WATER RESOURCES RESEARCH, VOL. 49, 2253–2273, doi:10.1002/wrcr.20161, 2013

Estimating epistemic and aleatory uncertainties during hydrologic modeling: An information theoretic approach

Wei Gong,[1] Hoshin V. Gupta,[2] Dawen Yang,[1] Kumar Sricharan,[3] and Alfred O. Hero III[3]

Received 19 December 2012; revised 22 February 2013; accepted 22 February 2013; published 30 April 2013

[1] With growing interest in understanding the magnitudes and sources of uncertainty in hydrological modeling, the difficult problem of characterizing model structure adequacy is now attracting considerable attention. Here, we examine this problem via a model-structure-independent approach based in *information theory*. In particular, we (a) discuss how to assess and compute the information content in multivariate hydrological data, (b) present practical methods for quantifying the uncertainty and shared information in data while accounting for heteroscedasticity, (c) show how these tools can be used to estimate the best achievable predictive performance of a model (for a system given the available data), and (d) show how model adequacy can be characterized in terms of the magnitude and nature of its *aleatory uncertainty* that cannot be diminished (and is resolvable only up to specification of its density), and its *epistemic uncertainty* that can, in principle, be suitably resolved by improving the model. An illustrative modeling example is provided using catchment-scale data from three river basins, the Leaf and Chunky River basins in the United States and the Chuzhou basin in China. Our analysis shows that the aleatory uncertainty associated with making catchment simulations using this data set is significant (~50%). Further, estimated epistemic uncertainties of the HyMod, SAC-SMA, and Xinanjiang model hypotheses indicate that considerable room for model structural improvements remain.

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

1. Introduction

1.1. Background

[2] The extant literature discusses three sources of hydrological model uncertainty: data uncertainty, parameter uncertainty, and model structure uncertainty. With the advent of physically based distributed models, the problem of estimating predictive uncertainty soon came to the forefront [e.g., *Beven*, 1989; *Beven and Binley*, 1992]. Early research focused primarily on parameter and output data uncertainties while ignoring other sources [e.g., *Sorooshian and Dracup*, 1980; *Kuczera*, 1982; *Beven and Binley*, 1992; *Freer et al*., 1996; *Gupta et al*., 1998; *Duan and Schaake*, 2002; *Vrugt et al*., 2003]. With further developments in modeling, interest in the other sources of uncertainty has grown [*Vrugt et al*., 2005; *Ewen et al*., 2006], leading to explorations of input uncertainties due to errors in the forcing data [*Kavetski et al*., 2006a, 2006b; *Vrugt et al*., 2008; *Li et al*., 2012] and of the inadequacies of model structure [*Butts et al*., 2004; *Clark et al*., 2008; *Fenicia et al*., 2008]. Recently, formal Bayesian approaches have been proposed for the quantification of model uncertainty [*Bates and Campbell*, 2001; *Thiemann et al*., 2001; *Kaheil et al*., 2006], and, particularly, for the explicit integration of multiple sources of uncertainty [*Kavetski et al*., 2006a, 2006b; *Ajami et al*., 2007; *Marshall et al*., 2007; *Zhang et al*., 2009].

[3] Although data and parameter uncertainties can be characterized relatively easily in terms of probability distributions, bounds, or limits, the treatment of structure uncertainty requires special attention [*Gupta et al*., 2012]. The most popular method to date is probably Bayesian multimodel averaging (BMA) [*Hoeting et al*., 1999; *Neuman*, 2003a, 2003b], which uses multiple structures to characterize the uncertainty in our knowledge of the mechanics of underlying hydrological processes [*Butts et al*., 2004; *Georgakakos et al*., 2004; *Ajami et al*., 2006; *Duan et al*., 2007]. Other methods seek not just to characterize model structure uncertainty but to also improve the structure of the hydrological model; examples include time-variable parameter methods such as the state-dependent parameter (SDP) estimation method [*Young et al*., 2001; *Young and Ratto*, 2009], the recursive prediction error (RPE) approach [*Lin and Beck*, 2007], and the time-dependent parameters approach [*Reichert and Mieleitner*, 2009]. Recently, a new method called the Bayesian estimation of structure (BESt) approach [*Bulygina and Gupta*, 2009, 2010, 2011] has been proposed to resolve the underlying structure of the model via data assimilation conducted on the raw data.

[1]State Key Laboratory of Hydroscience and Engineering, Tsinghua University, Beijing, China.
[2]Department of Hydrology and Water Resources, The University of Arizona, Tucson, Arizona, USA.
[3]Department of Electrical Engineering and Computer Science (EECS), University of Michigan, Ann Arbor, Michigan, USA.

Corresponding author: W. Gong, State Key Laboratory of Hydroscience and Engineering, Tsinghua University, 100084 Beijing, China. (gongwei2012@bnu.edu.cn)

0043-1397/13/10.1002/wrcr.20161

[4] One major problem faced by all such methods is lack of knowledge about the best achievable performance (BAP) of a model for a particular system given the available data (note that performance can only be meaningfully and precisely assessed in the context of *a specific system and data set*). Previous investigations into model structure uncertainty have all been based on the adoption of one or more prior structural hypotheses. However, since the "true" model structure is unknown (and the concept itself is arguably suspect—see discussion by *Gupta et al*. [2012]), any uncertainty analysis based on prior assumptions regarding the model structural hypotheses lacks the specification of a proper reference frame. To address this issue, we use Shannon's definition of *information* to develop a method that can potentially characterize the inherent degree of explanatory power present in a given data set. This provides a measure of BAP for a given system and data set, which provides a meaningful benchmark against which the power of a model structure hypothesis can be assessed, even if the structural form of the "best" possible model remains unknown.

[5] Once the BAP of a model (*for a system given the available data*) can be characterized, we can then address two distinct categories of model uncertainty, (1) *aleatory uncertainties* (AU) that are only resolvable up to the specification of their probability densities and (2) *epistemic uncertainties* (EU) that can (in principle) be suitably resolved by a structurally adequate model (one that properly expresses the nonaleatory relationships present in the given data set). As such, characterization of the BAP establishes the size and nature of aleatory uncertainties associated with the currently available data set, which cannot be reduced by any possible model hypothesis regarding relationships expressed by the existing data (it is of course generally difficult to know whether new/additional data might help to reduce these aleatory uncertainties further). Meanwhile the difference between the achieved performance of a specific model hypothesis and the BAP (by any possible model hypothesis) establishes the size and nature of epistemic uncertainty that can be reduced by improvements to the existing model structure. Based on these concepts, random observational errors associated with the input/output data can be categorized into AU (along with any inherent stochasticity in the explanatory relationships connecting the input and output data), while parameter uncertainty and model structural inadequacies (including systematic errors in the observational data) can be categorized into EU.

1.2. Goal, Objectives, and Scope

[6] The goal of this paper is to examine how we can quantify the extent to which information about the catchment-scale rainfall-runoff process is expressed by a hydrological data set, and to assess how much of that information is actually expressed by a hydrological model used to simulate the behavior of that system. Similar to other methods of uncertainty analysis, our approach provides estimates of the distribution of system output simulation errors. However, it goes further by also quantifying, explicitly, the amount of information about the system outputs that is expressed by the available catchment data and by the model—in other words, how much of the uncertainty in the system output can be reduced by use of the information contained in the available catchment data and by the use of a model structural hypothesis that relates system input to outputs via a state-space representation.

[7] Our primary objectives, therefore, are to evaluate: (a) How much information is required to generate "reasonably accurate" simulations of the system outputs (e.g., runoff) (we call this the "information required"); (b) How much information is contained in the available system input-state-output data ("information available"); and (c) To what degree the available information is correctly expressed by a given model hypothesis ("information expressed"). We can then compute two kinds of uncertainty (aleatory and epistemic) from these three kinds of information as follows:

[8] (1) AU = information required – information available.

[9] (2) EU = information available – information expressed.

[10] Note that AU expresses the information gap caused by random errors in the observation data or by inherent randomness in the hydrological processes linking the system inputs, state variables, and outputs; this uncertainty cannot be removed by improving the model. Similarly, EU expresses the information gap caused by model structural inadequacies; this uncertainty can potentially be reduced by improvements to the model hypothesis, possibly including adjustment factors to account for systematic bias in the observational data (as was done by *Kavetski et al*. [2006a, 2006b] and *Vrugt et al*. [2008]).

[11] Concepts and tools from information theory have been used in hydrology and water resources for decades [*Singh*, 1997, 2000]. Close to the focus of this paper, information theory has been applied for model evaluation and uncertainty analysis as far back as the 1970s [e.g., *Amorocho and Espildor*, 1973; *Chapman*, 1986; *Weijs et al*., 2010a, 2010b; *Weijs and van de Giesen*, 2011; *Abebe and Price*, 2003; *Pokhrel and Gupta*, 2010]. However, none of the studies have provided a comprehensive method for qualifying the overall information content of a data set (the information available as mentioned earlier). One possible reason for this lack may be that whereas the data sets of contemporary interest tend to be of high dimension, conventional methods available for estimating "information content" are limited to a variable dimension not exceeding three. To mitigate this problem, we draw upon some very recent advances in information theory to develop an appropriate methodology that is suitable to our goals. With this method, we are able to meet the primary objectives mentioned earlier.

[12] The paper is organized as follows. We first lay out principles of *information theory* that are relevant to the problem of quantifying BAP of a model for a system given the available data and to the problem of characterizing model structural uncertainty. Section 3 presents the computational methods necessary for effective and efficient estimation of the information metrics discussed in section 2. In section 4, we employ a case study to demonstrate how aleatory uncertainty in an available data set and epistemic uncertainty of a model hypothesis can be quantitatively estimated, providing information useful for establishing limits to achievable model performance and for assessing model structural adequacy. Finally (in section 5), we discuss some

interesting implications of our results and make suggestions for future work.

2. Concepts of Information Theory and Their Relevance to Hydrological Modeling

2.1. Principles of Information Entropy and Mutual Information

[13] Entropy is the basic principle of information theory proposed by *Shannon* [1948]. Shannon's definition of information entropy for a variable that takes on a discrete set of values is

$$H_{\text{discrete}}(X) = -\sum_{x \in X} p(x) \log_b p(x), \tag{1}$$

where the negative sign ensures that $H(X)$ is positive (since $\log_b p(x)$ is negative). This is the only possible definition that ensures that H satisfies three properties:

[14] (1) H is continuous in $p(.)$.

[15] (2) If all elements $p(.)$ are equally probable, then H is a monotonic increasing function of the number of elements.

[16] (3) H is additive [*Shannon*, 1948].

