Parameter-dependent convergence bounds and complexity measure for a class of conceptual hydrological models

Saket Pande, Luis A. Bastidas, Sandjai Bhulai and Mac McKee

ABSTRACT

We provide analytical bounds on convergence rates for a class of hydrologic models and consequently derive a complexity measure based on the Vapnik–Chervonenkis (VC) generalization theory. The class of hydrologic models is a spatially explicit interconnected set of linear reservoirs with the aim of representing globally nonlinear hydrologic behavior by locally linear models. Here, by convergence rate, we mean convergence of the empirical risk to the expected risk. The derived measure of complexity measures a model's propensity to overfit data. We explore how data finiteness can affect model selection for this class of hydrologic model and provide theoretical results on how model performance on a finite sample converges to its expected performance as data size approaches infinity. These bounds can then be used for model selection, as the bounds provide a tradeoff between model complexity and model performance on finite data. The convergence bounds for the considered hydrologic models depend on the magnitude of their parameters, which are the recession parameters of constituting linear reservoirs. Further, the complexity of hydrologic models not only varies with the magnitude of their parameters but also depends on the network structure of the models (in terms of the spatial heterogeneity of parameters and the nature of hydrologic connectivity).

Key words | complexity, convergence bounds, hydrological model identification, inverse problems, probabilistic and statistical methods, stochastic processes

Saket Pande (corresponding author)
Department of Water Management,
Delft University of Technology,
Delft,
The Netherlands
E-mail: *s.pande@tudelft.nl*

Luis A. Bastidas
Mac McKee
Department of Civil and Environmental
    Engineering and Utah Water Research
    Laboratory,
Utah State University,
Logan, Utah,
USA

Sandjai Bhulai
Department of Mathematics,
VU University,
Amsterdam,
The Netherlands

INTRODUCTION

In this paper we present simple conceptual water balance models and then derive a complexity measure of such hydrologic models and assess the complexity of hydrologic responses; estimate a bound on its convergence rates; and discuss its applicability and extensions with examples.

Hereinafter, by convergence rate we mean convergence of the empirical risk to the expected risk when calibrating hydrologic models (using the definitions of Vapnik & Chervonenkis 1971). The empirical risk is a measure of the deviation of the modeled output from the observed output for a given dataset (a measure of prediction error on a given sample, such as mean absolute error) and the expected risk is the expectation of the empirical risk. These two

quantities are further defined in the section on 'Parameter-dependent complexity measure and convergence bound for a simple one-reservoir model'.

The model presented, although simple, is widely used as a component of many hydrologic models, as it conceptualizes a storage–discharge relationship and consequently the evolution of soil moisture over space and time in a similar manner (Burnash 1995). The motivation behind the choice of this simple conceptualization is to elucidate the link between parameters driving storage–discharge relationships, model complexity, and prediction performance of such models.

Apart from its contribution to statistical learning theory applications in hydrologic sciences (see Schoups *et al.*

(2008) and Pande *et al.* (2009) for some initial work in this direction), this paper estimates a complexity measure for models with memory and its representation in terms of model parameters (that also define the memory). Also, in a manner distinct to others (Bartlett & Kulkarni 1998 and references therein; Meir 2000), the convergence bounds presented here are in terms of the model parameters and are tight due to the parametric specification of the model space. A key result for hydrologic applications is that complexity, for hydrologic models, does not only depend on the magnitude (in addition to the number) of parameters but also on the structure of the models. We formally establish the relationship between model complexity and model parameters and structure (such as hydrologic connectivity, Wang & Waymire 1991). This relationship provides insights into the complexity of hydrologic response. We introduce a quantitative definition of the complexity of the rainfall–runoff process and describe its implications for decentralized systems, such as decentralized agriculture production systems (which function without an organized center or authority), which depend on hydrologic responses.

This paper thus contributes to hydrological model uncertainty assessment and provides a theoretical basis for the application of complexity regularized parameter estimation of hydrological models. Through the study of convergence bounds we mathematically formalize finite sample performance of hydrological models in the context of the Vapnik–Chervonenkis (VC) generalization theory. Our results formally reveal how model complexity trades off with available information and how hydrological model complexity becomes irrelevant as sample size goes to infinity. We also quantify complexity of a class of hydrological models. Although the theory that is presented is applicable for a simple class of interconnected linear reservoir models, we consider this step as a first in the direction of quantifying complexity of state-of-the-art hydrologic models. The analytical bounds (and its derivation) allow geometric interpretation of the notion of complexity and how it affects model performance. This situation also allows insights into quantification of complexity for other hydrological models.

Yet another interesting finding is that model complexity depends on the structure of hydrological model, which for a spatially explicit hydrological model includes network topology and spatial heterogeneity as well as the magnitude of parameter fields. If a model is a close approximation of underlying processes, the complexity of the underlying processes can be said to be driven by its biogeophysical properties by implication. Further the proofs underlying the lemmas and theorems suggest a close connection between complexity measure and model output space. Given that model output space embodies the nature of model response to input forcing and if the model is a close approximation of reality, our interpretation broadly defines complexity of underlying processes as how it responds to exogenous forcing (governed by its biogeophysical properties).

BACKGROUND

The concepts underlying many hydrological models originate from applying the Boussinesq flow equation (BE), which is derived from the continuity equation along with Darcy's law (Lacey *et al.* 1982). Several approximations of the BE have been used to model ground water flow under different boundary and initial conditions (Brutsaert & Ibrahim 1966). These results have motivated its use to model subsurface flows (Beven 1981; Paniconi *et al.* 2003), bank storage (Govindaraju & Koelliker 1994), and surface water body–aquifer interaction (Pulido-Velazquez *et al.* 2007). The solution to the BE (outflow), under certain conditions, can be represented by a linear reservoir (Brutsaert & Nieber 1977) or by an infinite collection of linear reservoirs connected in series (Pulido-Velazquez *et al.* 2007). Nonetheless, if solutions of the BE are to be used, extensive datasets are needed to describe its coefficients (if they are not calibrated).

As reconciliation, hydrologic responses are conceptualized (e.g. Gupta & Sorooshian 1983; Savenije 2009) by certain classes of functions; such as the collection of interconnected linear reservoirs used here. The nonlinearity of hydrologic response due to within-catchment heterogeneity has been explored by a combination of linear reservoirs connected in parallel (e.g. Harman *et al.* 2009), as an alternative to hydraulic theory. A truncated series of linear reservoirs (connected in series) as an approximation to the solution of a linearized BE has also been employed to

simulate surface water body-aquifer interactions (e.g. Pulido-Velazquez *et al.* 2007). By extension, hillslope responses can be approximated by a linearized BE while channel flows are approximated by linear reservoir models. Thus catchment-scale response to rainfall can be conceptualized by interconnected linear reservoir models with reservoir network topology ascribed by channel network topology and geophysical properties (that affects the spatial distribution of hill slopes and its approximation). Its parameters are then 'effective' rather than physically based, and need to be calibrated (Savenije 2009). A class of models of interconnected linear reservoir models is therefore not unrealistic to describe more complex physically based models and that the study of complexity of a linear reservoir model is one of the fundamental steps to study complexity of state-of-the-art hydrological models.

Several methodologies exist that estimate parameters (inverse problem), providing either unique parameter estimates (when using gradient-based algorithms, or global search algorithms such as SCE-UA (Duan *et al.* 1992)) or its distribution (such as MOSCEM-UA (Vrugt *et al.* 2003a)). We note that ill-posed problems lead to unreliable parameter estimates while non-convex optimization (minimization) problems (with non-convex hydrological models as is generally the case) lead to non-unique parameter estimates. However, such observations are theoretical. In practice, parameter estimation algorithms are designed either to provide a single parameter set (such as gradient-based algorithms, SCE-UA (Duan *et al.* 1992)) or a distribution of parameter sets (such as MOSCEM-UA (Vrugt *et al.* 1993)) as a solution irrespective of the nature of the underlying optimization problem. For example, global search algorithms such as SCE-UA are less efficient than gradient-based optimizers when the problem is convex while neither of these two algorithms may be useful when the problem is ill posed as the resulting solutions would be highly unreliable (due to complex model identification problems). While parameter solutions to non-convex optimization problems have been intensely studied resulting in global search algorithms, the study of ill-posed problems is still in its infancy in hydrological modeling. Problems are ill posed as a result of a mismatch between model complexity and available data (Vapnik 2002) and this is the topic of this paper.

Broadly these strategies aim at selecting a model or a subset of models (e.g. Beven & Binley 1992; Gupta *et al.* 1998) from a model space. However, few have explored the effect of data finiteness on model selection (e.g. Ye *et al.* 2005, 2008; Pande *et al.* 2009). Most inference methods are conditional on a data set that is used via different sampling algorithms to arrive at a posterior parameter distribution (e.g. Vrugt *et al.* 2003a, b; van Griensven & Meixner 2007).

By data finiteness we imply any data size smaller than infinite and we employ it to describe the finite sample performance of a model (e.g. in terms of mean absolute deviation of model prediction from the observed). The law of large numbers dictates convergence of performance of any model on any finite data to its performance on infinite data sets (generated from the same underlying but unknown ergodic process). In this paper we provide a stronger law of large numbers in the form of a bound on convergence rates (for example, the result of Lemma 2) that describes 'how' finite sample model performance converge to infinite sample model performance as a function of sample size and model complexity. In doing so we also describe how model performance improves with increasing (but finite) data size.

In this paper, as its central motivation, we explore how data finiteness affects model selection for a class of hydrologic models defined by interconnected linear reservoirs. This class of models attempts to conceptualize within-catchment heterogeneities, where each linear reservoir represents a subbasin. We provide theoretical results on how a model performance on a finite sample converges to its expected performance as the data size approaches infinity. These bounds can then be used for model selection, akin to a regularized solution to an inverse problem (Elayyan & Isakov 1997).

Our convergence results are based on the Vapnik–Chervonenkis generalization theory (Vapnik 1982). Bounds of convergence, for various classes of functions, have been extensively studied (e.g. Blumer *et al.* 1989; Bartlett 1998; Bartlett & Kulkarni 1998; Lugosi & Nobel 1999; Pontil 2003; Meir 2000). Innovative statistical tools such as Support Vector Machines (SVMs) are based on these bounds (Vapnik 2002; Han *et al.* 2007), which essentially describe how empirical risk, that is a measure of how a model's

performance on finite sample, converges to its expectation. These concepts are also closely linked to ε-optimal model selection problems wherein Probably Approximately Optimal (PAO) or Probably Approximately Correct (PAC) (Valiant 1984) models are selected based on convergence bounds (Haussler *et al.* 1991; Kearns & Schapire 1994; Fong 1995; Alon *et al.* 1997).

Convergence bounds that explicitly account for the tradeoff between a measure of model complexity (e.g. via covering number; Cucker & Smale 2001) and performance on a finite data size are of particular interest to the hydrologic community. If a complexity measure of hydrologic models can be ascribed to their structure, which in turn may be ascribed (via conceptualizations) to various sources of within-basin heterogeneities, then data needs for process conceptualization can be ascribed to the complexity of the underlying hydrological processes. Understanding such a tradeoff constitutes the key to robust model selection in conceptual hydrological modeling.

It is important to mention that bounds for hydrological models need to be estimated afresh because available convergence bounds generally rely on the assumption that the residuals are independently and identically distributed (i.i.d.) (Vapnik 2002). Hydrological model responses have temporal memory, thus disobey the i.i.d. assumption. The particular class of hydrologic models, considered in this paper, allows us to obtain tighter convergence bounds than those currently available for a class of functions with memory (see for example Bartlett & Kulkarni 1998; Meir 2000). It also provides an opportunity to study these bounds in terms of parameters and structure (hydrologic connectivity and parameter heterogeneity) of spatially explicit hydrological models.

