Hydrol. Earth Syst. Sci. Discuss., 12, 3945–4004, 2015
www.hydrol-earth-syst-sci-discuss.net/12/3945/2015/
doi:10.5194/hessd-12-3945-2015

Hydrology and Earth System Sciences
Open Access
Discussions

Hydrological model parameter dimensionality is a weak measure of prediction uncertainty

S. Pande[1], **L. Arkesteijn**[1], **H. Savenije**[1], **and L. A. Bastidas**[2]

[1]Department of Water Management, Delft University of Technology, Delft, the Netherlands
[2]ENERCON Services Inc., Pittsburgh Office, Murrysville, PA, USA

Received: 19 March 2015 – Accepted: 24 March 2015 – Published: 16 April 2015

Correspondence to: S. Pande (s.pande@tudelft.nl)

Published by Copernicus Publications on behalf of the European Geosciences Union.

Abstract

This paper shows that instability of hydrological system representation in response to different pieces of information and associated prediction uncertainty is a function of model complexity. After demonstrating the connection between unstable model representation and model complexity, complexity is analyzed in a step by step manner. This is done measuring differences between simulations of a model under different realizations of input forcings. Algorithms are then suggested to estimate model complexity. Model complexities of the two model structures, SAC-SMA (Sacramento Soil Moisture Accounting) and its simplified version SIXPAR (Six Parameter Model), are computed on resampled input data sets from basins that span across the continental US. The model complexities for SIXPAR are estimated for various parameter ranges. It is shown that complexity of SIXPAR increases with lower storage capacity and/or higher recession coefficients. Thus it is argued that a conceptually simple model structure, such as SIXPAR, can be more complex than an intuitively more complex model structure, such as SAC-SMA for certain parameter ranges. We therefore contend that magnitudes of feasible model parameters influence the complexity of the model selection problem just as parameter dimensionality (number of parameters) does and that parameter dimensionality is an incomplete indicator of stability of hydrological model selection and prediction problems.

1 Introduction

Reconciling models with observations is often ill-conditioned, especially when single performance measures, such as mean square errors, are used to infer models (Gupta et al., 2008). This ill-condition is often attributed to our attempt to extract higher dimensional information (about the model) from a single dimension of information given by the measure. It is therefore often recommended to select hydrological models either using multiple signatures of hydrological response or multiple objectives, the idea be-

HESSD
12, 3945–4004, 2015

ing to constrain the model selection exercise (Gupta et al., 2008). Different signatures (Sawicz et al., 2014) or multiple objectives (Gupta et al., 1998) are different measures of closeness, which when orthogonal, provide complementary pieces of information to select a better constrained model (Sivapalan et al., 2003; Winsemius et al., 2006). The constraints imposed on the model selection exercise in effect may condition the problem well.

But is the issue of ill-conditionness limited to the discourse of the number of measures used? Can we say something about the nature of conditionness first before addressing the question of how it can be ameliorated by, for example, the use of multiple signatures or objectives? A definition of ill-conditioness and the consequences of an ill-conditioned hydrological model problem are therefore needed. Renard et al. (2010) are the first to formally introduce the notion of ill-posedness in hydrological modeling and emphasized the importance of prior specification in correcting or properly conditioning ill-posed model selection problems. Their approach appears to have been motivated by the issue of non-identifiability, that not all parameters of interest are often decipherable from limited rainfall–runoff information (Beven, 2006). We ask an equivalent question and attempt to formalize what ill-conditionness means: what happens when an ill conditioned model is selected to represent the underlying hydrological system? Since it fails to exploit interesting information in the data, there is uncertainty in system representation (Gupta et al., 2008; Gupta and Nearing, 2014). Should not this uncertainty in assessing structure deficiency depend on the class of model structures which are used to assess deficiencies? The characteristics of uncertainty in system representation can then identify the consequences of ill-conditioned model selection problem and hence define ill-conditioned model selection.

We characterize uncertainty in hydrological system representation as composed of non-uniqueness and instability in system representation. Non-uniqueness in system representation (of the underlying processes) is synonymous to equifinality (Savenije, 2001). Meanwhile, instability refers to inconsistency in process representation as more information on the underlying processes is available. That is, a set of models that ap-

HESSD
12, 3945–4004, 2015

pears to be more representative on smaller pieces of information is not the best representation as more pieces of information are brought to bear (Arkesteijn and Pande, 2013). The instability of model representation is not specific to the case when multiple measures define different pieces of information as elucidated by Gupta et al. (2008). Instability may exist even when using a single measure of performance but when the information content increases as the amount of available data increases. This is equivalent to suggesting that information content of smaller datasets of similar lengths is dissimilar. Instability can then be rephrased as a changing set of good system representations (models) of underlying processes when different data sets of similar lengths are used since different data sets of similar lengths may have dissimilar information content. This may especially be the case when data size is small and noisy, assuming that the observations are samples from a probability distribution defined by the underlying processes. The small data sets suffer from sampling uncertainty.

In other words, different systems representations may appear to be suitable on different realizations. This may also partly explain how equifinal models may distinguish themselves when additional pieces of information are provided. Equifinal models on one set of data, assuming use of a single measure of performance, may no longer be equifinal on another set of data (or on another piece of information) if the two data sets contain different information. This paper demonstrates that instability of a given model over different realizations of data can be understood and controlled by what we term as model output space. Ill-conditionness of model identification can then be corrected by constraining the extent of model output space. We call the extent of model output space as the measure of complexity since its regularization would lead to a stabler representation of underlying processes (Vapnik, 1982; Arkesteijn and Pande, 2013; Pande et al., 2009).

If an unstable system representation (model) is used for model prediction on yet unseen data (or on another realization of underlying processes), its instability directly translates into uncertainty in its prediction. Instability in model representation can also be seen as poor representative of underlying processes since that selected model will

HESSD
12, 3945–4004, 2015

not be a good representation of the underlying processes on another realization. Here by prediction we mean model simulation of a variable of interest conditioned by certain future values of input (forcing) variables. The regularization of the model selection problem by complexity, which corrects for the instability in system representation, then ameliorates prediction uncertainty. Complexity controlled model selection selects a model that predicts future values of a variable of interest with least uncertainty amongst the set of competing models (Pande et al., 2009, 2012).

The Bayesian treatment of prediction uncertainty and model complexity is through its specification of a marginal likelihood function of a hydrological model structure. The marginal likelihoods of hydrological model structures are often approximated by measures such as AIC (Akaike Information Criterion), BIC (Bayesian Information Criterion) and KIC (Kashyap Information Criterion) (Ye et al., 2008; Marshall et al., 2005). These measures therefore embody Bayesian interpretation of model prediction uncertainty.

Less complex hydrological models are often preferred for stable system representaiton (Pande et al., 2009; Schoups et al., 2008). Low computational complexity of simulations of models is also often desired (Keating et al., 2010; Young, 2003). We here however only explore the concept of model complexity in context of stable system representation. Often models with low parameter dimensionality (i.e. less number of parameters) are considered less complex and hence are associated with low prediction uncertainty. Whether this is always the case remains to be explored.

We follow an alternate to Bayesian, i.e. frequentist, approach (Montanari and Brath, 2004) to model complexity to explore whether parameter dimensionality is the only indicator of model complexity, instability in system represention and hence prediction uncertainty. One strength of a frequentist approach is the ease with which unstable model representation can be geometrically interpreted (Pande et al., 2009, 2012; Arkesteijn and Pande, 2013; Gupta et al., 2008). It also makes less restrictive assumptions. After illustrating the context in which model complexity has been defined, i.e. in context of unstable model representation of underlying processes and prediction uncertainty, we explore the question of whether a hydrological model with more parameters is more

HESSD
12, 3945–4004, 2015

complex or less complex in context of its influence on stability of system representation and hence prediction uncertainty. Within this context, we test the hypothesis that model complexity also depends on the magnitude of parameters that define constitutive relationships and model architecture.

The paper is organized as follows. Section 2 on methodology provides the theory, the models structures, datasets and the algorithms used. The theory first explores the connection between unstable process representation and model complexity and then provides justification for complexity regularized model selection to ameliorate instability in system representation. It then follows up with how hydrological model complexity may be calculated. Algorithms for estimating complexity of an arbitrary hydrological model is then presented and the data sets to be used are introduced. Finally the two model structures, SAC-SMA (Sacramento Soil Moisture Accounting model) and SIXPAR (Six Parameter model), are introduced. Section 3 presents and discusses the results. Here complexities of the two model structures are estimated to demonstrate the applicability of the algorithms. Then parameter ranges of SIXPAR are varied in a controlled manner to demonstrate the effect of the magnitude of parameters on model complexity, in particular in comparison with complexity computed for SAC-SMA model structure. Finally Sect. 4 concludes.

2 Methodology

2.1 Unstable system representation and model complexity

We now illustrate that instability of a given model over different realizations of data can be measured by what we term as model output space. Thereupon we demonstrate that ill-conditionedness of model structure identification is corrected by constraining (in a certain fashion, to be deliberated upon further) the extent of model structure output space (that is a union of output spaces of models that comprise a model structure).

HESSD
12, 3945–4004, 2015

In order to do so, we first define what we mean by model output space. If N is the sample size, the model output space is defined in a N-dimensional space. It is a collection of all model outputs that are obtained for all possible N-dimensional input forcings that underlying input processes may generate. Let distance in this space be measured by a metric such as mean of absolute deviations or by any other measure of similarity. It is however required that the measure of similarity obeys the conditions of being a metric (see Appendix A for further details). Figure 1 illustrates the concept of model output space.

We define instability of a given model by the variability in the differences between its outputs over two different realizations of data. A model then is more unstable if it tends to have larger differences between model simulations for any given pair of data realizations. Such a definition is sufficient to encapsulate the notion of inconsistency in process representation by a model. This is because it is quite likely that a highly unstable model that appears to be a suitable representation of the underlying system on one piece of information may not be a suitable representation on another or more pieces of information. Figure 1 illustrates this concept further.

Let $\boldsymbol{T}$ represent a set of observed output values for different realizations of input forcing. For illustration purposes we have assumed $N = 2$ in Fig. 1, hence we have a 2 dimensional space in which the output space is defined. Let $\boldsymbol{o}_1 = (o_{11}, o_{12})$ represent one observed output value for a given input forcing. Let $\boldsymbol{p}_1$ be the simulation of a model parameterized by θ_1 corresponding to the same input forcing. A collection of such simulations over all possible input forcings define the output space $\boldsymbol{M}(\theta_1)$ of the model. Let $\boldsymbol{o}_2$ and $\boldsymbol{p}_2$ be another pair of points in sets $\boldsymbol{T}$ and $\boldsymbol{M}(\theta_1)$ corresponding to another realization of input forcing. Let $\boldsymbol{A}$, $\boldsymbol{B}$, $\boldsymbol{C}$ and $\boldsymbol{D}$ be the vectors connecting the 4 points (see Fig. 1). Let $\|.\|$ measure the magnitude of vectors and define the metric used. For example $\|\boldsymbol{A}\| = d(\boldsymbol{p}_2, \boldsymbol{o}_2)$, where $d(\boldsymbol{p}_2, \boldsymbol{o}_2)$ is a metric that measures the nearness between two points in the model output space, for example mean absolute error or any other measure that satisfies the conditions of being a metric (see Appendix A). Thus $\|\boldsymbol{A}\|$ and $\|\boldsymbol{C}\|$ measure the similarity of model representations of the output to the ob-

HESSD
12, 3945–4004, 2015

served values for two different input forcings. Meanwhile $\|\boldsymbol{B}\|$ measures the closeness of two model representations themselves and $\|\boldsymbol{D}\|$ measures the closeness of the two observed time series.

Using the triangle inequality, see Appendix B, it can then be shown that $|\|\boldsymbol{A}\| - \|\boldsymbol{C}\|| \leqslant \|\boldsymbol{B}\| + \|\boldsymbol{D}\|$. Thus the deviation in performance of a model over two different information sources is bounded by $\|\boldsymbol{B}\|$ that measures how large is the model output space. If we now consider another model parameterized by θ_2 that belongs to the same model structure as the model parameterized by θ_1, we can define a model structure (here a model structure is defined as composed of of models corresponding to parameter sets θ_1 and θ_2) output space that is a union of model output spaces $\boldsymbol{M}(\theta_1)$ and $\boldsymbol{M}(\theta_2)$ (Fig. 2). One can thus obtain model structure output spaces for arbitrary model structures.

We now consider a case of nested model structures $\wedge_1$ and $\wedge_2$ such that all process representations possessed by $\wedge_2$ are also possessed by $\wedge_1$ but not vice versa (Fig. 3). This is to elucidate the role of the size of model structure output space in controlling the uncertainty in representing underlying processes. For an observed data point let $\boldsymbol{o}_1$ be an observation of underlying processes and let $\boldsymbol{p}_1^1$ and $\boldsymbol{p}_1^2$ be the *best* model representations provided by two model structures $\wedge_1$ and $\wedge_2$. The two simulations $\boldsymbol{p}_1^1$ and $\boldsymbol{p}_1^2$ correspond to models parameterised by θ^{1*} and θ^{2*}, obtained from model structures $\wedge_1$ and $\wedge_2$ respectively, which are most similar to $\boldsymbol{o}_1$ in simulations. Since $\wedge_2$ is nested within $\wedge_1$, if $\boldsymbol{p}_1^1$ is not the same as $\boldsymbol{p}_1^2$ then $\boldsymbol{p}_1^1$ is closer to the observation $\boldsymbol{o}_1$. However, if one observes another realization $\boldsymbol{o}_2$ of the underlying processes, the performance of model parameterized by θ^{1*} has more possibility to vary than the performance of model parameterized by θ^{2*}, since the output space of $\wedge_2$ lies nested inside the output space of $\wedge_1$. If $\boldsymbol{p}_2^2$ is the response provided by θ^{2*} to the input forcing corresponding to $\boldsymbol{o}_2$, the response provided by θ^{1*} to the same input forcings may vary widely, such as $\boldsymbol{p}_2^1$ or $\boldsymbol{p}_2^{\prime 1}$, in terms of its distance from $\boldsymbol{o}_2$. This possibility of more variable response

HESSD
12, 3945–4004, 2015

to the same input forcing emerges from the larger output space of $\wedge_1$ in comparison to $\wedge_2$.

We illustrate this further through a synthetic case study. Appendix C describes the set up in detail. 100 pairs of synthetic data sets, corrupted by input and output noises, are generated from a simple single linear reservoir model. Two nested model structures are then considered. These are model structures defined by linear reservoir models ($\wedge_2$) and by two reservoir models ($\wedge_1$). In the case of the latter, each of the two reservoirs are linear reservoirs. The top reservoir feeds the lower reservoir via percolation as well as produces runoff. Meanwhile the lower reservoir produces only runoff. It is evident that ($\wedge_1$) is more flexible than ($\wedge_2$) and therefore intuitively has more propensity to produce unstable system representations. The differences $|\|\boldsymbol{A}\| - \|\boldsymbol{C}\||$ are calculated for each of the 100 pairs and kernel density estimates of $\Pr(|\|\boldsymbol{A}\| - \|\boldsymbol{C}\||\geqslant\epsilon)$ are produced. Similarly $\Pr(\|\boldsymbol{B}\|\geqslant\epsilon)$ is estimated. Both these probability of exceedences are plotted in Fig. 4.

