Hydrol. Earth Syst. Sci. Discuss., 11, 2555–2582, 2014
www.hydrol-earth-syst-sci-discuss.net/11/2555/2014/
doi:10.5194/hessd-11-2555-2014

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. H. G. 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: 3 February 2014 – Accepted: 21 February 2014 – Published: 3 March 2014

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

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

Abstract

This paper presents evidence that model prediction uncertainty does not necessarily rise with parameter dimensionality (the number of parameters). Here by prediction we mean future simulation of a variable of interest conditioned on certain future values of input variables. We utilize a relationship between prediction uncertainty, sample size and model complexity based on Vapnik–Chervonenkis (VC) generalization theory. It suggests that models with higher complexity tend to have higher prediction uncertainty for limited sample size. However, model complexity is not necessarily related to the number of parameters. Here by limited sample size we mean a sample size that is limited in representing the dynamics of the underlying processes. Based on VC theory, we demonstrate that model complexity crucially depends on the magnitude of model parameters. We do this by using two model structures, SAC-SMA and its simplification, SIXPAR, and 5 MOPEX basin data sets across the United States. We conclude that parsimonious model selection based on parameter dimensionality may lead to a less informed model choice.

1 Introduction

Less complex hydrological models are often preferred either due to low computational cost of simulations of such models (Keating et al., 2010; Young, 2003) or to ameliorate overfitting of models on observed data (Pande et al., 2009; Schoups et al., 2008). We here explore the concept of model complexity in context of the latter. Overfitting, which leads to highly uncertain model predictions on future unseen (input or forcing) data, is especially severe when observed data size on which the model is selected is limited in representing the underlying process dynamics. Here by prediction we mean model simulation of a variable of interest conditioned by certain future values of input (forcing) variables. Often models with low parameter dimensionality (i.e. less number of

HESSD
11, 2555–2582, 2014

parameters) are considered less complex and hence are associated with low prediction uncertainty. Whether this is always the case remains to be explored.

We do not suggest that model complexity is the sole determinant of prediction uncertainty. Its trade-off with model performance on an observed dataset determines the prediction uncertainty of the selected model. One can envisage a case wherein a model has significantly higher model complexity than another model yet the tradeoff between model performance and complexity may deem the more complex model with lower prediction uncertainty. Nonetheless, our analysis suggests that parameter dimensionality is a weak measure of model complexity and hence a weak measure of prediction uncertainty under similar finite sample prediction performances of competing models. Thus complexity controlled model selection ensures that the selected model predicts future values of a variable of interest with least uncertainty amongst the set of competing models (Pande et al., 2009, 2012). The need for complexity controlled model selection also arises in cases of ill-posed problems (Vapnik, 1982; Arkesteijn and Pande, 2013; Schoups et al., 2008) where complexity control acts as a "stabilizer". Thus the estimation of model complexity is paramount, especially when the sample size is small.

The Bayesian treatment of prediction uncertainty is through its specification of the likelihood function. It specifies the probability with which the observed values of a variable of interest are generated by a hydrological model. The marginal likelihood of a hydrological model structure (representing a class of hydrological models that represent same processes and hence have the same parameter dimensionality) is obtained when it is integrated over the prior distribution of model parameters, which then measures the prediction uncertainty of the model structure (Marshall et al., 2005).

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) (Schoups et al., 2008; Ye et al., 2008; Marshall et al., 2005). These measures therefore embody Bayesian interpretation of model prediction uncertainty. Ye et al. (2008) compared AIC, BIC and KIC measures. They showed that KIC is a finite sample version of BIC and depends on the Hessian

HESSD
11, 2555–2582, 2014

of the likelihood function at the optimum. Meanwhile BIC is a function of parameter dimensionality.

Bayesian inference chooses a model (and a model structure) from a set of competing models (from the set of corresponding model structures) such that the value of a Bayesian criterion is maximized (or prediction uncertainty in choosing a model structure is minimized). A choice of a model based on BIC trades off the log likelihood value with parameter dimensionality while KIC trades off the log likelihood value with the determinant of its Hessian at the parameter value that maximizes the likelihood. The relationship between prediction uncertainty and model structure complexity is in the tradeoff between the best performance that a structure can provide and its complexity (measured either by parameter dimensionality or by the determinant of the Hessian). The Hessian term in KIC acts as a measure of complexity because it measures the curvature of the likelihood function in the neighbourhood of an optimum, which in turn depends on the relationships embedded within the model structure. Further, under certain specification and regularity conditions, the Hessian is also equivalent to Fisher's Information matrix (Davidson and MacKinnon, 2004) that measures the variance of estimated parameters. Hence, the Information matrix is also implicitly used as a measure of complexity (Jakeman and Hornberger, 1993) in Bayesian inference.