The paper is organized as follows: the sections entitled 'A simple hydrological model (one-reservoir model)' and 'Parameter-dependent complexity measure and convergence bound for a simple one-reservoir model' introduce an estimation of complexity and convergence bounds for a simple linear reservoir model; then the section covering the journey from a single reservoir model to a model of interconnected reservoirs does the same for a spatially explicit model of interconnected linear reservoir models; the next section provides applications and extensions of the approach, such as implications (and applications) for decentralized systems, such as agriculture production systems, that depend on hydrologic responses; finally conclusions are presented.

A SIMPLE HYDROLOGIC MODEL (ONE-RESERVOIR MODEL)

The hydrologic model used here, defines a linear storage that transforms effective precipitation (input) to discharge or outflow (output) as a linear function of storage. Effective precipitation, that is, precipitation minus evaporation and transpiration, updates the amount of soil moisture over time. This moisture availability is represented by storage, which in turn is released as streamflow. For additional details on hydrologic models, readers are referred to Burnash (1995). The model obeys the following conservation of mass equation:

$$\frac{\mathrm{d}S(t)}{\mathrm{d}t} = -Q(t) + u(t) \tag{1}$$

where $S(t)$ is the state variable (soil moisture or storage) at the end of time interval t, $Q(t)$ is the outflow or discharge, and $u(t)$ is the effective precipitation.

Here we make some additional assumptions:

Assumption A: (1) The outflow $Q(t)$ is linearly related to the soil moisture $S(t)$ as $Q(t) = kS(t)$, where $k \in (0, 1)$ is a runoff or recession coefficient (a parameter).

Assumption A: (2) The storage capacity is never reached, i.e., $S(t) < S_{\max}$, where $S_{\max}$ is the storage capacity of the reservoir.

Assumption A: (3) The effective precipitation $u(t) < C_{\max}$ is constant over discrete time intervals Δt with $S(t)$ observed at the end of such time intervals. $C_{\max}$ defines an upper bound on effective precipitation. The mathematical expectation of precipitation is small compared with $C_{\max}$, i.e., $E(u(t)) \ll C_{\max}$. Finally, $u(t)$ is independently and identically distributed over time.

Assumption 1 describes the storage–discharge relationship of a linear reservoir model. A linear reservoir model is the building block of the class of models of interconnected linear reservoir models that we study in this paper. Assumption 2 conceptualizes dryland areas where

water stored in the subsurface (in both the saturated and unsaturated zones) rarely exceeds the subsurface capacity to store water. We acknowledge that this assumption is strict and limiting. While this assumption can be relaxed, we delay it for brevity reasons. Assumption 3 suggests that input forcing or effective precipitation (actual precipitation minus actual evaporation for a single reservoir model) is bounded from above. It also describes that storage at each time step is the value at the end of that time step. We also assume that effective precipitation on an average is small compared with maximum possible precipitation. Finally we assume that effective precipitation at point in time is not correlated with effective precipitation at previous time steps. Low autocorrelation is generally observed for time series at daily scale (Guenni & Bardossy 2002), thus the assumption may not be restrictive when the temporal scale of the problem is daily or finer.

We choose $\Delta t = 1$ and therefore fix our model resolution at the scale over which $u(t)$ is uniform. For sufficiently large t and under Assumption A (1–3), the solution for $S(\mathbf{u}; t)$ is (the solution to a linear ordinary differential equation of order 1 with constant coefficients) where $\mathbf{u} = \{u(t)\}_t$ is a vector of input forcings:

$$S(\mathbf{u}; t) = \sum_{\tau=1}^{t} u(\tau) \int_{\tau-\Delta\tau}^{\tau} e^{-k(t-\tau')} d\tau' + e^{-kt} S(0)$$

$$= \frac{d(k)}{k} \sum_{\tau=1}^{t} u(\tau) e^{-k(t-\tau)}$$

Here, $d(k) = 1 - e^{-k\Delta\tau} = 1 - e^{-k}$.

From Assumption A (1) and calculating the total outflow, $Q(\mathbf{u}; t)$, during time interval $[t - \Delta t, t]$, we have:

$$Q(\mathbf{u}; t) = kS(\mathbf{u}; t) = d(k) \sum_{\tau=1}^{t} u(\tau) e^{-k(t-\tau)} \tag{2}$$

From here on we ignore $\mathbf{u}$ as an argument of Q or related quantities when the role of $\mathbf{u}$ need not be emphasized. Our model Equation (2), for the total outflow, defines a convolution of past input series while the convolution depends on parameter k. If we choose a coefficient m (indicative of the process memory) that defines an

ε-approximation of the outflow: $Q_m(t)$ for $\varepsilon > 0$, we have

$$|Q(t) - Q_m(t)| = d(k) \sum_{\tau=m+1}^{\infty} u(\tau) e^{-k(t-\tau)} \le \varepsilon \tag{2b}$$

Note that the following inequality holds from (2b)

$$\big\| |Q_m - EQ_m| - |Q - EQ| \big\| \le 2\varepsilon$$

As the following hold by triangle inequality:

$$|Q_m - EQ_m| \le |Q_m| + |EQ_m|,$$
$$|Q - EQ| \le |Q| + |EQ|$$

and

$$|Q_m| - |Q| \le |Q_m - Q| \le \varepsilon$$
$$|EQ_m| - |EQ| \le |EQ_m - EQ| \le \varepsilon.$$

For a sufficiently large t and $u(\tau) \le C_{\max}$, i.e., Assumption A (3), m obeys the inequality:

$$e^{-k(m+1)} \le \frac{\varepsilon}{C_{\max}} \tag{3}$$

If $0 < d < t - m$, then the following holds:

$$
\begin{bmatrix} Q_m(t) \\ Q_m(t-1) \\ \vdots \\ Q_m(t-d) \end{bmatrix} = d(k) \begin{bmatrix} 1 & e^{-k} & \cdots & 0 & 0 \\ 0 & 1 & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & e^{-k(m-1)} & e^{-km} \end{bmatrix} \begin{bmatrix} u(t) \\ u(t-1) \\ \vdots \\ u(t-d-m) \end{bmatrix}
$$
$$\tag{4a}$$

If $\mathbf{q}$, $\mathbf{V}$, and $\mathbf{u}$ are $(d+1) \times 1$, $(d+1) \times (d+m+1)$, and $(d+m+1) \times 1$ matrices respectively, then Equation (4a) can be represented by,

$$\mathbf{q} = \mathbf{V}\mathbf{u} \tag{4b}$$

Thus, any $(d+1)$ dimensional (outflow) response of a model to a $(d+m+1)$ dimensional input of effective precipitation always lies in the span defined by the columns

of the matrix $\mathbf{V}$. The following Lemma 1 characterizes one of its properties used later.

Consider an example with $d = 2$ and $m = 1$ (consequently an appropriate choice of k such that Equation (3) is satisfied). Then we obtain from Equation (4a)

$$\begin{bmatrix} Q_m(t) \\ Q_m(t-1) \end{bmatrix} = d(k) \begin{bmatrix} 1 & e^{-k} & 0 \\ 0 & 1 & e^{-k} \end{bmatrix} \begin{bmatrix} u(t) \\ u(t-1) \\ u(t-2) \end{bmatrix}$$

The rows on the left-hand side of the above equation represent one of the two prediction dimensions (corresponding to the y_1 and y_2 dimensions in Figure 1 below, representing a model output vector $\{\tilde{y}_1, \tilde{y}_2\}$). Each column of the corresponding $\mathbf{V}$ matrix represents a basis vector. These columns then define the span of the corresponding hydrologic model (parameterized by k). The span then maps the input forcings $\mathbf{u}$ to what we call model output space or Q-space. Thus by model output space (or Q-space as referred to in Figure 1), we mean how arbitrary input forcing, i.e., a vector with elements $u(t)$, $u(t-1)$ and $u(t-2)$, are transformed by the model span (note the difference between model output space and model span). The shape and size of this span then completely characterizes a model's response for arbitrary input forcings with no persistence or autocorrelation. Also a comparison between any two models can then be performed in terms of its output space irrespective of the nature of the input forcing
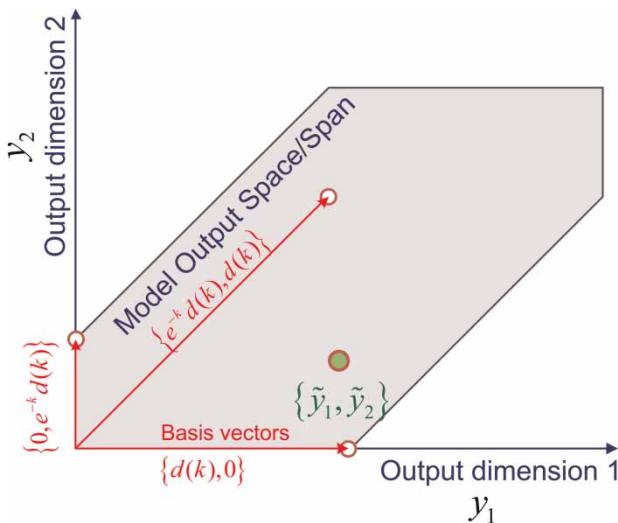
(though under the assumption that both the models face input forcings that are independently sampled from the same distribution at each time step). However in cases when input forcing is autocorrelated, model output space comprises the memory effect of both the input forcing as well as the model itself. Thus complexity quantification of models based solely on model output space in the presence of autocorrelated input forcing is not possible. Model span (comprised of model basis vectors) that solely represents memory effect of a model can then be used to cross compare two different models. The concept of model span alongside model output space can be used to decompose complexity of model response into complexity of model and complexity of input forcing.

It is then intuitive to expect that the extent of the Q-space, described in Figure 1, defines how flexible a model is in terms of how an arbitrary input forcing is transformed into output. In the following we quantify (the order of) the extent of hydrological models (either a linear reservoir or interconnected set of linear reservoir models). Lemma 1 provides an upper bound in the range of $|Q_m(t) - E(Q_m(t))|$. This upper bound also holds for the range of $|Q_m(t) - E(Q_m(t))|/N$ and is a measure of complexity that affects model performance. The quantity $|Q_m(t) - E(Q_m(t))|/N$ measures the distance between metric $Q_m(t)$ and $E(Q_m(t))$ in N-dimensional output space with ℓ_1-norm as the metric. It range thus measures the extent of model output space.

We also formalize a relationship between the extent of model output space (or Q-space), which we call model complexity, and model's prediction uncertainty.

Let $q_t = |Q_m(t) - E(Q_m(t))|$.

Lemma 1:   (**Upper Bound on the range of q_t**): Let $a \le q_t \le b$, $\forall t = 1, \ldots, N + m$ and $N = d + 1$. Then the following holds with probability $u - l$ for $0 < l < u < 1$,

$$r = |b - a| \le \sqrt{D} \left[\sqrt{\ln \frac{2}{l}} - \sqrt{\ln \frac{2}{u}} \right]$$

where $D = \dfrac{1 - e^{-k}}{1 + e^{-k}} \left(1 - \dfrac{\varepsilon^2}{C_{max}^2} \right) C_{max}^2$

Appendix A provides the proof.

Figure 1 │ Description of model output space and measure of complexity.

This result is used in Lemma 2 in the section entitled 'Parameter-dependent complexity measure and convergence bound for a simple one-reservoir model' to establish a bound on the rate of convergence for $Q_m(t)$. We later show, through a corollary, that the rate of convergence of the empirical error to the expected error (for such models) depends on the volume of the model span resulting from **V**. This corollary therefore connects the geometric interpretation of model span to model performance. As can also be seen in Figure 1, the volume of the model output space depends on the basis vectors that are columns of **V**.