Let $\mathcal{E}$ be some event and let $\Pr(\mathcal{E})$ define the probability of occurance of that event. We first note that $\Pr(\|\boldsymbol{B}\|\geqslant\epsilon)$ is larger for two reservoir model structure $\wedge_1$ than for single reservoir model structure $\wedge_2$ for nearly all $\epsilon\geqslant 0$. Let $\mathbb{E}[\|\boldsymbol{B}\|]$ be the expected value of $\|\boldsymbol{B}\|$ over multiple realizations of data. If the extent of a model structure output space is measured by $\mathbb{E}[\|\boldsymbol{B}\|]$, i.e. what is the distance between two arbitrary model simulations in expected sense, we note that the extent of $\wedge_1$ is larger than $\wedge_2$. This is because $\mathbb{E}[\|\boldsymbol{B}\|] = \int_0^\infty \Pr(\|\boldsymbol{B}\|\geqslant\epsilon)\mathrm{d}\epsilon$. Thus the distance between any two simulations is expected to be larger for $\wedge_1$ than for $\wedge_2$ since Fig. 4b demonstrates that $\Pr(\|\boldsymbol{B}\|\geqslant\epsilon)$ is larger for $\wedge_1$ for nearly all $\epsilon\geqslant 0$. The extent of model output space as measured by $\mathbb{E}[\|\boldsymbol{B}\|]$ may be able to distinguish between model structures in terms of stability in system representation. We later provide further motivation for why it can be used as a measure to control for instability in system representation and, by doing so, we provide the context for defining it a measure of model complexity.

Imagine uncertainty in process representation as the possibility of more variable responses to the same input forcing. In the case of nested model structures, it is due to larger size of structure output space. Hence it is a measure of structural complexity

HESSD
12, 3945–4004, 2015

since larger complexity leads to higher possibility of more variable responses. The definition of complexity is intuitive since structure output space of $\wedge_1$ is larger than $\wedge_2$ because it has model concepts not in $\wedge_2$. Hence it is more complex. Thus, uncertainty in system representation can be controlled by controlling for model structure complexity, at least when nested model structures are considered.

Figure 4a also demonstrates that deviation in performance of system representations from model structure $\wedge_2$ is often larger than $\wedge_1$, to the extent that $\Pr(|\|\boldsymbol{A}\| - \|\boldsymbol{C}\||\geqslant\epsilon)$ is larger for nearly all $\epsilon\geqslant 0$. Thus, process representations from model structure $\wedge_2$ is expected to be more unstable than $\wedge_1$. The similarity in the ordering of complexity and instability thus suggest that constraining the complexity of model structures can control instability in representation of underlying processes. Further, model structure complexity is a measure of instability in process representation in the sense that larger model structure complexity implies larger possibility of unstable process representation (or higher uncertainty in process representation).

One can now flip the notion of uncertainty in process representation by considering the variability in system representations when a modeler has the liberty to select a new representation as new information in the form of another realization of observations comes to fore. Since $\wedge_1$ is more complex than $\wedge_2$, the variation, in *best* representations, over different realizations of observations, obtained from $\wedge_1$ is larger than when they are obtained from $\wedge_2$. This is because $\wedge_2$ is nested within $\wedge_1$ and this leads to the possibility of larger variation in distances between best model representations and observations for the latter. This is also illustrated in Fig. 5. One realization of observations $\boldsymbol{o}_1$ results in a selection of models corresponding to θ^{1*} and θ^{2*} from model structures $\wedge_1$ and $\wedge_2$ respectively. However for another realization $\boldsymbol{o}_2$, the model structure $\wedge_2$ retains the same model representation while the model structure $\wedge_1$, owing to its more flexible structure, allows the selection of another model representation $\tilde{\theta}^1$. Since the model structure $\wedge_2$ is nested within $\wedge_1$, model representations chosen from $\wedge_1$ would at least be as unstable as those chosen from $\wedge_1$, if not more, but never less. The figure therefore illustrates that a more complex model structure results in a more

HESSD
12, 3945–4004, 2015

unstable representation of the underlying processes. Thus it is necessary to control the complexity of a model selection problem in a certain fashion if a "stable" process representation is desired.

Following the synthetic case study presented in Appendix C, Fig. 6 demonstrates the variability in *best* system representations from the two model structures. Figure 6a plots the kernel density estimate of variability in process representations from $\wedge_1$ over 100 data pair realizations while Fig. 6b plots the pairwise kernel density estimate of the same for $\wedge_1$. It is evident from Fig. 6 that $\wedge_1$ offers more flexibility to accomodate sample variability since it has higher complexity, especially by the tradeoff between k_3 and k_1. One can observed this behavior by noting that bivariate densities of θ^{1*} often have higher values of k_1 and lower values of k_3 when compared with the bivariate densities of $\widetilde{\theta}^1$. The parametric variation offered by $\wedge_2$ is rather limited as witnessed by the cumulative density functions of θ^{2*} and $\widetilde{\theta}^2$.

2.2 Complexity regularized model selection

2.2.1 Abstract parameterization

Both Figs. 4 and 6 suggest that controlling for the complexity in a model selection exercise may *stabilize* the representation of underlying processes. This is akin to "correcting" the ill-posedness (Vapnik, 1982) of model selection problem by constraining the complexity of the model structures used. This is equivalent to regularized model selection problem.

Let a vector $\boldsymbol{y}^0 = \{y^0(1), y^0(2), \ldots, y^0(N)\}$ define the set of observations of a variable of prediction interest such as streamflow. It represents a realization of observations $\boldsymbol{o}$. Similarly, let forcing be represented by $\boldsymbol{x} = \{x_1, x_2, \ldots, x_N\}$ where x_1 may not be univariate, though assumed here to be univariate for simplicity without any loss of generality. Further let a model from a model structure $\wedge$ be represented by a parameter set θ that for given forcing $\boldsymbol{x}$ simulates $\boldsymbol{y}(\boldsymbol{x};\theta) = \{y(t,\boldsymbol{x};\theta)\}_{t=1,\ldots,N}$. The prediction variable

HESSD
12, 3945–4004, 2015

thus represents $\boldsymbol{p}$. Let $\xi_N(\boldsymbol{y}^0, \boldsymbol{x}; \theta)$ be defined as empirical risk that measures the performance of the model in terms of deviations of its predictions from the observed, for example by mean absolute error,

$$\xi_N(\boldsymbol{y}^0, \boldsymbol{x}; \theta) = \frac{\sum_{t=1}^{N} |y(t, x; \theta) - y^0(t)|}{N}. \quad (1)$$

This represents $\|\boldsymbol{p} - \boldsymbol{o}\|$, where we have assumed mean of absolute deviations as the metric.

Let us now reformulate the definition of a model structure wherein its internal architecture of how various subsystem representations are connected as well as its parameters can both be defined by an abstract parameter set α. That is, both a model structure, for e.g. $\wedge$, and a model from the structure, for e.g. θ, are parameterized by α. Consider the linear and the two linear reservoirs model (as discussed in the previous section). The linear reservoir model has only 1 parameter, i.e. the recession parameter $k \in [0, 1]$ (dimension: $[1/T]$). Meanwhile, the two reservoir model has 3 recession parameters $\boldsymbol{k} = \{k_1, k_2, k_3\} \in [0, 1]^3$ (dimensions: $[1/T]^3$). If we now define the abstract parameter set $\alpha = \{\alpha_1, \alpha_2, \alpha_3\} \in [0, 1]^3$ then we can describe both the model structures. Model structure of a single reservoir model can be described by the set $\wedge_1 = \{0\} \times \{0\} \times [0, 1]$, in which case α_1 and α_2 is restricted to 0 while α_3 is allowed to vary between [0,1]. The two reservoir model structure can be described by the same parameter α, which is less constrained and belong to $[0, 1] \times [0, 1] \times [0, 1]$. Thus such a representation not only distinguishes between two model structures in terms of its different subsystem architecture (one vs. two reservoir model structure) but also distinguishes in terms of its parameter magnitudes. For example, in this representation a two reservoir model structure defined by $\{\alpha : \alpha \in [0, 1] \times [0.5, 1] \times [0, 1]\}$ that only permits fast flow from the second reservoir is different from a model structure $\{\alpha : \alpha \in [0, 1] \times [0, 1] \times [0, 1]\}$ that does not restrict the nature of flow from the second reservoir. Equation (1) can be reformulated in terms of α as,

HESSD
12, 3945–4004, 2015

$$\xi_N(\boldsymbol{y}^0, \boldsymbol{x}; \alpha) = \frac{\sum_{t=1}^{N} |y(t, x; \alpha) - y^0(t)|}{N}. \quad (2)$$

This represents $\|\boldsymbol{p} - \boldsymbol{o}\|$.

By doing this we no longer distinguish between a model and a model structure and allow models to seamlessly change their model structure by changing α. Then a model and its corresponding model structure is represented by $y(t, \boldsymbol{x}; \alpha)$. We suppress $t, \boldsymbol{x}$ and represent a model by $y(\alpha)$. Further, since a distinction between a model and a model structure has been dissolved by using α, any compact set of αs can now be called a model structure.

2.2.2 A continuum of model structures defined by complexity

Let $\Phi(y(\alpha))$ be the complexity (here the extent of model output space) of the model $y(\alpha)$. We now note that the model output of any hydrological model is continuous in its parameters. Further, the extent of model output space is continuous in model outputs (the extent of one model output space is smaller than another if two simulations of the former are closer than the latter for any given pair of input forcing). Therefore, $\Phi(y(\alpha))$ is continuous in α. In other words, a set $\wedge = \{\alpha : \Phi(y(\alpha)) \leqslant c\}$ is compact and defines a model structure. By extension, we can obtain a sequence of model structures $\wedge_m$ using the inequality $\Phi(\wedge) \leqslant c_m$ for a sequence of c_m, where $m = 1, 2, \ldots, j \ldots$ What this says is that if the difference between any two upper bounds on $\Phi(y(\alpha))$ is small, the corresponding model structures are similar. Based on our construction, we note that a model structure here is a result not just of the architecture of how various model components are interconnected but also how they are parameterised. Thus model structures with different architecture and parameterization may be deemed similar.

A model structure is then nested within another model structure if the complexity of the former is smaller than the latter. Formally, $\wedge_1 = \{\alpha : \Phi(y(\alpha)) \leq c_1\}$ is nested within $\wedge_2 = \{\alpha : \Phi(y(\alpha)) \leq c_2\}$ if $c_1 \leq c_2$. A continuum of model structures may therefore be

HESSD
12, 3945–4004, 2015

obtained by a sequence of c. The nomenclature "continuum of model structures" has also been invoked elsewhere (Farmer et al., 2003; Gupta et al., 2008).

This is interesting because a definitive statement on structure complexity based on parameter ranges or parameter dimensionality, i.e. without knowing their complexity in advance, can only be made if the corresponding structures are nested. For a given model structure, such a statement can only be made if parameter ranges of one are a subset of another. But the effect of parameter dimensionality on model complexity, jointly with parameter magnitudes is not always clear. This is because the abstract parameters corresponding to parameter dimensionality and their interaction with other "real" parameters is not evident. The effect of parameter magnitudes on model complexity is also not clear. Hence, complexities of model structures and their effect on prediction uncertainties may be counterintuitive. For example, a model structure that has higher number of parameters than another may be less complex than the other for certain parameter ranges. This is where the number of parameters and parameter magnitudes jointly effect model complexity, uncertainty in process representation and consequent prediction uncertainty.

Consider the example of the linear reservoir model structure of Appendix C. In this case one can state that a model structure with $k \in [0, 0.5]$ is less complex than a model structure with $k \in [0, 1]$. However, no statement can yet be made on how the complexity of the model structure $k \in [0, 0.5]$ fares with complexity of model structure with $k \in [0.5, 1]$. Now if we consider the 3 parameter model structure in Appendix C alongside the single reservoir model structure, one can still state that a single reservoir model structure with $k \in [0, 1]$, i.e. with $\alpha \in [0, 1] \times \{0\} \times \{0\}$, is less complex than the 3 reservoir model structure with $\boldsymbol{k} \in [0, 1]^3$, i.e. with $\alpha \in [0, 1]^3$. This is because the former structure is a subset of the latter. However one cannot state anything about structure complexities of the two model structures with $\alpha \in [0, 1] \times \{0\} \times \{0\}$ and $\alpha \in [0, 0.5] \times [0, 0.3] \times [0, 0.5]$ respectively unless their complexities are computed. This is because we cannot say that one model structure is nested within the other.

HESSD
12, 3945–4004, 2015

2.2.3 Stable system representation and top-down modelling approach

Let us now revisit the definition of a stable system representation: a problem of system representation is stable if for two realizations of observations that are δ-close, corresponding selected system representations are ϵ-close such that as δ becomes small so does ϵ. Here by ϵ-close one means that distance between two system representations is not larger than ϵ. Intuitively, it means that the problem of model selection is bounded such that the selected representations do not differ dramatically for two different realizations of data. Thus a model selection process is stable if the models (or model structures) selected on similar realizations of observations are similar. Now note that the demands of stable model selection are two two-fold: the need for a good representation of the underlying processes and the need to have a bounded representation, i.e. no two representations are drastically different when confronted with similar observations. Since the complexity measure expressed in the form of $\Phi(y(\alpha)) \leqslant c_m$ ensures that model structures corresponding to $\Phi(y(\alpha)) \leqslant c_m$ are similar if two values of c_m are similar, complexity measure acts as a natural constraint to ensure stable model selection. Thus two objectives need to be considered, (i) maximize finite sample performance by minimizing ξ_N, which ensures that a good model on a given sample is selected and (ii) obey a constraint on model complexity for some value of c_m, say c^*, which ensures that model complexity is controlled for. Such a model selection problem can be posed as,

$\min_\alpha \xi_N(\boldsymbol{y}^0, \boldsymbol{x}; \alpha)$

s.t.

$\Phi(y(\alpha)) \leqslant c^*$.

The above can alternatively be written as

$$\Xi(\alpha_N^*) = \min_\alpha \ \xi_N(\boldsymbol{y}^0, \boldsymbol{x}; \alpha) + \lambda^* \Phi(y(\alpha)). \quad (3)$$

HESSD
12, 3945–4004, 2015

Here $\lambda^* \geq 0$ still has to be estimated. This is often done on a set of observations that is independent of the observations used to estimate models. Thus the choice of λ^* depends on the model structures used and the underlying hydrological system since its estimation is based on observations. It gauges how ill-posed is the problem of system representation and how tightly should the model selection problem be controlled by complexity so that it can be *stabilized*.

Since the measure of complexity, in the context defined here, "stabilizes" system representation, a complexity regularized model selection problem yields least uncertain system representation over future unseen data. If the representation is used to predict system behavior, such a representation also has least predictive uncertainty. It is in this sense that complexity controls predictive uncertainty if the problem of identifying system representation is regularized by complexity, i.e. it controls predictive uncertainty by "stabilizing" the problem of system representation. In other words, complexity is a measure of predictive uncertainty since higher complexity of system representation of underlying processes leads to more unstable system representation, which in turn implies higher predictive uncertainty.

An approach wherein additional process representations are added in a stepwise manner, or a top down approach, increases complexity in a stepwise manner (Farmer et al., 2003; Buttsa et al., 2004; Bai et al., 2009). Additional complexity with more detailed or additional process representations trades off with the accuracy with which the processes are represented. Thus more complexity may be acceptable when it sufficiently improves the representation of the underlying processes. Equation (3) describes this tradeoff. The multiplier $\lambda^* \geq 0$ is the minimum amount of improvement in system representation that is desired in order for a unit increment in model complexity.