Bayesian model selection criteria incorporate a measure of model complexity that is not necessarily the parameter dimensionality of the model. It also provides a relationship between model prediction uncertainty and complexity. However, the Bayesian measure of complexity is evaluated at the likelihood optimum, making it application specific (in particular with respect to the variable of prediction interest). For example, KIC incorporates the Hessian of the loglikelihood function as a measure of complexity that is evaluated at the parameter values that maximizes the likelihood function. This limits its use in a comparison of model complexities that are independent of applications.

An alternative, frequentist, approach to model selection has been proposed (Pande et al., 2009, 2012; Arkesteijn and Pande, 2013) that makes less restrictive assumptions. However, this then entails a model selection criterion that may be conservative

HESSD
11, 2555–2582, 2014

in nature. We use an extension of Vapnik–Chervonenkis generalization theory for hydrological model selection (see Pande et al., 2012; Arkesteijn and Pande, 2013) that allows a representation of prediction uncertainty as a tradeoff between model performance on a given sample of data and its complexity. It suggests that a model, out of a set of competing models with similar prediction performances on a given sample of data, has the largest prediction uncertainty if it has the largest model complexity. The measure of complexity is independent of variables of prediction interest though it is dependent on the values of forcing data. It is in this sense that the proposed measure of complexity is independent of the observed values of prediction interest whereas Bayesian measures of complexity are not.

We demonstrate that a hydrological model with more parameters may be less complex in context of its influence on prediction uncertainty. We show that model complexity depends on the magnitude of parameters as well. Similar conclusions have been drawn elsewhere for regression problems, where it has been shown that model complexity is a function of the magnitude of model parameters. For example, Bartlett (1998) and Vapnik and Chapelle (2000) find that the complexity of ANNs and SVMs are not only dependent on the dimensionality of the regressors but it also crucially depends 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).

In order to demonstrate that the same holds for hydrological model structures, we use two model structures, SAC-SMA and its simplification, SIXPAR. We estimate its complexity on daily rainfall and potential evaporation data sets of 5 MOPEX basins across the United States (Duan et al., 2006). We demonstrate that depending on the magnitude of its parameters, SAC-SMA can be less complex than SIXPAR even though the former has more parameters.

SAC-SMA (Sacramento Soil Moisture Accounting Model) is a complex model with two upper zone reservoirs and three lower zones reservoirs and a nonlinear percolation conceptualization. Meanwhile SIXPAR is a conceptual simplification of SAC-SMA with one upper and one lower reservoir and retains the percolation process concept. When

HESSD
11, 2555–2582, 2014

the parameter ranges corresponding to upper and lower reservoirs in the two model structures are made equivalent (i.e. when the total upper and lower zone capacities of the two structures are the same and the corresponding recession parameters have the same geometric means) SAC-SMA is indeed found to be more complex than SIXPAR.

This experiment controls the effect of parameter magnitude on model complexity and confirms what is intuitive that SACSMA is conceptually more complex than SIXPAR. However, when the parameter ranges are allowed to vary, we find a significant influence of parameter magnitudes on the complexity of the models. For example, the parameter ranges when storage capacities of SIXPAR are smaller and recession parameters larger than SAC-SMA, we find the former to be more complex than the latter.

The paper is organized as follows. Section 2 provides the theory, the models structures, datasets and the algorithms used. Section 3 presents and discusses the results. Finally Sect. 4 concludes.

2 Methodology

2.1 Prediction uncertainty

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. Similarly let forcings be represented by $\boldsymbol{x} = \{x_1, x_2, \ldots, x_N\}$ where x_1 may not be univariate, assumed here to be univariate for simplicity without any loss of generality. Further let a model be represented by a parameter set α that for given forcing $\boldsymbol{x}$ predicts $\boldsymbol{y} = \{y(t, \boldsymbol{x}; \alpha)\}_{t=1,\ldots,N}$. Let $\xi_N(\boldsymbol{y}^0, \boldsymbol{x}; \alpha)$ be defined as empirical risk that measures the performance of the model in terms of deviations of a model's predictions from the observed, for example by mean absolute error,

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

HESSD
11, 2555–2582, 2014

Finally we define the expectation of the empirical risk, $E[\xi_N(\boldsymbol{y}^0, \boldsymbol{x}; \alpha)]$ as

$$E[\xi_N(\boldsymbol{y}^0, \boldsymbol{x}; \alpha)] = \int \xi_N(\boldsymbol{y}^0, \boldsymbol{x}; \alpha) P(\boldsymbol{y}^0, x) \mathrm{d}\boldsymbol{x}\boldsymbol{y}$$

Thus, the expected risk is the expectation of empirical risk over the underlying unknown probability distribution that specifies the stochasticity of the underlying processes.