We note here that model selection is a task of differentiating between different model classes and in this work we only consider one simple class of models. In which case when the problem is expanded to include additional model classes, it may not be possible justify model complexity based on output alone. Thus the concept of model output space, or the Q-space may appear to not provide a rigorous and clear measure of model complexity.

The concept of model output space can be extended to distinguish between model classes with different model structures/mathematical description by defining a general class of models. This general class of models is a collection of models each of which is identified by an abstract parameter set. We call this parameter set abstract because it describes both the model structure as well as (real) parameters corresponding to a given model structure. The Q-space or model output space can similarly be defined for each element of this class of models as well.

For example, basis vectors for interconnected linear reservoir models (one class of models) can be obtained as in the section entitled 'From a single reservoir model to a model of interconnected reservoirs' and an intuitively nonlinear basis function would need to be quantified for nonlinear models (another class of models). Both of these types of basis vectors define the model span of the elements of respective model classes (which characterizes the model output space or Q-space) in the output space. Here by output space we mean a positive real space with sample size as its dimensionality. While the shape of model span differs for these different model classes, they are defined in the same N (sample size) dimensional space. Thus the two model classes can be differentiated based on the shape and size of span of constituting model elements in the same N-dimensional space.

We present the concept of Q-space or model output space as geometric interpretation of the proof of Lemma 1 (provided in the appendix). Lemma 1 is a fundamental building block of analytical convergence bounds provided in this paper. Therefore, the concept of Q-space as measure of complexity is equivalent to the notion of complexity that effect rate of convergence as presented in this paper.

Other well established criterion (such as AIC and KIC) may as well be used to measure complexity and its effect on model performance. These measures are Bayesian and are complimentary to the measure of complexity presented in this paper (which is based on a frequentist approach to complexity and prediction uncertainty).

From hereon we assume $C_{\max} = 1$ and $\varepsilon/C_{\max} \ll 1$ for ease of exposition.

PARAMETER-DEPENDENT COMPLEXITY MEASURE AND CONVERGENCE BOUND FOR A SIMPLE ONE-RESERVOIR MODEL

Let $\mathbf{Z} = \{y(t), u(t)\}_{t<d+m+1}$ be a $d+m+1 \times 2$ matrix defining a given input-output data set, where $y(t)$ represents the observed outflow (measured by stream gauges for example) and $\mathbf{u} = \{u(t)\}_{t=1,\cdots,N}$ represents the input sequence.

Further, let $N = d+1$ and the empirical risk be defined as,

$$\xi_Z(k) = \frac{\sum_{t=1}^{N} |y(t) - Q(\mathbf{u}; t)|}{N} = \frac{\sum_{t=1}^{N} \ell_k(y, \mathbf{u}; t)}{N}$$

Let $Q(\mathbf{u}; t)$ represent modeled outflow for a given input sequence $\mathbf{u}$.

Assumption B: For some $\eta > 0$, let $|\ell_k(y, \mathbf{u}; t) - E[\ell_k(y, \mathbf{u}; t)]| \le \eta |Q(\mathbf{u}; t) - E[Q(\mathbf{u}; t)]|$ for any admissible observed output $\mathbf{y} = \{y(t)\}$ and input $\mathbf{u} = \{u(t)\}$ sequences.

Assumption B is a standard assumption, suggesting that $|\ell_k(y, \mathbf{u}; t) - E[\ell_k(y, \mathbf{u}; t)]|$ is of the same order of magnitude as $|Q(\mathbf{u}; t) - E[Q(\mathbf{u}; t)]|$. This assumption also implies that variance in prediction residuals ($\ell_k(y, \mathbf{u}; t)$) is dominated by variance in output of prediction models ($Q(\mathbf{u}; t)$).

We define convergence of the empirical error of a model to its expected error as the convergence of $\xi_Z(k)$ to its expectation $E(\xi_Z(k))$ (the 'expected' risk).

The upper bound on its rate is obtained by a bound for $\xi_Z(k)$, which in turn is obtained from the bounds on the rate of convergence for ε-approximation of Q, Q_m, (as previously defined, $|Q_m(\mathbf{u}; t) - Q(\mathbf{u}; t)| \leq \varepsilon$).

Lemma 2: (**Bound on the rate of convergence for Q_m**): Let assumptions A and B hold. Further, let $N = d + 1$, m define the memory coefficient of a model parameterized by k, Q_m define the ε-approximate model for outflow Q in (2), and $\mathbf{u} = \{u(t)\}_{t=1,\ldots,N}$ be any arbitrary input sequence. Let $C_{\max} = 1$, $\varepsilon \ll 1$ and $h = (1 - e^{-k})/(1 + e^{-k})$. Then for $u - l$ sufficiently close to 1 with $0 < l < u < 1$,

$$\Pr\left(\sum_{t=1}^{N}\left|\frac{Q_m(\mathbf{u}; t) - E[Q_m(\mathbf{u}; t)]}{N}\right| > \gamma\right)$$

$$\leq \exp\left(-\frac{\gamma^2 N}{2h[\sqrt{\ln 2/l} - \sqrt{\ln 2/u}]^2}\right)$$

Proof of the Lemma is provided in Appendix A.

In the above, h can be considered as a *measure of complexity* of the simple hydrologic model. It depends on the dominant vector in the set of basis vectors defining the span emerging from $\mathbf{V}$. This also defines the order of magnitude of its volume. This can also be observed in Figure 1, wherein the dominant column vector of $\mathbf{V}$ is $d(k)\begin{bmatrix} e^{-k} \\ 1 \end{bmatrix}$.

It also determines the order of magnitude of the major axis of circumscribing ellipsoid.

Corollary 2: (**Upper bound on the volume of span defined by V**): Consider a span defined by columns of $\mathbf{V}$ in N-dimensional space. Let $C_{\max} = 1$, $\varepsilon \ll 1$, and $h = (1 - e^{-k})/(1 + e^{-k})$. Its volume is then bounded by $V(k) \propto (h/2)^{N/2}$.

Proof of the corollary is provided in Appendix A.

Volume of the span defined by $\mathbf{V}$ also defines the volume of output space defined by $Q_m(\mathbf{u}; t)$ for any arbitrary $\mathbf{u}$. The span of $\mathbf{V}$ allows geometric interpretation of how parameters of a model (conceptualizing the underlying physics, here k) shape the transformation of forcing variables (here $\mathbf{u}$) into observed variables (here $Q_m(\mathbf{u}; t)$). The transformation of any possible forcing to

model output is the model span in Figure 1. The transformation is an underlying process of interest (here a simple mass balance). The bound on span volume is linked to results obtained in Lemma 2 through h. In Lemma 2, h has an interpretation of complexity while in Corollary 2, it defines a geometric entity (volume) describing the nature of the process being modeled. This link therefore quantifies one aspect of the complexity of the underlying process modeled by a simple linear reservoir model.

Lemma 3: (**Bound on the rate of convergence for Q**): Let assumptions A and B hold. Further, let $N = d + 1$, m define the memory coefficient of a model parameterized by k, Q_m defines an ε-approximate model for the outflow Q in (2), and $\mathbf{u} = \{u(t)\}_{t=1,\ldots,N}$ be any arbitrary input sequence.

Then,

$$\Pr\left(\frac{\sum_{t=1}^{N}|Q(\mathbf{u}; t) - E[Q(\mathbf{u}; t)]|}{N} > \gamma + 2\varepsilon\right)$$

$$\leq \Pr\left(\frac{\sum_{t=1}^{N}|Q_m(\mathbf{u}; t) - E[Q_m(\mathbf{u}; t)]|}{N} > \gamma\right)$$

Proof of the lemma is provided in Appendix A. This lemma builds upon Lemma 2 and uses the probability bound derived for ε-approximate streamflow $Q_m(\mathbf{u}; t)$ in the latter to derive a probability bound for $Q(\mathbf{u}; t)$ by using inequality (2b).

Theorem 1: (**Bound on the rate of convergence for ξ_Z**): Let Assumptions A and B hold. Further let $N = d + 1$, m defines the memory (or recession) coefficient of a model parameterized by k, $\xi_Z(m, k)$ define the empirical error, $\mathbf{u} = \{u(t)\}_{t=1,\ldots,N}$ be any arbitrary input sequence, corresponding y as observed output sequence, and $\mathbf{Z} = \{y(t), u(t)\}_{t=1,\ldots,N}$. Let $C_{\max} = 1$, $\varepsilon \ll 1$ and $h = (1 - e^{-k})/(1 + e^{-k})$. Then for $u - l$ sufficiently close to 1 with $0 < l < u < 1$,

$$\Pr(|\xi_Z(k) - E[\xi_Z(k)]| > \delta) \leq \exp\left(-\frac{[\delta/\eta]^2 N}{2h\left[\sqrt{\ln 2/l} - \sqrt{\ln 2/u}\right]^2}\right)$$

Proof of Theorem 1 is provided in Appendix A. This theorem builds upon Lemma 3 to link the probability bound for

streamflow with the probability bound for prediction error. A key message of this theorem is that the expected prediction error $E[\xi_Z(k)]$ is a function of empirical prediction error (i.e., prediction error on finite sample) $\xi_Z(k)$, and a tradeoff of sample size N with model complexity h. This is interesting because the absolute deviation between empirical and expected error, $|\xi_Z(k) - E[\xi_Z(k)]|$, is closely linked to the probability of choosing a suboptimal model when selection is based on empirical error, $\xi_Z(k)$. By suboptimality we mean selecting a model with minimum empirical error, $\xi_Z(k)$, but higher expected error $E[\xi_Z(k)]$ than other competing models. As the probability bound (the right-hand side (RHS) of the above inequality) depends on a tradeoff between sample size N and complexity h, the probability of picking a suboptimal model is high when the choice set is a collection of highly complex models. This theorem also formalizes the notion that highly complex models tend to overfit on small sample size. Models with large absolute deviation between empirical and expected error, $|\xi_Z(k) - E[\xi_Z(k)]|$, will have a large propensity to overfit. This is so because its average performance (in terms of prediction error) over many repeated samples (essentially $E[\xi_Z(k)]$) can be quite different from the performance that is observed on one sample ($\xi_Z(k)$). If this model also happens to have been selected (based on low $\xi_Z(k)$ by chance), it can have poor performance on other samples of similar size. The theorem suggests that propensity to overfit increases with model complexity.

FROM A SINGLE RESERVOIR MODEL TO A MODEL OF INTERCONNECTED RESERVOIRS

For a model with more than one reservoir its span defined by the corresponding matrix **V**, has more columns. As Lemma 2 depends on Lemma 1 to define bounds on convergence rates and complexity therefore the bounds can be readily obtained for the case of a model with interconnected reservoirs.

Consider the structure of interconnected reservoirs in the form of a network with nodes and links (in terms of a pattern of reservoir connections, Figure 2). These connections converge to one node, representing the outlet reservoir. Each such node represents a reservoir and the

links represent connections between such reservoirs. The recession coefficients define the strength of these links. Further, let each reservoir cover an equal area in terms of amount of precipitation received. Figure 2 describes an example description of such a model. The numbers in parentheses denote the order of the link (from the outlet).