Thus λ^* measures the tradeoff between improvement in model performance and corresponding increase in model complexity. This allows a formal framework to assess how much additional model complexity is warranted, especially in a top down modelling approach. This is because it also suggests that a more complex model is not selected if it provides "really bad" system representation (Farmer et al., 2003; Son and

HESSD
12, 3945–4004, 2015

Sivapalan, 2007). Thus model complexity may be increased in a step-wise manner till model performance begins to decline.

2.2.4 Continuum of models and model complexity: parameter magnitude vs. dimensionality

The continuum of model structures is an important construct since it dissolves the distinction between model architecture and model parameters. Model structures can be defined based on constraints on model parameters or model outcomes or both, ofcourse not excluding the case when structures are induced by different architectures. Then complexity of such structures, now defined as a set of abstract parameters, can be defined as the combined (union of) extent of output spaces of models corresponding to the parameters.

Since model complexity does not distinguish model architecture from model parameter magnitudes (by using abstract parameters), one can assess the relative effect of model architecture over parameter magnitude on model complexity. Again, we can do so because the concept of model complexity presented here depends on how a model transform input forcings to model simulations. This depends both on the architecture and strengths of constitutive relationships.

The effects of number of parameters (as a result of model architecture) and magnitude of parameters (as a result of the strength of constitutive relatioships) on model complexity can be decomposed. This can be done by estimating model complexity of two model structures when their parameter ranges are "equivalent" and then fixing the model architecture and varying parameter magnitudes. Equivalent parameter ranges ensure that two model architectures have, for example, similar water storage capacities and water residence times but are different in model architectures. Meanwhile variation in parameter mangnitudes for the same model architecture provides model structures that differ in storage capacities and residence times. Overall model complexity can then be thought of as the combined effect of architecture and parameter magnitudes.

HESSD
12, 3945–4004, 2015

2.3 Estimation of model complexity

Section 2.1 suggests that $\mathbb{E}[\|\boldsymbol{B}\|]$ is able to distinguish between model (structures) in terms of stability in process representation and can serve as a measure of model complexity (the extent of model output space) in the context of stabilizing system representation. We however note that it is one statistic of the distribution $\Pr(\|\boldsymbol{B}\| \geqslant \epsilon)$. A distributed measure of complexity may well be desired but we leave an investigaton of this for future research. Here we demonstrate how $\mathbb{E}[\|\boldsymbol{B}\|]$ can be estimated in a step by step manner (see also Arkesteijn and Pande, 2013). By doing so we also explain the algorithms presented in Sect. 2.4.

First we note that $\mathbb{E}[\|\boldsymbol{B}\|]$ is the expected difference in a model's simulations for two realizations of observations. We now translate what it means for an arbitrary hydrological represented by $y(\alpha)$.

Definition 1: Let $\mathbb{E}[\|\boldsymbol{B}\|] = \mathbb{E}[\|\boldsymbol{y}(\boldsymbol{x}_1;\alpha) - \boldsymbol{y}(\boldsymbol{x}_2;\alpha)\|]$, where $\|\boldsymbol{y}(\boldsymbol{x};\alpha) - \mathbb{E}[\boldsymbol{y}(\boldsymbol{x};\alpha)]\| = \sum_{t=1,N} \frac{|y(t,\boldsymbol{x};\alpha) - \mathbb{E}[\boldsymbol{y}(\boldsymbol{x};\alpha)]|}{N}$. Thus we assume that the mean of absolute deviations is the metric used.

The statistic provided in Definition 1 is similar to $\mathbb{E}[\|\boldsymbol{y}(\boldsymbol{x};\alpha) - \mathbb{E}[\boldsymbol{y}(\boldsymbol{x};\alpha)]\|]$, which also measures variation in simulations of a model parameterized by α. We will use the latter to represent $\mathbb{E}[\|\boldsymbol{B}\|]$.

Also, note that the expectation is obtained by taking the average of $\|\boldsymbol{y}(\boldsymbol{x};\alpha) - \mathbb{E}[\boldsymbol{y}(\boldsymbol{x};\alpha)]\|$ over a large number, say M, of realizations of observations, i.e.

$$\mathbb{E}[\|\boldsymbol{y}(\boldsymbol{x};\alpha) - \mathbb{E}[\boldsymbol{y}(\boldsymbol{x};\alpha)]\|] = \lim_{M\to\infty} \sum_{k=1}^{M} \frac{\|\boldsymbol{y}(\boldsymbol{x}_k;\alpha) - \mathbb{E}[\boldsymbol{y}(\boldsymbol{x}_k;\alpha)]\|}{M}.$$

This is because a very large set of observations of size N' can be divided into very large M subsets of observations of size N such that $N' = MN$. The above thus allows us to estimate complexity $\mathbb{E}[\|\boldsymbol{B}\|]$ by estimating $\|\boldsymbol{y}(\boldsymbol{x}_k;\alpha) - \mathbb{E}[\boldsymbol{y}(\boldsymbol{x}_k;\alpha)]\|$ on $k = 1,\ldots,M$ sets of observations of size N. Also, $M \to \infty$ is indicative of a very large M. Often, M may not be required to be large if variation in $\|\boldsymbol{B}\|$ asymptotes after some finite value of

HESSD
12, 3945–4004, 2015

M, whereupon $\mathbb{E}[\|\boldsymbol{B}\|]$ can be estimated with high confidence. The estimation of model complexity as presented here thus rests on estimating $\mathbb{E}[\|\boldsymbol{y}(\boldsymbol{x};\alpha) - \mathbb{E}[\boldsymbol{y}(\boldsymbol{x};\alpha)]\|]$.

Definition 2: Let us denote $\mathbb{E}[\|\boldsymbol{y}(\boldsymbol{x};\alpha) - \mathbb{E}[\boldsymbol{y}(\boldsymbol{x};\alpha)]\|]$, that measures the complexity of a model parameterised by α, by $\widetilde{\gamma}$.

Then, by definition, the probability that $\mathbb{E}[\|\boldsymbol{y}(\boldsymbol{x};\alpha) - \mathbb{E}[\boldsymbol{y}(\boldsymbol{x};\alpha)]\|] \geqslant \gamma$ is 1 when $\gamma \leqslant \widetilde{\gamma}$ and 0 otherwise for all $\gamma \geqslant 0$. This is because $\Pr(\mathbb{E}[\|\boldsymbol{y}(\boldsymbol{x};\alpha) - \mathbb{E}[\boldsymbol{y}(\boldsymbol{x};\alpha)]\|] \geqslant \gamma) = \Pr(\widetilde{\gamma} \geqslant \gamma)$, which is equal to 1 when $\gamma \leqslant \widetilde{\gamma}$. It is equal to 0 otherwise. Thus,

$$\Pr(\mathbb{E}[\|\boldsymbol{y}(\boldsymbol{x};\alpha) - \mathbb{E}[\boldsymbol{y}(\boldsymbol{x};\alpha)]\|] \geqslant \gamma) = \begin{cases} 1 & \text{if } \gamma \leqslant \widetilde{\gamma} \\ 0 & \text{otherwise} \end{cases}. \tag{4}$$

We now show that $\mathbb{E}[\|\boldsymbol{y}(\boldsymbol{x};\alpha) - \mathbb{E}[\boldsymbol{y}(\boldsymbol{x};\alpha)]\|]$ can be expressed as $\lim_{N' \to \infty} \sum_{N'} \frac{|y(t,\boldsymbol{x};\alpha) - \mathbb{E}[y(t,\boldsymbol{x};\alpha)]|}{N'}$, where $N' = MN$.

$$\begin{aligned} \mathbb{E}[\|\boldsymbol{y}(\boldsymbol{x};\alpha) - \mathbb{E}[\boldsymbol{y}(\boldsymbol{x};\alpha)]\|] &= \lim_{M \to \infty} \sum_{k=1,..,M} \frac{\|\boldsymbol{y}(\boldsymbol{x}_k;\alpha) - \mathbb{E}[\boldsymbol{y}(\boldsymbol{x}_k;\alpha)]\|}{M} \\ &= \lim_{M \to \infty} \sum_{k=1,\ldots,M} \frac{1}{M} \sum_{t=1,\ldots,N} \frac{|y(t,\boldsymbol{x}_k;\alpha) - \mathbb{E}[y(t,\boldsymbol{x}_k;\alpha)]|}{N} \\ &= \lim_{N' \to \infty} \sum_{N'} \frac{|y(t,\boldsymbol{x};\alpha) - \mathbb{E}[y(t,\boldsymbol{x};\alpha)]|}{N'}. \end{aligned} \tag{5}$$

From equation system (5) we note that

$$\Pr(\mathbb{E}[\|\boldsymbol{y}(\boldsymbol{x};\alpha) - \mathbb{E}[\boldsymbol{y}(\boldsymbol{x};\alpha)]\|] \geqslant \gamma) = \Pr\left(\lim_{N' \to \infty} \sum_{N'} \frac{|y(t,\boldsymbol{x};\alpha) - \mathbb{E}[y(t,\boldsymbol{x};\alpha)]|}{N'} \geqslant \gamma\right). \tag{6}$$

The argument of the Right Hand Side (RHS) of Eq. (6) therefore also measures complexity, i.e. $\mathbb{E}[\|\boldsymbol{B}\|]$, since the argument of Left Hand Side (LHS) measure it as per definition 1. We now note as a consequence of Proposition 1.1.1 of Ross (1996) that

HESSD
12, 3945–4004, 2015

$$\Pr\left(\lim_{N'\to\infty}\sum_{N'}\frac{|y(t,\boldsymbol{x};\alpha)-\mathbb{E}[y(t,\boldsymbol{x};\alpha)]|}{N'}\geqslant\gamma\right)$$
$$=\lim_{N'\to\infty}\Pr\left(\sum_{N'}\frac{|y(t,\boldsymbol{x};\alpha)-\mathbb{E}[y(t,\boldsymbol{x};\alpha)]|}{N'}\geqslant\gamma\right). \quad (7)$$

Equation (7) states that the limit of the probability is the same as the probability of the limit. Readers are referred to the Supplement of (Arkesteijn and Pande, 2013) for additional details.

Definition 3: Let $P_{N,\gamma}$ be defined as $\Pr\left(\sum_N\frac{|y(t,\boldsymbol{x};\alpha)-\mathbb{E}[y(t,\boldsymbol{x};\alpha)]|}{N}\geqslant\gamma\right)$.

We now estimate $P_{N,\gamma}$, since its argument contains the measure of complexity as per definition 2 and Eqs. (6) and (7). How the measure of complexity is extracted from the argument is now demonstrated.

For this we first invoke Markov's Lemma, which states that for any $X\geqslant 0$ and $t>0$ the following inequality holds,

$$\Pr(X\geqslant 0)\leqslant\frac{\mathbb{E}[X^2]}{t^2}. \quad (8)$$

By substituting X by $\sum_N\frac{|y(t,\boldsymbol{x};\alpha)-\mathbb{E}[y(t,\boldsymbol{x};\alpha)]|}{N}$ in inequality (8), we obtain the following inequality,

$$P_{N,\gamma}=\Pr\left(\sum_N\frac{|y(t,\boldsymbol{x};\alpha)-\mathbb{E}[y(t,\boldsymbol{x};\alpha)]|}{N}\geqslant\gamma\right)\leqslant\frac{\mathbb{E}\left[\left(\sum_N|y(t,\boldsymbol{x};\alpha)-\mathbb{E}[y(t,\boldsymbol{x};\alpha)]|\right)^2\right]}{N^2\gamma^2}. \quad (9)$$

The inequality (9) can now be rearranged to obtained an expression for $P_{N,\gamma}N^2\gamma^2$, the motivation behind invoking Markov's inequality. We then obtain

$$P_{N,\gamma}N^2\gamma^2\leqslant\mathbb{E}\left[\left(\sum_N|y(t,\boldsymbol{x};\alpha)-\mathbb{E}[y(t,\boldsymbol{x};\alpha)]|\right)^2\right]. \quad (10)$$

HESSD
12, 3945–4004, 2015

Several points are in order based on inequality (10). The Right Hand Side is independent of γ. It is a sum of $\frac{N(N+1)}{2}$ non-negative numbers, thus it can be bounded by some function of N^2. Since the inequality is not strict, a maximum of the Left Hand Side, i.e. $P_{N,\gamma}N^2\gamma^2$, with respect to γ can be equated to the Right Hand Side. Thus $\max_\gamma P_{N,\gamma}N^2\gamma^2$ is a function only of N, while $P_{N,\gamma}$ is both a function of γ and N. Since the Right Hand Side is $\mathcal{O}(g(N^2))$ (a function $f(x) = \mathcal{O}(g(N^2))$ means that $|f(x)| \leq c|g(N^2)|$, where $c > 0$), we assume it to be a quadratic functon of form $f(h,N) = \beta_2 N^2 + \beta_1 N + \beta_0$ with $h = \{\beta_2, \beta_1, \beta_0\}$. We therefore have

$$\max_\gamma P_{N,\gamma}N^2\gamma^2 = f(h,N). \tag{11}$$

Let γ_N^* represent the γ that maximizes $P_{N,\gamma}N^2\gamma^2$, then $P_{N,\gamma_N^*} = \frac{f(h,N)}{N^2\gamma_N^{*2}} = \frac{1}{\gamma_N^{*2}}\left(\beta_2 + \frac{\beta_1}{N} + \frac{\beta_0}{N^2}\right)$. Finally, if we represent γ^* as the γ that maximizes $\lim_{N\to\infty} P_{N,\gamma}N^2\gamma^2$, then $P_{N,\gamma_N^*} \to \frac{\beta_2}{\gamma^{*2}}$ as $N \to \infty$.

We now show that $\gamma^* = \mathbb{E}[\|\boldsymbol{y}(\boldsymbol{x};\alpha) - \mathbb{E}[\boldsymbol{y}(\boldsymbol{x};\alpha)]\|] = \widetilde{\gamma}$ maximizes $P_{N,\gamma}N^2\gamma^2$ as $N \to \infty$. This is because of two reasons. First, as $N \to \infty$, $P_{N,\gamma} \to \Pr(\mathbb{E}[\|\boldsymbol{y}(\boldsymbol{x};\alpha) - \mathbb{E}[\boldsymbol{y}(\boldsymbol{x};\alpha)]\|] \geqslant \gamma)$ from Eqs. (7) and (6). But then $\Pr(\mathbb{E}[\|\boldsymbol{y}(\boldsymbol{x};\alpha) - \mathbb{E}[\boldsymbol{y}(\boldsymbol{x};\alpha)]\|] \geqslant \gamma)$ is either 1 (maximum value) or 0 (minimum value). The maximum value is achieved when $\gamma \leqslant \mathbb{E}[\|\boldsymbol{y}(\boldsymbol{x};\alpha) - \mathbb{E}[\boldsymbol{y}(\boldsymbol{x};\alpha)]\|]$ and the minimum value is achieved when $\gamma > \mathbb{E}[\|\boldsymbol{y}(\boldsymbol{x};\alpha) - \mathbb{E}[\boldsymbol{y}(\boldsymbol{x};\alpha)]\|]$ respectively (see Eq. 4). Meanwhile γ^2 is increasing in γ. Thus γ^* that maximizes $P_{N,\gamma}N^2\gamma^2$ as $N \to \infty$ is the maximum possible value of γ for which $\lim_{\to\infty} P_{N,\gamma} = 1$. This is $\gamma^* = \mathbb{E}[\|\boldsymbol{y}(\boldsymbol{x};\alpha) - \mathbb{E}[\boldsymbol{y}(\boldsymbol{x};\alpha)]\|] = \widetilde{\gamma}$, which is a measure of complexity.