[17] The base used for the logarithm is arbitrary. If $b = 2$, then $H(X)$ is measured in bits (common in information theory), whereas if $b = e$ it is measured in nats (common in thermodynamics). Any choice is acceptable since a change of base only changes the result by a constant; in this paper, we use base e. According to this definition, H_{discrete} can vary on $[0, \infty]$.

[18] Similarly, the definition of information entropy for a variable whose values are continuous is

$$H_{\text{continuous}}(X) = -\int_X f(x) \log f(x) \mathrm{d}x, \tag{2}$$

where $f(x)$ is the probability density function (pdf) of X, and the integral over the support set of X is assumed to exist. However, it is common for measurements of continuous variables to be expressed in discrete fashion due to quantization during the observational process. When a continuous variable is discretized using bin width $\mathbf{w}$, the relation between the resulting discrete entropy and the original continuous entropy can be expressed via the following equation, provided that $\mathbf{w}$ is not too large:

$$H_{\text{discrete}}(X) + \log(\mathbf{w}) \to H_{\text{continuous}}(X), \text{as } \mathbf{w} \to 0. \tag{3}$$

Note that the difference between the discrete and continuous versions of entropy is expressed by the bin-width factor $\log(\mathbf{w})$. Further, note that since $\log(\mathbf{w})$ is negative for $\mathbf{w} < 1$, $H_{\text{continuous}}$ can vary on $[\pm\infty]$.

[19] Generally speaking, information entropy quantifies the "extension" of a random variable. A variable having a larger range of extension, corresponding to larger uncertainty, will have larger entropy H, and a greater amount of information will required to characterize it. So, information entropy provides an alternative characterization of uncertainty to that provided by the commonly used standard deviation statistic σ, or the (say 95%) confidence interval. Further, it provides a more accurate characterization than σ, since the latter depends only on the second moment, whereas information entropy takes into account the effects of higher order moments. Note that some of the popularity of the Gaussian distribution may be explained by the fact that, of all continuous distributions having identical variance, the Gaussian distribution has the largest entropy (because its higher order moments are all zero), making it a relatively robust choice when the actual shape of the distribution is unknown; for more about basic concepts of entropy, please refer to "Elements of Information Theory" by *Cover and Thomas* [2006].

[20] Whereas information entropy characterizes the uncertainty in a variable, we are typically interested in knowing what one (uncertain) variable X can tell us about another (uncertain) variable Y; in other words, how much information is shared between the two; or in still other words, *how much of the uncertainty about Y can be reduced by knowing X*. Whereas, in the analysis of data, it is common to characterize the shared information between the two variables by means of the linear correlation statistic ρ_{XY}, such a characterization is only accurate when the relationship is linear. For more general cases, the shared information can be characterized using *mutual information* (MI), which quantifies the divergence between the joint pdf $\mathrm{p}(x,y)$ and the product $\mathrm{p}(x).\mathrm{p}(y)$ of the independent pdfs. For continuous variables, MI is defined as

$$I(X;Y) = \int \int f_{x,y}(x,y) \log \frac{f_{x,y}(x,y)}{f_x(x) f_y(y)} \mathrm{d}x \mathrm{d}y, \tag{4}$$

where $f_x(x)$ and $f_y(y)$ are marginals of the joint pdf $f_{x,y}(x,y)$. Naturally, if the joint density factors exactly into the product of its marginals, there is no shared information between the two variables and hence $I(X;Y) = 0$. More generally, the necessary and sufficient condition for independence of X and Y is that $I(X;Y) = 0$. Note that we can similarly define "discrete" MI, whose value is equivalent to continuous MI for sufficiently small bin width $\mathbf{w}$ [*Cover and Thomas*, 2006, equation (8.50)]. According to this definition, MI can vary on $[0, \infty]$.

[21] An intuitive feel for this can be obtained from the case where X and Y are jointly Gaussian with correlation coefficient ρ_{XY}, for which $I(X;Y) = \frac{1}{2} \log\left(1 - \rho_{XY}^2\right)$ [*Cover and Thomas*, 2006, equation (8.56), p. 252]. So, as $\rho_{XY} \to \pm 1$ we have $I(X;Y) \to \pm\infty$, meaning that the relationship offers infinitely large amounts of information and allows one to determine the value of one variable from the other *with an infinitely small level of precision*. Therefore, the R-statistic proposed by *Granger and Lin* [1994]:

$$R(X;Y) = \sqrt{1 - \exp(-2I(X;Y))} \tag{5}$$

is equivalent to the correlation coefficient ρ_{XY} when X and Y are bivariate normal, but is also meaningful in nonlinear and/or non-Gaussian cases. To obtain an intuitive feel, note that (for jointly Gaussian data) $|\rho_{XY}|$ values of [0.5, 0.8, 0.9, 0.95, and 0.99] correspond to $I(X;Y)$ values of [0.14, 0.51, 0.83, 1.16, and 1.96]. So, of practical importance is the fact that $I(X;Y) < 2$ as long as $|\rho_{XY}| < 0.9907$. Note, also, that in contrast with ρ_{XY}, the value of $I(X;Y)$ *does not*

depend on one-to-one nonlinear transformations of X and Y [*Granger and Lin*, 1994].

[22] Conveniently, because of the additive property of information, the value for MI can be computed from knowledge of the individual information entropies and the *joint entropies* of each variable. Because the log transformation converts multiplication and division of variables to addition and subtraction of their logarithms, the MI shared by two variables X and Y can be expressed as

$$I(X;Y) = H(X) + H(Y) - H(X,Y), \tag{6}$$

where $H(X,Y)$ is the joint information entropy given by

$$H(X,Y) = -\int\int f_{x,y}(x,y)\log f_{x,y}(x,y)\mathrm{d}x\mathrm{d}y\,. \tag{7}$$

Further, the extent to which uncertainty in knowledge of Y is reduced by knowledge of X is characterized by the conditional entropy relationship:

$$H(Y|X) = H(Y) - I(X;Y), \tag{8}$$

which expresses the fact that the amount of uncertainty reduction is precisely the MI $I(X;Y)$ shared by the two variables.

[23] In general, of course, it is the *information shared by multiple variables* that is of interest in hydrology, in which case we work with the multivariate joint density. In this case, by extension to equation (6) the *multivariate* MI shared by several variables is expressed by the relationship:

$$\begin{aligned} I(X_1,X_2,...,X_m;Y) = {} & H(X_1,X_2,...,X_m) + H(Y) \\ & - H(X_1,X_2,...,X_m,Y), \end{aligned} \tag{9}$$

which denotes the information that the several variables X_1, X_2, ..., X_m contain about variable Y, and where it is not required that the variables X_1, X_2, ..., X_m be mutually independent. Similar to equation (8) above, we now have

$$H(Y|X_1,X_2,...,X_m) = H(Y) - I(X_1,X_2,...,X_m;Y), \tag{10}$$

which expresses how knowledge of the several variables X_1, X_2,...,X_m helps to reduce uncertainty in knowledge of Y.

2.2. Characterizing Epistemic and Aleatory Uncertainty in Hydrological Modeling

[24] These concepts can now be used to express the different kinds of uncertainty in hydrological modeling. For simplicity, we consider first a *simple* system in which the values of system outputs Y are to be explained (estimated, simulated) via knowledge of available values of system inputs X.

[25] For any system for which we have input and output data, the true values of the system inputs X and outputs Y, the size and type of observation error associated with each, and the true stochastic nature of the relationships between X and Y are unknown. While we may have some conceptual ideas regarding system structure, all that is quantitatively and explicitly available to us are the observed values X_{obs} and Y_{obs}. As such, all the information about this system that is expressed by the available input/output data is contained in the joint and marginal distributions of X and Y. From the joint distribution, we can evaluate the information about output Y contained in the input X. Further, from the marginal distribution of Y we can evaluate the amount of information required to specify Y to within the precision defined by the measurement error process.

[26] The goal of uncertainty analysis is to establish uncertainty bounds on an estimate. More generally, we are interested in the distribution of residual errors between the model simulated values and the observations. In the framework of information theory, once all of the MI between the inputs X_{obs} and outputs Y_{obs} has been exploited, the uncertainty remaining in knowledge of Y_{obs} is quantified by the conditional information entropy $H(Y_{\mathrm{obs}}|X_{\mathrm{obs}})$, computation of which requires knowledge of the conditional distribution $f_{Y|X}(y|x)$ of the outputs Y_{obs} given the value of X_{obs}. Alternatively, equation (8) can be used to compute $H(Y_{\mathrm{obs}}|X_{\mathrm{obs}})$ from knowledge of the prior (unconditional) information entropy $H(Y_{\mathrm{obs}})$ and the mutual (shared) information entropy $I(X_{\mathrm{obs}};Y_{\mathrm{obs}})$. Figure 1 illustrates these relationships in probability space, showing the dependence between the prior distribution of Y_{obs}, the joint distribution of X_{obs} and Y_{obs}, and the posterior distribution of Y_{obs} given X_{obs}. If only the frequency distribution of the system outputs Y is available (and hence the relationship between X_{obs} and Y_{obs} is unknown), the prior uncertainty in knowledge of Y_{obs} is characterized by $f_Y(y)$, and its magnitude can be quantified in terms of its information entropy $H(Y_{\mathrm{obs}}) = -\int f_Y(y)\ln f_Y(y)\mathrm{d}x$. If the inputs X_{obs} corresponding to the outputs Y_{obs} are also known, and as long as some dependence relationship between X_{obs} and Y_{obs} exists, the conditional uncertainty in knowledge of Y_{obs} given X_{obs} is characterized by $f_{Y|X}(y|x)$ as shown in Figure 1, and the corresponding (reduced) information entropy is quantified as $H(Y_{\mathrm{obs}}|X_{\mathrm{obs}}) = -\int f_{Y|X}(y|x)\ln f_{Y|X}(y|x)\mathrm{d}x$. As illustrated

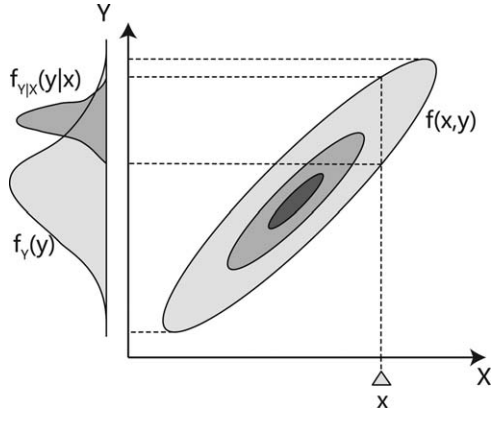

Figure 1. A simple example showing the uncertainty in ability to predict the value of variable Y before and after information regarding inputs X is added. After information about X is introduced, the uncertainty is significantly reduced. The unconditional and conditional information entropies of Y (before and after introducing X) and the MI shared between and X and Y can be measured by the methods presented in this paper.