The flow contributed by each reservoir to the outlet, at each instance of time, can then be characterized by the precipitation amount at the reservoirs and the set of recession coefficients along its path. Let i denote the order of a reservoir in the model, j identify a reservoir within the set of reservoirs of order i, k_{ij} denote the set of recession coefficients along the path of the ijth reservoir to the outlet, Q_{ij} denote its contribution to the total outflow at the outlet and let there be R reservoirs in total. Here, the order of a reservoir indicates the number of reservoirs between (and including) itself and the outlet. Then,

$$Q_{ij}(t) = \sum_{\tau_i \cdots \le \tau_2 \le \tau_1 \le t} d\left(k_{1j^*_{(1)}}\right) e^{-k_{1j^*_{(1)}}(t-\tau_1)}$$
$$\cdot d\left(k_{2j^*_{(2)}}\right) e^{-k_{2j^*_{(2)}}(\tau_1-\tau_2)} \cdots d(k_{ij}) e^{-k_{ij}(\tau_{i-1}-\tau_i)} . u(\tau_i)$$

where $j^*_{(\ell)}$ denotes the ℓth order reservoir in the set k_{ij}. Similar to a single reservoir case, we define memory m_{ij}
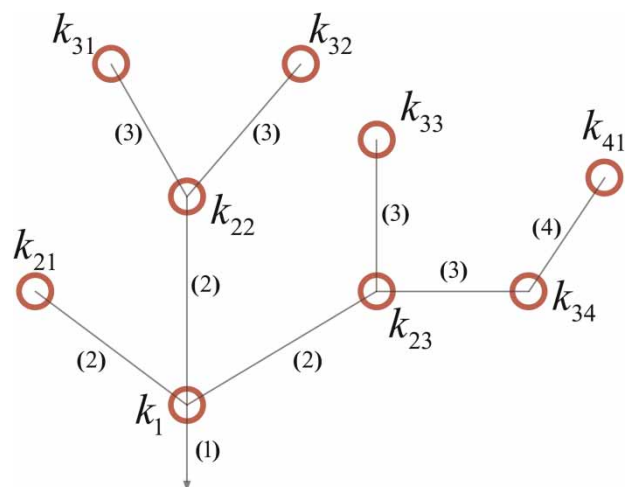

Figure 2 │ A model of interconnected reservoirs. The integers in parentheses indicate the order of the corresponding reservoir from the outlet, for example, the reservoir corresponding to k_{23} ($\{i,j\}$= {2,3}) is a second order reservoir as it has two reservoirs between (and including) itself and the outlet reservoir k_1 and third in the set of same order reservoirs (when counted clockwise).

corresponding to $\varepsilon_{ij} = \varepsilon/R$ approximation such that

$$|Q_{ij}(t) - Q_{ij,m_{ij}}(t)| \leq d(k_{ij}^{\min}) \sum_{\tau=t-m_{ij}+1}^{1} u(\tau)e^{-k_{ij}^{\min}(t-\tau)} \leq \frac{\varepsilon}{R} \qquad (6)$$

where $k_{ij}^{\min}$ is the minimum element of the set k_{ij}. Inequality (6) follows from (5) and $d(k)e^{-kx}$ being monotonically decreasing in k.

Finally, Equation (5) can be approximated in a similar fashion to (4b) by

$$\mathbf{q}_{ij} = \mathbf{V}_{ij}\mathbf{u}_{ij} \qquad (6b)$$

where the subscript ij identifies the corresponding reservoir.

Following Lemma 1, the sum of square of elements of any row of (equivalent) matrix $\mathbf{V}_{ij}$ corresponding to $Q_{ij,m_{ij}}(t)$ is $d^{2i}(k_{ij}^{\min})/(1 - e^{-2ik_{ij}^{\min}})(1 - (\varepsilon/(RC_{\max}))^{2i})$. The following lemma then follows.

Lemma 4: (**Bound on the rate of convergence for Q for a model of interconnected reservoirs**): Let Assumptions A and B hold. Further let $N = d + 1$, i denote the order of a reservoir in the model, j identify a reservoir within the set of reservoirs of order i, k_{ij} denote the set of recession coefficients along the path of the ijth reservoir to the outlet, $Q_{ij,m_{ij}}$ denote the contribution of the ijth reservoir to the total outflow at the outlet, R be the total number of reservoirs in the model, m define memory coefficient of a model, Q_m define ε-approximate model for outflow Q, and $\mathbf{u} = \{u(t)\}_{t=1,\dots,N}$ be any arbitrary input sequence. Let $h_{ij} = d^{2i}(k_{ij}^{\min})/(1 - e^{-2ik_{ij}^{\min}})(1 - (\varepsilon/(RC_{\max}))^{2i})$. Then for $u - l$ sufficiently close to 1 with $0 < l < u < 1$,

$$\Pr\left(\frac{\sum_{t=1}^{N}|Q(\mathbf{u};t) - E[Q(\mathbf{u};t)]|}{N} > \gamma + 2\varepsilon\right)$$

$$\leq \sum_{ij} \exp\left(-\frac{\gamma^2 N}{2R^2 C_{\max}^2 h_{ij}[\sqrt{\ln 2/l} - \sqrt{\ln 2/u}]^2}\right)$$

Proof of Lemma 4 is provided in Appendix A.

Note here again that the above bounds are for a spatially explicit hydrological model, which in effect incorporates the volume of its span through h_{ij}. The quantity h_{ij} depends on the slowest reservoir along the path from the ith order reservoir to the outlet, thereby distinguishing between hydrological responses based on upstream heterogeneity in hydrologic properties. This in turn quantifies the complexity of a rainfall–runoff process. The quantity h_{ij} is exponential in i, the order of reservoirs. Meanwhile, the bound is a sum over all R reservoirs. Thus the probability bounds also encapsulate the degree of convergence in spatial connectivity on complexity of hydrologic response.

Theorem 2: (**Bound on the rate of convergence for ξ_Z for a model of interconnected reservoirs**): Let Assumptions A and B hold. Further let $N = d + 1$, i denote the order of a reservoir in the model, j identify a reservoir within the set of reservoirs of order i, R be the total number of reservoirs in the model, m define the memory coefficient of a model, $\xi_Z(m, k)$ define the empirical error, $\mathbf{u} = \{u(t)\}_{t=1,\dots,N}$ be any arbitrary input sequence with y as observed output sequence, and $Z = \{y(t), u(t)\}_{t=1,\dots,N}$. Let $C_{\max} = 1$, $\varepsilon \ll 1$ and $h_{ij} = d^{2i}(k_{ij}^{\min})/(1 - e^{-2ik_{ij}^{\min}})(1 - (\varepsilon/(RC_{\max}))^{2i})$. Then,

$$\Pr(|\xi_Z(k) - E[\xi_Z(k)]| > \delta)$$

$$\leq \sum_{ij} \exp\left(-\frac{[\delta/\eta]^2 N}{2R^2 h_{ij}[\sqrt{\ln 2/l} - \sqrt{\ln 2/u}]^2}\right)$$

Proof: This can be shown in a manner similar to the proof of Theorem 1.

Theorem 2 extends the message of Theorem 1 to a spatially explicit model of interconnected linear reservoirs. Thus Theorem 2 can be seen as an extension of the probability bounds on prediction errors for one type of nonlinear hydrological model. The measure of complexity in Theorem 2 constitutes not only the effects of recession coefficients but also the spatial structure of hydrological models. The propensity to select a suboptimal model as well to overfit depends on the spatial structure of the underlying hydrology as envisaged by the nonlinear model of interconnected linear reservoirs.

APPLICATIONS AND EXTENSIONS

Examples

Corollary 2 links the volume of the span defined by columns of $\mathbf{V}$ in Equation (4b) with the complexity measure that appears in Lemma 1 with $\varepsilon \ll 1$, $C_{\max} = 1$. We empirically estimate the approximate diameter of the linear reservoir's model output space using a global optimization scheme called Shuffled Complex Evolution, SCE-UA (Duan *et al.* 1992), wherein a maximization of mean absolute deviation between any two model outputs ($Q(t)$) for a fixed parameter value is performed while searching over input data ($u(t)$ uniformly distributed between 0 and 1) with $N = 200$ (dimensionality). The SCE-UA, is a global search optimization method designed to handle difficult, nonlinear response surfaces encountered in the calibration of conceptual watershed models and has been widely used in the hydrologic community. A detailed description of the method appears in Duan *et al.* (1992). In summary the algorithm is a mix of the downhill simplex approach with some evolutionary optimization concepts, in which a 'population' of points is selected randomly from the feasible parameter space, is partitioned into several complexes (group of points), each of which is allowed to evolve independently but periodically shuffled to share information. At that point new complexes are formed. The evolution and shuffling are repeated until the specified convergence criteria are satisfied. In the present study 20 complexes of $2N + 1$ (where N is the sample size) points were used with a convergence criterion of 0.1% (change in objective function). The search is terminated after 100,000 iterations if an optimum value is not found.

We use SCE-UA to find $\max_{\mathbf{u}} \sum_{t=1}^{N} |Q_1(t) - Q_2(t)|/N$ where $\{Q_1(t)\}_{t=1,\ldots,N}$ and $\{Q_2(t)\}_{t=1,\ldots,N}$ are model output vectors for two different instantiations of input forcings $\mathbf{u} = \{u(t)\}_{t=1,\ldots,N}$ and maximum is taken with respect to random input instantiations. This provides an estimate of the maximal extent of model output space. We however add a note of caution here. We use SCE-UA algorithm to search a high dimensional (dimensionality $= N$) input data set such that the mean absolute error between any two model outputs for a particular parameter value is maximized. Given the high dimensionality of the model, SCE-UA may yield a local optima. Thus, the obtained extent of model output space may be lower than the global maxima.

We note that the maximum possible value of $\sum_{t=1}^{N} |Q(t) - EQ(t)|/N$ over different instantiations of input forcings is never smaller than half the absolute difference between any two input forcing instantiation. That is,

$$\max_{\mathbf{u}} \sum_{t=1}^{N} \frac{|Q(t) - EQ(t)|}{N} \geq \frac{1}{2} \max_{\mathbf{u}} \sum_{t=1}^{N} \frac{|Q_1(t) - Q_2(t)|}{N}$$

This follows with equality holding for distributions of $Q(t)$ that are symmetric around $EQ(t)$. In the experiment presented here, distribution of $Q(t)$ is symmetric around $EQ(t)$ as $\mathbf{u} = \{u(t)\}_{t=1,\ldots N+m}$ (input forcing) is independently and identically distributed (uniform distribution) with $EQ(t) = Eu(t)$. Thus for the input forcing used in this,

$$\max_{\mathbf{u}} \sum_{t=1}^{N} \frac{|Q(t) - EQ(t)|}{N} = \frac{1}{2} \max_{\mathbf{u}} \sum_{t=1}^{N} \frac{|Q_1(t) - Q_2(t)|}{N}$$

If the range of $\sum_{t=1}^{N} |Q_m(t) - EQ_m(t)|/N$ is approximately equal to $\max_{\mathbf{u}} \sum_{t=1}^{N} |Q(t) - EQ(t)|/N$, the former may as well replace the latter in the above equality to yield an approximate equality between the range of $\sum_{t=1}^{N} |Q_m(t) - EQ_m(t)|/N$ and $\max_{\mathbf{u}} \sum_{t=1}^{N} |Q_1(t) - Q_2(t)|/N$.

Lemma 1 provides *an upper bound* on the (magnitude of) range (with confidence $u - l$) of $|Q_m(t) - EQ_m(t)|$. It thus also provides an upper bound on the range of $\sum_{t=1}^{N} |Q_m(t) - EQ_m(t)|/N$. When $\varepsilon/C_{\max} \ll 1$, Lemma 1 provides an upper bound on the range of $\sum_{t=1}^{N} |Q(t) - EQ(t)|/N$. We estimate these bounds at confidence levels of 95, 90, 80, 75, and 70% by substituting $l = \{0.025, 0.05, 0.10, 0.125,$ and 0.15 resp.$\}$ and $u = \{0.925, 0.95, 0.90, 0.875,$ and 0.85 resp.$\}$. For the case when $\mathbf{u}$ is uniformly distributed between 0 and 1 with $E(u(t)) = 0.5$, $\sum_{t=1}^{M} |Q(t) - EQ(t)|/N$ can never be larger than 0.5. Thus the upper bounds on the range of $\sum_{t=1}^{N} |Q(t) - EQ(t)|/N$ can never be larger than 0.5, which we impose on the bounds obtained from Lemma 1.