Thus as N becomes large we note the following based on the arguments above: (i) $P_{N,\gamma_N^{*2}}$ becomes 1, and (ii) $P_{N,\gamma_N^{*2}}$ becomes $\frac{\beta_2}{\gamma_N^{*2}}$ and (iii) γ_N^{*2} becomes the measure of complexity $\mathbb{E}[\|\boldsymbol{y}(\boldsymbol{x};\alpha) - \mathbb{E}[\boldsymbol{y}(\boldsymbol{x};\alpha)]\|]^2$. These 3 points therefore suggest that hydrological model complexity can be estimated if we estimate β_2 since $\mathbb{E}[\|\boldsymbol{y}(\boldsymbol{x};\alpha) - \mathbb{E}[\boldsymbol{y}(\boldsymbol{x};\alpha)]\|]$ becomes $\sqrt{\beta_2}$ as N becomes large.

HESSD
12, 3945–4004, 2015

All we now have to do is estimate β_2 to estimate complexity, which in turn can be estimated based on Eq. (11). We study the behavior of $P_{N,\gamma} = \Pr(\sum_N \frac{|y(t,\boldsymbol{x};\alpha) - \mathbb{E}[y(t,\boldsymbol{x};\alpha)]|}{N} \geqslant \gamma)$ for a given model on synthetically generated data in order to estimate β_2. In particular we study the maximum of $P_N N^2 \gamma^2$ for various values of N and when it asymptotes we obtain the measure of model complexity, β_2.

We summarize the above arguments to estimate complexity based on expressions (11), (7), (5) and (4) in the following steps.

1. Let γ_N^* be the one that maximizes $P_{N,\gamma} N^2 \gamma^2$. Then from equality (11) $P_{N,\gamma^*} = \frac{f(h,N)}{N^2 \gamma_N^{*2}}$.

2. Let $\lim_{N\to\infty} \gamma_N^* = \gamma^*$. From inequality (7),

$$\lim_{N\to\infty} P_{N,\gamma_N^*} = \Pr\left(\lim_{N\to\infty} \sum_N \frac{|y(t,\boldsymbol{x};\alpha) - \mathbb{E}[y(t,\boldsymbol{x};\alpha)]|}{N} \geqslant \gamma^*\right).$$

3. From expression (5),

$$\Pr\left(\mathbb{E}[\|\boldsymbol{y}(\boldsymbol{x};\alpha) - \mathbb{E}[\boldsymbol{y}(\boldsymbol{x};\alpha)]\|] \geqslant \gamma^*\right) = \Pr\left(\lim_{N\to\infty} \sum_N \frac{|y(t,\boldsymbol{x};\alpha) - \mathbb{E}[y(t,\boldsymbol{x};\alpha)]|}{N} \geqslant \gamma^*\right).$$

4. From expression (4) we have $\Pr(\mathbb{E}[\|\boldsymbol{y}(\boldsymbol{x};\alpha) - \mathbb{E}[\boldsymbol{y}(\boldsymbol{x};\alpha)]\|] \geqslant \gamma)$ is either 0 or 1 for different values of γ. Since, from steps (1)–(3), γ^* maximizes $\lim_{N\to\infty} P_{N,\gamma} N^2 \gamma^2$, we require $\Pr\left(\mathbb{E}[\|\boldsymbol{y}(\boldsymbol{x};\alpha) - \mathbb{E}[\boldsymbol{y}(\boldsymbol{x};\alpha)]\|] \geqslant \gamma^*\right) = 1$ and $\gamma^* = \widetilde{\gamma} = \mathbb{E}[\|\boldsymbol{y}(\boldsymbol{x};\alpha) - \mathbb{E}[\boldsymbol{y}(\boldsymbol{x};\alpha)]\|]$.

5. From steps (4) to (2), we obtain $\lim_{N\to\infty} P_{N,\widetilde{\gamma}} = 1$.

6. From steps (5), (4), (2) and (1) we obtain $\lim_{N\to\infty} \frac{\beta_2 N^2 + \beta_1 N + \beta_0}{N^2 \widetilde{\gamma}^2} = \frac{\beta_2}{\widetilde{\gamma}^2} = 1$.

7. From steps (6) and (4), we obtain $\mathbb{E}[\|\boldsymbol{y}(\boldsymbol{x};\alpha) - \mathbb{E}[\boldsymbol{y}(\boldsymbol{x};\alpha)]\|] = \widetilde{\gamma} = \sqrt{\beta_2}$.

HESSD
12, 3945–4004, 2015

Thus complexity can be estimated by β_2. The parameter set h, that includes β_2, are estimated by regressing a quadratic function to $\max_\gamma P_{N,\gamma} N^2 \gamma^2$ that is numerically estimated for various values of N. Algorithms 1 and 2 peform this task. Using Eq. (4) we further define a measure of complexity, $F(h,N) = \frac{f(h,N)}{N^2}$, that is dependent on N such that $\lim_{N\to\infty} F(h,N) = \beta_2$. We call β_2 asymptotic complexity in this context.

2.4 Algorithms and data

The computation of model complexity requires a synthetically generated input forcing data set because $P_{N,\gamma}$ needs to be estimated in order to estimate $\max_\gamma P_{N,\gamma} N^2 \gamma^2$ for each N. This inturn requires the estimation of $\mathbb{E}[y(t,\boldsymbol{x};\alpha)]$, which can be estimated based on synthetically generated input forcing data.

We here note that a vector $\mathbb{E}[\boldsymbol{y}(\boldsymbol{x};\alpha)] = \{\mathbb{E}[y(t,\boldsymbol{x};\alpha)]\}_{t=1,\ldots,N}$ is desired that preserves the autocorrelation that a model simulation may bring. It also represents the expectation of a N-vector in the N-dimensional model output space. Here $\boldsymbol{x}$ is a N-dimensional input forcing, i.e. $\boldsymbol{x} = (x_1, x_2, \ldots, x_t, \ldots, x_N)$. For notational simplicity we have assumed x_t is a one-dimensional varaible. Also, note that the intention is to use it to estimate $\Pr(\|\boldsymbol{y}(\boldsymbol{x};\alpha) - \mathbb{E}[\boldsymbol{y}(\boldsymbol{x};\alpha)]\| \geq \gamma)$. Thus if have M realizations of input forcings, i.e. $\{\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_k, \ldots, \boldsymbol{x}_M\}$, we estimate the expectation of N-dimensional model simulations as

$$\mathbb{E}[\boldsymbol{y}(\boldsymbol{x};\alpha)] = \left\{ \sum_{k=1,\ldots,M} \frac{y(1,\boldsymbol{x}_k;\alpha)}{M}, \sum_{k=1,\ldots,M} \frac{y(2,\boldsymbol{x}_k;\alpha)}{M}, \ldots, \right.$$
$$\left. \sum_{k=1,\ldots,M} \frac{y(t,\boldsymbol{x}_k;\alpha)}{M}, \ldots, \sum_{k=1,\ldots,M} \frac{y(N,\boldsymbol{x}_k;\alpha)}{M} \right\}.$$

We now present an algorithm that computes the expectation operator on the synthetically generated data set (Arkesteijn and Pande, 2013). The input forcing basin datasets for this algorithm are obtained from the MOPEX data sets (Duan et al., 2006; Brooks

HESSD
12, 3945–4004, 2015

et al., 2011). 5 basins from different hydroclimatic regions are used. By doing so we test whether the ordering in terms of its complexity of various model structure set-ups changes with different data sets. Insensitivity of the ordering of structure complexities to the data sets used for input forcings is crucial for any robust statement about the role of parameter magnitudes in determining model complexity. Table 1 provides characteristics such as area, mean annual precipitation and evaporation and hydrologic ratios such as runoff ratio and dryness index, for the basins used in this study. Figure 7 displays them.

The algorithm is a resampler that block bootstraps time series from a given sample of data (Kundzewicz and Robson, 2004; Politis and Romano, 1994). Arkesteijn and Pande (2013) discuss that the weather resampler bootstraps blocks of wet/dry spell pairs where each block contains one wet/dry spell pair. The algorithm can be improved by increasing the number of contiguous wet/dry samples within each block. We use basin input forcing data set (of precipitation and potential evapotranspiration) and generate multiple realizations for the complexity, one for each sampled parameter. We also partially account for the sensitivity of complexity computation by permuting data at monthly scale in such a way that intra-annual autocorrelation in forcing time series is randomized. Sensitivity of complexity computation is also tested against multiple basins and different wet-dry spell identification by choosing basins from different regions of the US (Fig. 7).

Algorithm 1:

1. Extract daily precipitation and potential evapotranspiration data for a basin.
2. Identify a block of contiguous wet (a set of contiguous days with positive precipitation) and dry (a set of contiguous days with zero precipitation) spell pairs for each month: determine the amount and length of spell pairs and attach an identifier to each spell.
3. Construct a one month sample for each month: conditioned on a selected month, randomly sample (with replacement) blocks of spell pairs, along with potential

HESSD
12, 3945–4004, 2015

evaporation values for the same days, across different years for the same month, appending these blocks till the total length of the sequence exceeds 30 days.

4. Go to step 3 for other months until all 12 months of a year have been sampled.
5. Permute the months (if correlation between months is to be removed), while maintaining the order of sequences within each month, to create one year sample.
6. Repeat steps 4 and 5 to create a realization of input forcings at daily time steps with $\overline{N}$ datapoints.
7. Go to step 6 until M realizations of $\overline{N}$ datapoints are created.

The algorithm resamples forcing data from an observed dataset of a basin such that auto (and cross) correlation of the variables are preserved at certain scale. For each month, for example January, wet-dry spell pairs are identified and a resample for the month is generated by bootstrapping such pairs with replacement (i.e. the pairs are put back in the month and can be resampled again). A resample for a month is created once the total length of days resampled in such a manner is at least 30. Then if the auto-correlation is to be preserved at certain scale, for example at 3 month scale (called "Medium 4"), then the ordering of 3 month blocks of monthly (re-)samples is permuted. The "4" in "Medium 4" therefore represents the number of blocks in a year that need to be permuted. That is, the ordering of the set of 3-tuples {JFM,AMJ,JAS,OND} is permuted, where each letter stands for the beginning letter of a resampled month ("JFM" for January-February-March, "AMJ" for April-May-June, and so on). Thus a resample of forcing data for a year that preserves correlation at 3 month scale can be {AMJ, JFM, OND, JAS}. Repeating the process for multiple years thus re-samples (or stochastically generates) forcing data for multiple years and correlation is preserved at certain scale. The preservation of the entire seasonal cycle ("Complete"), of the monthly correlation at 6 month scale ("Medium2"), of the monthly auto-correlation at 3 month scale ("Medium 4") and of no month to month autocorrelation ("None") is currently allowed.

HESSD
12, 3945–4004, 2015

Using the weather resampler, $M = 2000$ sequences of $\overline{N} = 5000$ datapoints for daily precipitation and potential evaporation are obtained. For each realization, input forcings of smaller sample sizes $N = 200 : 50 : \overline{N}$ are obtained by sampling its first N data points. Since SIXPAR model structure does not explicitly incorporate evaporation (see Supplement), the precipitation data used for SIXPAR is assumed to be equal to a maximum of the precipitation minus potential evaporation and zero.

Once multiple realizations of input forcing data have been generated (resampled), Algorithm 2 computes the complexity of models for a sampled parameter set (Arkesteijn and Pande, 2013) as outlined in the previous section. It uses the M realizations of input forcings to first estimate expected value of model simulations of size $\overline{N}$, i.e. $\mathbb{E}[\boldsymbol{y}(\boldsymbol{x};\alpha)]$ and then estimate probabilities of exceedences for $\gamma = 0 : \overline{\gamma}$, where $P_{N,\gamma} = \Pr(\sum_N \frac{|y(t,\boldsymbol{x};\alpha) - \mathbb{E}[y(t,\boldsymbol{x};\alpha)]|}{N} \geqslant \gamma)$. These are the steps involved in step 1 of Algorithm 2.

Algorithm 2:

1. For each parameter set of a model structure set up, estimate $P_{N,\gamma}$, for a given value of N and γ using M samples of data set of size N, obtained from Algorithm 1.
2. Estimate the maximum $\widetilde{f}(N)$ of $P_{N,\gamma} N^2 \gamma^2$ with respect to γ for each N. Let the maximizing γ be $\gamma^N_{\max}$.
3. Repeat steps 1 and 2 for $N = 200 : 50 : \overline{N}$.
4. Determine the set of coefficients $h = \{\beta_2, \beta_1, \beta_0\}$ of $f(h,N) = \beta_2 N^2 + \beta_1 N + \beta_0$ that fits data points $\{\widetilde{f}(N), N = 200 : 50 : \overline{N}\}$. The set of coefficients h defines the model complexity.
5. Repeat step 1–4 to estimate complexity for different parameter sets of a model structure.

HESSD
12, 3945–4004, 2015

In total 500 parameter sets are sampled from each range presented in Tables 2 and 3.

2.5 Model structures and parameter ranges

2.5.1 SAC-SMA and SIXPAR model structures

The two model structures that are used are SAC-SMA (Sacramention Soil Moisture Accounting model) and SIXPAR (Six Parameter model). SAC-SMA is a complex hydrological model structure with a two layer reservoir architecture and a nonlinear percolation conceptualization. The two upper zone reservoirs represent a free water zone and a tension water zone, wherein the former controls the percolation to the lower zones while the tension water zone mainly controls the evaporation and feeds the free water zone. The percolation is a nonlinear complex function of demand from the lower reservoirs and available supply of water from the upper zone reservoirs. Both the upper and lower zones also control the outflows. The SIXPAR model structure, which is a conceptual simplification of the SAC-SMA model with one upper and lower zone, evaporation and the concept of tension water zones but retains the complex conceptualization of percolation. These models are run at daily time steps using input forcing from selected basins (in Table 1). Additional details on the models can be found elsewhere (Burnash, 1995; Duan et al., 1992; Arkesteijn and Pande, 2013).The code used and further explanation for SIXPAR is provided in the Supplement.

2.5.2 Parameter ranges as model structures

Table 2 provides the "reference" parameter ranges for SAC-SMA. Table 3 provides various parameter ranges of SIXPAR, including so called "reference" ranges and "equivalent" ranges. The model structures and various parameter ranges that govern parameter magnitudes of models that are sampled from these ranges allow us to study the (decomposed) effect of structure architecture and parameter magnitudes on computed

HESSD
12, 3945–4004, 2015

complexity. We note that these two effects are mixed when arbitrary (here called "reference") parameter ranges of SAC-SMA and SIXPAR are considered. However the effect of structure architecture (and the role of the number of parameters) on complexity emerges when we control the ranges of the parameters. This is when we have "equivalent" parameter ranges for the two model structures.

The parameter range of SIXPAR model structure is made "equivalent" to the "reference" parameter range of SAC-SMA model structure by ensuring that (i) the upper bounds on the reservoir capacities of the two layers of SIXPAR is equal to the sum of upper bounds on the reservoir capacities of the corresponding layers of SAC-SMA model structure and (ii) the corresponding lower and upper bounds of the recession parameter ranges of SIXPAR model structure are the geometric means of the corresponding lower and upper bounds of the SAC-SMA recession parameters. In terms of abstract parameters, this would then mean that the set of αs (abstract parameters) corresponding to the SIXPAR model structure are a subset of αs corresponding to the SAC-SMA model structure. Hence SIXPAR model structure would be nested within SAC-SMA structure in "abstract" sense.