by Figure 1, the conditional distribution $f_{Y|X}(y|x)$ expresses the amount of uncertainty in knowledge of Y_{obs} that cannot be reduced by exploiting the information contained in the inputs X_{obs}. The magnitude of this uncertainty can be quantified in terms of the information entropy $H(Y_{\text{obs}}|X_{\text{obs}})$, and, since all available explanatory information in the data set has been exploited, we define this remaining uncertainty as the aleatory uncertainty (for this system given this data set):

$$\text{AU} = H(Y_{\text{obs}}|X_{\text{obs}}) = H(Y_{\text{obs}}) - I(Y_{\text{obs}};X_{\text{obs}}). \quad (11)$$

[27] This relationship is further illustrated in Figure 2, where the leftmost bar $H(Y_{\text{obs}})$ indicates, in terms of information entropy, the magnitude of the prior uncertainty in knowledge of Y_{obs} (when X_{obs} is not available), and the middle bar $I(X_{\text{obs}};Y_{\text{obs}})$ indicates the total amount of explanatory information about Y_{obs} contained in the raw data set $\{X_{\text{obs}},Y_{\text{obs}}\}$. Because the explanatory information provided by X_{obs} will, in general, be incomplete, we will have $I(X_{\text{obs}};Y_{\text{obs}}) < \infty$, and it will be impossible to obtain a perfect estimate having infinitely small variation in y of the conditional distribution $f_{Y|X}(y|x)$. Of course, this incompleteness will be due to missing explanatory variables, data error, and/or inherent (unresolvable) randomness in the underlying $X \rightarrow Y$ relationship. In the case of the former, this may be a problem either of conceptual inadequacy (lack of knowledge, or simplifications, regarding variables involved in the relationship), or simply of unavailability of (or inability to obtain) such data. In this paper, we define AU in a practical sense, as being *conditional on the information available in the given data set*, and therefore expressing the total uncertainty associated with all of these sources. Of course, it is possible that additional explanatory variables (new data not yet included in X_{obs}) may help to further reduce this uncertainty, but we do not consider that situation further in this paper.

[28] Knowing the conditional distribution of Y_{obs} given X_{obs}, uncertainty bounds on Y for given X can be easily established. And, for example, in the common situation where the error distribution is assumed to be homoscedastic (the error variance does not change with the magnitude of the variable), or even identical (the type and parameters of the distribution do not change with the magnitude of the variable), the confidence interval can readily be computed. Take the situation where the conditional distribution of Y given X is assumed to follow an exponential power distribution (as used in *Thiemann et al*. [2001] and *Vrugt et al*. [2003]), namely the generalized Gaussian distribution [*Nadarajah*, 2005]. In this case, the conditional information entropy can be written as

$$H(Y|X) = \frac{1}{s} - \ln\left[\frac{s}{2\alpha\Gamma(1/s)}\right], \quad (12)$$

where s is the shape parameter and α is the scale parameter. If the shape parameter $s = 2$, the generalized Gaussian distribution simplifies to the standard normal distribution, and the standard deviation σ can be inversely computed as

$$\begin{aligned} \alpha &= \left[\sqrt{\pi}\exp\left(\frac{1}{2} - H(Y|X)\right)\right]^{-1}, \\ \sigma &= \alpha \cdot \left[\frac{\Gamma(3/s)}{\Gamma(1/s)}\right]^{1/2} \approx 0.7071\alpha. \end{aligned} \quad (13)$$

So by assumption of a suitable distributional form for the conditional uncertainty in Y, the estimated aleatory uncertainty can be converted into a confidence interval as required by traditional uncertainty analysis.

[29] Now consider that a structural hypothesis $Y = f(X)$ regarding the relationships between system inputs and outputs is posed. From the data $\{X_{\text{obs}}\}$ and the assumed model equations, we can compute the model-based estimate $\{Y_{\text{sim}}\}$ of system output $\{Y_{\text{obs}}\}$. Further, the *mutual information* $I(Y_{\text{sim}};Y_{\text{obs}})$ quantifies the total amount of explanatory

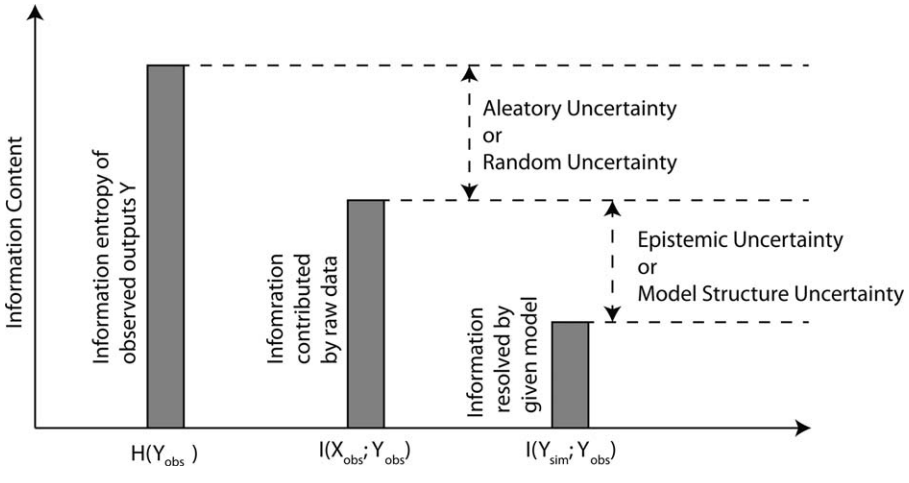

Figure 2. Illustration of the concepts discussed in section 3.2. The leftmost bar indicates the amount of information $H(Y_{\text{obs}})$ required to provide estimates of the system output. The middle bar indicates the amount of such information $I(X_{\text{obs}};Y_{\text{obs}})$ contained in the available system input variables X_{obs}. The rightmost bar indicates the amount of explanatory information $I(Y_{\text{sim}};Y_{\text{obs}})$ contained in the estimates Y_{sim} obtained by use of the model hypothesis $Y = f(X)$. The difference between length of the left and middle bars represents unresolvable AU. The difference between length of the middle and right bars represents potentially resolvable EU.

information about $\{Y_{obs}\}$ contained in the model-based estimate $\{Y_{sim}\}$, and hence in the set of explanatory quantities $\{X_{obs}, f(.)\}$. If the model hypothesis $f(.)$ completely captures the essence of the $X \rightarrow Y$ relationship, then we will obtain $I(Y_{sim};Y_{obs}) = I(X_{obs};Y_{obs})$. In general, however, we can expect that $I(Y_{sim};Y_{obs}) < I(X_{obs};Y_{obs})$ due to limitations in the model hypothesis, as shown by the rightmost bar in Figure 2. The difference:

$$\mathrm{EU} = I(X_{obs}; Y_{obs}) - I(X_{sim}; Y_{obs}) \tag{14}$$

expresses the amount of epistemic uncertainty associated with our system (as specified), that can (in principle) be suitably resolved by improving the model structural hypothesis. This characterizes what we define here as model structure inadequacy, which can be resolved by a combination of improved system conceptualization, and/or improved model structural equations [*Gupta et al*., 2012].

2.3. Data Processing Inequality

[30] Before moving on, one important point should be clearly made, that a model hypothesis in the form of a functional relationship between the available inputs and outputs cannot add information to the system, it only acts to transform available information from one form to another. This immutable fact is expressed by a theorem called the data processing inequality, which demonstrates that no method of data manipulation, whether a model or other form of transformation, can improve upon the amount of MI contained between the available system input and output variables. In other words, this expresses a law of conservation of information; data processing methods do not generate information, and the only contributing source of information is the raw data.

[31] The data processing inequality is defined on an invertible Markov chain. Suppose random variables A, B, C form an invertible Markov chain $A \leftrightarrow B \leftrightarrow C$ where C is conditionally dependent on B but conditionally independent of A (A and C are connected only via B). The data processing inequality can be written as

$$I(A;B) \geq I(A;C). \tag{15}$$

(For the proof and more details, see section 2.8 in *Cover and Thomas* [2006]). In the context of our hydrological modeling problem, suppose that A is the observed output Y_{obs}, B is the observed input X_{obs}, and C is the simulated output Y_{sim} via a model $Y = f(X)$, where $f(.)$ can be either deterministic or stochastic. In this context, the data processing inequality can be expressed as

$$I(X_{obs}; Y_{obs}) \geq I(Y_{sim}; Y_{obs}), \tag{16}$$

which means that the information about the output data expressed by the input data is always larger (or equal) than the information expressed by the same input data after it has been processed through a model.

[32] In the case study presented below, we demonstrate how the quantities $H(Y_{obs})$, $I(X_{obs};Y_{obs})$ and $I(Y_{sim};Y_{obs})$ can be computed in the context of rainfall-runoff modeling of a catchment, and from which the resolvable epistemic and unresolvable aleatory uncertainties (as defined above) can be estimated, thereby providing a quantifiable assessment of model structural inadequacy.

3. Numerical Computation of Information-Based Indices

3.1. Computation of Multivariate Information Entropy

[33] To compute accurate estimates of information offered by observed input and output data, an effective and efficient estimator of *multivariate information entropy* $H(X_1, X_2, \ldots X_m)$ is required. Estimating information entropy for low-dimensional data is quite easy, but for high-dimensional data is recognized to be a difficult problem [*Hero et al*., 2002]. The most widely used methods belong to a class called "plug-in estimators," which first compute (i.e., estimate) the joint pdf of the variables, and then compute *information entropy* and MI by direct application of the definition (equations (1), (2), and (4)). Examples of this approach in hydrological research include the kernel-based estimator [*Sharma*, 2000], bin counting method [*Ruddell and Kumar*, 2009], and the average-shifted histogram method [*Fernando et al*., 2009].

[34] However, because of the curse of dimensionality [*Bellman and Corporation*, 1957], estimating the joint pdf of a high-dimensional data set ($\boldsymbol{m} > 4$) is particularly difficult, and hence plug-in methods are only applicable when there are a small number of variables to be analyzed. A number of "nonplugin" estimators have been proposed to try to mitigate the curse of dimensionality; these include the *Rényi entropy estimator* [*Hero et al*., 2002] and the *Shannon entropy estimator* [*Leonenko et al*., 2008]. While these methods have been shown to work well in many applications, our preliminary investigations found that they may not be suitable for the kinds of hydrological modeling problems of interest here [*Gong*, 2012]. In particular, both of the aforementioned approaches were unable to provide consistent estimates of entropy for the synthetic hydrologic modeling study reported in section 3.2. The problem of how to compute *multivariate information entropy* for data sets such as ours remains, therefore, an open problem.

