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 3-C1), 0.74 ≤ 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 3-C2), 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, i.e. all three fluxes can vary widely. Even for the strict threshold, there are 1270 acceptable model runs with a moderately wide range (≥ 50% variation) of possible flux contributions as shown in Figure 3-D1 (compare this case with Figure 3-B1). 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.

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 3-A1), acceptable fluxes are remarkably scattered—occupying about 25% of the flux map plane, shown in Figure 4-A. 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 4-B). 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 4-C). 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.

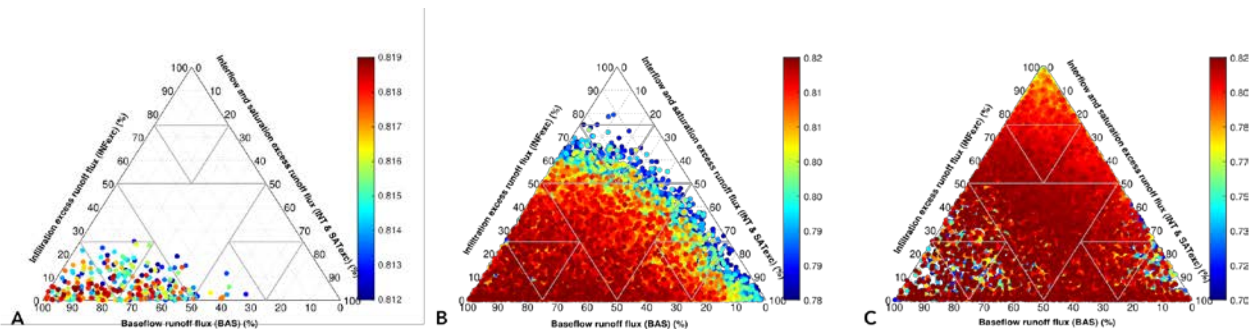


Figure 4. Class IV flux maps for very strict threshold (0.99×SCE-NSE=0.80 ≤ Ensemble-NSE=0.80 ≤ SCE-NSE) with 287 equifinal simulations (A), strict threshold (0.95×SCE-NSE=0.77 ≤ Ensemble-NSE ≤ SCE-NSE) with 69,090 equifinal simulations (B), and relaxed threshold (0.85×SCE-NSE=0.69 ≤ Ensemble-NSE ≤ SCE-NSE)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

5 Discussion

5.1 Flux mapping and model evaluation: flux space vs. 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 behaviour that flux mapping provides, cannot (easily) be characterised by common model evaluation tools, like dotty plots and parameter distributions. The dotty plot is a tool to visualise 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 characterise the model behaviour, particularly its internal dynamics and process-representation; hence the flux map is a valuable complement to these tools. For instance, the differences in model behaviours 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).