In order to study the effect of parameter magnitudes on model complexity, we restrict our attention to the "reference" ranges of SIXPAR and contrain the paramater ranges in three ways. The three parameter ranges are called (i) "High recession", (ii) "Low recession", and (iii) "High storage/Low recession". These correspond to the "reference" parameter ranges for SIXPAR except that (i) corresponds to the case where the lower bounds of the recession ranges for the two layers are higher than the means of the corresponding "reference" ranges, (ii) corresponds to the case where the upper bounds of the recession ranges are lower than the means of the corresponding "reference" ranges and (iii) corresponds to the case where the means of storage capacities are larger than the means of the corresponding "reference" ranges and where the recession ranges are the same as in (ii). The three parameters ranges define three different model structures. The "High recession" and "Low recession" model structures are nested within the "reference" model structure of SIXPAR, while the "Low recession" SIXPAR model

HESSD
12, 3945–4004, 2015

structure is nested within "High storage/Low recession" model structure. Finally, both the "Low recession" and "High storage/Low recession" model structures are nested wthin the "equivalent" SIXPAR model structure.

The complexities of SIXPAR model structures for "reference", "equivalent", and (i)–(iii) ranges are computed on the selected hydrological data sets of MOPEX basins (see Table 1) and compared with the SAC-SMA model structure complexities computed on the same basins for its "reference" parameter range. The complexities of the model structures corresponding to the specified ranges are computed using Algorithm 2. It uses resampled basin scale potential evaporation and precipitation data using Algorithm 1.

3 Results

The Algorithm 2 provides complexity computations for each of the two structures for the parameter sets sampled from ranges defined in Tables 2 and 3. The algorithm uses input forcing realizations resampled by Algorithm 1 from input forcings of the selected MOPEX basins. The parameters are sampled using Latin Hypercube Sampling. As a result, Algorithm 2 provides a collection of $\{\beta_1, \beta_2, \beta_3\}$ corresponding to parameter sets that are sampled from a specified range for each model structure. Note that the algorithm computes one set of $\{\beta_1, \beta_2, \beta_3\}$ corresponding to one sampled parameter set. A corresponding distribution of $F(h,N) = \frac{f(h,N)}{N^2} = \beta_2 + \frac{\beta_1}{N} + \frac{\beta_0}{N^2}$ as a function of N can therefore be obtained. Figure 8 demonstrates the variation of 50th percentile values of $F(h,N)$ (over the 500 parameters sampled from "equivalent" parameter ranges) with N for the SIXPAR model structure using data from basin "NC".

The different curves correspond to different month permutations (step 5 of Algorithm 1) of the resampled input forcing data set. We note that the estimation of the curve is insensitive to the type of permutation in step 5 of Algorithm 1. We further note that $F(h,N)$ declines with increasing N and reaches an asymptote for large N. Since $F(h,N)$ is a function of complexity, represented by "h", and N, the value of $F(h,N)$ at large

HESSD
12, 3945–4004, 2015

N (when $F(h,N)$ asymptotes and becomes insensitive to N) reveals the measure of complexity (β_2). The asymptotic value of $F(h,N)$ is used to compare the complexity of different model structure set-ups. We here note that $F(h,N)$ already asymptotes around 2500 data points. Since Algorithm 1 resamples daily datasets, this means that $N = 2500$ is as large as $N \rightarrow \infty$ with regards to computing asymptotic complexity. It also means that when $N = 2500$ datapoints are enough to obtain a representation of the underlying processes that is not influenced by sampling uncertainty. In other words, this sample size is large enough to accurately reveal how unstable is the representation of underlying processes by SIXPAR model structure that is "equivalent" to SAC-SMA.

Figure 9 demonstrates that the asymptotic complexity for parameter ranges of SAC-SMA sampled from its "reference" ranges (Table 2) appears to be less complex than the asymptotic complexity of SIXPAR when sampled from its "reference" ranges (Table 3). This may appear counterintuitive since SIXPAR model structure is a conceptual simplification of SAC-SMA. However, similar conclusions have been drawn elsewhere for regression problems, where it has been shown that model complexity is both a function of magnitude and dimensionality of model parameters. For example, Bartlett (1998) and Vapnik and Chapelle (2000) find that the complexity of ANNs (Artificial Neural Networks) and SVMs (Support Vector Machines) are not only dependent on the dimensionality of the regressors but also crucially depend on the magnitude of the parameters. Ridge regression also regularizes the linear regression problem by penalizing the magnitude of the parameters (Marquardt and Snee, 1975).

Based on our construct of a continuum of model structures, which does not distinguish between model structures and parameter magnitudes, it may be possible that effect of parameter magnitudes on model complexity may compensate for the effect of structure architecture. We know that higher parameter dimensionality as a result of more complicated structure architecture leads to higher complexity. Thus magnitudes of parameters sampled from "reference" parameter range for SIXPAR compared to "reference" parameter range for SAC-SMA must have some compensating effect to re-

HESSD
12, 3945–4004, 2015

duce model complexity to such an extent, inspite of higher parameter dimensionality. We now look for those possible effect of parameter magnitudes on model complexity.

Figure 10 further studies the effect of sampling SIXPAR parameters from various ranges in Table 3 on its complexity. It suggests that complexity is less sensitive to recession parameters at lower magnitudes than it is at higher magnitudes since the median complexity for "low recession" range is closer to median complexity for "reference" recession range than the median complexity for "high recession" range. Further, the model complexity increases when the magnitudes of the recession parameters are increased. Finally, an increase in reservoir storage capacities leads to a reduction in model complexity. This can be seen from the median complexities of box plots corresponding to "Low recession" and "HS/LR" (i.e. high storage with low recession). Finally "Equivalent" SIXPAR model structure has the lowest complexity. We note that parameters sampled for "Equivalent" SIXPAR tends to have high storage and low recession when compared to the parameters sampled for the "Reference" SIXPAR. It is this effect of sampling high storage capacities and low recession parameters that brings down the complexity of SIXPAR in its "Equivalent" version.

The figure therefore suggests that high values of recession coefficients, i.e. small residence times, and low storage capacities lead to high complexity. This is intuitive, models with smaller residence time and lower storage capacities are more sensitive to perturbation in input forcing and hence have higher possibility of leading us to an unstable systen representation.

Over all, this demonstrates that the magnitude of parameters appear to have an effect on model complexity. Figure 11 shows a comparative variation of computed complexity with sample size N for SAC-SMA and SIXPAR. Figure 11a shows the comparison between the two models when parameters are sampled from "reference" parameter ranges and Fig. 11b compares the two model structures when the parameters are sampled from "equivalent" parameter ranges. The y axis, $P_N = P_{N,\gamma^*_N}$, where γ^*_N is the one that maximizes $P_{N,\gamma} N^2 \gamma^2$ in Eq. (11). Then from equality (11), $P_{N,\gamma^*} = \frac{f(h,N)}{N^2 \gamma^{*2}}$ is an increasing function of model complexity as defined by the 3-tuple $\{\beta_2, \beta_1, \beta_0\}$.

HESSD
12, 3945–4004, 2015

Both the figures demonstrate that the differences in complexities of the two model structures are more evident for small sample sizes. In effect, this figure demonstrates the decomposed effect of parameter dimensionality and parameter magnitudes on model complexity. Figure 11a suggests that SIXPAR model structure is more complex, due to sampled parameters having higher recession values and lower reservoir storage capacities for all sample sizes N. Meanwhile Fig. 11b shows SAC-SMA is more complex for all sample sizes N when the sampled parameter sets of SIXPAR are from ranges that are "equivalent" to SAC-SMA parameter ranges. The change in complexity and ordering from Fig. 11b to Fig. 11a is due to the effect of parameter magnitudes. The comparison suggests that parameter magnitudes also play a role in model complexity and that parameter dimensionality is an incomplete measure of complexity and hence prediction uncertainty. Figure 12 presents the case again for the asymptotic complexities β_2 of "reference" SAC-SMA, "reference" SIXPAR and "equivalent" SIXPAR.

The complexities are computed using input forcings from historical dataset of "NC" basin using Algorithm 1. Is the conclusion that parameter magnitudes may have an effect on model complexity sensitive to the basin that is selected for resampling of input forcings? Figure 13 plots the asymptotic complexities for the same ranges of SIXPAR model structure on input forcing resampled from CA, IA, GA and ME MOPEX basins that are from different hydro-climatic regions of continental US (Table 1). We observe a similar pattern in asymptotic complexities with parameter ranges and hence with parameter magnitudes.

4 Discussion

The evidence from Figs. 12 and 13 suggest that (i) model complexity is increasing in parameter dimensionality when parameter magnitudes of two model structures are "equivalent" and (ii) model complexity depends on the magnitudes of model parameters irrespective (to a certain extent) of model parameter dimensionality. Since model complexity is linked to instability in process representation and hence predictive uncer-

HESSD
12, 3945–4004, 2015

tainty, it then follows that predictive uncertainty of a model structure need not be lower if it has lower number of parameters. A SIXPAR model in one application with lower number of parameters but with "high recession" parameter values may have higher predictive uncertainty than an application of SAC-SMA model that is parameterized from "reference" parameter ranges (given in Table 2).

An important implication for complexity controlled model selection is then that parameter range specification should be application dependent. The modelling of a fast catchment with shallow unsaturated or saturated zones requires high recession and low reservoir capacity ranges. Our results (though for SIXPAR but may be extended to other models as well) demonstrate that complexity and hence predictive uncertainty is more sensitive to these parameters ranges since model complexity is high for such recession and reservoir capacity values. Model selection should consider parameter magnitudes in addition to parametric dimensionality when modelling such catchments. On the other hand, model parameter dimensionality may be a sufficient criterion to select a model with low prediction uncertainty in modelling slower basins.

Figure 13 demonstrated that for a given specification of parameter range, the magnitudes of asymptotic complexities are different for different basins. This indicates the influence of basin specific wetness conditions since higher magnitude of input forcings leads to a larger model output space and hence larger magnitudes of estimated complexities. The hydrologic ratios of these basins presented in Table 1 may be indicative of this. The CA basin is extremely dry with low runoff ratio while IA basin is moderately dry with a moderate ratio of annual evaporation to annual potential evaporation $\left(\frac{E}{E_P}\right)$. The asymptotic complexities of CA are lower than those of IA for corresponding SIXPAR variants (Fig. 13). Yet their asymptotic complexities are a lot lower than those of the remaining 3 basins. Incidently these 3 basins have dryness indices $\frac{E_P}{P} < 1$. They also have higher $\left(\frac{E}{E_P}\right)$ ratios. Thus the last 3 basins are comparatively wetter. Had we normalized the input forcings (by substracting the mean and dividing by SD), the correlation structure in input forcings on model structure complexities would have been

HESSD
12, 3945–4004, 2015

revealed. A detailed analysis of such effect on computing model complexity and of its own interpretation of complexity is left for future research.

Farmer et al. (2003) noted that more complex model structures are needed for dry catchments. This is because the runoff response of these catchments is more sensitive to small perturbations in input forcing than wetter catchments. Dry catchments experience more disruption of connectivity than wet catchments. The notion of complexity as proposed in this paper is also defined as a measure of sensitivity of modelled responses to perturbations in input forcings. The paper formally builds the notion of complexity and measures it. The context of model complexity is how stable or unstable the model is to input pertubations. This then leads to instability of system representation and prediction uncertainty. We find that sensitivity of model outcomes to input perturbations does not just depend on how complicated the architecture of model structure is (for example the structure variants of Farmer et al. (2003), Bai et al. (2009) and others) but also on the magnitude of the parameters that define a model structure. Higher recession coefficients and smaller storage capacities result in response variability at finer/shorter time scales (assuming the input forcing remains the same).

This is not to say that more complex model structures are always unsuitable. A top-down modelling approach proposes to increase model structure complexity in a nested fashion, starting with model structures of lower complexity and increasing the complexity till system representation degrades. However the formalism presented here takes one step further. It allows the possibility of not always rejecting a more complex model, even if it has higher complexity, if the ratio of reduced performance with increased complexity is less than a certain threshold (λ^* in Eq. 3). This allows the possibility that a more complex model with poorer performance on one realization of observations may perform better on another realization. This "acceptability" threshold is derived from the information embedded within the observations of the underlying processes.

We note that the notion of complexity and stable process representation is not limited to the use of one performance metric. Any performance measure, such as based on flow duration curve or Nash–Shutcliffe efficiency, can be used as long as it is a valid

HESSD
12, 3945–4004, 2015

metric. Thus the results obtained are general and testable on a wide variety of empirical evidence on appropriate model complexity and process representation that has been documented so far. Further, the method to estimate complexity is independent of the type of hydrological model used. Hence it is applicable for conceptual models, physically based models, empirical models as well as data driven models. For example Arkesteijn and Pande (2013) estimated the complexities of flexible conceptual rainfall–runoff models and demonstrated the applicability of the theory for a class of linear regression models.

Finally, we here highlight one limitation of the approach. The notion of complexity control on prediction uncertainty is based on a triangle inequality, wherein prediction uncertainty is bounded by the measure of complexity. It thus rests on the idea that controlling the measure of complexity only avoids the possibility that variation in model performance over two different realizations of data is not large. In context of top-down modeling approach, if we gradually ease the control on complexity, i.e. make models more complex, variation in model performance gradually increases as well. However, if this increase in model complexity is guided by better system representation, the possible increase in variation of model performance may be compensated by better average model performance to a certain extent.

5 Conclusions

Model complexity is an important criterion in model selection. This paper showed that this is because instability in hydrological system representation and prediction uncertainty are functions of model complexity. After demonstrating the connection between unstable model representation and model complexity, it was shown in a step by step manner how this complexity can be estimated. This was based on measuring differences between simulations of a model under different realizations of model forcing. Algorithms were then suggested to estimate model complexity. Algorithm 1 was created to resample multiple realizations of input forcing data sets, and Algorithm 2 was

HESSD
12, 3945–4004, 2015

created to estimate complexity based on inequality (11) using resampled input dataset generated by Algorithm 1. Complexities of two model structures, SAC-SMA and SIX-PAR, were then computed using these algorithms.

The model complexities of the two model structures, SIXPAR and SAC-SMA were computed on resampled input data sets from basins that spanned across the continental US. The model complexities for SIXPAR were estimated for various parameter ranges. The range specifications included an "equivalent" range wherein the ranges were such that total soil moisture storage and recession parameters of SIXPAR were equivalent to the "reference" ranges of SAC-SMA, and other parameter ranges that constrained the recession parameters to be either at the higher or lower end of the reference range as well as the storage parameters towards the higher end of the reference range.

SIXPAR was found to be more complex than SAC-SMA model structure when "reference" ranges were used. However when both the model structures were applied using respective "equivalent" parameter ranges, SAC-SMA was found to be more complex, as expected. We further observed, on multiple basins data sets, that computed complexity of SIXPAR increased with lower storage capacity and/or higher recession coefficients. Thus a conceptually simple model structure, such as SIXPAR, can be more complex than an intuitively more complex model structure, such as SAC-SMA. We therefore concluded, with important implications for robust model selection, that magnitudes of feasible model parameters influence the complexity of the model selection problem just as parameter dimensionality does. Hence we recommend caution in thinking that parameter dimensionality is the only indicator of stablity of hydrological model selection and prediction problem.