[35] In this paper, we use an approach based on independent component analysis (ICA) to estimate high-dimensional entropy and demonstrate (section 3.2) that it provides consistent estimates under increasing levels of data measurement uncertainty. *Leiva-Murillo and Artes-Rodriguez* [2007] used ICA to maximize MI for feature extraction. In contrast, our approach decomposes the raw data into independent components, from which high-dimensional entropy can be estimated without having to compute the joint pdf. The method has two major steps: (1) use of ICA to decompose the high-dimensional data into independent variables [*Hyvarinen et al*., 2001] and (2) use of plug-in estimators to compute the Shannon entropy of the decomposed variables. It takes advantage of the chain rule of entropy [*Cover and Thomas*, 2006, Theorem 8.6.2]:

$$\begin{aligned} H(X_1, X_2, \ldots X_m) &= \sum_{i=1}^{m} H(X_i | X_{i-1}, X_{i-2}, \ldots X_1), \\ &\leq \sum_{i=1}^{m} H(X_i), \end{aligned} \tag{17}$$

which states that the multivariate information entropy $H(X_1, X_2, \ldots X_m)$ is equal to the sum of 1-D entropies of the variables under the (if and only if) condition that $(X_1, X_2, \ldots X_m)$ are independent. So, if the original data matrix $\mathbf{X} = [X_1, X_2, \ldots X_m]$ can be converted to a matrix of independent "signals" $\mathbf{S} = [S_1, S_2, \ldots S_m]$ then the entropy of $\mathbf{S}$ can be computed as $H(S) = \sum_{i=1}^{m} H(S_i)$. Further, if $\mathbf{X}$ can be related to $\mathbf{S}$ via a linear transformation $\boldsymbol{X^T} = \boldsymbol{AS^T}$, where $\mathbf{A}$ is the transform matrix, the entropy of the original data matrix $H(\mathbf{X})$ can be recovered using [*Cover and Thomas*, 2006, see corollary of Theorem 8.6.3, eq. (8.71)])

$$H(\mathbf{X}) = H(\mathbf{S}) + \log|\det(\mathbf{A})|. \tag{18}$$

To decompose the original data matrix $\mathbf{X}$ into independent signals $\mathbf{S}$, we can use the Fast ICA (FastICA) method proposed by *Hyvarinen and Oja* [1997]. In the case that $\mathbf{X}$ is jointly Gaussian, the decomposition into independent 1-D Gaussian components can be achieved via principle component analysis (PCA). However, when $\mathbf{X}$ is non-Gaussian, the PCA method is not adequate and the ICA method can be used instead. The FastICA method obtains an estimate of the matrix $\mathbf{A}$ by using a modified version of differential entropy (called negentropy) to measure the degree of non-Gaussianity of the data. Because only an approximate value for negentropy is required, the method is efficient at providing an estimate of $\mathbf{A}$, and hence, the approximate estimate of negentropy leads to a more accurate estimation of high-dimensional entropy, provided that the independent signals are correctly decomposed. For more information about ICA, please refer to *Hyvarinen and Oja* [1997, 2000] and *Hyvarinen et al*. [2001].

[36] In this paper, we implement the method as follows (note that steps 1 and 2 are exactly the same as in PCA):

[37] (1) **Center X:** Transform each column of $\mathbf{X}$ into a zero-mean vector by subtracting its mean.

[38] (2) **Whiten X:** Whiten each column $\mathbf{x}$ of $\mathbf{X}$, to obtain $\tilde{\boldsymbol{X}}$ such that each $\tilde{x}$ is uncorrelated and has unit variance. Eigen value decomposition (EVD) of the covariance matrix $\boldsymbol{XX^T}$ can be used to do whitening.

[39] (3) **Compute non-Gaussianity of $\tilde{\boldsymbol{X}}$:** Compute 1-D negentropy J(x) of each column of $\tilde{\boldsymbol{X}}$ using

$$J(x) \propto [E\{G(x)\} - E\{G(\nu)\}]^2, \tag{19}$$

where E is the expectation operator, $G(.)$ is a nonquadratic function, and ν is a Gaussian variable of zero mean and unit variance. $J(x)$is a nonnegative value and equal to zero if and only if x is Gaussian. If every column of $\tilde{\boldsymbol{X}}$ is Gaussian, stop here. If not, carry out FastICA as indicated in step 4.

[40] (4) **Implement FastICA to find S:** FastICA finds an "optimal" value for $\mathbf{A}$ (and hence $\mathbf{S}$) that maximizes non-Gaussianity, by making use of a fixed-point iteration scheme to perform this optimization effectively and efficiently.

[41] (5) **Compute $\boldsymbol{H(S)}$:** Compute the Shannon entropy $\boldsymbol{H(S_i)}$ of each independent signal using a bin-counting method and sum them to obtain $\boldsymbol{H(S)}$.

[42] (6) **Compute $\boldsymbol{H(X)}$:** Compute the Shannon entropy $\boldsymbol{H(X)}$ by adding $\log|\boldsymbol{A}|$ as shown in equation (19).

[43] In this paper we used the default MATLAB implementation of FastICA (http://research.ics.aalto.fi/ica/fastica/), where an initial $\mathbf{A}$ is chosen randomly and another selection for $\mathbf{A}$ is generated if the fixed-point iteration fails. We found FastICA to be very efficient, requiring only a few seconds of CPU time to decompose a 10-D data set having 10^4 samples (on a desktop PC). To compute 1-D *information entropy*, we used the optimal bin width $\mathbf{w} = 3.73\sigma k^{-1/3}$ proposed by *Scott* [2004], where σ is the standard deviation of each signal $\mathbf{s_i}$ and k is the number of samples. A test to evaluate accuracy and precision of the method under idealized conditions is shown in the appendices.

[44] Of course, the effectiveness of the approach outlined above depends on the ability to transform $\mathbf{X}$ into $\mathbf{S}$ such that the components of $\boldsymbol{S}$ are independent. Here we have assumed that this can be achieved via the linear transformation $\boldsymbol{X^T} = \boldsymbol{AS^T}$. If nonlinearity in the signal interdependence is sufficiently strong, it may not be possible to achieve a sufficient degree of independence in $\boldsymbol{S}$ and the estimate of *multivariate information entropy* will be positively biased. In the studies reported here, we investigate the degree of independence using scatterplots. In future we will extend the method to cases where the interdependence is more strongly nonlinear—for example, via nonlinear preprocessing transformations of data.

3.2. Synthetic Study to Investigate the Precision

[45] To evaluate the precision of our method for computing *mutual information entropy*, we adopted the recommendation of a reviewer and conducted a synthetic hydrological simulation study in which there is no model structural error (so that EU = 0) and where the nature and size of the observational uncertainties contributing to AU can be controlled. We used a parsimonious five-state-variable implementation of the *HyMod* model (Figure 3) [*Boyle*, 2000], in which evapotranspiration losses are computed using a nonlinear soil moisture accounting module [*Moore*, 1985], and two series of parallel linear tanks (three quick and one slow) control the rates of moisture drainage to the catchment outlet. The model, which has been used in many previous investigations [*Bulygina and Gupta*, 2009, 2010, 2011; *Vrugt et al*., 2003; *Wagener et al*., 2003; *Xu et al*., 2010], requires two input values, computes estimates of five state variables, and generates two system outputs. System-specific dynamical behavior of the model is achieved by selecting values for the five model parameters.

[46] The data used are from the humid Leaf River basin (1944 km^2) located north of Collins, Mississippi. Forty water years (October 1948 to September 1988) of daily mean areal precipitation P (mm/d), potential evapotranspiration PET (mm/d), and runoff Q (m^3/d) data are available from the Hydrology Lab of the U.S. National Weather Service. The basin has been widely used in many previous investigations of catchment modeling [*Bulygina and Gupta*, 2009, 2010, 2011; *Sorooshian et al*., 1993; *Thiemann et al*., 2001; *Vrugt et al*., 2003; *Wagener et al*., 2001]. To ensure sufficient data to compute stable entropy estimates (see appendices), we used the first 30 years (10,950 sample points).

[47] The simulation study was conducted as follows. First, assuming the available precipitation P and potential evapotranspiration PET data to be "error-free," we ran a simulation of the HyMod model with selected values for

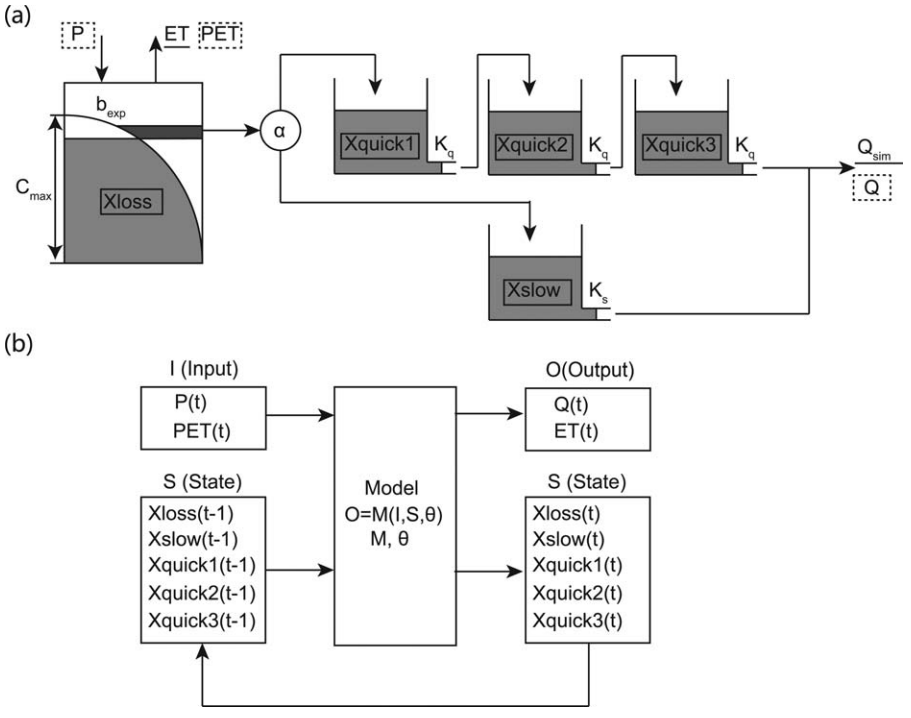

Figure 3. Conceptual representation of the HyMod model, showing (a) sketch of model structure and (b) system diagram. Input variables are precipitation P and potential evapotranspiration PET. State variables are soil moisture content Xloss, quick flow tank content Xquick1, Xquick2, Xquick3, and slow flow tank content Xslow. Model outputs are actual ET and runoff Q_{sim}. Parameters are soil moisture capacity $C_{\max}$, spatial variability index $b_{\exp}$, quick tank and slow tank leakage rates K_q and K_s, and quick-to-slow ratio α.

the model parameters (see Table 1) to generate an error-free sequence of "true" runoff values Q_{true}.