We then define prediction uncertainty as the following quantity for some $\gamma \geq 0$.

$$\Pr\left(|\xi_N(\mathbf{y}^0, \mathbf{x}; \alpha) - E[\xi_N(\mathbf{y}^0, \mathbf{x}; \alpha)]| > \gamma\right)$$

This quantity expresses the probability distribution of the deviation of model performance on a finite sample size from the case when the sample size is large. The quantity therefore expresses the uncertainty in performance, evaluated over finite sample, of a model.

Under certain conditions, it can be bounded by a function that is decreasing in N, γ, and increasing in a measure of complexity h (Cucker and Smale, 2002), i.e.

$$\Pr\left(|\xi_N(\mathbf{y}^0, \mathbf{x}; \alpha) - E[\xi_N(\mathbf{y}^0, \mathbf{x}; \alpha)]| > \gamma\right) \leq \Phi(N, \gamma, h) \tag{1}$$

Recently Arkesteijn and Pande (2013) have shown that inequality of type (1) for hydrological models can be formulated for any $\gamma \geq 0$ as:

$$\Pr\left(|\xi_N(\mathbf{y}^0, \mathbf{x}; \alpha) - E[\xi_N(\mathbf{y}^0, \mathbf{x}; \alpha)]| > \eta\gamma\right) \leq \frac{f(h,N)}{N^2\gamma^2} = \frac{F(h,N)}{\gamma^2} \tag{2}$$

where $\eta > 0$ is a given constant, $f(h,N) = \beta_2 N^2 + \beta_1 N + \beta_0$ and $h = \{\beta_2, \beta_1, \beta_0\}$ is a measure of model complexity. Further the function $f(h,N)$ is such that the following holds:

$$P_N = \Pr\left(\sum_{t=1}^{N} |y(t, \mathbf{x}; \alpha) - E[y(t, \mathbf{x}; \alpha)]| \geq N\gamma\right) \leq \frac{f(h,N)}{N^2\gamma^2} \tag{3}$$

HESSD
11, 2555–2582, 2014

2.2 SAC-SMA and SIXPAR model structures

The two model structures that are used are SAC-SMA and SIXPAR. SAC-SMA is a complex 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 evapotranspiration 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 flows. The SIXPAR model structure, which is a conceptual simplification of the SAC-SMA model with one upper and lower zone, excludes the evapotranspiration and the concept of tension water zones but retains the complex conceptualization of percolation. Additional details on the models can be found elsewhere (Burnash, 1995; Duan et al., 1992; Arkesteijn and Pande, 2013).

Table 1 provides the "reference" parameter ranges for SAC-SMA. Table 2 provides the various parameter ranges of SIXPAR, including so called "reference" ranges and "equivalent" ranges. The reason behind the terms is explained in the below. The complexity of models corresponding to 500 points sampled from these ranges are computed and compared.

We note that the complexity of SAC-SMA is computed for parameters sampled from the ranges presented in Table 1 and is annotated as "reference" since the computed complexity of SIXPAR for all the parameter ranges shown in Table 2 is compared to the reference complexity of SAC-SMA. The case with naive parameter ranges of SIXPAR is also called "reference" since the parameter ranges are prescribed without design. Another set of parameter ranges is called "equivalent" because (i) the upper bounds on the reservoir capacities of the two layers is equal to the sum of upper bounds on the reservoir capacities of the corresponding layers for SAC-SMA and (ii) the corresponding lower and upper bounds of the recession parameter ranges are the geometric means of corresponding lower and upper bounds of the SAC-SMA recession parameters. The

HESSD
11, 2555–2582, 2014

other three parameter range types are (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 complexities of SIXPAR model structures for “reference”, “equivalent”, and (i)–(iii) ranges are computed on hydrological data sets of MOPEX basins and compared with the SAC-SMA model structure complexity 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 evapotranspiration and precipitation data using Algorithm 1. Both these algorithms are provided in the next section and are based on the theory presented in Arkesteijn and Pande (2013). The data sets used are also described in Sect. 2.3.