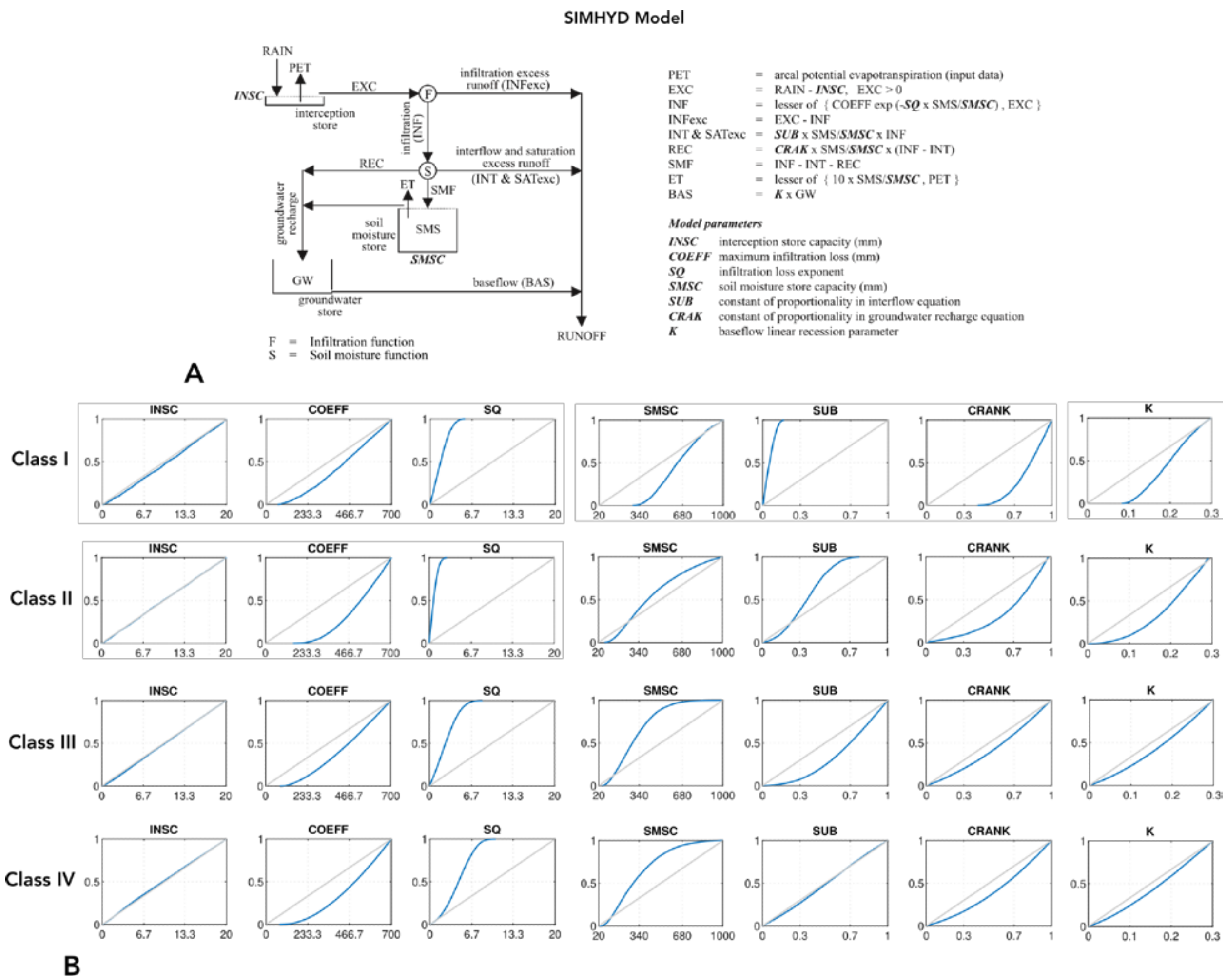


Figure 5. (A) Schematic illustration of the SIMHYD model structure together with the description of the model's internal fluxes and parameters (derived from Peel et al., 2000). (B) Marginal (cumulative) distribution of model parameters (grey is the prior and blue is the posterior distributions) corresponding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

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 5-B) are almost identical except for parameter SUB, which controls the INT & SATexc flux; while Figure 3 clearly shows 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 (towards Class IV) the BAS flux becomes less constrained. That said, while BAS flux is (almost) equally variable for both Classes III and IV (varying between 0 to 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, model structure, etc.) 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 behaviour into the flux space, i.e. 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, *modelling uncertainties*—the interplay between data information content, modelling framework and model structure, performance metrics, and modeller's decisions/understanding—are lumped into and evaluated within the parameter space. Flux mapping can illustrate the impact of *modelling uncertainties*, regardless of the source and nature, on model internal behaviour and process-representation.

In this study, flux maps could be interpreted as a new visualisation 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, summarise, and visualise model internal behaviour. 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 modelling 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 multiple working hypotheses based on model process-representation.

Furthermore, runoff flux space has lower dimensions than parameter space. For instance, in the case of SIMHYD, visualising 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 realisations are the product of perturbing model parameters, other components of the model could also be used for generating flux maps, e.g. 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. *Modelling 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, i.e. 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 3A) 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 modelling setups (given modeller's judgements/decisions) are in (dis)agreement with each other and to what extent; and eventually pose an essential question of conceptual modelling, i.e. how well model output (and internal behaviour) 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—i.e. 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; i.e., 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 behaviour 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 evapotranspiration (ET) process representation within three conceptual rainfall-runoff models, by comparing the simulated actual ET (AET) with measurements. They observed some unrealistic behaviour 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 behaviour/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 modelling scenarios (as MWH) could then be evaluated in terms 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, e.g. salt load (Nathan & Mudgway, 1997), 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 modeller'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, i.e. 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 *modelling* *uncertainties* on model behaviour; 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 emphasise 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) behaviour of their corresponding conceptual models under all *modelling 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 modelling 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 modelling 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 modelling artefact due to *modelling uncertainties*. For instance, to what extent a constrained or space-filling flux map is due to catchment characteristics or due to modelling setup. What occurred in reality in a particular period would be mapped to a single unknown *true point* on the flux map. Model-equifinality, e.g. flux equifinality, results in a point cloud which does not necessarily encompass the *true point* (i.e. reality). What we can hope for is to refine our hypotheses while encompassing the *truth* by improving model realism—by improving model structures, evaluation schemes, 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 *behaviour* under *modelling uncertainties*) and hypothesis generating, could be further extended to other domains. It could also be used to improve understanding/modelling in data-poor catchments/regions (Davtalab et al., 2017) and in regional generalization (also known as regionalisation 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 modelling concerned with understanding/modelling open complex systems in the face of uncertainties. Real-world processes, 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 e.g. data assimilation(Alvarez-Garreton et al., 2014;

Teweldebrhan et al., 2018a) and flood forecasting (Alvarez-Garreton et al., 2015), catchment classification (Kelleher et al., 2015; Knoben et al., 2018; Sawicz et al., 2014), socio-hydrological 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 modelling (Billari et al., 2006; Madani et al., 2014), sustainability transitions (de Haan et al., 2016; Moallemi & Köhler, 2019) and exploratory modelling (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 modelling/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 modelling. 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 behaviour 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, i.e. 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 modelling dealing with conceptual modelling of open complex systems under uncertainty. We argued that equifinality plays a central role in scientific modelling, particularly within the paradigm of multiple working hypotheses.