Figure 3 shows that the (analytical) upper bound on the range of $\sum_{t=1}^{N} |Q(t) - EQ(t)|/N$ is never smaller than the
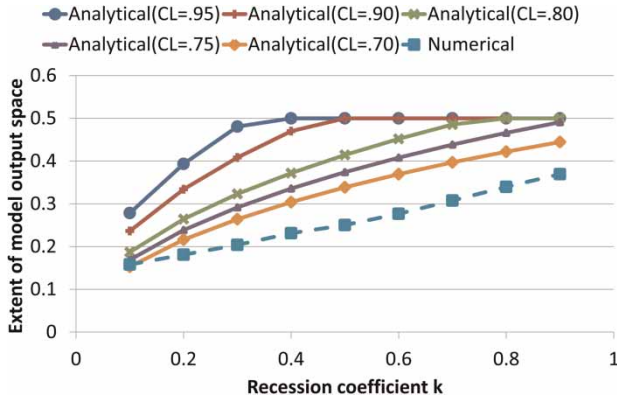
Figure 3 | A comparison of analytical upper bound on the range of $\sum_{t=1}^{N} |Q(t) - EQ(t)|/N$ and empirical estimate of the diameter of linear reservoir model output space. The input forcing ($u(t)$) is uniformly distributed and lies between 0 and 1, i.e., $C_{max} = 1$. In the legend, CL is confidence level.

numerical estimation of $\max_{\mathbf{u}} \sum_{t=1}^{N} |Q_1(t) - Q_2(t)|/N$ for all confidence levels considered. Both the numerical and analytical estimates increase with increasing recession coefficient suggesting that complexity of the underlying rainfall–runoff process increases with faster transformation of rainfall into runoff. Further, the analytical bounds provided in Lemma 1 provide upper bounds on the empirically-derived diameter of modeling space, the latter of which in turn can be used to quantify the complexity of arbitrary hydrological models.

We add a note of caution here. We use SCE-UA algorithm to search for a high dimensional input data set such that the mean absolute error between any two model output for a particular parameter value is maximized. The estimated model output space diameter is sensitive to the choice of SCE-UA parameters and may as well be sensitive to the choice of solver itself (SCE-UA). Further how closely should optimization based output space diameter match with derivation based on Lemma 1 needs further deliberation. We postpone its investigation to a later study.

As another example we consider a spatially distributed model with recession parameters k that are spatially constant. Further let $R = 3$, $\varepsilon = 0$, and consider two configurations of the reservoirs as shown in Figures 4(a) and (b). Note that different structures (such as the two configurations here) also conceptualize geomorphologic influence on hydrologic flows in natural systems. Figure 4(c) show how the bounds (the RHS of the probability inequality
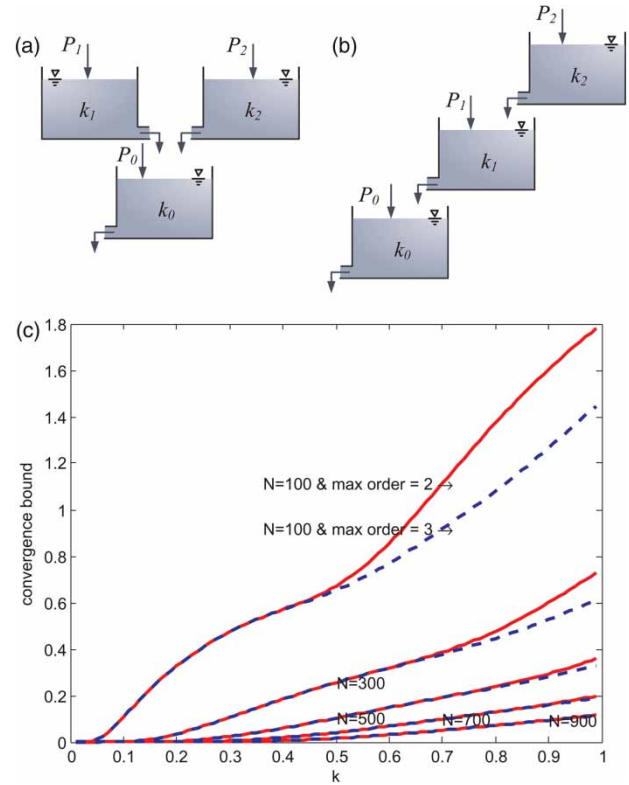

Figure 4 | (a) and (b) show the two configurations with $R = 3$, (a) has a maximum order of 2 while (b) has a maximum order of 3. (c) shows how the convergence bound (the RHS of the probability inequality in Theorem 2) for these two models performs when the recession coefficient is spatially constant (i.e., $k_1 = k_2 = k_3 = k$), k is varied over a range [0.01, 0.99] and N in $(\delta/\eta)^2(N/2R^2)$ is varied {100, 300, 500, 700, 900} {$\delta/\eta = 0.1$, $R = 3$). P_i denotes effective precipitation (input) going into the store i. In (c), 'Max order $= 2$' refers to the model in (a) with maximum order of reservoir $= 2$ and 'Max order $= 3$' refers to the model in (b).

in Theorem 2) perform as k and $(\delta/\eta)^2(N/2R^2)$ is varied over a range (with k varied between 0 and 1).

Figure 4(c) also shows that for larger values of N, the bounds of both models converge and approach 0. For any value of N, the model in Figure 4(a) with the lower order has looser bounds than the model in Figure 4(b), at least for larger values of k. Further, the bounds loosen with increasing values of recession coefficient in both models. Loosening of the bounds indicate larger complexity (as also observed in Figure 3). As the bounds are a function of N and complexity, the latter (for fixed N) increases with increasing k. Similarly the model in Figure 4(a) (with maximum reservoir order of 2) displays higher complexity than the model in Figure 4(b) (with maximum reservoir order of 3), especially for lower values of N and large values of k.

This example shows that the spatial structure of a model (which has a clear hydrologic meaning as it is an interpretation of the predominant physical processes in a particular catchment) can affect the convergence bounds, and therefore its parameter estimation. Further, these bounds are also a function of the recession parameters through complexity.

Global nonlinear hydrologic behavior representation by locally linear behavior

Nonlinear hydrologic behavior at catchment scale can be represented through interconnected linear reservoir models (Uhlenbrook *et al.* 2004; Clark *et al.* 2009). Nonlinearity of catchment response is represented by parameter heterogeneity of constituting linear reservoirs. Such a representation is akin to representation of nonlinear functions by piecewise linear functions. Each linear reservoir represents local behavior and its parameters model dominant physics at that scale (catchment response is then a combination of hillslope responses and flow through stream network). Complexity of catchment response is driven by complexity of constituting local behaviors, which in turn is driven by its parameters.

Consider a two-dimensional linear groundwater flow equation (a linearized BE) (Pulido-Velazquez *et al.* 2007),

$$\frac{\partial}{\partial x}\left(T_x \frac{\partial h}{\partial x}\right) + \frac{\partial}{\partial y}\left(T_y \frac{\partial h}{\partial y}\right) + w = S\frac{\partial h}{\partial t}$$

where $h = h(x, y, t)$ is the hydraulic head, $S(x, y)$ is the storage coefficient, $w(x, y, t)$ is net recharge, and T_x, T_y are transmissivity coefficients along the x and y directions (depth to water table is the z-direction). These transmissivity coefficients are assumed independent of hydraulic head h (under the assumption that saturated thickness is significantly larger than the fluctuations in hydraulic head h). These coefficients are therefore product of location specific hydraulic conductivity and saturated thickness. Depending on spatial discretization of the problem, data on hydraulic properties and saturated thickness is needed before a solution to the linearized BE can be obtained (unless it is calibrated). The data requirement increases with the resolution of spatial discretization.

Its analytical solution for lateral flow from an aquifer to surface water body can be expressed as (Pulido-Velazquez *et al.* 2007),

$$\tilde{Q}(t) = \sum_{i=1}^{\infty} \alpha_i V_i(t) = \sum_{i=1}^{\infty} \alpha_i b_i (1 - e^{-\alpha_i \Delta t})\left(\sum_{\tau=1}^{t} R(\tau) e^{-\alpha_i \tau \Delta t}\right)$$

Here, α_i is a 'discharge coefficient' of ith reservoir and b_i is a fraction of total stress $R(\tau)$ (exogenous forcing on to the system) applied to it. A nonlinear (groundwater flow) process represented by the linear groundwater flow equation is combination of linear reservoirs connected in series.

Linear reservoir models presented in the section entitled 'A simple hydrologic model (one reservoir model)', when connected in a spatially explicit manner, can therefore represent highly nonlinear process such as above. This is evident by defining α_i as k_i, $(1 - e^{-\alpha_i \Delta t})$ as $d(k_i)$, $b_i R(\tau)$ as $u_i(\tau)$ and $\Delta t = 1$. Complexity of the process above depends on the constituting coefficients as well as connectivity (which here is in series). The probability bounds presented in the above section quantify such a relationship, as shown in the sections defining the model and the following section on the parameters.

Why complexity?

The shape and size of the model output space (defined by columns of V in the case of a single linear reservoir) governs the flexibility of hydrologic response under stochastic forcings. Its size has been defined here as complexity (via Corollary 2). Such a behavior is valid for models that are physics based, such as the one presented in the previous subsection of the paper, that are closest in representing underlying flow processes in porous medium (such as a soil matrix). Given that coefficients (or parameters in conceptual models) quantifies complexity, a quantification of (rather than qualitative) the nature of processes complexity emerges as a result. This quantification is a unique contribution of this paper.

Model selection that best identifies the underlying process is governed by both the complexity of the underlying process (manifesting itself in available information) and

complexity of the set of models available (one of which is finally selected). This is elicited in Theorems 1 and 2). Consider Theorem 1 (and the corresponding definitions in section on parameters),

$$\Pr(|\xi_Z(k) - E[\xi_Z(k)]| > \delta|) \leq \exp\left(-\frac{[\delta/\eta]^2 N}{2h[\sqrt{\ln 2/l} - \sqrt{\ln 2/u}]^2}\right)$$

Without loss of generality, let $\eta = 1$ and let $H = h[\sqrt{\ln 2/l} - \sqrt{\ln 2/u}]^2$. By equating the RHS to χ, we can state the following with probability of at least $1 - \chi$:

$$E[\xi_Z(k)] \leq \xi_Z(k) + \sqrt{\frac{2H}{N}\ln\left(\frac{1}{\chi}\right)}$$

For a given set of models defined by a set of possible values of k, and given the available information on the underlying process (embodied in data Z defined in the parameters section), the best available model from the set can be selected by minimizing the RHS of the above inequality. Such a minimization also formalizes Occam's razor principle. Occam's principle of parsimony has the following form: 'given two explanations of the data, all other things being equal, the simpler explanation is preferable' (Blumer *et al.* 1987). In other words, choose the simplest hypothesis that is consistent with the sample data (Blumer *et al.* 1987). A hypothesis chosen based on this principle is the best predictor of future observations with high probability (which has been proved here).

Finally, the complexity of the model selected to represent the underlying processes also has implications for assessing the impact of hydrologic response on human systems. We further elaborate this aspect of complexity in the following section.