2.3 Data and algorithms

The computation of model complexity for any given parameter set requires input forcing data set. We note in Sect. 2.1 that the expectation operator is computed on this data set as described in Algorithm 1 below. The algorithm is obtained from Arkesteijn and Pande (2013). The input forcing basin datasets for the computation of model complexity is obtained from the MOPEX data sets (Duan et al., 2006; Brooks 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

HESSD
11, 2555–2582, 2014

determining model complexity. Table 3 provides information on the basins used in this study and Fig. 1 displays them.

The computation of the expectation operator in Eq. (3) and hence the computation of model complexity depends the data of the input used. The computation of the expectation operator is based on 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 United States (Fig. 1).

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 evapotranspiration 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 another months until all 12 months of a year.

HESSD
11, 2555–2582, 2014

5. Permute the months (if correlatiqon 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. 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.

Using the weather resampler, $M = 2000$ sequences of $\overline{N} = 5000$ datapoints for daily precipitation and potential potential evapotranspiration 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 the evapotranspiration processes, the precipitation data used for SIXPAR is assumed to be equal to a maximum of the precipitation minus the evapotransipration and zero.

HESSD
11, 2555–2582, 2014

Once multiple realizations of input forcing data have been generated (resampled), Algorithm 2 computes the complexity of models for a sampled parameter set. The algorithm is obtained from Arkesteijn and Pande (2013). This is based on the theory presented in Sect. 2.1. In total 500 parameter sets are sampled from each range presented in Tables 1 and 2.

Algorithm 2:

1. For each parameter set of a model structure set up, estimate the Left Hand Side (LHS) probability in inequality Eq. (3), for a given value of N and γ using M samples of data set of size N, obtained form Algorithm 1.
2. Estimate the maximum $\tilde{f}(N)$ of $P_N N^2 \gamma^2$ with respect to γ for each N. Let the maximizing γ be $\gamma_{\max}^N$.
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 $\{\tilde{f}(N), N = 200 : 50 : \overline{N}\}$. 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.

3 Results and discussions

The Algorithm 2 based on input forcing realizations resampled by Algorithm 1 provides complexity computations for each of the two structures for the parameter sets sampled from ranges defined in Tables 1 and 2. The parameters are sampled using Latin Hypercube Sampling. Figure 2 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".

HESSD
11, 2555–2582, 2014

The different curves correspond to different month permutations (step 5) of the re-sampled 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, indicating the convergence of P_N 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 N (when $F(h,N)$ asymptotes and becomes insensitive to N) reveals the measure of complexity. Since, $F(h,N)$ increases with complexity (Arkesteijn and Pande, 2013), the asymptotic value of $F(h,N)$ can be used to compare the complexity of different model structure set-ups.

However, due to the approximation of $f(h,N)$ by a quadratic function in N in Eq. (2), we can directly estimate this complexity. In particular we estimate the asymptotic complexity, which is the asymptotic value of $F(h,N)$. It is the coefficient β_2 of the N^2 term of $f(h,N)$ (see Eq. 2). Figure 3 demonstrates that the asymptotic complexity for parameter ranges of SAC-SMA sampled from its "reference" ranges (Table 1) appears to be less complex than the asymptotic complexity for SIXPAR when sampled from its "reference" ranges (Table 2). This may appear counterintuitive since SIXPAR model structure is a conceptual simplification of SAC-SMA. However, the evidence from other regression models emphasizes the contribution of the magnitude of parameters in addition to that of parameter dimensionality to model complexity (Marquardt and Snee, 1975; Bartlett, 1998; Vapnik and Chapelle, 2000). This is further explored for SIXPAR model structure in the following analysis.

Figure 4 further studies the effect of sampling SIXPAR parameters from various ranges in Table 2 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.

HESSD
11, 2555–2582, 2014

This demonstrates that the magnitude of parameters appear to affect the complexity of a model. Figure 5 shows a comparative variation of computed complexity with sample size N for SAC-SMA and SIXPAR. Figure 5a shows the comparison between the two models when parameters are sampled from "reference" parameter ranges and Fig. 5b compares the two model structures when the parameters are sampled from "equivalent" parameter ranges. The y axis, P_N, is an increasing function of model complexities (from Eq. 3).

Both the figures demonstrate that the differences in complexities of the two model structures are more evident for small sample sizes. Figure 5a suggests that SIXPAR model structure is more complex, due to higher recession parameters and lower reservoir storage capacities for all sample sizes N. Meanwhile Fig. 5b shows SAC-SMA is more complex for all sample sizes N when the parameter ranges of SIXPAR are "equivalent" to SAC-SMA. Thus the comparison suggests that parameter magnitude plays a dominant role on model complexity and that parameter dimensionality is only a weak measure of complexity. Figure 6 presents the case again for the asymptotic complexities of "reference" SAC-SMA, "reference" SIXPAR and "equivalent" SIXPAR.