HESSD
12, 3945–4004, 2015

Appendix A: Metric space

For any two vectors $\boldsymbol{u}$ and $\boldsymbol{v}$ in a N-dimensional space $\boldsymbol{X}$, a real valued function that measures the distance between the two vectors, $d : \boldsymbol{X} \times \boldsymbol{X} \rightarrow \mathbb{R}_+$ is a metric if the following 3 conditions are satisfied (Chiang, 1984, p.73):

1. $d(\boldsymbol{u},\boldsymbol{v}) = 0 \quad \text{for } \boldsymbol{u} = \boldsymbol{v}$
2. $d(\boldsymbol{u},\boldsymbol{v}) = d(\boldsymbol{v},\boldsymbol{u}) \quad \text{for } \boldsymbol{u} \neq \boldsymbol{v}$
3. $d(\boldsymbol{u},\boldsymbol{v}) \leqslant d(\boldsymbol{u},\boldsymbol{w}) + d(\boldsymbol{w},\boldsymbol{v}) \quad \text{for } \boldsymbol{w} \neq \boldsymbol{u},\boldsymbol{v}.$

Appendix B: Triangle inequality

The triangle inequality states that the sum of the magnitudes of any two sides of a triangle is greater than the third (see condition 3 of Appendix A). Consider Fig. 1 and let the vectors be such that $\|\boldsymbol{A}\| \geqslant \|\boldsymbol{C}\|$. First consider the triangle formed by vectors $\boldsymbol{B},\boldsymbol{C}$ and $\boldsymbol{E}$ (not shown in Fig. 1). Then $\|\boldsymbol{B}\| + \|\boldsymbol{C}\| \geqslant \|\boldsymbol{E}\|$. Similarly $\|\boldsymbol{E}\| + \|\boldsymbol{D}\| \geqslant \|-\boldsymbol{A}\|$. Thus, $\|\boldsymbol{B}\| + \|\boldsymbol{C}\| + \|\boldsymbol{D}\| \geqslant \|\boldsymbol{A}\|$ or $\|\boldsymbol{A}\| - \|\boldsymbol{C}\| \leqslant \|\boldsymbol{B}\| + \|\boldsymbol{D}\|$.

Appendix C: Synthetic data sets

Two sets of synthetic data sets $\{r_1(t), r_2(t)\}_{t=1,\ldots,N}$ of size $N = 50$ each are generated 100 times. Synthetic effective rainfall is generated by $\widetilde{r}_j(t) = (\omega_1 \leq 0.85)\cdot 0 + (\omega_1 \geqslant 0.85)\cdot \omega_2, t = 1,\ldots,N$ and $j = 1,2$. Here ω_1 and ω_2 are random numbers generated from a uniform distribution and lie between 0 and 1. The synthetic rainfall is then corrupted by multiplicative noise, $r_j(t) = \widetilde{r}_j(t) \cdot e^{\mathrm{Log}(0.1)\cdot \upsilon_1}$, where $\upsilon_1 \sim \mathcal{N}(0,1)$.

HESSD
12, 3945–4004, 2015

Each pair of synthetic rainfall is run through a linear reservoir model, given by the following mass balance equation, to produce synthetic streamflow $\widetilde{q}_j(t), t = 1,\ldots,N$.

$$\frac{dS}{dt} = r(t) - \widetilde{q}_j(t),$$
$$\widetilde{q}_j(t) = \widetilde{k}S(t),$$
$$S(1) = 0,$$
$$t = 1,\ldots,N;\ j = 1,2$$

$\widetilde{k} = 0.2$ is chosen for generating the synthetic streamflow. $\widetilde{q}_j(t)$ is further corrupted by heteroskedastic noise with a SD of 0.10, to finally generate a pair of synthetic runoff $\{o_1(t), o_2(t)\}_{t=1,\ldots,N}$. That is, $o_j(t) = (1 + 0.10 \cdot \upsilon_2) \cdot \widetilde{q}_j(t), t = 1,\ldots,N, j = 1,2$. Here, $\upsilon_2 \sim \mathcal{N}(0,1)$.

Each pairs of data sets $\{\boldsymbol{r}_j, \boldsymbol{o}_j\}_{j=1,2}$, where $\boldsymbol{r}_j = \{r_j(t)\}_{t=1,\ldots,N}$ and $\boldsymbol{o}_j = \{o_j(t)\}_{t=1,\ldots,N}$, are then used to calculate $\|\boldsymbol{A}\|$ and $\|\boldsymbol{C}\|$ for 2 model structures $\wedge_1$ and $\wedge_2$.

Model structure $\wedge_2$ is a two reservoir model structure that is defined in the following. It represents a two layer hydrological model structure with the second layer fed by percolation from the first layer. The outflow from both the layers is assumed to be linear in its respective soil moisture storages. Percolation from the first layer to the second layer is also assumed to be linear in top layer soil moisture.

$$\frac{dS_1}{dt} = r(t) - k_1 S_1(t) - k_2 S_1(t),$$
$$\frac{dS_2}{dt} = k_2 S_1(t) - k_3 S_2(t),$$
$$p_j^{2,\boldsymbol{k}}(t) = k_1 S_1(t) + k_3 S_2(t),$$
$$S_1(1) = 0, S_2(1) = 0,$$
$$t = 1,\ldots,N,$$
$$\boldsymbol{k} = \{k_1, k_2, k_3\} \in [0,1]^3$$

HESSD
12, 3945–4004, 2015

Model structure $\wedge_1$ is a single reservoir model structure defined by the following equation:

$$\frac{dS}{dt} = r(t) - p^{1,k}(t),$$
$$p^{1,k}(t) = kS(t),$$
$$S(1) = 0,$$
$$t = 1, \ldots, N; j = 1, 2,$$
$$k \in [0, 1].$$

Model structure $\wedge_1$ is nested within model structure $\wedge_2$ because when the parameters of $\wedge_2$ are restricted such that $\boldsymbol{k} \in [0,1] \times \{0\} \times \{0\}$, it becomes model structure $\wedge_1$. By definition $\wedge_2$ is more complex than $\wedge_1$, since the former would have larger output space than the latter. Model structure $\wedge_2$ is more flexible than $\wedge_1$ since it can model both fast and slow flow by routing flows through top layers and top and bottom layers respectively.

To study the effect of larger complexity on instability in underlying process representation, *best* approximations from the two model structures are obtained on one dataset and then the performances of the corresponding approximations are measured on the second dataset from the pair of synthetic noisy data set generated above. This is done as follows. Consider the pair of synthetic data sets, $\{\boldsymbol{r}_j, \boldsymbol{o}_j\}_{j=1,2}$. The model structures are confronted with $\{\boldsymbol{r}_1, \boldsymbol{o}_1\}$ and *best* representations, parameterized by θ_1^{1*} and θ_1^{2*}, are obtained from $\wedge_1$ and $\wedge_2$ respectively. Here, $\theta_1^{1*} = \boldsymbol{k}^*$ and $\theta_1^{2*} = k^*$ and provide us with model simulations $\boldsymbol{p}_1^1$ and $\boldsymbol{p}_1^2$ respectively. The best representations from each model structure are then used to simulate streamflow on the input forcings of the second data set, $\boldsymbol{r}_2$. Thus simulations $\boldsymbol{p}_2^1$ and $\boldsymbol{p}_2^2$ are respectively obtained on another dataset. The distances $\|\boldsymbol{p}_2^m - \boldsymbol{o}_2\|$ and $\|\boldsymbol{p}_1^m - \boldsymbol{o}_1\|$ then represent $\|\boldsymbol{A}\|$ and $\|\boldsymbol{C}\|$ for a model structure $\wedge_m, m = \{1,2\}$. We let $\|\bullet\|$ to be the mean of absolute values, for e.g. $\|\boldsymbol{p}_1^m - \boldsymbol{o}_1\| = \sum_{t=1}^{N} \frac{|p_1^m(t) - o_1(t)|}{N}$. This choice of metric implies mean absolute errors when

HESSD
12, 3945–4004, 2015

distances between a model simulation and observed values are measured. Similarly, $\|\boldsymbol{p}_1^m - \boldsymbol{p}_2^m\|$ provides us an estimate of $\|\boldsymbol{B}\| = \sum_{t=1}^{N} \frac{|p_1^m(t) - p_2^m(t)|}{N}$.

The above process is repeated 100 times. By doing so, we obtain 100 values of $|\|\boldsymbol{A}\| - \|\boldsymbol{C}\||$ and $\|\boldsymbol{B}\|$. Kernel density estimation is then used to estimate the probability of exceedences for $|\|\boldsymbol{A}\| - \|\boldsymbol{C}\||$ and $\|\boldsymbol{B}\|$ for the two model structures, i.e. $\Pr(|\|\boldsymbol{A}\| - \|\boldsymbol{C}\||\geqslant\epsilon)$ and $\Pr(\|\boldsymbol{B}\|\geqslant\epsilon)$ respectively.

Note that $\|\boldsymbol{C}\|$ is the performance of the best representation of underlying processes (inferred from one data set) on another realization of data. The probability $\Pr(|\|\boldsymbol{A}\| - \|\boldsymbol{C}\||\geqslant\epsilon)$ thus quantifies the instability in process representation over different realizations of data. A hydrological model structure that provides a less stable representation of underlying processes has higher $\Pr(|\|\boldsymbol{A}\| - \|\boldsymbol{C}\||\geqslant\epsilon)$ for all possible values of $\epsilon\geqslant 0$. This reiterates the definition of instability in process representation. A less stable process representation leads to larger deviations in model performances over any two realizations of data. Meanwhile the probability $\Pr(\|\boldsymbol{B}\|\geqslant\epsilon)$ represents the variation in the extent of model structure output space. It essentially represents how large is the output space of a model structure.

The stability of process representation can be reinterpreted in terms of the variability in *best* model representations obtained on two different realizations of data. A model structure that provides more unstable representation also has more variability in *best* model representations obtained on two different realizations of data. In order to interpret instability in such a manner, we obtain best model representations θ_j^{m*} on each of the two data sets, $\{\boldsymbol{r}_j, \boldsymbol{o}_j\}_{j=1,2}$ and from model structure $\wedge_m, m = 1,2$. Since we generate 100 such dataset pairs, we have 100 pairs of best representations from each model structure. Kernel densities are then estimated for $\Pr\left(\theta_j^{m*}\right)$ in order to study how variability in best representations of the underlying processes differs across two nested model structures.

The Supplement related to this article is available online at doi:10.5194/hessd-12-3945-2015-supplement.

HESSD
12, 3945–4004, 2015

Author contributions. S. Pande wrote the manuscript. L. Arkesteijn and S. Pande ran the models and critically analyzed the results. H. Savenije and L. A. Bastidas contributed to hydrological interpretation of model complexity. All authors reviewed the work.

Acknowledgements. We express our deep gradtitude to Andras Bardossy and to two anonymous reviewers for their critical review of a previous verison of the manuscript. The authors are also grateful to the comments of Shervan Gharari and Mehdi Moayeri that helped shape this version of the paper.

References

Arkesteijn, L. and Pande, S.: On hydrological model complexity, its geometrical interpretations and prediction uncertainty, Water Resour. Res., 49, 7048–7063, doi:10.1002/wrcr.20529, 2013. 3948, 3949, 3962, 3964, 3967, 3968, 3970, 3971, 3979

Bai, Y., Wagener, T., and Reed, P.: A top-down framework for watershed model evaluation and selection under uncertainty, Environ. Modell. Softw., 24, 901–916, 2009. 3960

Bartlett, P. L.: The sample complexity of pattern classification with neural networks: The size of the weights is more important than the size of the network, IEEE T. Inform. Theory, 44, 525–536, 1998. 3974

Beven, K. J.: A manifest for the equifinality thesis, J. Hydrol., 320, 18–36, 2006. 3947

Brooks, P. D., Troch, P. A., Durcik, M., Gallo, E., and Schlegel, M.: Quantifying regional scale ecosystem response to changes in precipitation: not all rain is created equal, Water Resour. Res., 47, W00J08, doi:10.1029/2010WR009762, 2011. 3967, 3989, 3998

Burnash, R. J. C.: The NWS river forecast system-catchment modelling, in: Computer Models of Watershed Hydrology, edited by: Singh, V. P., Water Resource Publications, Highlands Ranch, Colorado, USA, 311–366, 1995. 3971

Buttsa, M. B., Paynea, J. T., Kristensenb, M., and Madsen, H.: An evaluation of the impact of model structure on hydrological modelling uncertainty for streamflow simulation, J. Hydrol., 298, 242–266, 2004. 3960

Cavanaugh, J. E. and Neath, A. A.: Generalizing the derivation of the Schwarz information criterion, Commun. Stat. Theory, 28, 49–66, 1999.

Cucker, F. and Smale, S.: On the mathematical foundations of learning, B. Am. Math. Soc., 39, 1–49, 2002.

HESSD

12, 3945–4004, 2015

Davidson, R. and MacKinnon, J. G.: Econometric Theory and Methods, Oxford University Press, New York, 1–760, 2004.
Duan, Q., Sorooshian, S., and Gupta, V.: Effective and efficient global optimization for conceptual rainfall–runoff models, Water Resour. Res., 28, 1015–1031, 1992. 3971
Duan, Q., Schaake, J., Andréassian, V., Franks, S., Goteti, G., Gupta, H. V., Gusev, Y. M., Habets, F., Hall, A., Hay, L., Hogue, T., Huang, M., Leavesley, G., Liang, X., Nasonova, O. N., Noilhan, J., Oudin, L., Sorooshian, S., Wagener, T., and Wood, E. F.: Model Parameter Estimation Experiment (MOPEX): an overview of science strategy and major results from the second and third workshops, J. Hydrol., 320, 3–17, doi:10.1016/j.jhydrol.2005.07.031, 2006. 3967, 3989, 3998
Farmer, D., Sivapalan, M., and Jothityangkoon, C.: Climate, soil, and vegetation controls upon the variability of water balance in temperate and semiarid landscapes: downward approach to water balance analysis, Water Resour. Res., 39, 1035, doi:10.1029/2001WR000328, 2003. 3958, 3960
Gelman, A., Jakulin, A., Pittau, M. G., and Yu-Sung, S.: A weakly informative default prior distribution for logistic and other regression, Ann. Appl. Stat., 2, 1360–1383, 2008.
Gupta, H. V. and Nearing, G. S.: Debates on water resources: using models and data to learn – a systems theoretic perspective on the future of hydrological science, Water Resour. Res., 50, 5351–5359, doi:10.1002/2013WR015096, 2014. 3947
Gupta, H. V., Sorooshian, S., and Yapo, P. O.: Towards improved calibration of hydrologic models: multiple and non-commensurable measures of information, Water Resour. Res., 34, 751–763, 1998. 3947
Gupta, H. V., Wagener, T., and Liu, Y.: Reconciling theory with observations: elements of a diagnostic approach to model evaluation, Hydrol. Process., 22, 3802–3813, doi:10.1002/hyp.6989, 2008. 3946, 3947, 3948, 3949, 3958
Jakeman, A. J. and Hornberger, G. M.: How much complexity is warranted in a rainfall–runoff model?, Water Resour. Res., 29, 2637–2649, 1993.
Kass, R. E. and Raftery, A. E.: Bayes factors, J. Am. Stat. Assoc., 90, 773–795, 1995.
Keating, E. H., Doherty, J., Vrugt, J. A., and Kang, Q.: Optimization and uncertainty assessment of strongly nonlinear groundwater models with high parameter dimensionality, Water Resour. Res., 46, W10517, doi:10.1029/2009WR008584, 2010. 3949
Kundzewicz, C. W. and Robson, A. J.: Change detection in hydrological records: a review of the methodology, Hydrolog. Sci. J., 49, 7–19, 2004. 3968