Acknowledgments, Samples, and Data

The authors gratefully acknowledge the support of the University of Melbourne and Australian Government in carrying out this research; Sina Khatami supported by Melbourne International Research and Fee Remission Scholarships (MIRS and MIFRS), Murray Peel 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 dataset 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*(1), 378-390. doi:doi: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*. doi:doi: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*(1), W01403. doi: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. doi: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. *Hydrol. Earth Syst. Sci., 19*(4), 1659-1676. doi: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*(7), 6135-6153. doi:doi:10.1002/2013WR014898

Beck, M. B. (1987). Water quality modeling: A review of the analysis of uncertainty. *Water resources research, 23*(8), 1393-1442. doi:10.1029/WR023i008p01393

Beck, M. B. (2002). *Environmental foresight and models: a manifesto*: Elsevier.

Bennett, N. D., Croke, B. F. W., Guariso, G., Guillaume, J. H. A., Hamilton, S. H., Jakeman, A. J., . . . Andreassian, V. (2013). Characterising performance of environmental models. *Environmental Modelling & Software, 40*, 1-20. doi: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. doi:doi:10.1002/wat2.1311

Bethke, C. M. (1992). The question of uniqueness in geochemical modeling. *Geochimica et Cosmochimica Acta, 56*(12), 4315-4320. doi:http://dx.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. doi:http://dx.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. doi:http://dx.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.
Beven, K. (2012b). *Rainfall-runoff modelling: the primer* (2nd ed.). Chicheseter, UK: John Wiley & Sons, Ltd.
Beven, K. (2016). Facets of uncertainty: epistemic uncertainty, non-stationarity, likelihood, hypothesis testing, and communication. *Hydrological Sciences Journal, 61*(9), 1652-1665. doi: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. doi:10.1002/hyp.3360060305
Beven, K., & Binley, A. (2014). GLUE: 20 years on. *Hydrological Processes, 28*(24), 5897-5918. doi: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*(11), W11801. doi: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. doi:10.1002/hyp.7963
Beven, K. J., Aspinall, W. P., Bates, P. D., Borgomeo, E., Goda, K., Hall, J. W., . . . Watson, M. (2017). Epistemic uncertainties and natural hazard risk assessment. 2. What should constitute good practice? *Nat. Hazards Earth Syst. Sci. Discuss., 2017*, 1-25. doi: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*(12), W00B16. doi:10.1029/2007WR006726
Blöschl, G., & Sivapalan, M. (1995). Scale issues in hydrological modelling: A review. *Hydrological Processes, 9*(3-4), 251-290. doi:10.1002/hyp.3360090305
Boughton, W. C. (2007). Effect of data length on rainfall–runoff modelling. *Environmental Modelling & Software, 22*(3), 406-413. doi:http://dx.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*(12). doi:doi: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. doi: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*(5). doi:doi: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), 242-266. doi: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. doi: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. doi: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. doi: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).

Chiew, F. H. S., Stewardson, M. J., & McMahon, T. A. (1993). Comparison of six rainfall-runoff modelling approaches. *Journal of Hydrology, 147*(1), 1-36. doi: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. doi: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*(10), W10510. doi: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*(9).

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*(11).

Clark, M. P., Nijssen, B., Lundquist, J. D., Kavetski, D., Rupp, D. E., Woods, R. A., . . . Rasmussen, R. M. (2015a). A unified approach for process-based hydrologic modeling: 1. Modeling concept. *Water resources research, 51*(4), 2498-2514. doi:doi:10.1002/2015WR017198

Clark, M. P., Nijssen, B., Lundquist, J. D., Kavetski, D., Rupp, D. E., Woods, R. A., . . . Marks, D. G. (2015b). A unified approach for process-based hydrologic modeling: 2. Model implementation and case studies. *Water resources research, 51*(4), 2515-2542. doi:10.1002/2015wr017200

Clark, M. P., Slater, A. G., Rupp, D. E., Woods, R. A., Vrugt, J. A., Gupta, H. V., . . . Hay, L. E. (2008). Framework for Understanding Structural Errors (FUSE): A modular framework to diagnose differences between hydrological models. *Water resources research, 44*(12). doi:doi:10.1029/2007WR006735

Crochemore, L., Perrin, C., Andréassian, V., Ehret, U., Seibert, S. P., Grimaldi, S., . . . Paturel, J.-E. (2015). Comparing expert judgement and numerical criteria for hydrograph evaluation. *Hydrological Sciences Journal, 60*(3), 402-423. doi: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), 05017011. doi:doi:10.1061/(ASCE)HE.1943-5584.0001525