Figure 7 plots the asymptotic complexities for the same ranges of SIXPAR model structure for CA, IA, GA and ME MOPEX basins (Table 3). We observe a similar pattern in asymptotic complexities with parameter ranges and hence with parameter magnitudes. For a given specification of parameter range, the magnitude of asymptotic complexities is different for different basins. This may indicate the influence of basin specific correlation structure in input forcings on the estimation of model structure complexities. It appears that the correlation structure in the input forcings of GA and ME basins (in comparison with CA and IA) is similar in effect (on model complexity) to low recession values. This is indicative of longer term memory in rainfall and/or evaporation of GA and ME basins than CA and IA basins. A detailed analysis of such effect on computing model complexity and of its own interpretation of complexity is left for future research.

HESSD
11, 2555–2582, 2014

The evidence from Figs. 6 and 7 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. In Sect. 2.1 we defined prediction uncertainty as the probability with which empirical risk deviates from its expectation (Left Hand Side of inequality Eq. 2). The RHS of inequality Eq. (2) that bounds this probability of error is a function of complexity discussed previously. It then follows from the inequality 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 1).

An important implication for complexity controlled model selection is 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 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. 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 model with low prediction uncertainty in modelling slower basins.

4 Conclusions

Model complexity is an important criterion in model selection, since prediction uncertainty (here defined as the probability with which empirical risk deviates from its expectation) is a function of model complexity (see inequality Eq. 2). The inequality suggests that a model (out of a set of competing models with similar performance in predicting a variable of interest on limited sample) predicts the values of an output variable

HESSD
11, 2555–2582, 2014

of interest with higher uncertainty if it has higher model complexity. In this paper the complexity of two model structures, SAC-SMA and SIXPAR, was computed using two different algorithms. Algorithm 1, was created to resample multiple realizations of input forcing data sets, and Algorithm 2 was created to estimate complexity based on inequality Eq. (3) using resampled input dataset generated by Algorithm 1.

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 conterminous United States. The model complexity for SIXPAR were estimated for various parameter ranges. The range specifications included "equivalent" 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.

For "reference" ranges, ranges obtained from literature, SIXPAR was found to be more complex than SAC-SMA model structure. 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 the choice of parameter ranges influences model complexity as well and that other measures that solely use parameter dimensionality as a measure, may be weak determinants of prediction uncertainty.

HESSD
11, 2555–2582, 2014

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. 2557, 2558, 2559, 2561, 2562, 2563, 2564, 2566, 2567

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. 2559, 2567

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. 2563, 2575, 2576

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. 2562

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

Cucker, F. and Smale, S.: On the mathematical foundations of learning, B. Am. Math. Soc., 39, 1–49, 2002. 2561

Davidson, R. and MacKinnon, J. G.: Econometric Theory and Methods, Oxford University Press, New York, 750, 2004. 2558

Duan, Q., Sorooshian, S., and Gupta, V.: Effective and efficient global optimization for conceptual rainfall–runoff models, Water Resour. Res., 28, 1015–1031, 1992. 2562

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.: The 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. 2559, 2563, 2575, 2576

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.

Jakeman, A. J. and Hornberger, G. M.: How much complexity is warranted in a rainfall–runoff model?, Water Resour. Res., 29, 2637–2649, 1993. 2558

Kass, R. E. and Raftery, A. E.: Bayes factors, J. Am. Stat. Assoc., 90, 773–795, 1995.

HESSD
11, 2555–2582, 2014

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. 2556
Kundzewicz, C. W. and Robson, A. J.: Change detection in hydrological records a review of the methodology, Hydrolog. Sci. J., 49, 7–19, 2004. 2564
Marquardt, D. W. and Snee, R. D.: Ridge regression in practise, Am. Stat., 29, 3–20, 1975. 2559, 2567
Marshall, L., Nott, D., and Sharma, A.: Hydrological model selection: a Bayesian alternative, Water Resour. Res., 41, W10422, doi:10.1029/2004WR003719, 2005. 2557
Pande, S., McKee, M., and Bastidas, L. A.: Complexity-based robust hydrologic prediction, Water Resour. Res., 45, W10406, doi:10.1029/2008WR007524, 2009. 2556, 2557, 2558
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. 2557, 2558, 2559
Politis, D. and Romano, J.: The stationary bootstrap, J. Am. Stat. Assoc., 89, 1303–1313, 1994. 2564
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. 2556, 2557
Slate, E. H.: Parameterizations for natural Exponential families with quadratic functions, J. Am. Stat. Assoc., 89, 1471–1482, 1994.
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. 2557
Vapnik, V. and Chapelle, O.: Bounds on error expectation for support vector machines, Neural Comput., 12, 2013–2036, 2000. 2559, 2567
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. 2557
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. 2556
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
11, 2555–2582, 2014

Table 1. SAC-SMA model structure parameter ranges used in the study.