[48] Next, to simulate the observational process, we corrupted these values of P, PET, and Q_{true}. by adding random measurement errors. For PET, we assumed the errors to be Gaussian, zero-mean, and homoscedastic with standard deviation σ. However, because errors in P and Q are typically heteroscedastic, we first transformed each variable using a Box-Cox power transformation $\left(Z = \left(Y^{\lambda} - 1\right)/\lambda\right)$, then added Gaussian zero-mean homoscedastic random errors having standard deviation σ and finally back-transformed the error-corrupted variables to the original space using the inverse Box-Cox transformation. To be consistent with previous research [e.g., *Misirli et al*., 2003; *Bulygina and Gupta*, 2010], we use $\lambda = 0.3$ to transform both variables (although we deviate from *Bulygina and Gupta* [2010] who actually use $\lambda = 0$ for precipitation (10% uncertainty), which stretches out the region of very small precipitation values < 1.0 mm in an undesirable manner). The result is a synthetic set of "observed" time series P_{obs}, PET_{obs}, and Q_{obs} having random "observational" errors of size characterized by the magnitude of σ.

Table 1. Parameters of the HyMod Model Used for Simulation Study

Parameters Name and Unit	$C_{\max}$ (mm)	$b_{\exp}$	α	K_s (d)	K_q (d)
Parameter value	224.4	0.261	0.861	0.0033	0.465

[49] To evaluate the performance of our method of estimating multivariate MI, we proceeded as follows. First, we computed the MI $I(\text{Input}_{\text{obs}}; Q_{\text{obs}})$ actually present in the "error-corrupted" data (where $\text{Input}_{\text{obs}} = \{P_{\text{obs}}, \text{PET}_{\text{obs}}\}$). Because this is a synthetically generated data set, where the model is a complete representation of the mapping from inputs to outputs, if the data were to contain no measurement errors $(\sigma = 0)$, the MI $I(\text{Input}_{\text{obs}}; Q_{\text{obs}}) \to \infty$. However, because $\sigma > 0$ the data set is of finite length, and information about the initial state of the system is not given by the input-output data, the value of $I(\text{Input}_{\text{obs}}; Q_{\text{obs}})$ will, in practice, be finite. As a practical approach we followed *Hsu et al*. [2002], who demonstrated that, for this catchment, the information contained in three previous time steps of the input and output data is sufficient to enable accurate predicts of runoff at the current time step. So we defined the input data set to be $\text{Input}^{*}_{\text{obs}} \sim P_{\text{obs}}(t-1)$, $P_{\text{obs}}(t-2), P_{\text{obs}}(t-3), \text{PET}_{\text{obs}}(t-1), \text{PET}_{\text{obs}}(t-2)$, $\text{PET}_{\text{obs}}(t-3), Q_{\text{obs}}(t-1), Q_{\text{obs}}(t-2), Q_{\text{obs}}(t-3)$ where the three lagged values of Q_{obs} function as substitutes for information regarding the current state of the system (thereby avoiding need to use an infinitely long time series of P and PET input data values; see *Box and Jenkins* [1976]). With this approach, $\text{Input}^{*}_{\text{obs}}$ has nine columns of values and can be used to compute an estimate of the MI offered by the raw data $I(\text{Input}^{*}_{\text{obs}}; Q_{\text{obs}})$ using the ICA method described earlier.

[50] Next, we emulate the normal practice of model-based estimation by forcing HyMod with the error-corrupted input data P_{obs} and PET_{obs} to obtain "simulated runoff" Q_{sim}. From this, we computed the bivariate mutual

Table 2. The Input-Output Variables Used to Estimate BAP (Observed Data) and to Evaluate Actual Model Performance (Model Simulation)

	Input	Output
Observed data	$Q_{obs}(t-1), Q_{obs}(t-2), \ldots$ $P(t-1), P(t-2), \ldots$ $\mathrm{PET}(t-1), \mathrm{PET}(t-2), \ldots$	$Q_{obs}(t)$
Model simulation	$P(t)$, $\mathrm{PET}(t)$, $\mathrm{State}(t-1)$	$Q_{sim}(t)$

information $I(Q_{sim}; Q_{obs})$ that quantifies the degree of similarity between Q_{sim} and Q_{obs} and indicates the degree to which knowledge of Q_{sim} can reduce our uncertainty in the knowledge of Q_{obs}. Because Q_{sim} is obtained by processing the error-corrupted inputs P_{obs} and PET_{obs} through the catchment model equations, and since the model is a "perfect" representation of the catchment system, the value of $I(Q_{sim}; Q_{obs})$ is completely controlled by the "size" of the random observational errors and by the way in which the Markovian state transition process modifies these errors. The former is characterized by the magnitude of σ specified earlier, so that as $\sigma \to \mathrm{C}$ we have $I(Q_{sim}; Q_{obs}) \to \infty$, and as $\sigma \to \infty$, we have $I(Q_{sim}; Q_{obs}) \to 0$). The nature of the state transition process is, of course, determined by the model structural hypothesis. For clarity, the variables used in estimating $I(\mathrm{Input}^{*}_{obs}; Q_{sim})$ and $I(Q_{sim}; Q_{obs})$ are shown in Table 2.

[51] The results are shown in Figure 4 for different sizes σ of the random observational error. As should be expected, the estimated values of $I(\mathrm{Input}^{*}_{obs}; Q_{obs})$ and $I(Q_{sim}; Q_{obs})$ both decrease with increasing σ, indicating that the proposed estimator is behaving correctly. However, whereas both values should be similar, the value computed for *multivariate mutual information* $I(\mathrm{Input}^{*}_{obs}; Q_{obs})$ is smaller than that for $I(Q_{sim}; Q_{obs})$, indicating that our procedure underestimates MI when applied to the higher-dimensional (10-D) data set (accuracy and precision of estimator for the 2-D case is confirmed in the Appendix A). This implication is that, when applied to the real-world studies in section 4, our procedure will underestimate the total amount of shared information, so that estimates of AU will be positively biased (larger than actual) and estimates of EU will be negatively biased (smaller than actual).

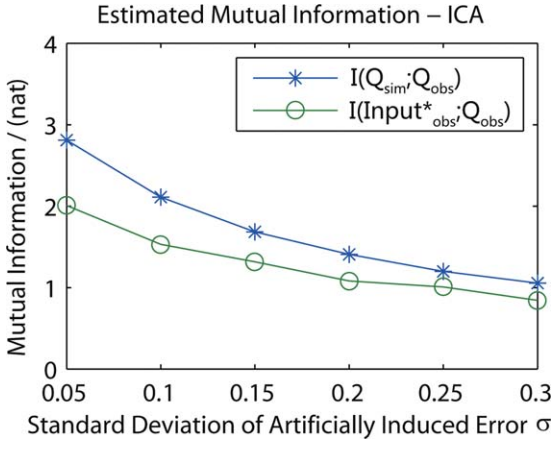

Figure 4. Synthetic hydrological modeling assessment of procedure for estimating MI under conditions of progressively increasing σ: The line with stars shows $I(\mathrm{Input}^{**}_{obs}; Q_{obs})$ estimated from the data and the line with circles shows $I(Q_{sim}; Q_{obs})$ computed from the model results.

[52] Some of the causes for this bias can be seen by examining the scatterplots between pairs of "independent" components of the data set obtained using FastICA decomposition (Figure 5); the range of each signal has been normalized to the range (0,1). We see that the signals are nearly one-to-one independent (confirmed in the Appendix B by computing the MI between each pair of signals). However, the various (noncircular) patterns indicate that the joint distribution is highly non-Gaussian. In addition, the patterns indicate the presence of a significant number of outliers that can cause instability in the ICA estimator.

4. Case Study

4.1. Introduction

[53] In the following investigation, we examine the usefulness of the concepts and methods developed in this paper for assessing (a) the simulation ability inherent in a catchment-scale rainfall runoff data set and (b) how much of that ability is encapsulated in a given model hypothesis. Our purpose is both illustrative and exploratory—to illustrate the potential power of methods based on information theoretic analysis and to explore the problems that we might encounter. We therefore compare the performance of three hydrological models on three catchments. The comparison across catchments shows that the BAP is controlled by the characteristics of the catchment and the nature of random observational errors, while the comparison across models indicates the relative strengths of alternative model structural hypotheses.

4.2. Data, Models Used, and Model Calibration Procedure Employed

[54] The three catchments used for this case study are the Leaf and Chunky River basins in the Southern United States, and the Suichuanjiang River basin in Southern China (gauged at the Chuzhou hydrologic station); see Table 3 for basic catchment information. The Leaf and Chunky River basins are located next to each other and have similar landform (relatively flat) and climatic (humid) characteristics, as well as similar observational standards. In contrast, the Suichuanjiang River basin is located in a mountainous region, has a greater degree of spatial variability and range of elevations, and uses a different observational standard.

[55] The three spatially lumped catchment-scale hydrological models used for this case study are HyMod (used earlier in the synthetic study), the Sacramento Soil Moisture Accounting (SAC-SMA) model used by the U.S. National Weather Service for flood forecasting [*Burnash et al*., 1973; *Brazil*, 1988], and the Xinanjiang model widely used in southern China [*Zhao et al*., 1980]. Although HyMod is structurally very simple, parsimonious, and designed mainly for preliminary testing of model identification procedures, the SAC-SMA and Xinanjiang models are structurally more sophisticated and were specifically designed for operational hydrological forecasting (for local conditions specific to the country of origin). In general, we would expect the SAC-SMA and Xinanjiang model structural hypotheses to be superior to that provided by HyMod.