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*(Supplement C), 279-292. doi:https://doi.org/10.1016/j.envsoft.2016.05.019

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*(6), 4770-4781. doi:doi:10.1002/2014WR016416

Duan, Q., Sorooshian, S., & Gupta, V. (1992). Effective and efficient global optimization for conceptual rainfall-runoff models. *Water resources research, 28*(4), 1015-1031. doi: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. doi: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. doi: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*. doi:doi:10.1002/hyp.13367

Euser, T., Winsemius, H. C., Hrachowitz, M., Fenicia, F., Uhlenbrook, S., & Savenije, H. H. G. (2013). A framework to assess the realism of model structures using hydrological signatures. *Hydrol. Earth Syst. Sci., 17*(5), 1893-1912. doi: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*(11). doi: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. doi:10.1002/hyp.9726

Flyvbjerg, B., Skamris Holm, M. K., & Buhl, S. L. (2006). Inaccuracy in Traffic Forecasts. *Transport Reviews, 26*(1), 1-24. doi: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*(**5**), 3392-3408. doi:doi: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*(3), 1820-1846. doi: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. *Hydrol. Earth Syst. Sci., 18*(12), 4839-4859. doi: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*(4), 2253-2273. doi: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. doi: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*(1), 435-454. doi: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. doi:http://dx.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*(6), 5351-5359. doi:doi: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. doi:doi: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. doi: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. doi:10.1029/WR019i001p00269

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. doi: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. doi: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*. doi:10.1029/2018ms001526

Holländer, H. M., Blume, T., Bormann, H., Buytaert, W., Chirico, G. B., Exbrayat, J. F., . . . Flühler, H. (2009). Comparative predictions of discharge from an artificial catchment (Chicken Creek) using sparse data. *Hydrol. Earth Syst. Sci., 13*(11), 2069-2094. doi: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., . . . Gascuel-Odoux, C. (2014). Process consistency in models: The importance of system signatures, expert knowledge, and process complexity. *Water resources research, 50*(9), 7445-7469. doi: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. doi: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*(4), 233. doi: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. doi: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*(11). doi: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*(3), W03407. doi: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. *Hydrol. Earth Syst. Sci., 21*(7), 3325-3352. doi: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*(6), 4109-4136. doi:doi: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.

Khatami, S., Peel, M. C., Peterson, T. J., & Western, A. W. (under review). Equifinality, Uncertainty, Multiple Working Hypotheses, Philosophy of Science, and Hydrological Systems *Water resources research.*

Khazaei, B., Khatami, S., Alemohammad, S. H., Rashidi, L., Wu, C., Madani, K., . . . Aghakouchak, A. (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. doi: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. *Hydrol. Earth Syst. Sci. Discuss., 20*(1), 299-328. doi: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*(7), 5088-5109. doi:doi: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. doi: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*(10), 5430-5452. doi:doi:10.1002/2015JD024482

Konikow, L. F., & Bredehoeft, J. D. (1992). Ground-water models cannot be validated. *Advances in water resources*, *15*(1), 75-83. doi:http://dx.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. doi: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. doi: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. doi: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), 93-103. doi: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. doi: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).

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. doi:http://dx.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. doi: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*(Supplement C), 271-287. doi: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. doi: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*. doi: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*. doi: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. doi:http://dx.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. doi: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. doi:http://dx.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. doi:doi: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*(1), 524-538. doi:10.1002/2014WR015895

Newman, A. J., Clark, M. P., Craig, J., Nijssen, B., Wood, A., Gutmann, E., . . . Arnold, J. R. (2015). Gridded Ensemble Precipitation and Temperature Estimates for the Contiguous United States. *Journal of Hydrometeorology, 16*(6), 2481-2500. doi: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. doi:http://dx.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. doi: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. doi: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. *Hydrol. Earth Syst. Sci., 11*(5), 1633-1644. doi: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. doi:http://dx.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*(4), 2993-3009. doi: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*(4), 3010-3029. doi: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. doi: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. doi: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. doi:https://doi.org/10.1016/j.jhydrol.2011.11.055

Quine, W. V. (1975). On Empirically Equivalent Systems of the World. *Erkenntnis (1975-), 9*(3), 313-328.

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*(5), W05521. doi:10.1029/2009WR008328

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*. doi: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. *Hydrol. Earth Syst. Sci.*, *22*(8), 4583-4591. doi:10.5194/hess-22-4583-2018

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

Sawicz, K. A., Kelleher, C., Wagener, T., Troch, P., Sivapalan, M., & Carrillo, G. (2014). Characterizing hydrologic change through catchment classification. *Hydrol. Earth Syst. Sci.*, *18*(1), 273-285. doi: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. doi:10.1002/hyp.10972

Schaefli, B., & Gupta, H. V. (2007). Do Nash values have value? *Hydrological Processes, 21*(15), 2075-2080. doi: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. *Hydrol. Earth Syst. Sci.*, *15*(2), 635-646. doi:10.5194/hess-15-635-2011

Seibert, J. (2001). On the need for benchmarks in hydrological modelling. *Hydrological Processes*, *15*(6), 1063-1064. doi:doi: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), 23-21-23-14. doi:10.1029/2001WR000978

Sivakumar, B. (2000). Chaos theory in hydrology: important issues and interpretations. *Journal of Hydrology, 227*(1–4), 1-20. doi:http://dx.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.