HESSD
12, 3945–4004, 2015

Marquardt, D. W. and Snee, R. D.: Ridge regression in practise, Am. Stat., 29, 3–20, 1975. 3974
Marshall, L., Nott, D., and Sharma, A.: Hydrological model selection: a Bayesian alternative, Water Resour. Res., 41, W10422, doi:10.1029/2004WR003719, 2005. 3949
Montanari, A. and Brath, A.: A stochastic approach for assessing the uncertainty of rainfall–runoff simulations, Water Resour. Res., 40, W01106, doi:10.1029/2003WR002540, 2004. 3949
Pande, S., McKee, M., and Bastidas, L. A.: Complexity-based robust hydrologic prediction, Water Resour. Res., 45, W10406, doi:10.1029/2008WR007524, 2009. 3948, 3949
Pande, S., Bastidas, L. A., Bhulai, S., and McKee, M.: Parameter dependent convergence bounds and complexity measure for a class of conceptual hydrological models, J. Hydroinform., 14, 443–463, doi:10.2166/hydro.2011.005, 2012. 3949
Politis, D. and Romano, J.: The stationary bootstrap, J. Am. Stat. Assoc., 89, 1303–1313, 1994. 3968
Renard, B., Kavetski, D., Thyer, M., Kuczera, G., and Franks, S.: Understanding predictive uncertainty in hydrologic modeling: the challenge of identifying input and structural errors, Water Resour. Res., 46, W05521, doi:10.1029/2009WR008328, 2009. 3947
Ross, S. M.: Stochastic Processes, 2nd Edn., John Wiley and Sons, Inc., New York, 1996. 3963
Savenije, H. H. G.: Equifinality, a blessing in disguise?, Hydrol. Process., 15, 2835–2838, 2001. 3947
Sawicz, K. A., Kelleher, C., Wagener, T., Troch, P., Sivapalan, M., and Carrillo, G.: Characterizing hydrologic change through catchment classification, Hydrol. Earth Syst. Sci., 18, 273–285, doi:10.5194/hess-18-273-2014, 2014. 3947
Schoups, G., van de Giesen, N. C., and Savenije, H. H. G.: Model complexity control for hydrologic prediction, Water Resour. Res., 44, W00B03, doi:10.1029/2008WR006836. 2008. 3949
Silberstein, R. P.: Hydrological models are so good, do we still need data?, Environ. Modell. Softw., 21, 1340–1352, 2006.
Sivapalan, M., Blöschl, G., Zhang, L., and Vertessy, R.: Downward approach to hydrological prediction, Hydrol. Process., 17, 2101–2111, doi:10.1002/hyp.1425, 2003. 3947
Slate, E. H.: Parameterizations for natural Exponential families with quadratic functions, J. Am. Stat. Assoc., 89, 1471–1482, 1994.

HESSD
12, 3945–4004, 2015

Son, K. and Sivapalan, M.: Improving model structure and reducing parameter uncertainty in conceptual water balance models through the use of auxiliary data, Water Resour. Res., 43, W01415, doi:10.1029/2006WR005032, 2007. 3960

Tierney, T. and Kadane, J. B.: Accurate approximations for posterior moments and marginal densities, J. Am. Stat. Assoc., 81, 82–86, 1986.

Vapnik, V.: Estimation of Dependencies Based on Empirical Data, Springer Verlag, New York, 1982. 3948, 3955

Vapnik, V. and Chapelle, O.: Bounds on error expectation for support vector machines, Neural Comput., 12, 2013–2036, 2000. 3974

Winsemius, H. C., Savenije, H. H. G., Gerrits, A. M. J., Zapreeva, E. A., and Klees, R.: Comparison of two model approaches in the Zambezi river basin with regard to model reliability and identifiability, Hydrol. Earth Syst. Sci., 10, 339–352, doi:10.5194/hess-10-339-2006, 2006. 3947

Ye, M., Meyer, P. D., and Neuman, S. P.: On model selection criteria in multimodel analysis, Water Resour. Res., 44, W03428, doi:10.1029/2008WR006803, 2008. 3949

Young, P.: Top-down and data-based mechanistic modelling of rainfall–flow dynamics at the catchment scale, Hydrol. Process., 17, 2195–2217, doi:10.1002/hyp.1328, 2003. 3949

Young, P. C.: Hypothetico-inductive data-based mechanistic modeling of hydrological systems, Water Resour. Res., 49, 915–935, doi:10.1002/wrcr.20068, 2013.

HESSD
12, 3945–4004, 2015

Table 1. Basins used in this study. $\overline{P}$ = Mean Annual Precipitation, $\overline{E}_P$ = Mean Annual E_P. $\overline{P}$ and $\overline{E}_P$ are calculated using data from the period 1948–1970. The hydrologic ratios $\frac{Q}{P}$ [–], $\frac{E}{E_P}$ [–] and $\frac{E_P}{P}$ [–] are ratios of annual runoff to annual precipitation, annual evaporation to annual potential evaporation and annual potential evaporation to annual precipitation (dryness index) respectively. Data obtained from Duan et al. (2006) and Brooks et al. (2011).

Site Id	Area [km^2]	$\overline{P}$ [mm yr^{-1}]	$\overline{E}_P$ [mm yr^{-1}]	$\frac{Q}{P}$ [–]	$\frac{E}{E_P}$ [–]	$\frac{E_P}{P}$ [–]	Code
03451500	945.00	1491	820	0.59	0.96	0.43	NC
11138500	281.00	380	1334	0.05	0.32	2.94	CA
05479000	1308.00	711	977	0.20	0.60	1.31	IA
02228000	2790.00	1215	1132	0.33	0.75	0.89	GA
01060000	141.00	1100	N/A	0.47	0.85	0.62	ME

HESSD
12, 3945–4004, 2015

Table 2. SAC-SMA model structure parameter ranges used in the study.

Parameters	Description	"Reference"	Parameter	Description	"Reference"
UZTWM [mm]	Upper zone tension water capacity	1–150	UZWFM [mm]	Upper zone free water capacity	1–150
UZK [day^{-1}]	Fractional daily upper zone free water withdrawal rate	0.1–0.5	PCTIM [–]	Minimum impervious area	0–0.1
ADIMP [–]	Additional impervious area	0–0.4	RIVA [–]	Riparian vegetation area	0
ZPERC [–]	Maximum percolation rate	1–250	REXP [–]	Exponent for percolation equation	1–5
LZTWM [mm]	Lower zone tension water capacity	1–1000	LZFSM [mm]	Upper zone free water capacity	1–1000
LZFPM [mm]	Lower zone primary free water capacity	1–1000	LZSK [day^{-1}]	Fractional daily supplemental withdrawal rate	0.01–0.25
LZPK [day^{-1}]	Fractional daily primary withdrawal rate	0.0001–0.025	PFREE [–]	Fraction of percolated water directly to lower zone free storages	0.0–0.6
RSERV [–]	Fraction of lower zone free water not transferred to lower zone tension water	0.3	SIDE [–]	Ratio of non-channel baseflow to channel baseflow	0.0

HESSD
12, 3945–4004, 2015

Table 3. SIXPAR model structure parameter ranges used in the study.

Parameter	Description	"Reference"	"High recession"	"Low recession"	"High storage/ Low recession"	"Equivalent"
UM [mm]	Upper zone reservoir capacity	0–50	0–50	0–50	1–300	1–300
UK [day^{-1}]	Upper zone recession parameter	0–1	0.75–1.00	0.10–0.25	0.10–0.25	0–0.5
BM [mm]	Lower zone reservoir capacity	0–50	0–50	0–50	1–3000	0–3000
BK [day^{-1}]	Lower zone recession coefficient	0–1	0.75–1.00	0.001–0.005	0.001–0.005	0–0.07906
Z [–]	Percolation parameter 1	0–1	1–250	1–250	1–250	0–1
X [–]	Percolation parameter 2	0–10	1–5	1–5	1–5	0–10

HESSD
12, 3945–4004, 2015

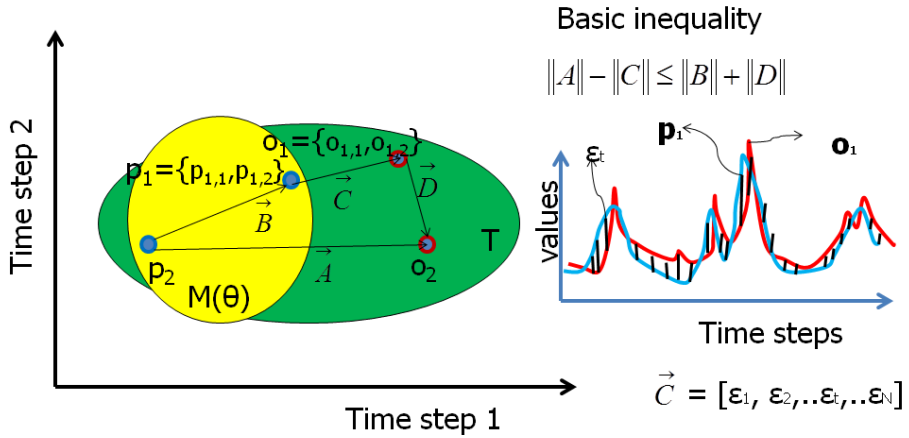

Figure 1. An illustration of model output space. Consider a model parameterized by θ, say $M(\theta)$. Difference between observed and simulated streamflows (for e.g. $\boldsymbol{o}_1$ and $\boldsymbol{p}_1$), say of size N, defines a N-dimensional vector $\boldsymbol{C}$. $N = 2$ is considered for illustration purposes. The magnitude of this vector may represent a measure of model performance, such as Mean Absolute Error. A similar vector $\boldsymbol{A}$ may be obtained for another realization of the pair $(\boldsymbol{o}_2, \boldsymbol{p}_2)$. The N-vectors $(\boldsymbol{p}_1, \boldsymbol{p}_2)$ then define two simulation points in the mode output space. The distance between them is indicated by $\boldsymbol{B}$. Repeating realizations of such pairs then populates the model output space while an expectation of $\boldsymbol{B}$ over these realizations then measures the span of the model output space

HESSD
12, 3945–4004, 2015

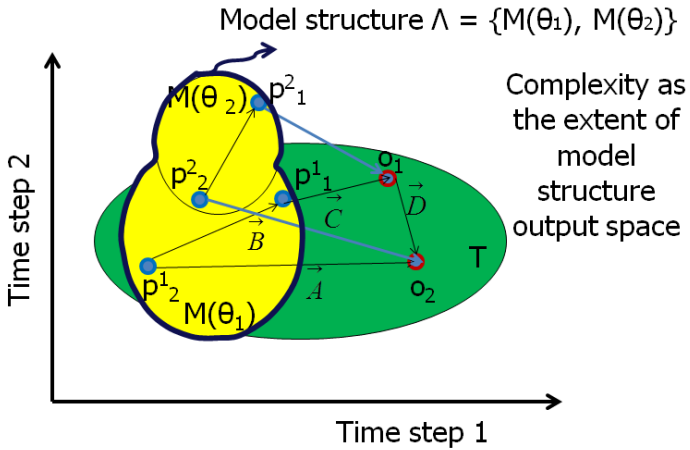

Figure 2. An illustration of model structure output space. Here a model structure is thought of as a collection of model parameter sets. Here consider such a model structure, $\wedge = \{M(\theta_1), M(\theta_2)\}$. Model simulations corresponding to observations $(\boldsymbol{o}_1, \boldsymbol{o}_2)$ for $M(\theta_1)$ and $M(\theta_2)$ are indicated by pairs $\left(\boldsymbol{p}_1^1, \boldsymbol{p}_2^1\right)$ and $\left(\boldsymbol{p}_1^2, \boldsymbol{p}_2^2\right)$ respectively. These pairs populate the model output spaces correspoding to models $M(\theta_1)$ and $M(\theta_2)$ respectively. Since the model structure is defined as a combination of the two models, the corresponding model structure output space is the union of constituting model output spaces.

HESSD
12, 3945–4004, 2015

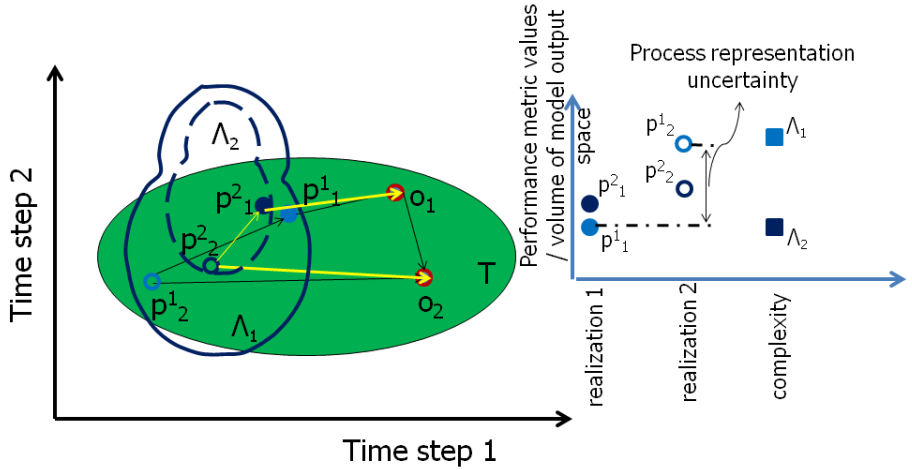

Figure 3. Instability in system representation. Consider two model structures, $\wedge_2$ that is nested in $\wedge_1$. The larger the model output space of $\wedge_1$ leads to higher possibility of differences between any two simulations, shown by $\boldsymbol{p}_1^1$ and $\boldsymbol{p}_2^1$, than $\wedge_2$ as shown by $\boldsymbol{p}_1^2$ and $\boldsymbol{p}_2^2$ for the same input forcings. This implies higher instability in system representation offered by $\wedge_2$.

HESSD
12, 3945–4004, 2015

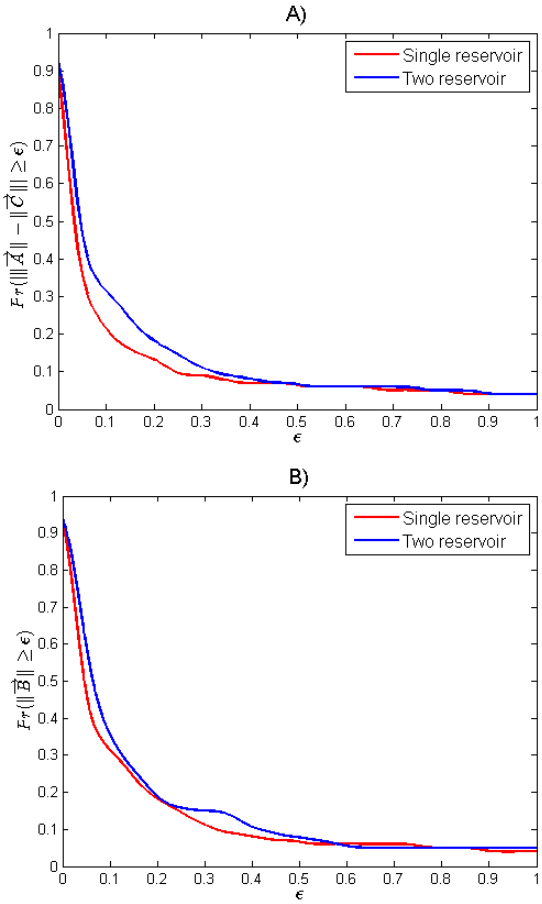

Figure 4. Estimation of $\Pr(|\|\boldsymbol{A}\| - \|\boldsymbol{C}\||\geq\epsilon)$ and $\Pr(\|\boldsymbol{B}\|\geq\epsilon)$ based on 100 synthetic data pairs. Two nested model structures are considered, i.e. single reservoir and two reservoir model structure. See Appendix C for additional details.