Parameters	"Reference"	Parameter	"Reference"
UZTWM [mm]	1–150	UZWFM [mm]	1–150
UZK [day^{-1}]	0.1–0.5	PCTIM [–]	0–0.1
ADIMP [–]	0–0.4	RIVA [–]	0
ZPERC [–]	1–250	REXP [–]	1–5
LZTWM [mm]	1–1000	LZFSM [mm]	1–1000
LZFPM [mm]	1–1000	LZSK [day^{-1}]	0.01–0.25
LZPK [day^{-1}]	0.0001–0.025	PFREE [–]	0.0–0.6
RSERV [–]	0.3	SIDE [–]	0.0

HESSD
11, 2555–2582, 2014

Table 2. SIXPAR model structure parameter ranges used in the study.

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

HESSD
11, 2555–2582, 2014

Table 3. Basins used in this study. MAP = Mean Annual Precipitation, MAPET = Mean Annual PET. MAP and MAPET are calculated using data from the period 1948–1970. Data obtained from Duan et al. (2006) and Brooks et al. (2011).

Site Id	Area [km^2]	MAP [$mm\,yr^{-1}$]	MAPET [$mm\,yr^{-1}$]	Code
03451500	945.00	1491	820	NC
11138500	281.00	380	1334	CA
05479000	1308.00	711	977	IA
02228000	2790.00	1215	1132	GA
01060000	141.00	1100	N/A	ME

HESSD
11, 2555–2582, 2014

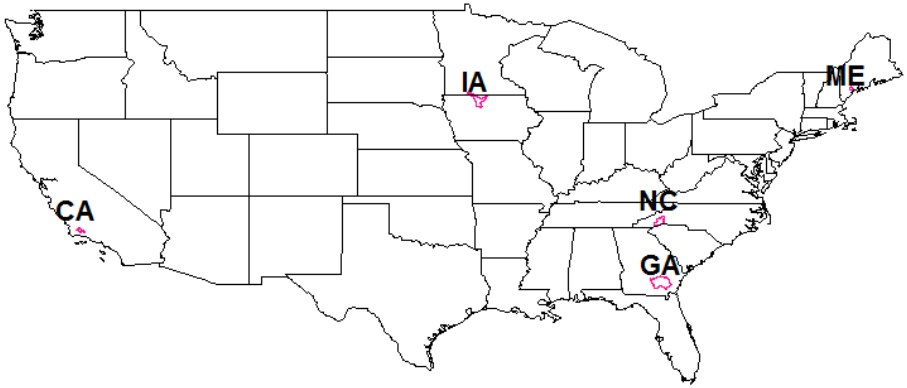

Fig. 1. A selection of basins across the United States spanning different hydro-climatic regions. Data obtained from Duan et al. (2006) and Brooks et al. (2011).

HESSD
11, 2555–2582, 2014

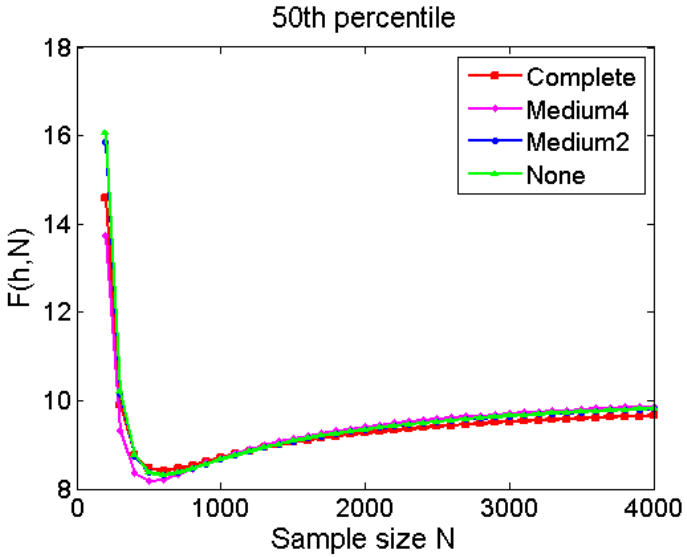

Fig. 2. Complexity curves for 50th percentile values of different month permutations.