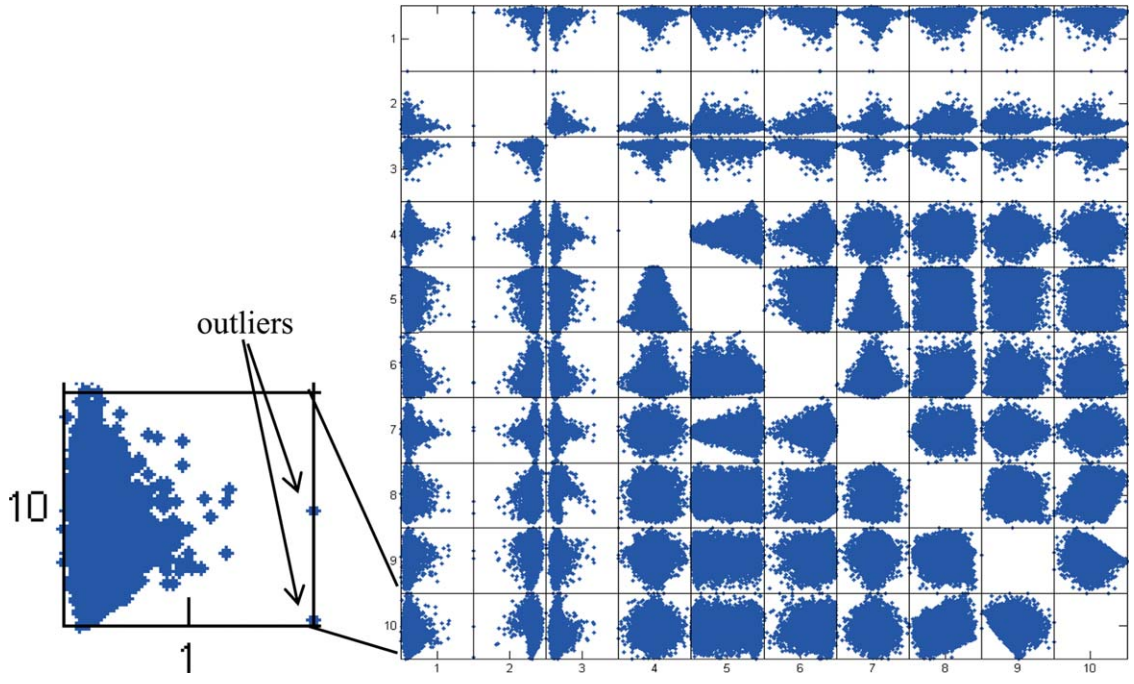

Figure 5. One-to-one scatter plot between each independent signal component of the hydrological data set. Figure 5 shows (1) degree of independence between each component, (2) non-Gaussianity due to various patterns, and (3) outliers causing instability of ICA.

[56] In this study, we examined six model-catchment combinations, deemed to be sufficient for the purposes of this study. The HyMod and SAC-SMA models were both applied to the Leaf River and Chunky River basins, while the HyMod and Xinanjiang models were applied to the Chuzhou basin. In each case, the models were calibrated to the observed catchment runoff data (see Appendix C) using the well-established Shuffled Complex Evolution Metropolis algorithm (SCEM-UA) single-criterion maximum likelihood calibration approach proposed by *Vrugt et al*. [2003]. All of the available data were used for model calibration (we are not concerned, in this study, with issues of temporal nonstationarity), and the first 65 days were used as a warm-up period to minimize uncertainty caused by incorrect initialization of the state variables.

4.3. Stabilizing the Estimate of High-Dimensional Mutual Information

[57] Having calibrated the models, we next proceeded to estimate the aleatory and epistemic uncertainties for each model-catchment case. For each catchment, we progressively increased the number of lagged previous time steps used to compute$I\left(\text{Input}^{*}_{\text{obs}};Q_{\text{obs}}\right)$ until a nonincreasing estimate of AU was obtained. However, during ICA decomposition of the data sets, we encountered a stability problem (see Appendix D) caused by outliers corresponding to time steps when the observed values of P, PET, and Q are small. Investigation revealed that this problem could be significantly reduced by adding small random errors to the raw data, indicating the likely cause to be quantization of the data during the observational process (such quantization is more significant when the observed values are small). Therefore, to stabilize the computations, we added a very small random Gaussian error ($\sigma = 10^{-3}$) to the data.

4.4. Estimates of BAP for the Basins

[58] The estimated values of BAP for the three basins, as expressed by $I\left(\text{Input}^{**}_{\text{obs}};Q_{\text{obs}}\right)$, are shown in Figure 6. The x axis corresponds to the number, n, of lagged previous time steps used to construct $I\left(\text{Input}^{**}_{\text{obs}}\right)$. For each case, we obtain 10 independent estimates of $I\left(\text{Input}^{**}_{\text{obs}};Q_{\text{obs}}\right)$, by randomly varying the initial value of the matrix A in the ICA analysis, and record the mean and standard deviation of the 10 estimates; this provides information regarding the stability and precision of the results. Figure 6a shows the mean estimated value of BAP as expressed by $I\left(\text{Input}^{**}_{\text{obs}};Q_{\text{obs}}\right)$, and Figure 6b shows the standard deviation of the estimate.

[59] The results shown in Figure 6a indicate that BAP can be expected to vary with landform type and observational precision of a catchment. In the case of the relatively similar Leaf and Chunky River basins, the MI content of their input-output data sets are remarkable similar (~1.6 nats)

Table 3. Basic Information About the Chosen Basins

	Country	Area (km^2)	Elevation (m)	Annual Rainfall (mm)	Landform	Data Used
Leaf River	USA	1924	60–185	1492	Plain	1948–1978
Chunky River	USA	956	66–199	1447	Plain	1948–1978
Chuzhou	China	289	380–1550	1550	Mountain	1980–2000

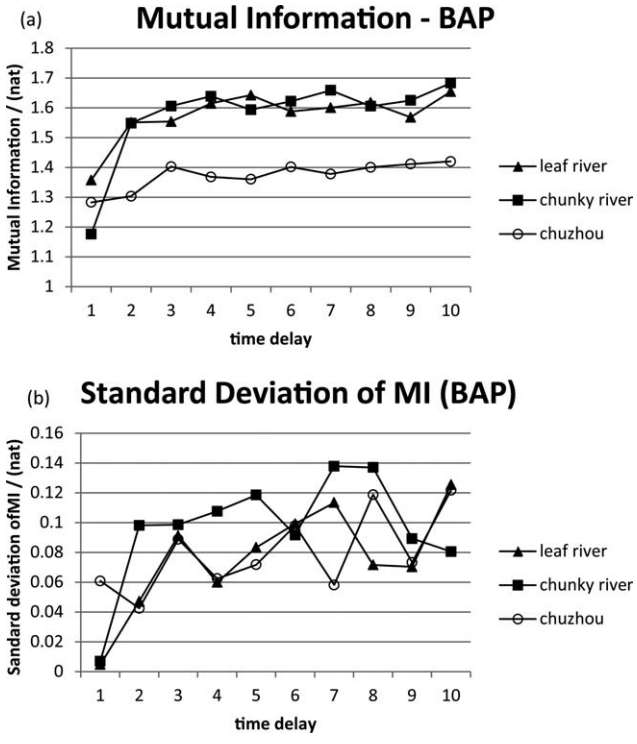

Figure 6. BAP for the different basins. (a) Average MI from 10 estimates. (b) The standard deviation of the 10 estimates. BAP for the Leaf and Chunky basins is similar and higher than for the Chuzhou, while standard errors of the estimates are similar for all three cases.

and quite high (close to the practical achievable limit of ~2.0 nats; see section 2), whereas for the Suichuanjiang River basin, the MI content is smaller (~1.4 nats). As such, the BAP by a hydrological model can be expected to be better for the U.S. basins, provided that the information in the data is properly exploited by the model hypothesis. In all three basins, the estimate of BAP increased quickly with the number of time lags used, began to stabilize around $n = 3$, and converged by about $n = 6$ days; these results are consistent with, and extend upon, the findings of *Hsu et al*. [2002], who modeled the Leaf River basin using an artificial neural network approach. The results shown in Figure 6b indicate that the standard error of the estimate is relatively small compared to the mean value (coefficient of variation of around 5%–6%) and that the value is similar for all three basins.

4.5. Evaluation of the Structural Adequacy of Different Model Hypotheses

[60] The estimated values of model performance for the six model-catchment cases, as expressed by $I(Q_{\text{sim}}; Q_{\text{obs}})$, are shown in Table 4. Also shown are the values of three other commonly used metrics, the Nash Sutcliffe efficiency (NSE), the linear correlation coefficient ρ between Q_{sim} and Q_{obs}, and the *R*-statistic. Of course, all of these metrics are interrelated—the NSE is approximately equal to ρ^2 [see *Gupta et al*., 2009; *Gupta and Kling*, 2011], and the MI $I(Q_{\text{sim}}; Q_{\text{obs}})$ can be interpreted as a generalized correlation coefficient (that provides a more accurate assessment of relationship strength when the data distribution is non-Gaussian and the relationship is nonlinear). Further, as shown in *Granger and Lin* [1994], the *R*-statistic is precisely related to MI and becomes equivalent to ρ when the data are jointly bivariate Gaussian (see equation (5) and discussion in section 2). No matter which statistic is used, it is clear that the structurally more complex SAC-SMA and Xinanjiang models are superior to HyMod.

[61] Figure 7 shows scatterplots of the simulated versus observed runoff for each of the six model-catchment cases. A consistent general tendency toward underestimation bias at higher flow levels is shown in Figure 7, which is expected when calibrating models using squared error type metrics [see *Gupta et al*., 2009]. However, the bias is consistently quite severe for HyMod and much less so for SAC-SMA and Xinanjiang. While this degree of bias could potentially be corrected by postprocessing [*Seo et al*., 2006], the correlation coefficient, *R*-statistic, and MI metrics clearly show how much of the lack of correspondence (the scatter) between simulated and observed runoff cannot be easily removed by such methods.