Implications for sustainable allocation at basin scale

Consider a simple example wherein there are two agents (upstream $= 1$, downstream $= 2$) residing in two contiguous subbasins (constituting a basin), that utilize water, $e_1(t)$, $e_2(t)$, for income generation, $\sum_{t=1}^{T} F_i(e_i(t), S_i(t))$. Assume that $F_i(e_i(t), S_i(t))$ is concave and increasing in the first

argument while convex and decreasing in the second argument. Hydrologic behavior dictates flow from upstream to downstream agents as the function of the upstream agent's soil moisture conditions. Here, hydrologic behavior is modeled by k_1 and k_2, which best approximate it in the sense of the previous section. For simplicity, we represent the basins by linear reservoir models, with store levels $S_1(t), S_2(t)$ such that the allocation solution is sustainable for the two basins taken together over a certain T period (under stochastic rainfall conditions $u_i(t), i = 1, 2$). Sustainable allocation solution can be decentralized (Lyon & Pande 2006) by the marginals (Lagrange multipliers $\mu_{i,t}, i = 1, 2$) of the following program:

$$W(\mathbf{u}, \mathbf{k}) = \max_{e,S} \sum_{i=1}^{2} \sum_{t=1}^{T} F_{i,t}(e_i(t), S_i(t))$$

st,
$$S_1(t+1) - S_1(t) = u_1(t) - e_1(t) - k_1 S_1(t) \qquad (\mu_{1,t})$$
$$S_2(t+1) - S_2(t) = u_2(t) - e_2(t) - k_2 S_2(t) + k_1 S_1(t) \qquad (\mu_{2,t})$$
$$S_i(T) - S_i(1) = 0, i = 1, 2$$

From first order conditions for $1 < t < T$ with respect to $S_i(t+1)$, $i = 1, 2$, we have

$$\mu_2(t) - \mu_2(t-1) = k_2 \mu_2(t) - \frac{\partial F_{2,t}}{\partial S_{2,t}} \qquad (7a)$$

$$\mu_1(t) - \mu_1(t-1) = k_1(\mu_1(t) - \mu_2(t)) - \frac{\partial F_{1,t}}{\partial S_{1,t}} \qquad (7b)$$

The partial derivatives in (7a) and (7b) depend on the stochasticity of $u_i(t)$, $i = 1, 2$, while $\mu_{i,t}$, $i = 1, 2$ describes the evolution of prices that can decentralize such an allocation solution. Equation (7a) is similar to a linear reservoir storage soil moisture evolution equation with stochastic input $-\partial F_{2,t}/\partial S_{2,t}$, while Equation (7b) is similar to a spatially distributed soil moisture evolution equation with stochastic input $-\partial F_{1,t}/\partial S_{1,t}$.

The T-dimensional span of the downstream agent's prices $\{\mu_{2,t}, t = 1, ..., T\}$ is determined by k_2 and its volume can be bounded using Lemma 1. Further, this volume also defines the 'complexity' of downstream prices, which

is due to the complexity of its underlying hydrologic response. Using the convergence bounds estimated in Lemma 3 (with variables renamed), this volume can be related to the flexibility of (or potential volatility in) the downstream agent's prices that are feasible for a range of stochastic input $u_i(t)$, $i = 1$, 2 (through $-\partial F_{2,t}/\partial S_{2,t}$). Similarly the complexity in hydrologic response can be related to the potential volatility in the upstream agent's prices using the convergence bounds described in Lemma 4 (with variables renamed).

The above simple example can be generalized for a sub-basin with arbitrary hydrologic connectivity (the way various subbasins interconnect) and within-subbasin nonlinearity. Complexity and potential volatility of subsequent prices can then similarly be extended.

DISCUSSION AND CONCLUSIONS

In this paper we introduced a quantitative measure of complexity that is applicable to hydrological models. The measure was based on the Vapnik–Chervonenkis generalization theory that relates model complexity to sample size and predictor error. We showed through a simple example and sequences of lemmas and theorems that this measure has geometric interpretation, and thereby allowed more intuitive insights into the theory presented. In particular, we showed that the complexity measure depends on the magnitude of model parameters (fast reservoirs are more complex than slow reservoirs) as well as model structure (parallel reservoir configurations are more complex than in series with the same number of reservoirs). By extension, if hydrologic models are assumed to represent underlying hydrological behavior closely, we argued that the complexity of hydrologic response depends on upstream hydrologic connectivity and heterogeneities (for example on heterogeneity in soil properties). This paper also estimated the convergence bounds, first for a simple single linear reservoir model and then a conceptual spatially explicit hydrologic model. By convergence we meant the rate (with increasing sample size) with which prediction error calculated on finite sample (empirical error/risk) converges to the mean of prediction errors over repeated samples of same size (expected error/risk). The convergence bounds on such predictive performance were shown to be a function of model parameters and structure as a consequence of its dependence on complexity measure. These bounds are a function of model complexity. Based on this, we argued that the complexity of modeled hydrologic response controls robust model identification and that complexity depends on the properties of modeled hydrologic processes. We also discussed its applications and extensions with examples.

We note that the results derived are not applicable when the class of models is changed from interconnected linear reservoir models to another class of models. However, the results and in particular the derivation of these results as well as its geometric interpretation are useful in deriving convergence bounds for an arbitrary class of models. For example, we note from geometric interpretation of Q-space for a linear reservoir model that the basis vectors define the span, which in turn define its complexity (and thus convergence bounds). One may qualitatively extrapolate this notion and look for nonlinear basis functions that can describe the span of nonlinear reservoir models. Research efforts may therefore be directed at finding nonlinear basis functions for particular hydrologic models. Complexity quantification of nonlinear models can then follow in a spirit similar to this paper. Comparison between any two state-of-the-art hydrological models can be made in terms of their nonlinear basis functions.

We also note that in deriving the bounds on rate of convergences, we considered the effect of memory on the (variance of) model output at each time step (considered in Lemma 1) but ignored the effect of memory on correlation between model outputs at two time steps (ignored in Lemma 2 by assuming independence of q_t between any two time instances). However this does not affect the conclusion that complexity increases with quickness of runoff response to rainfall. By Chebyshev's inequality (Boucheron *et al.* 2004), we have

$$\Pr\left\{\sum_t q_t \geq N\gamma\right\} \leq \frac{E[(\sum_t q_t)^2]}{N^2\gamma^2}$$

where $q_t = |Q_m(t) - E(Q_m(t))|$ as in Lemma 1.

It can then be shown for i.i.d. input forcing $u(\text{t})$, $\varepsilon/C_{\max} \ll 1$, and $C_{\max} = 1$ that

$$E\left\{ \left(\sum_t q_t \right)^2 \right\} \approx \left(\text{ND} + 2\frac{e^{-k}}{1 + e^{-k}} \right) \sigma^2(u)$$

where N is the sample size, $h = (1 - e^{-k})/(1 + e^{-k})$ is the complexity measure that we use (as defined in Lemma 1), $\sigma^2(u)$ is the variance of input forcing and k is the recession coefficient. Thus,

$$\Pr\left\{ \frac{\sum_t q_t}{N} \geq \gamma \right\} \leq \frac{(h + 2(e^{-k}/N(1 + e^{-k})))\sigma^2(u)}{N\gamma^2}$$

and for not too small N,

$$\Pr\left\{ \frac{\sum_t q_t}{N} \geq \gamma \right\} \leq \frac{h\sigma^2(u)}{N\gamma^2}$$

The left-hand side (LHS) is the probability that we bound in Lemma 2 and the above suggests that bound on probability should tighten with increasing N but loosen with h the complexity measure that we have proposed. As h increases with increasing value of k, convergence bound weaken with increasing k. Using Lemma 3 onwards, it then demonstrates again that complexity of rainfall–runoff processes increases with the quickness of the response. However, improvement of results presented in the lemmas of this paper is left for future work.

We here studied models that omit thresholding behavior (Liebe *et al.* 2009) in hydrologic behavior, i.e., models which conceptualize linear storage–discharge relationship. However, this is not a limitation as nonlinear basin response can be conceptualized through a distribution of interconnected linear reservoir models (Harman *et al.* 2009).

In future research, we intend to pursue numerical estimation of complexity for state-of-the-art hydrologic models based on the bounds (and the concept of complexity as the extent of model span) presented in this paper. We also intend to investigate how the shape of model output space as exemplified in Figure 1 can be used to describe model uncertainty and how it is linked to resilience of a model to perturbations to input forcings. Yet another interesting extension of the concepts presented here can be its implications for decentralized water resource management.

REFERENCES

Alon, N., Ben-David, S., Cesa-Bianchi, N. & Haussler, D. 1997 Scale-sensitive dimensions, uniform convergence, and learnability. *J. Assoc. Comp. Mach.* **44** (4), 615–631.

Bartlett, P. L. 1998 The sample complexity of pattern classification with neural networks: the size of the weights is more important than the size of the network. *IEEE Trans. Inf. Theory* **44** (2), 525–536.

Bartlett, P. L. & Kulkarni, S. R. 1998 The complexity of model classes, and smoothing noisy data. *Systems Control Lett.* **34** (3), 133–140.

Beven, K. J. 1981 Kinematic subsurface stormflow. *Wat. Resour. Res.* **17** (5), 1419–1424.

Beven, K. J. & Binley, A. M. 1992 The future of distributed models: model calibration and uncertainty prediction. *Hydro. Proc.* **6**, 279–298.

Blumer, A., Ehrenfeucht, A., Haussler, D. & Warmuth, M. K. 1987 Occam's Razor. *Inf. Process. Lett.* **24**, 377–380.

Blumer, A., Ehrenfeucht, A., Haussler, D. & Warmuth, M. K. 1989 Learnability and the Vapnik–Chervonenkis dimension. *J. Assoc. Comp. Mach.* **36** (4), 929–965.

Brutsaert, W. & Ibrahim, H. A. 1966 Research note on the first and second linearization of the Boussinesq equation. *Geophys. J. R. Astron. Soc.* **11**, 549–554.

Brutsaert, W. & Nieber, J. 1977 Regionalized drought flow hydrographs from a mature glaciated plateau. *Water Resour. Res.* **13** (3), 637–643.

Boucheron, S., Lugosi, G. & Bousquet, O. 2004 Concentration inequalities. *Advanced Lectures on Machine Learning, Lecture Notes in Computer Science, 2004*, **3176/2004**, 208–240.

Burnash, R. J. C. 1995 The NWS River forecast system-catchment modelling. In: *Computer Models of Watershed Hydrology* (V. P. Singh, ed.), Water Resource Publications, Highlands Ranch, CO, USA, pp. 311–366.

Clark, M. P., Rupp, D. E., Woods, R. A., Tromp-van Meerveld, H. J., Peters, N. E. & Freer, J. E. 2009 Consistency between hydrological models and field observations: linking processes at the hillslope scale to hydrological responses at the watershed scale. *Hydrol. Proc.* **23**, 311–319.

Cucker, F. & Smale, S. 2001 On the mathematical foundations of learning. *Bull. Amer. Math. Soc.* **39** (1), 1–49.

Duan, Q., Sorooshian, S. & Gupta, V. 1992 Effective and efficient global optimization for conceptual rainfall-runoff models. *Water Resour. Res.* **28** (4), 1015–1031.

Elayyan, A. & Isakov, V. 1997 On an inverse diffusion problem. *SIAM J. Appl. Math.* **57** (6), 1737–1748.

Fong, P. 1995 A quantitative study of hypothesis selection. In *Proceedings of the 12th International Conference on Machine Learning (ICML-95), 226–234*. Morgan Kaufmann Publishers, Inc (http://citeseer.ist.psu.edu/viewdoc/summary?doi=10.1.1.20.6423)

Goldstein, M. 1975 Some remarks on probability inequalities for sums of bounded convex random variables. *J. App. Prob.* **12**, 155–158.

Govindaraju, R. S. & Koelliker, J. K. 1994 Applicability of linearised Boussinesq equation for modeling bank storage under uncertain aquifer parameters. *J. Hydrol.* **157**, 349–366.