HESSD
11, 2555–2582, 2014

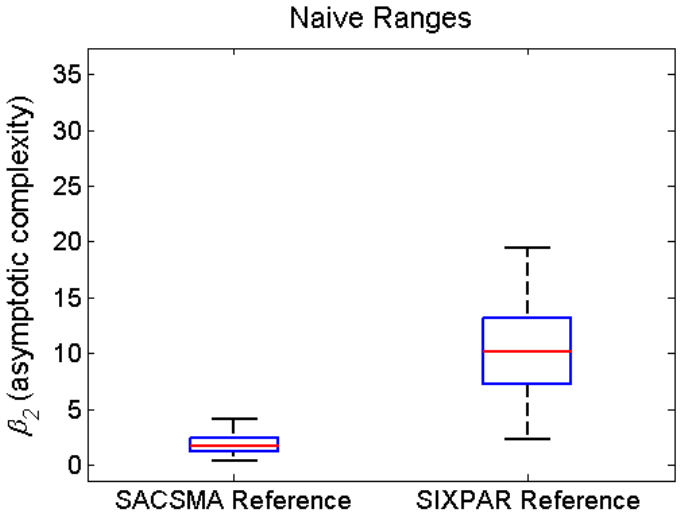

Fig. 3. Asymptotic complexity using reference ranges for SAC-SMA and SIXPAR model structure.

HESSD
11, 2555–2582, 2014

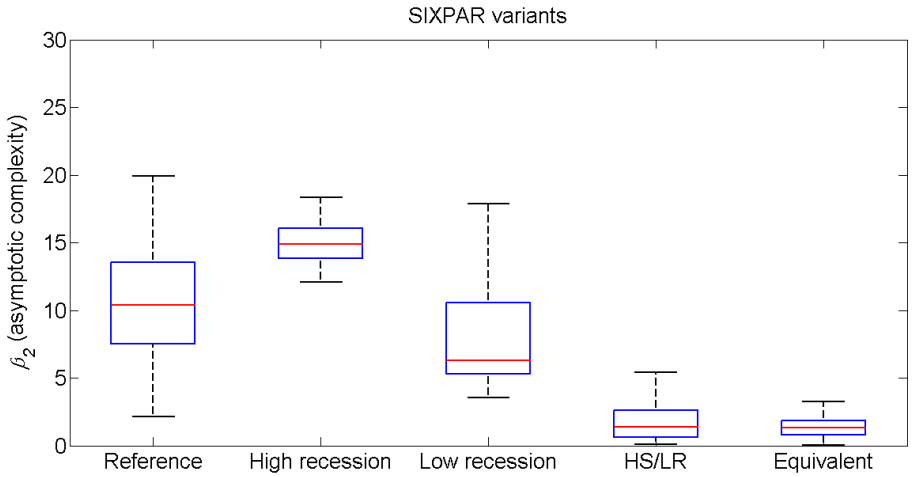

Fig. 4. Asymptotic complexity using different parameters ranges for SIXPAR model structure.

HESSD
11, 2555–2582, 2014

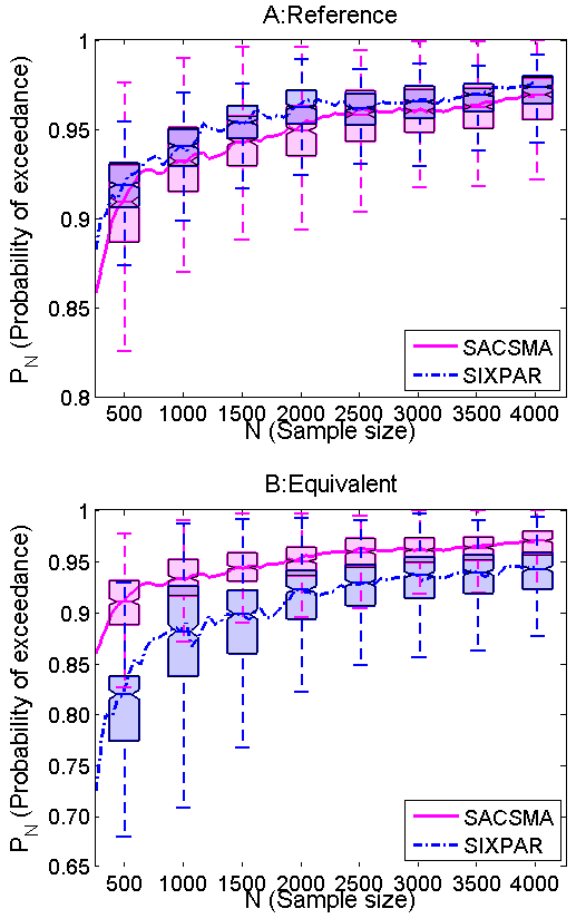

Fig. 5. Variation of computed complexity with sample size *N* for SAC-SMA and SIXPAR. **(A)** Reference parameter ranges and **(B)** equivalent parameter ranges.