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. doi: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. doi:doi: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: Biogeosciences*, *113*(G4), G04010. doi:10.1029/2008JG000757

Teweldebrhan, A., Burkhart, J., Schuler, T., & Xu, C.-Y. (2018a). Improving the Informational Value of MODIS Fractional Snow Cover Area Using Fuzzy Logic Based Ensemble Smoother Data Assimilation Frameworks. *Remote Sensing, 11*(1), 28.

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

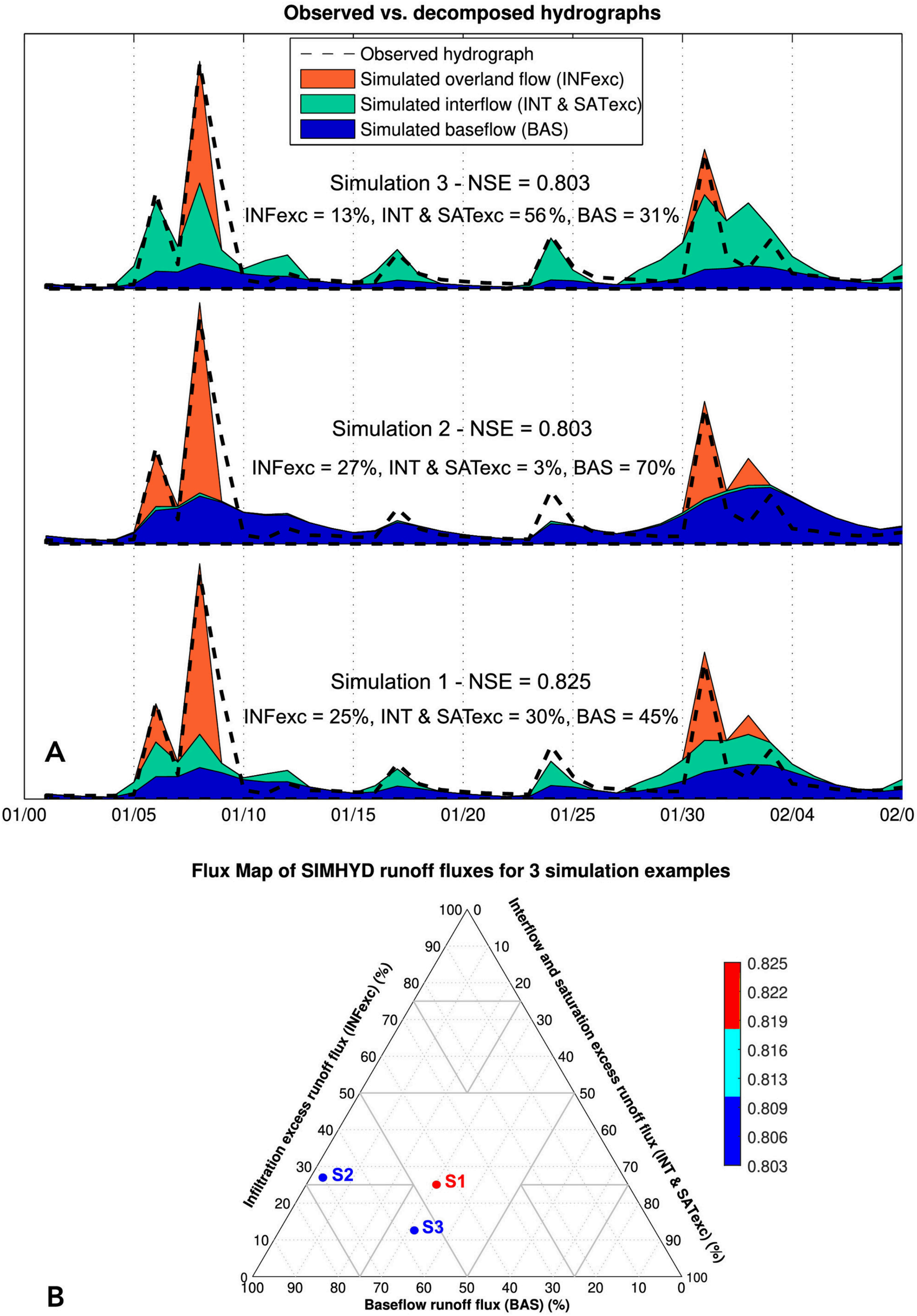
Tian, Y., Nearing, G. S., Peters-Lidard, C. D., Harrison, K. W., & Tang, L. (2016). Performance Metrics, Error Modeling, and Uncertainty Quantification. *Monthly Weather Review*, *144*(2), 607-613. doi:10.1175/mwr-d-15-0087.1