Guenni, L. & Bardossy, A. 2002 A two step disaggregation method for highly seasonal monthly rainfall. *Stochastic Environ. Res. Risk Assessment* **16** (3), 188–206.

Gupta, V. K. & Sorooshian, S. 1983 Uniqueness and observability of conceptual rainfall-runoff parameters percolation process examined. *Water Resour. Res.* **19**, 269–276.

Gupta, H. V., Sorooshian, S. & Yapo, P. O. 1998 Toward improved calibration of hydrologic models: multiple and noncommensurable measures of information. *Water Resour. Res.* **34** (4), 751–763.

Han, D., Chan, L. & Zhu, N. 2007 Flood forecasting using support vector machines. *J. Hydroinformatics* **9** (4), 267–276.

Harman, C. J., Sivapalan, M. & Kumar, P. 2009 Power law catchment-scale recessions arising from heterogeneous linear small-scale dynamics. *Water Resour. Res.* **45**, W09404.

Haussler, D., Kearns, M., Littleston, N. & Warmuth, M. K. 1991 Equivalence of models for polynomial learnability. *Inform. Comput.* **95**, 129–161.

Hoeffding, W. 1963 Probability inequalities for sums of bounded random variables. *J. Amer. Statist. Assoc.* **58**, 13–30.

Kearns, M. J. & Schapire, R. 1994 Efficient distribution-free learning of probabilistic concepts. *J. Comput. Syst. Sci.* **48** (3), 464–497.

Lacey, A. A., Ockendon, J. R. & Tayler, A. B. 1982 'Waiting Time' solutions of a nonlinear diffusion equation. *SIAM J. Appl. Math.* **42** (6), 1252–1264.

Liebe, J. R., van de Giesen, N., Andreini, M., Walter, M. T. & Steenhuis, T. S. 2009 Determining watershed response in data poor environments with remotely sensed small reservoirs as runoff gauges. *Water Resour. Res.* **45**, W07410.

Lugosi, G. & Nobel, A. B. 1999 Adaptive model selection using empirical complexities. *Ann. Statist.* **27** (6), 1830–1864.

Lyon, K. S. & Pande, S. 2006 The role of costate variable in stochastic renewable resource model. *Nat. Resour. Model.* **19** (1), 45–66.

Meir, R. 2000 Nonparametric time series prediction through adaptive model selection. *Mach. Learn.* **39**, 5–34.

Pande, S., McKee, M. & Bastidas, L. A. 2009 Complexity-based robust hydrologic prediction. *Water Resour. Res.* **45**, W10406.

Paniconi, C., Troch, P. A., van Loon, E. E. & Hilberts, A. G. J. 2003 Hillslope-storage Boussinesq model for subsurface flow and variable source areas along complex hillslopes: 2. Intercomparison with a three-dimensional Richards equation model. *Water Resour. Res.* **39** (11), 1317.

Pontil, M. 2003 A note on different covering numbers in learning theory. *J. Complexity* **19**, 665–671.

Pulido-Velazquez, D., Sahuquillo, A., Andreu, J. & Pulido-Velazquez, M. 2007 An efficient conceptual model to simulate surface water body-aquifer interaction in conjunctive use management models. *Water Resour. Res.* **43**, W07407.

Savenije, H. H. G. 2009 The art of hydrology. *Hydrol. Earth Syst. Sci.* **13**, 157–161.

Schoups, G., van de Giesen, N. C. & Savenije, H. H. G. 2008 Model complexity control for hydrologic prediction. *Water Resour. Res.* **44**, W00B03.

Uhlenbrook, S., Roser, S. & Tilch, N. 2004 Hydrological process representation at the meso-scale: the potential of a distributed, conceptual catchment model. *J. Hydrol.* **291**, 278–296.

Valiant, L. G. 1984 A theory of the learnable. *Commun. ACM* **27** 11, 1134–1142. http://doi.acm.org/10.1145/1968.1972.

van Griensven, A. & Meixner, T. 2007 A global and efficient multi-objective auto-calibration and uncertainty estimation method for water quality catchment models. *J. Hydroinform.* **9** (4), 277–291.

Vapnik, V. 1982 *Estimation of Dependencies based on Empirical Data*. Springer Verlag, New York.

Vapnik, V. 2002 *The Nature of Statistical Learning Theory*, 2nd edition. Springer-Verlag, New York.

Vapnik, V. & Chervonenkis, A. 1971 On uniform convergence of relative frequencies of events to their probabilities. *Theory Probab. Appl.* **16** (2), 264–280.

Vrugt, J. A., Gupta, H. V., Bastidas, L. A., Bouten, W. & Sorooshian, S. 2003a Effective and efficient algorithm for multiobjective optimization of hydrologic models. *Water Resour. Res.* **39** (8), 1214.

Vrugt, J. A., Gupta, H. V., Bouten, W. & Sorooshian, S. 2003b A shuffled complex evolution metropolis algorithm for optimization and uncertainty assessment of hydrologic model parameters. *Water Resour. Res.* **39**, 1201.

Wang, S. X. & Waymire, E. C. 1991 A large deviations rate and central limit theorem for Horton Ratios. *SIAM J. Discrete Math.* **4** (4), 575–588.

Ye, M., Neuman, S. P., Meyer, P. D. & Pohlmann, K. 2005 Sensitivity analysis and assessment of prior model probabilities in MLBMA with application to unsaturated fractured tuff. *Water Resour. Res.* **41**, W12429.

Ye, M., Meyer, P. D. & Neuman, S. P. 2008 On model selection criteria in multimodel analysis. *Water Resour. Res.* **44**, W03428.

APPENDIX A: PROOFS

Proof of Lemma 1:   We first use Hoeffding's inequality (Hoeffding 1963) to obtain the following,

$$\Pr(q_t \geq \eta) \leq 2 \exp\left(\frac{-\eta^2}{C_{\max}^2 \sum_{j=1}^{m+1} w_j^2}\right) \tag{A1}$$

where $w_j = d(k)e^{-(j-1)k}$. Heoffding inequality bounds the rate of convergence of mean of a finite number of random numbers to its expected value. The above inequality follows as

$$q_t = |Q_m(t) - E(Q_m(t))|$$
$$= \left| \sum_{j=t}^{t-(m+1)} w_j u(j) - E\left(\sum_{j=t}^{t-(m+1)} w_j u(j) \right) \right|,$$

$u(j) \leq C_{\max}$ and the random numbers are $w_j u(j)$.

Further,

$$\sum_{j=1}^{m+1} w_j^2 = d^2(k)(1 + e^{-\tilde{k}} + e^{-2\tilde{k}} + \cdots + e^{-m\tilde{k}})$$

where $\tilde{k} = 2k$.

$$\sum_{j=1}^{m+1} w_j^2 = \frac{1 - e^{-2(m+1)k}}{1 - e^{-2k}}(1 - e^{-k})^2$$
$$= \frac{1 - e^{-k}}{1 + e^{-k}}(1 - e^{-2(m+1)k}) \tag{A2}$$

As from Equation (3), we have

$$e^{-k(m+1)} \approx \frac{\varepsilon}{C_{\max}},$$

$$\sum_{j=1}^{m+1} w_j^2 = \frac{1 - e^{-k}}{1 + e^{-k}}\left(1 - \frac{\varepsilon^2}{C_{\max}^2}\right) = D/C_{\max}^2$$

Substituting (A2) in (A1), we obtain

$$\Pr(q_t \geq \eta) \leq 2 \exp\left(\frac{-\eta^2}{D}\right) \tag{A3}$$

Consider two events A and B. Let $\bar{B}$ be as an event complement to event B such that $\Pr(\bar{B}) = 1 - \Pr(B)$. Further let $A \cap \bar{B} = \varnothing$, i.e., there is no overlap between the two events. Let $0 < l < u < 1$ such that $\Pr(A) \geq 1 - u$, $\Pr(B) \geq 1 - l$ and that $\Pr(B) \geq \Pr(A)$. We assume that $\Pr(B) - \Pr(A) \geq u - l$. Then

$$\Pr(A \cup \bar{B}) = \Pr(A) + \Pr(\bar{B}) = \Pr(A) + 1 - \Pr(B)$$

$$\Rightarrow 1 - \Pr(A \cup \bar{B}) = \Pr(B) - \Pr(A) \geq u - l$$

From the above, we can then say with A or $\bar{B}$ *never* happen with probability of at least $u - l$.

Now let event A be $q_t < \eta_A$ such that

$$\Pr(q_t < \eta_A) > 1 - u \tag{A4}$$

Similarly let $\bar{B}$ be $q_t > \eta_B$ such that

$$\Pr(q_t > \eta_A) \leq l \tag{A5}$$

We now use (A3) to express u and l in terms of η_B and η_A in (A5) and (A4),

$$2 \exp\left(\frac{-\eta_B^2}{D}\right) = l \Rightarrow \eta_B = \sqrt{D \ln\left(\frac{2}{l}\right)}$$

$$2 \exp\left(\frac{-\eta_A^2}{D}\right) = u \Rightarrow \eta_A = \sqrt{D \ln\left(\frac{2}{l}\right)}$$

and state with probability of at least $u - l$ that q_t lies in the interval $[\sqrt{D \ln (2/u)}, \sqrt{D \ln (2/l)}]$.

Thus with probability of at least $u - l$, for $l < u$, we have

$$r = |b - a| \leq \sqrt{D}\left[\sqrt{\ln\frac{2}{l}} - \sqrt{\ln\frac{2}{u}}\right] \qquad \square$$

Proof of Lemma 2:   Let $q_t = |Q_m(t) - E(Q_m(t))|$, $S = \sum_{t=1}^{N} q_t$. We now apply a modified result of Theorem 1 of Goldstein (1975) (given in the Appendix B) for a convex

function $f(X_t) = |X_t| \overset{\text{def}}{=} q_t$ (subscript i in appendix replaced by t) with

1. $X_t = Q_m(t) - EQ_m(t),$

   $\mu_t = \mu = E(Q_m(t) - EQ_m(t)) = 0.$

2. $r_t = r \ \forall t,$ where

   $r = |b - a| \leq 2\sqrt{h}[\sqrt{\ln 2/l} - \sqrt{\ln 2/u}],$

   $h = (1 - e^{-k})/(1 + e^{-k})$ obtained from D Lemma 1 and given that $C_{\max} = 1$, $\varepsilon \ll 1$,

3. $\zeta_t = \zeta = \dfrac{-a}{b-a}, \rho_t = \rho = \dfrac{-f(a)}{f(b) - f(a)} \ \forall t,$

where, $a = \min(Q_m(t) - EQ_m(t))$, $b = \max(Q_m(t) - EQ_m(t))$ $f(a) = |\min(Q_m(t) - EQ_m(t))|$, $f(b) = |\max(Q_m(t) - EQ_m(t))|$ and min and max are with respect to input forcing such that these lower and upper bounds are never violated for any input forcing $u(t)$.