HESSD
12, 3945–4004, 2015

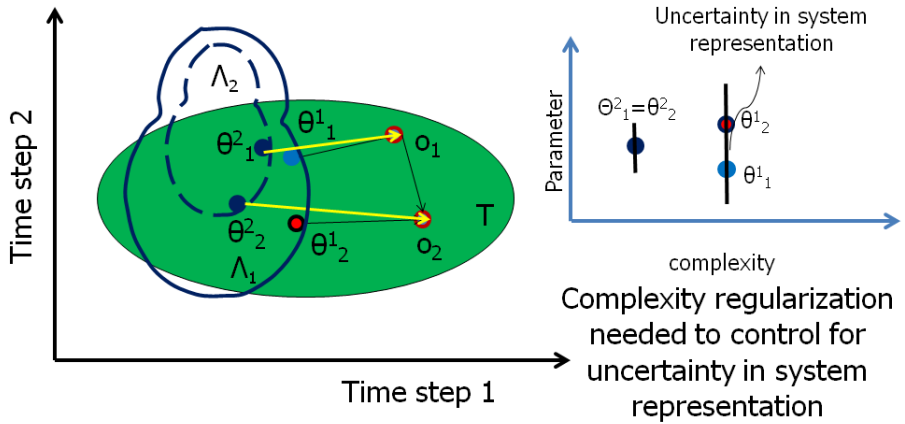

Figure 5. Instability in system representation over different realizations of observations. Consider two model structures, $\wedge_2$ that is nested in $\wedge_1$ as in Fig. 3. Let $M\left(\theta_1^1\right)$ and $M\left(\theta_2^1\right)$ be best system representations offered by model structure $\wedge_1$ based on two observations $\boldsymbol{o}_1$ and $\boldsymbol{o}_2$ respectively. Since the model structure $\wedge_2$ has smaller model output space and given that hydrological model outputs are continuous in their parameters, best system representations $M\left(\theta_1^2\right)$ and $M\left(\theta_2^2\right)$ corresponding to $\boldsymbol{o}_1$ and $\boldsymbol{o}_2$ offered by $\wedge_2$ are closer to each than those offered by $\wedge_1$. Here we assume $\theta_1^2 = \theta_2^2$ and suggest that larger possibility of variation in best model representation implies higher instability in system representation offered by $\wedge_1$.

HESSD
12, 3945–4004, 2015

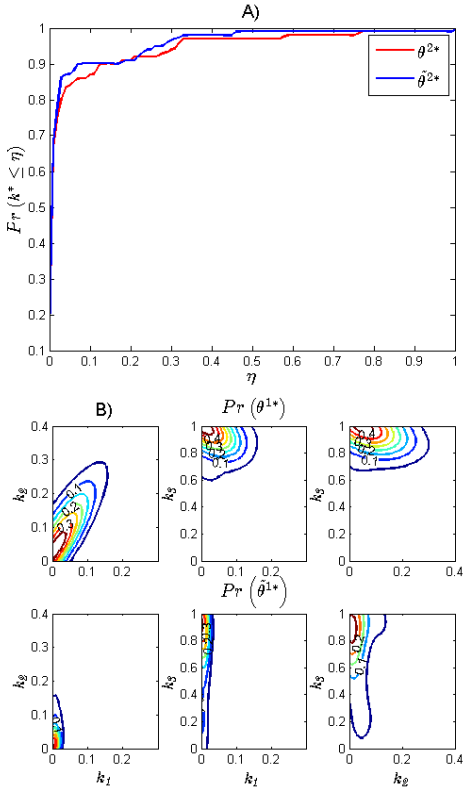

Figure 6. Kernel density estimation of variability in system representations selected from **(a)** $\wedge_2$ and **(b)** $\wedge_1$ over 100 pairs of data realizations from the data generating process provided in Appendix C. While Fig. 6b plots the pairwise kernel density estimate of the same for $\wedge_1$. θ^{k*} and $\tilde{\theta}^{k}$ represent two models selected from model structure $\wedge_k$ on two realizations of data. Here $\wedge_2$ is a single reservoir model structure while $\wedge_1$ is a 3 parameter two-reservoir model structure. See Appendix C for additional details.

HESSD
12, 3945–4004, 2015

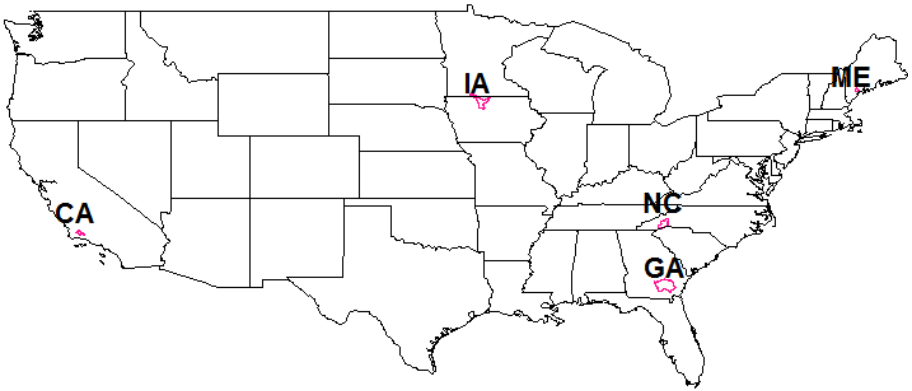

Figure 7. A selection of basins across the US spanning different hydro-climatic regions. Data obtained from Duan et al. (2006) and Brooks et al. (2011).

HESSD
12, 3945–4004, 2015

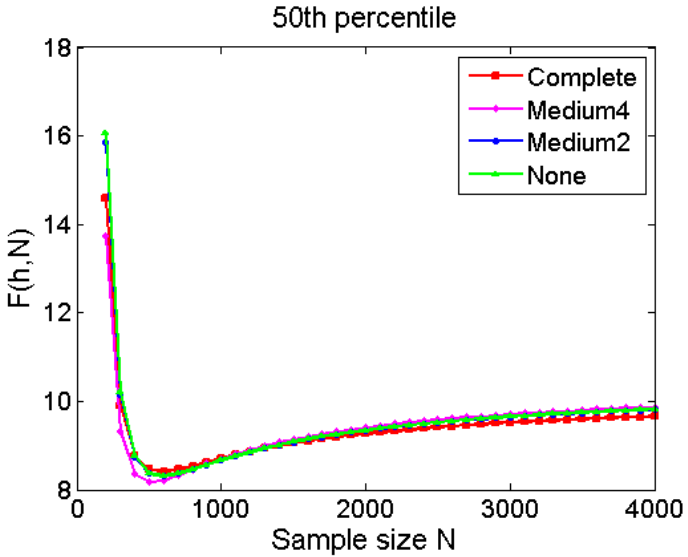

Figure 8. Complexity curves for 50th percentile values of different month permutations. Here "Medium 4" means that 4 blocks of 3 month tuples were permuted to create data resamples.

HESSD
12, 3945–4004, 2015

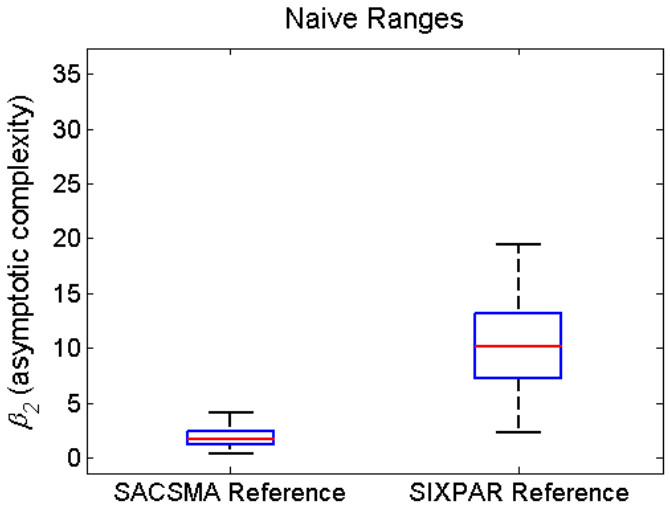

Figure 9. Asymptotic complexity using reference ranges for SAC-SMA and SIXPAR model structure.

HESSD
12, 3945–4004, 2015

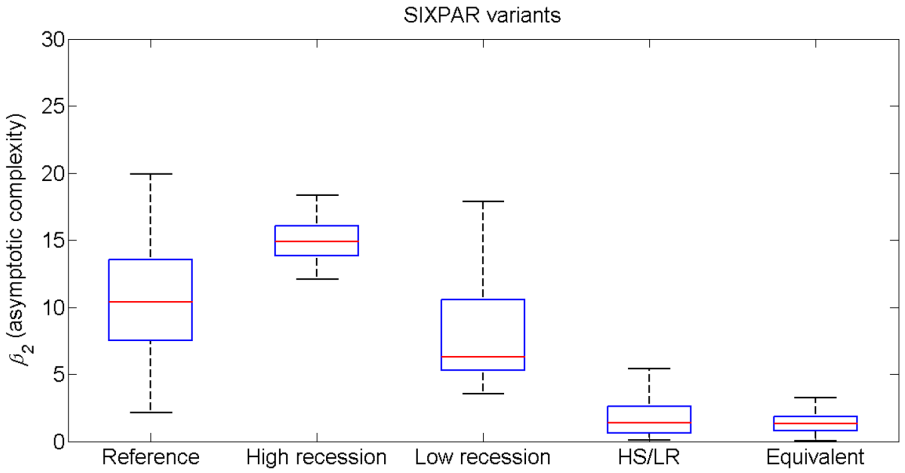

Figure 10. Asymptotic complexity using different parameters ranges for SIXPAR model structure.

HESSD
12, 3945–4004, 2015

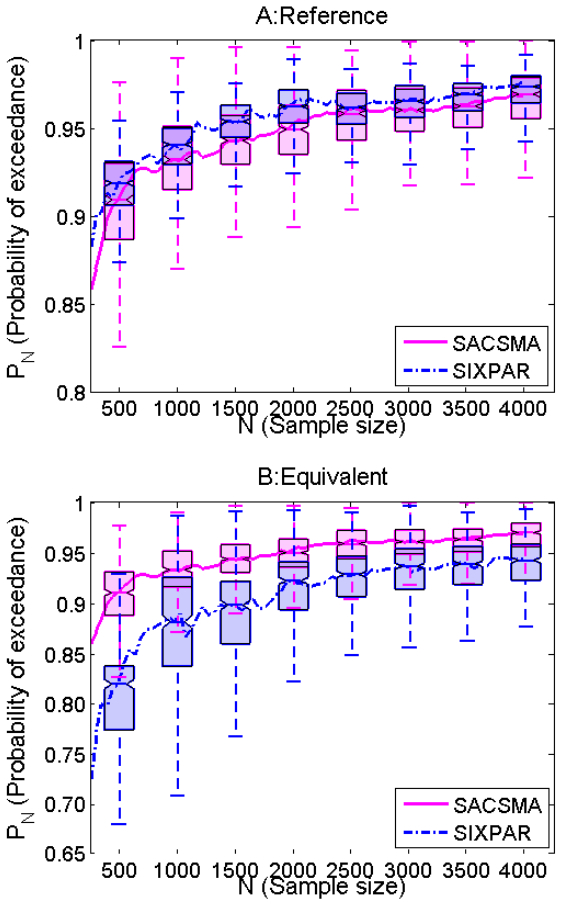

Figure 11. Variation of computed complexity with sample size *N* for SAC-SMA and SIXPAR. **(a)** Reference parameter ranges and **(b)** equivalent parameter ranges.

HESSD
12, 3945–4004, 2015

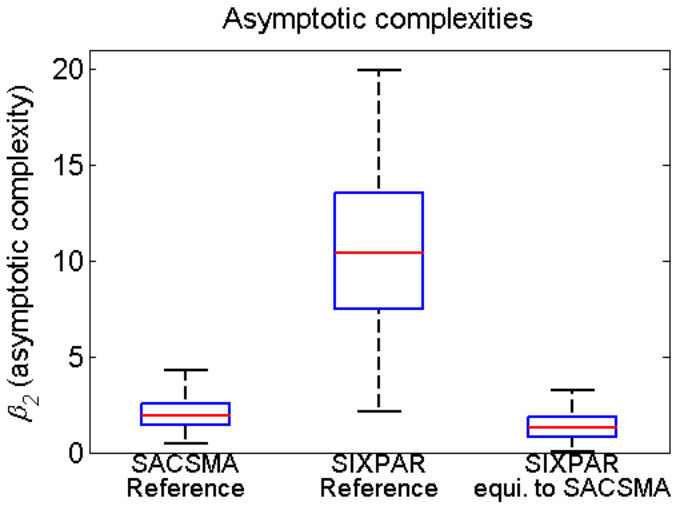

Figure 12. Asymptotic complexities of "reference" SAC-SMA, "reference" SIXPAR and "equivalent" SIXPAR.

HESSD
12, 3945–4004, 2015

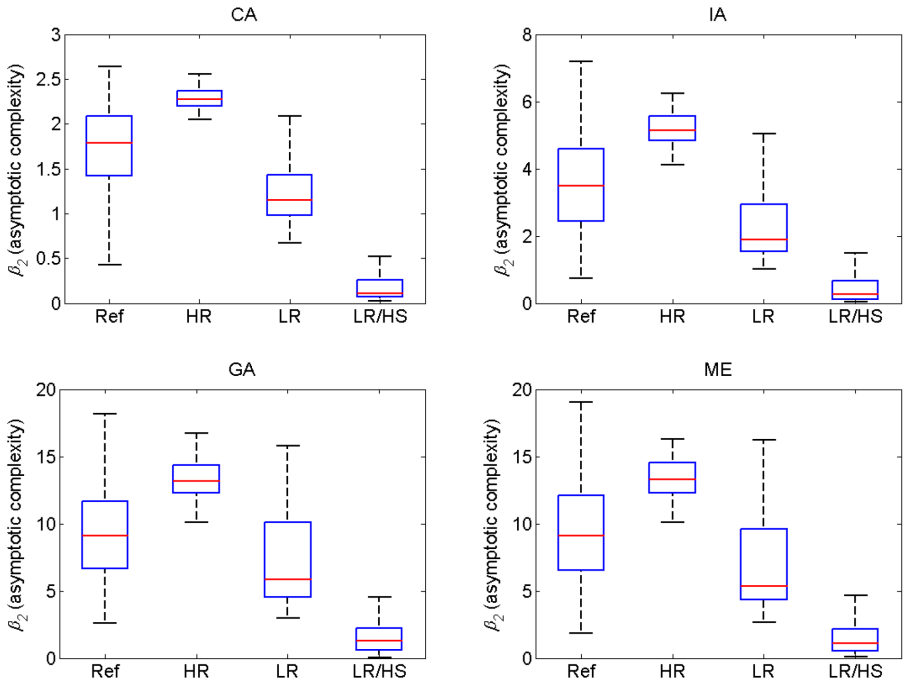

Figure 13. Asymptotic complexities of SIXPAR model structures for multiple basins across the counterminous US (CA, IA, GA, ME; see Table 3) and for various parameter ranges as described in Table 2 (Ref = "Reference", HR = "High recession", LR = "Low recession", LR/HS = "Low recession/High storage").

HESSD
12, 3945–4004, 2015