Tolson, B. A., & Shoemaker, C. A. (2008). Efficient prediction uncertainty approximation in the calibration of environmental simulation models. *Water resources research, 44*(4). doi:10.1029/2007wr005869

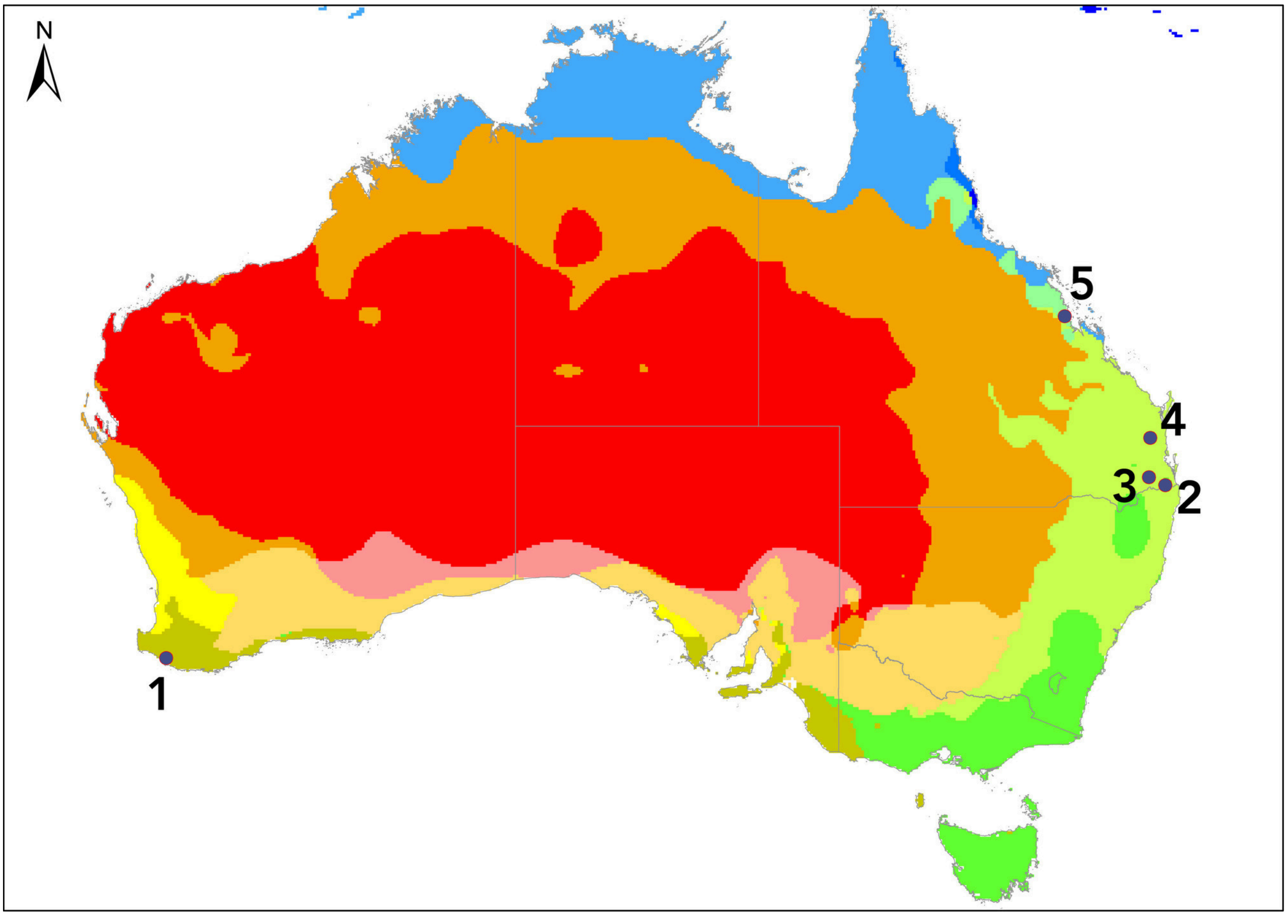
Turner, M. (2012). *Hydrologic Reference Station Selection Guidelines*. Retrieved from Melbourne, Australia: http://www.bom.gov.au/water/hrs/media/static/papers/Selection_Guidelines.pdf

van Emmerik, T. H. M., Li, Z., Sivapalan, M., Pande, S., Kandasamy, J., Savenije, H. H. G., . . . Vigneswaran, S. (2014). Socio-hydrologic modeling to understand and mediate the competition for water between