Further we note that

$a < 0$, $b > 0$, which implies that

$f(a) = |a| = -a,$
$f(b) = |b| = b$

Thus,

$b - a = |b| + |a| \geq |b| - |a| = f(b) - f(a),$ which implies that

$$\zeta - \rho = \frac{-a}{b-a} - \frac{-f(a)}{f(b) - f(a)} = \frac{f(a)}{b-a} + \frac{f(a)}{f(b) - f(a)}$$
$$\leq \frac{2f(a)}{f(b) - f(a)} = \frac{2f(a)}{r}$$
(A5)

Then,

$$\sum_{t=1}^{N} (\zeta_t - \rho_t) r_t = Nr(\zeta - \rho) < 2Nf(a)$$
$$= 2N|\min(Q_m(t) - EQ_m(t))|$$

Further as $\min Q_m(t) = 0$,

$|\min(Q_m(t) - EQ_m(t))| = |\min Q_m(t) - EQ_m(t)|$
$= |-EQ_m(t)| = |EQ_m(t)|$

Also, $\varepsilon/C_{\max} \ll 1 \Rightarrow Q_m(t) \to Q(t).$

Finally note that

$$EQ(t) = d(k) \sum_{\tau=1}^{t} e^{-k(t-\tau)} Eu(\tau) = E(u)d(k) \sum_{\tau=-\infty}^{t} e^{-k(t-\tau)} = Eu(\tau)$$

Thus,

$$\sum_{t=1}^{N} (\zeta_t - \rho_t) r_t = 2Eu(t)N.$$

Using the conclusions of (1)–(3) and applying Theorem 1 of Goldstein (1975) we have,

$$\Pr\left(\sum_{t=1}^{N} \left|\frac{Q_m(\mathbf{u}; t) - E[Q_m(\mathbf{u}; t)]}{N}\right| > \gamma\right)$$
$$\leq \exp\left(-\frac{(\gamma - 2Eu)^2 N}{2h\left[\sqrt{\ln(2/l)} - \sqrt{\ln(2/u)}\right]^2}\right)$$

and for $\gamma \gg E(u(t))$,

$$\Pr\left(\sum_{t=1}^{N} \left|\frac{Q_m(\mathbf{u}; t) - E[Q_m(\mathbf{u}; t)]}{N}\right| > \gamma\right)$$
$$\leq \exp\left(-\frac{\gamma^2 N}{2h\left[\sqrt{\ln(2/l)} - \sqrt{\ln(2/u)}\right]^2}\right)$$

Note that we here implicitly assumed that $|Q_m(\mathbf{u}; t) - E[Q_m(\mathbf{u}; t)]|$ is independently and identically distributed which may lead to an inaccurate upper bound on convergence.

Proof of Corollary 2: For any $k < N$, the Euclidean norm of any column of corresponding $\mathbf{V}$ is no larger than

$$\sqrt{h} = \sqrt{d^2(k)(1 + e^{-\tilde{k}} + e^{-2\tilde{k}} + \cdots + e^{-m\tilde{k}})}.$$

The span can therefore be circumscribed by N-sphere of radius $\sqrt{h}/2$. Thus, volume of the span defined by $\mathbf{V}$ is always bounded by the volume of N-sphere $V(k) \propto (h/4)^{N/2}$.

Proof of Lemma 3:   As

$$||Q_m - EQ_m| - |Q - EQ|| \leq 2\varepsilon$$

$$\frac{\sum_{t=1}^{N} |(Q(\mathbf{u};t) - E[Q(\mathbf{u};t)])|}{N} > \gamma + 2\varepsilon$$

$$\Rightarrow \frac{\sum_{t=1}^{N} |(Q_m(\mathbf{u};t) - E[Q_m(\mathbf{u};t)])|}{N} > \gamma$$

The inequality in the lemma then follows.   □

Proof of Theorem 1:   From Assumption B,

$$|\ell_k(y, \mathbf{u};t) - E[\ell_k(y, \mathbf{u};t)]| \leq \eta |Q(\mathbf{u};t) - E[Q(\mathbf{u};t)]|.$$

By triangle inequality we have

$$\left| \frac{\sum_{t=1}^{N} \ell_k(y, \mathbf{u};t)}{N} - \frac{\sum_{t=1}^{N} E[\ell_k(y, \mathbf{u};t)]}{N} \right|$$

$$\leq \sum_{t=1}^{N} \left| \frac{\ell_k(y, \mathbf{u};t) - E[\ell_k(y, \mathbf{u};t)]}{N} \right|$$

$$\leq \eta \sum_{t=1}^{N} \frac{|Q(\mathbf{u};t) - E[Q(\mathbf{u};t)]|}{N}$$

Thus,

$$\eta\theta \leq \left| \frac{\sum_{t=1}^{N} \ell_k(y, \mathbf{u};t)}{N} - \frac{\sum_{t=1}^{N} E[\ell_k(y, \mathbf{u};t)]}{N} \right|$$

$$\Rightarrow \eta\theta \leq \eta \sum_{t=1}^{N} \frac{|Q(\mathbf{u};t) - E[Q(\mathbf{u};t)]|}{N}$$

Finally we note that if there are two events *A* and *B* such that $A \Rightarrow B$ then $\Pr(A) \leq \Pr(B)$. This is because whenever event *A* occurs, *B* occurs. However whenever *B* occurs, *A* need not occur ($\because B \nRightarrow A$). Thus probability of occurrence of *A* is never larger than probability of occurrence of *B*. Let event *A* be

$$\eta\theta \leq \left| \frac{\sum_{t=1}^{N} \ell_k(y, \mathbf{u};t)}{N} - \frac{\sum_{t=1}^{N} E[\ell_k(y, \mathbf{u};t)]}{N} \right|$$

and event *B* be

$$\eta\theta \leq \eta \sum_{t=1}^{N} \frac{|Q(\mathbf{u};t) - E[Q(\mathbf{u};t)]|}{N}$$

It then follows that,

$$\Pr\left(\left| \frac{\sum_{t=1}^{N} \ell_k(y, \mathbf{u};t)}{N} - \frac{\sum_{t=1}^{N} E[\ell_k(y, \mathbf{u};t)]}{N} \right| > \eta\theta \right)$$

$$\leq \Pr\left(\frac{\sum_{t=1}^{N} |Q(\mathbf{u};t) - E[Q(\mathbf{u};t)]|}{N} > \theta \right)$$

Finally for $\delta = \eta\theta$, and applying Lemma 2 with $\gamma = \delta/\eta - 2\varepsilon$ with $\varepsilon \ll 1$,

$$\Pr(|\xi_Z(k) - E[\xi_Z(k)]| > \delta) \leq \exp\left(-\frac{[\delta/\eta]^2 N}{2h[\sqrt{\ln 2/l} - \sqrt{\ln 2/u}]^2} \right)$$

Proof of Lemma 4:
Let $\hat{Q} = (\sum_{t=1}^{N} |Q_m(\mathbf{u};t) - E[Q_m(\mathbf{u};t)]|)/N$, where

$$Q_m(\mathbf{u};t) = \sum_{ij} Q_{ij,m_{ij}}(\mathbf{u};t).$$

As

$$\gamma \leq \hat{Q} \leq \sum_{ij} \sum_{t=1}^{N} \left| \frac{Q_{ij,m_{ij}}(\mathbf{u};t) - E[Q_{ij,m_{ij}}(\mathbf{u};t)]}{N} \right|$$

$$\frac{\sum_{t=1}^{N} |Q_{ij,m_{ij}}(\mathbf{u};t) - E[Q_{ijm_{ij}}(\mathbf{u};t)]|}{N} \geq \frac{\gamma}{R}$$

for at least one *ij*. This is so because its complement in the following cannot hold if $\gamma \leq \hat{Q}$ holds. This is shown in the following, if

$$\frac{\sum_{t=1}^{N} \left| Q_{ij,m_{ij}}(\mathbf{u};t) - E[Q_{ij,m_{ij}}(\mathbf{u};t)] \right|}{N} \geq \frac{\gamma}{R} \quad \text{for all } ij$$

$$\Rightarrow \sum_{ij} \frac{\sum_{t=1}^{N} |Q_{ij,m_{ij}}(\mathbf{u};t) - E[Q_{ij,m_{ij}}(\mathbf{u};t)]|}{N} \geq \sum_{ij} \frac{\gamma}{R} = \gamma$$

But,

$$\sum_{ij} \frac{\sum_{t=1}^{N} |Q_{ij,m_{ij}}(\mathbf{u};t) - E[Q_{ij,m_{ij}}(\mathbf{u};t)]|}{N} \geq \hat{Q} \geq \gamma$$

a contradiction.

Let E be the event that $\gamma \leq \hat{Q}$ holds and F_{ij} be the event that

$$\frac{\sum_{t=1}^{N} |Q_{ij,m_{ij}}(\mathbf{u};t) - E[Q_{ij,m_{ij}}(\mathbf{u};t)]|}{N} \geq \frac{\gamma}{R}$$

holds.

Then,

$$E \Rightarrow \bigcup_{ij} F_{ij}$$

or,

$$\Pr(E) \leq \Pr\left(\bigcup_{ij} F_{ij}\right)$$

Further note that

$$\Pr\left(\bigcup_{ij} F_{ij}\right) = \sum_{ij} \Pr(F_{ij}) - \Pr\left(\bigcap_{ij} F_{ij}\right) \leq \sum_{ij} \Pr(F_{ij})$$

Therefore, we have

$\Pr(E) \leq \Sigma_{ij} \Pr(F_{ij})$, or

$$\Pr(\hat{Q} \geq \gamma) \leq \sum_{ij} \Pr\left(\frac{\sum_{t=1}^{N} |Q_{ij,m_{ij}}(\mathbf{u};t) - E[Q_{ij,m_{ij}}(\mathbf{u};t)]|}{N} \geq \frac{\gamma}{R}\right).$$

From Lemma 2,

$$\Pr\left(\frac{\sum_{t=1}^{N} |Q_{ij,m_{ij}}(\mathbf{u};t) - E[Q_{ij,m_{ij}}(\mathbf{u};t)]|}{N} \geq \frac{\gamma}{R}\right)$$

$$\leq \exp\left(-\frac{\gamma^2 N}{2R^2 C_{\max}^2 h_{ij}[\sqrt{\ln 2/l} - \sqrt{\ln 2/u}]^2}\right)$$

Finally given that

$$\frac{\sum_{t=1}^{N} |Q(\mathbf{u};t) - Q_m(\mathbf{u};t)|}{N}$$

$$\leq \sum_{ij} \frac{\sum_{t=1}^{N} |Q_{ij}(\mathbf{u};t) - Q_{ij,m_{ij}}(\mathbf{u};t)|}{N} \leq \sum_{ij} \varepsilon/R = \varepsilon$$

and following Lemma 3, we have

$$\Pr\left(\frac{\sum_{t=1}^{N} |Q(\mathbf{u};t) - E[Q(\mathbf{u};t)]|}{N} > \gamma + 2\varepsilon\right)$$

$$\leq \Pr\left(\frac{\sum_{t=1}^{N} |Q_m(\mathbf{u};t) - E[Q_m(\mathbf{u};t)]|}{N} > \gamma\right)$$

$$\leq \sum_{ij} \exp\left(-\gamma^2 N/2R^2 C_{\max}^2 h_{ij}\left[\sqrt{\ln\frac{2}{l}} - \sqrt{\ln\frac{2}{u}}\right]^2\right) \qquad \square$$

APPENDIX B

Theorem 1 of Goldstein (1975):   If $X_1, X_2, ..., X_n$ are independent random variables such that $a_i \leq X_i \leq b_i, i = 1, ..., n$, and f is a continuous convex function then, if $\delta > \max_{1 \leq i \leq n} |(\rho_i - \zeta_i)r_i|$,

$$P\{S \geq n\delta\} \leq \exp\left\{-2\frac{(n\delta - v_n)^2}{w_n}\right\}$$

where $w_n = \sum_{i=1}^{n} r_i^2$, $v_n = \sum_{i=1}^{n} (\zeta_i - \rho_i)r_i$, $S = \sum_{i=1}^{n} f(X_i)$, $r_i = f(b_i) - f(a_i)$ with $\zeta_i = (\mu_i - a_i)/(b_i - a_i)$, $\rho_i = (-f(a_i)/f(b_i) - f(a_i))$ and $\mu_i = EX_i$. $\qquad \square$