HESSD
11, 2555–2582, 2014

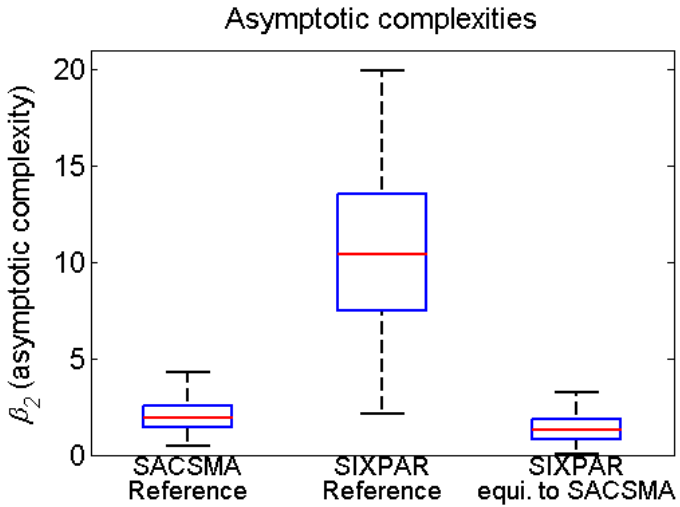

Fig. 6. Asymptotic complexities of "reference" SAC-SMA, "reference" SIXPAR and "equivalent" SIXPAR.

HESSD
11, 2555–2582, 2014

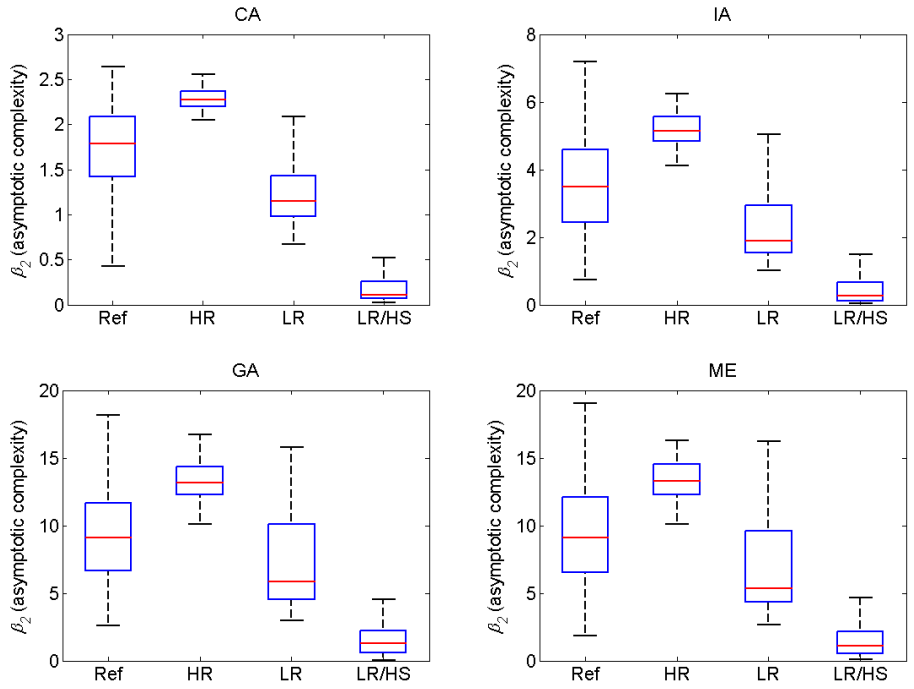

Fig. 7. Asymptotic complexities of SIXPAR model structures for multiple basins across the counterminous United States (CA, IA, GA, ME; see Table 3) and for various parameter ranges as described in Table 2 (Ref = "Reference", HR = "High recession", LR = "Low recession", LR/HS = "Low recession/High storage").

HESSD
11, 2555–2582, 2014