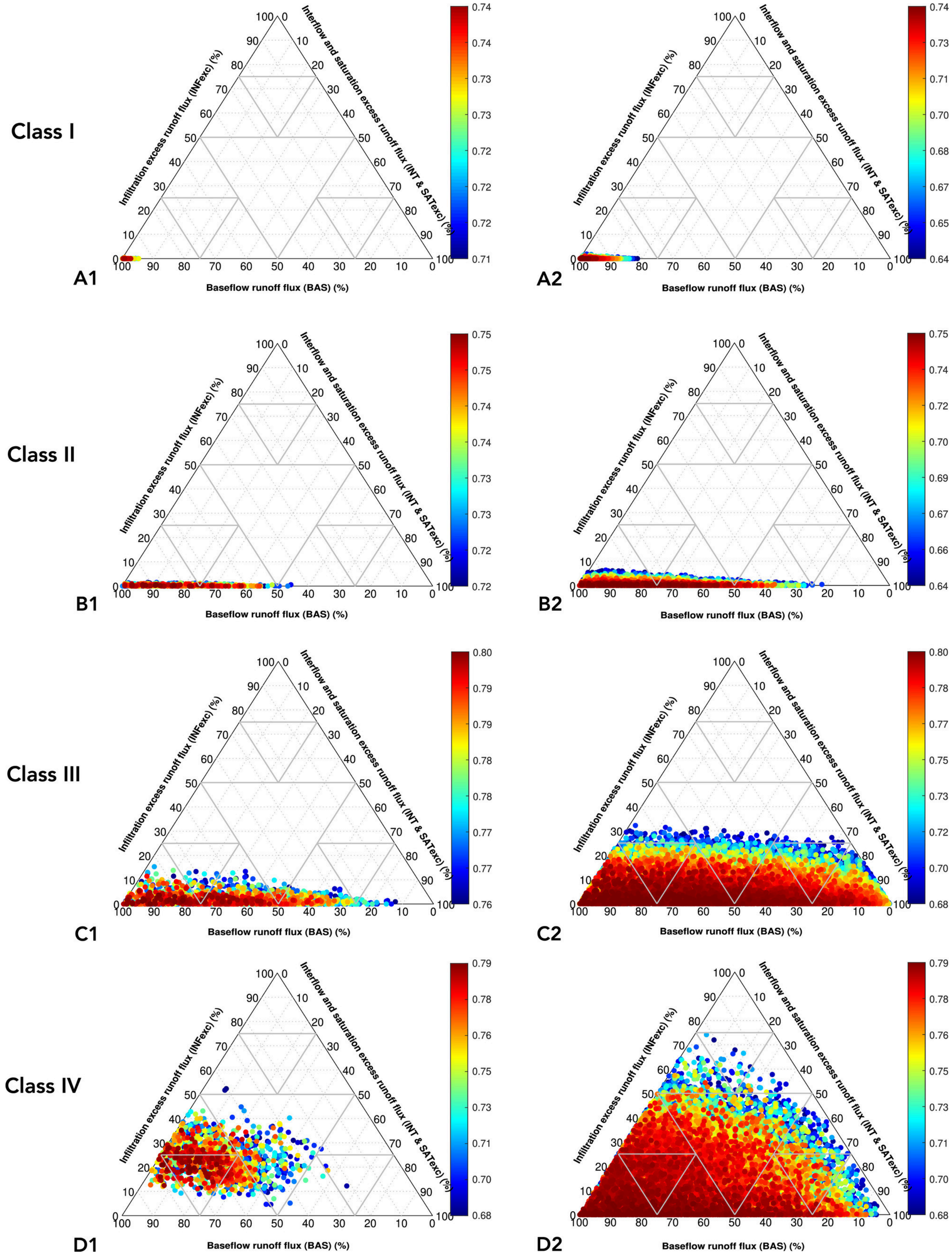
agriculture development and environmental health: Murrumbidgee River basin, Australia. *Hydrol. Earth Syst. Sci., 18*(10), 4239-4259. doi:10.5194/hess-18-4239-2014
van Ruijven, B., de Vries, B., van Vuuren, D. P., & van der Sluijs, J. P. (2010). A global model for residential energy use: Uncertainty in calibration to regional data. *Energy, 35*(1), 269-282. doi:https://doi.org/10.1016/j.energy.2009.09.019
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. doi: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*(12), W00B09. doi: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. doi:10.1080/02626667.2017.1356926
Westerberg, I. K., & McMillan, H. K. (2015). Uncertainty in hydrological signatures. *Hydrol. Earth Syst. Sci., 19*(9), 3951-3968. doi: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*(3), 1847-1865. doi:doi: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. doi:doi: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. doi: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*(12). doi:doi: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. doi: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. doi:http://dx.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*(9). doi:doi:10.1029/2007WR006716
Zin, I. (2002). *Incertitudes et ambiguïté dans la modélisation hydrologique.* (Thèse de Doctorat), Grenoble, France.