[62] Table 5 and Figure 8 show the results of the information analysis. The estimated aleatory uncertainty is quite high (44%–53%), indicating that the available data do not contain the information required to achieve streamflow simulation at the desired level of precision (finer than the bin width)—bear in mind, however, that the synthetic study reported in section 3 suggests that these values of AU are positively biased, and the actual shared information is likely somewhat higher. Further, AU for the U.S. basins is smaller than for the Chinese basin, consistent with higher quality of the former. Meanwhile, the epistemic uncertainty unresolved by the different model hypotheses varies from 12% (Xiananjiang for Chuzhou) to 45% (HyMod for Leaf), indicating the extent to which each of the model structural hypotheses has been able to exploit the potentially usable information in the data. For the two U.S. basins, the SAC-SMA model hypothesis has exploited ~57% of the available information and the epistemic uncertainty remains around 43%; this provides a quantitative measure of model structural inadequacy. While the SAC-SMA performance is from 2% to 10% better than that of HyMod, there appears to be considerable room for performance enhancements (by model structural improvements and/or by removal of

Table 4. Performance of Different Models for Each Basin

		NSE	Correlation Coefficient ρ	*R*-Statistic	$I(Q_{\text{sim}}; Q_{\text{obs}})$ (nat)
Leaf River	HyMod	0.6685	0.8275	0.8837	0.7591
	SAC-SMA	0.8436	0.9198	0.9164	0.9158
Chunky River	HyMod	0.7290	0.8549	0.9122	0.8921
	SAC-SMA	0.8066	0.9031	0.9183	0.9268
Chuzhou	HyMod	0.7579	0.8951	0.9378	1.0583
	Xinanjiang	0.8528	0.9443	0.9555	1.2212

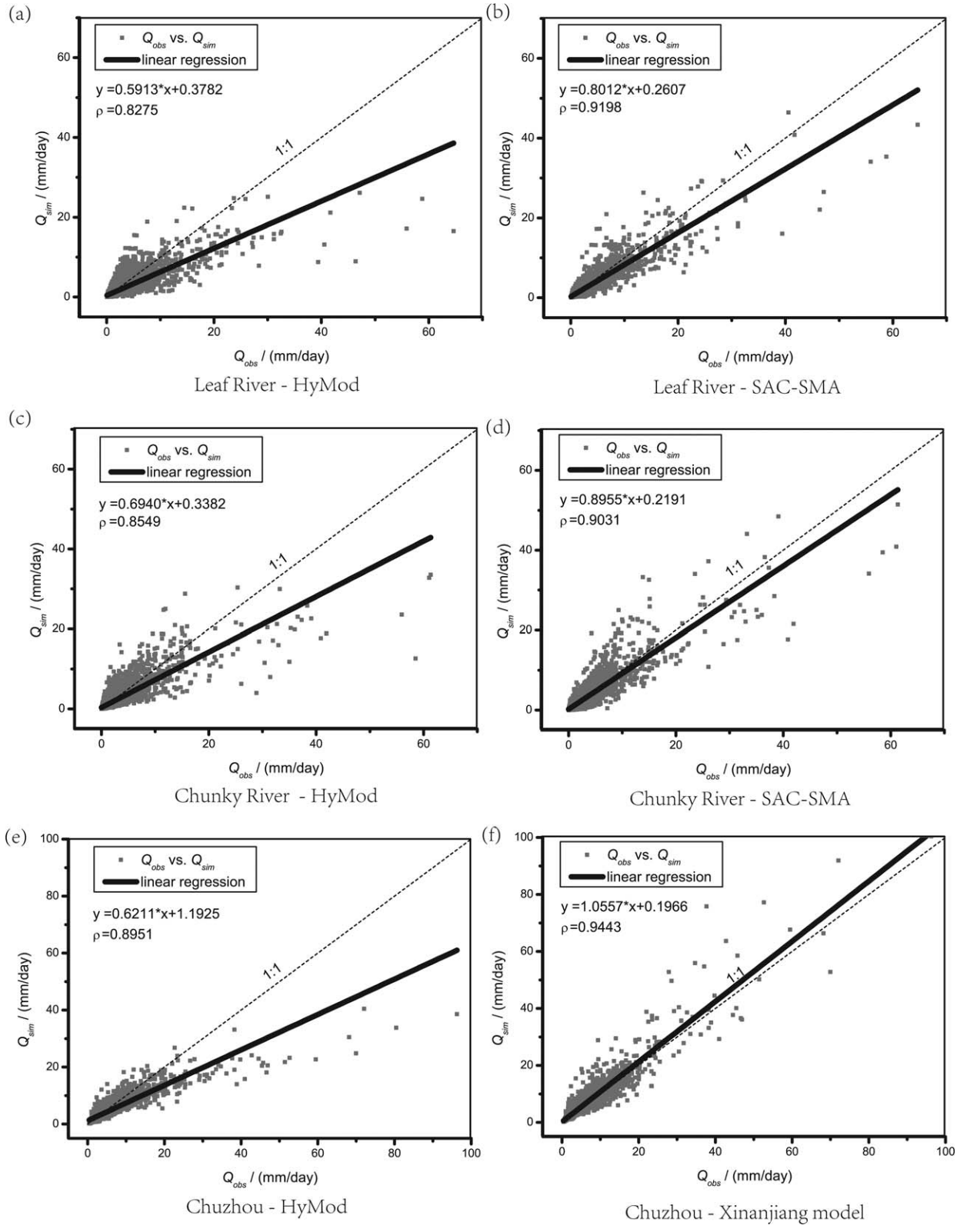

Figure 7. Scatterplots of simulated and observed runoff. (a) Leaf River-HyMod, (b) Leaf River-SAC-SMA, (c) Chunky River-HyMod, (d) Chunky River-SAC-SMA, (e) Chuzhou-HyMod, (f) Chuzhou-Xinanjiang model.

Table 5. Analysis of Information Expressed by the Observed Data and Exploited by the Models (Unit: nat)[a]

	Predictive Uncertainty $H(Q_{obs})$	Estimated *MI* in Data $I\left(\text{Input}^{**}_{obs}; Q_{obs}\right)$	Aleatory Unc. $(AU) = H(Q_{obs}) - I\left(\text{Input}^{**}_{obs}; Q_{obs}\right)$	Model	MI Expressed by Model $I(Q_{sim}; Q_{obs})$	Epistemic Unc. (EU) $I\left(\text{Input}^{**}_{obs}; Q_{obs}\right) - I(Q_{sim}; Q_{obs})$
Leaf	2.8870	1.6054	1.2816 (44% of H)	HyMod	0.7591	0.8463 (53% of Data MI)
				SAC-SMA	0.9158	0.6896 (43%)
Chunky	3.0410	1.6385	1.4025 (46%)	HyMod	0.8921	0.7464 (45%)
				SAC-SMA	0.9268	0.7117 (43%)
Chuzhou	2.9591	1.4012	1.5579 (53%)	HyMod	1.0583	0.3429 (24%)
				Xinanjiang	1.2212	0.1800 (12%)

[a]The value for $I\left(\text{Input}^{**}_{obs}; Q_{obs}\right)$ shown is the average obtained by varying the number of lags, n, from 6 to 10 days (see Figure 6).

systematic biases in the data—see discussion in the next paragraph). In contrast, the Xinanjiang model has exploited a relatively large fraction (about 87%) of the information available for the Chinese basin (and is 12% better than HyMod), so that the epistemic uncertainty is relatively small (12%), indicating much less room for performance improvements in the context of this basin.

[63] Figure 9 shows comparisons of the observed and simulated time series for selected periods of the data set. The shaded 95% "confidence intervals" around the observed data (computed using equation (13)) indicate the region within which a model simulation must fall (95% of the time) to have reached the BAP for that catchment (given the available data set); it is interesting to note that these intervals are quite narrow. Figures 9a and 9b indicate that the SAC-SMA and HyMod models have trouble matching both the peak flows and the behavior of the quick recession (the observed quick recession is much faster). Similarly, Figure 9c shows that the Xinanjiang and HyMod models are both unable to match the peaks. While the inability to properly simulate the quick recession may be caused by a combination of inadequate model structure and model calibration, the inability of all three models to match the peak flows is more likely due to underestimation biases in the higher intensity precipitation data—biases that cannot be removed by model structural improvements alone. In other words, because the catchment models are constrained to maintain water balance, when there are biases in the observed precipitation data, such models are unable to match the performance achievable by purely data-based modeling (indicated by $I\left(\text{Input}^{**}_{obs}; Q_{obs}\right)$).

5. Conclusions and Discussion

[64] The ability to quantitatively estimate the approximate magnitude of aleatory and epistemic uncertainties can add considerable power to a modeling analysis. By estimating the aleatory uncertainty for a given data set, one has access to a benchmark that defines the best model performance that is achievable by fully exploiting the information available in a particular data set. Given that the model structural hypothesis cannot add any new information over and above what is already contained in the data (due to the data-processing inequality), to do better than this benchmark one must provide new information in the form of additional explanatory variables. However, given data measurement error, and possible inherent stochasticity in the system (absence of a deterministic relationship between the system variables), it will generally be impossible to reduce the aleatory uncertainty to zero.

[65] The data-processing inequality makes it clear that, to the extent that the model does not bring in new information that is not encoded in the input-output data, the model structural hypothesis (including the conceptual and mathematical structure) does not increase the amount of information available for making estimates of the system outputs. Instead, *the role of the model hypothesis is to correctly describe the sequence of operations by which the information moves from the inputs to the outputs*. The methods described in this paper can help to quantify the best model performance possible. To improve upon that performance, it will be necessary to acquire observations regarding the "correct" set of input variables required to explain the dynamics of the system outputs, and to do so with sufficient precision. Note, however, that when using a physically based representation that is constrained to preserve water balance, such a model hypothesis will be unable to achieve the BAP limit when the data contain systematic biases, unless additional information in the form of how to correct such biases is provided (either in the form of data or in the form of a bias-correction model structural hypothesis).

[66] So, given that the data-processing inequality refers to the properties of transformations expressed in the form of a Markov chain, one may validly ask how the implications of this should be understood in the context of hydrological modeling. For example, is it really true that the physical knowledge embodied via a model structural hypothesis does not provide any additional information? In this context, it should be noted that the concept of "physical knowledge" typically used by hydrologists consists of two kinds—"relationships" and "data." To the extent that the "physical knowledge" brought to bear is a "relationship," it is in fact a *procedure that processes data*, and, as such (as expressed by the data-processing inequality), does not provide any new information—the equations process the information, but are unable to generate any new information out of the void (just as they cannot generate energy or mass unless this is explicitly expressed as data in the formulation). However, if the "physical knowledge" is a kind of "data," it will in fact provide additional information, that is then processed by the "relationship" (the model). For example, in the modeling of hill-slope hydrological processes, the values of slope length, slope gradient,