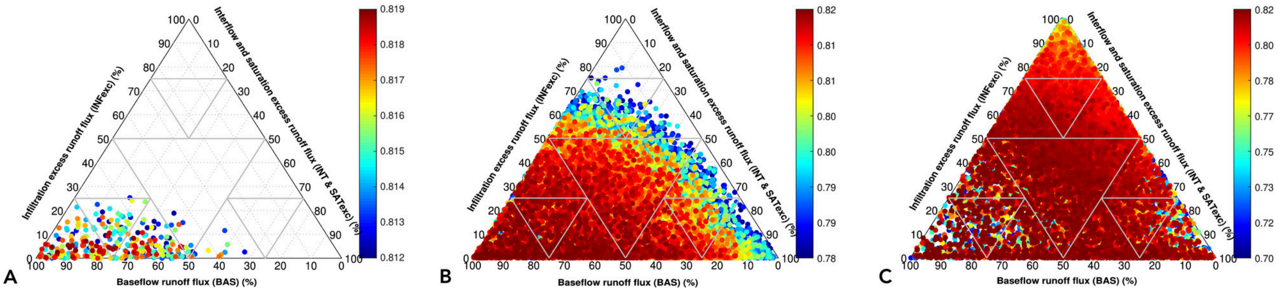
2018wr023750-f01-z-.eps



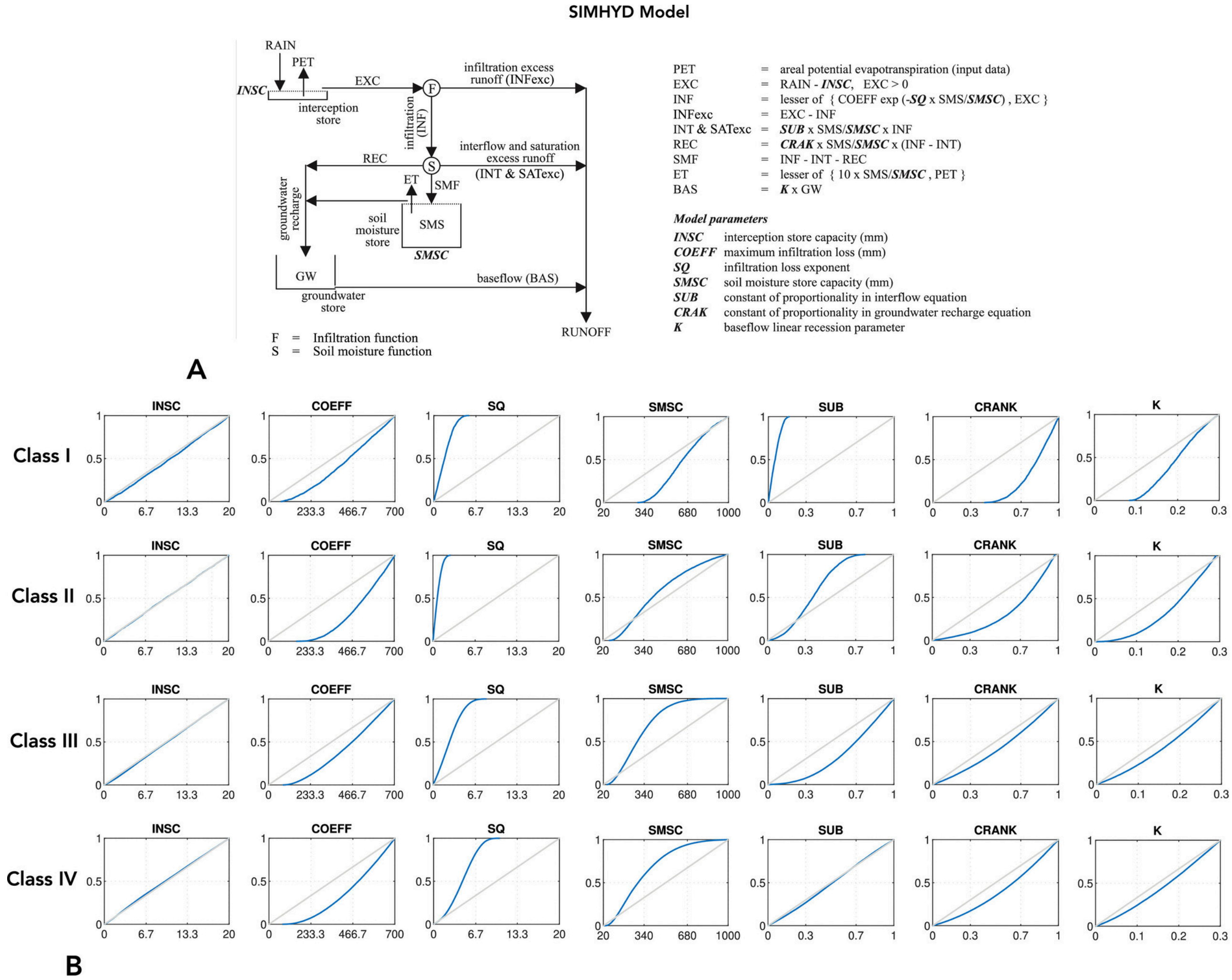
2018wr023750-f02-z-.eps



2018wr023750-f03-z-.eps



2018wr023750-f04-z-.eps



2018wr023750-f05-z-.eps