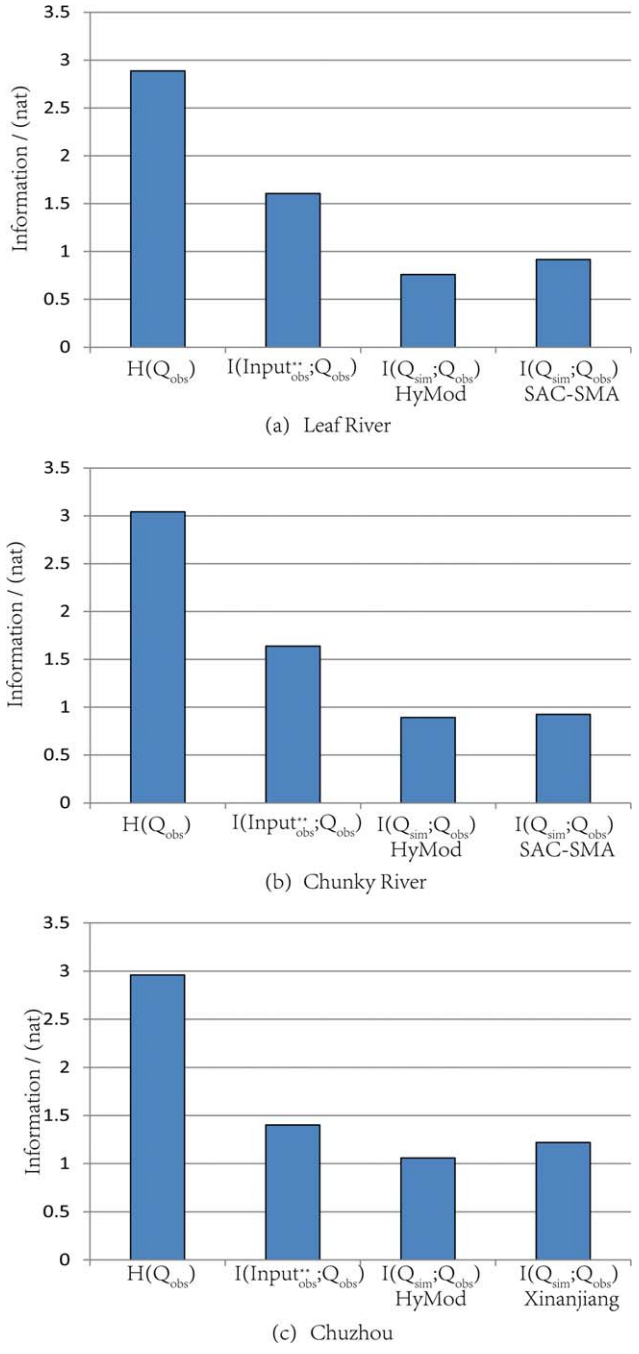

Figure 8. Information analysis for the (a) Leaf, (b) Chunky, and (c) Chuzhou basins. The leftmost bar indicates the amount of information $H(Q_{\text{obs}})$ required to predict streamflow, middle-left bar indicates predictive information $I(\text{Input}_{\text{obs}}^{**}; Q_{\text{obs}})$ contained in the data, and two bars on the right indicate predictive information $I(Q_{\text{sim}}; Q_{\text{obs}})$expressed by the models.

vegetation coverage, roughness, and other parameters, consist of "data" that do provide additional information, while the equations that describe the relationships among them are actually "relationships." In the case of physically based models, "prior knowledge" (data) can be provided in the form of additional information and used to reduce the requirements for model calibration. It will be interesting, in future work, to study how much additional information such physically based data can add to the model hypothesis, and how this can be used to evaluate achievable model performance in the context of making predictions in ungauged basins (PUB) [*Sivapalan et al*., 2003].

[67] In our real-data investigation, we have used an empirical (approximate) approach to estimate the aleatory uncertainty; this approach assumes that all of the information required to estimate runoff at the next time step can be expressed by the past 1–10 time steps of the variables *Q, P*, and PET. In reality, the number of lags expressing the memory of the system may be much larger (even infinite). However, our study shows that, for practical purposes and given the noise in the data, approximately three to six past lags seems to be sufficient. While more detailed analysis could be done to establish exactly how many lags should be applied to each variable—this would not advance the goals of the present study. Interestingly, we achieve this conclusion without recourse to any hypothesis regarding the (linear or nonlinear) form of the model *structural* hypothesis and obtain an answer consistent with that obtained by *Hsu et al*. [2002] using a sophisticated nonlinear (artificial neural network) modeling approach applied to the same data.

[68] Further, as shown in Figure 9, clearly there is considerable room to improve the explanatory power of the model. These results can be compared with those of *Vrugt et al*. [2008] and *Beven et al*. [2011], which indicate significant uncertainty in the input-output data over the entire time range (without data assimilation). Similarly, our results indicate that the aleatory uncertainty, including that arising from random input/output observational errors, is quite significant. Note also that the effects of data errors on the ability to make one time step ahead estimates via conceptual water-balance modeling can be exacerbated due to (a) accumulation of errors in the state estimates via error propagation and (b) the water-balance constraint imposed by fundamental principles of physically based modeling. To reduce uncertainties due to systematic biases in the data, data correction methods must be employed (e.g., BATEA) [*Kavetski et al*., 2006a, 2006b], and to reduce uncertainties due to accumulation of state estimation error, data assimilation can be performed [*Nearing et al*., 2013]. It will be interesting to use the methods discussed in this paper to evaluate the power of data correction and data assimilation methods in controlling error accumulation and reducing estimation uncertainty.

[69] We acknowledge that there are several limitations to this present study. For example, the stability and bias of the method developed to estimate MI in high-dimensional data sets needs to be further assessed, and more work is needed to ensure that the method is capable of handling the kinds of situation encountered in hydrological studies. In particular, there is a pressing need to extend the ICA method to handle significant nonlinearities that will most surely exist in data interdependencies. Further, while some recent literature has explored the partitioning of total uncertainty into its constituent sources—input, parameter, initial state and structure—we have shown only how to partition the total uncertainty into its aleatory and epistemic components. However, quantitative identification of the amount of information loss during modeling, and detection of where it is lost (input, parameter

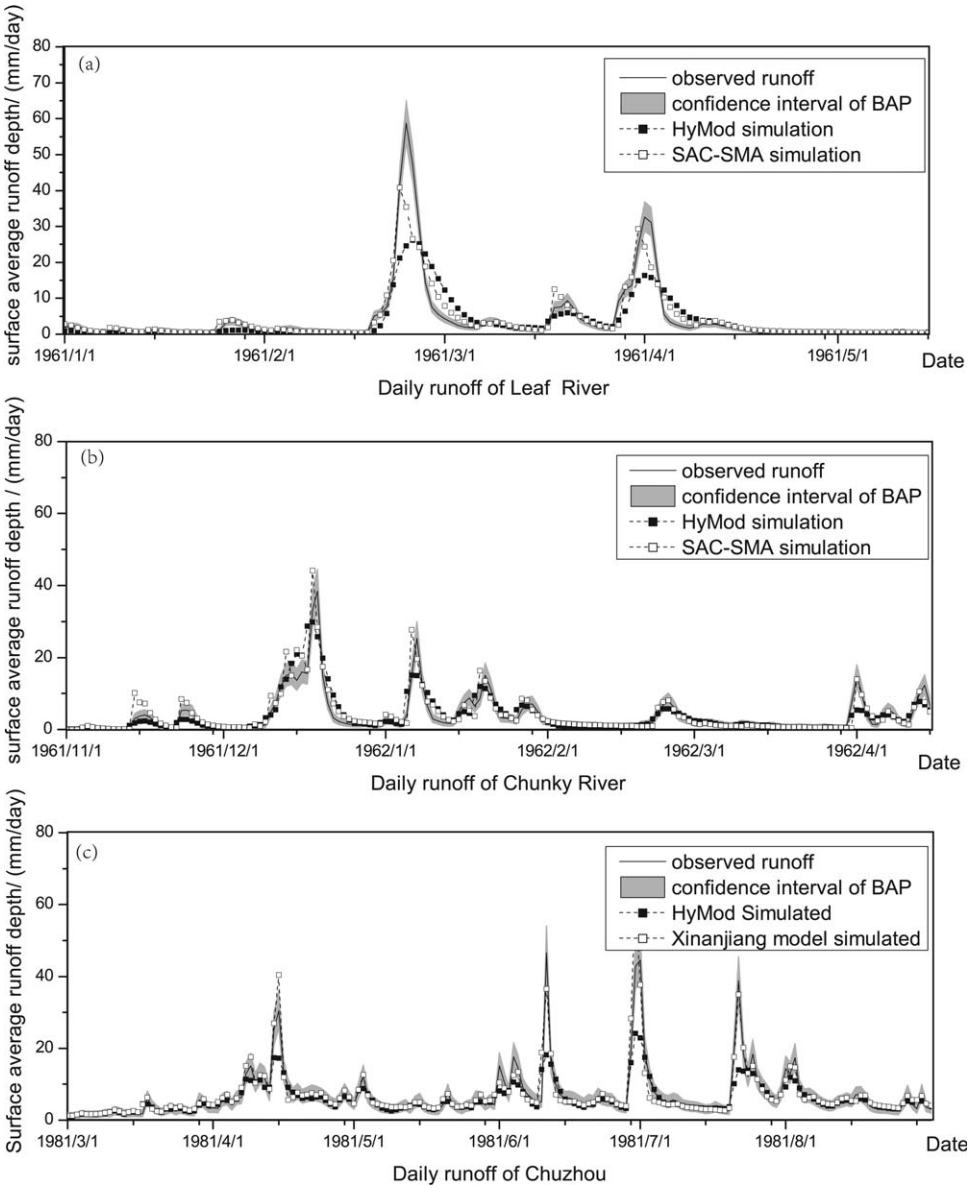

Figure 9. Comparison of observed and model simulated hydrographs for selected periods. Observed daily runoff, confidence interval of BAP, and simulated daily runoff of (a) Leaf River, (b) Chunky River, (c) Chuzhou, are shown in these figures. Observed values are connected by a solid line, around which the 95% confidence intervals are shown: a model simulation must fall within these intervals (95% of the time) to reach the BAP (given the available data).

or structure) can help lead to model structural improvements. As a generalized correlation coefficient, MI is sensitive to correlation but blunt to system bias. To give an appropriate assessment of model performance, it is necessary to use several metrics in combination and to examine scatterplots (Figure 7) and time-series plots (Figure 9) of the simulated and observed outputs. Finally, in physically based modeling, other available sources of information may include observations of basin characteristics such as geomorphology, land use and land cover, soil texture, and leaf area index. The problems of how to characterize the information contributed by basin characteristics, and of how to estimate best achievable model performance for ungaged basins are interesting research topics. We leave the issue of how *information theory* can be applied to such issues for future work.

[70] In summary, this paper has presented an exploration of an *information theoretic* approach to quantifying and characterizing the information content in hydrological data, with a view to (a) quantifying the information required to achieve an estimate with a desired level of precision, (b) providing a baseline for achievable model performance, and (c) establishing to what degree the available information been correctly expressed by a given model hypothesis. In ongoing work, we are exploring how these tools can be used to help in diagnosis and correction of model structural errors, and will report these results in due course. As always, we encourage and invite dialog and collaboration on this and other aspects of model identification. The computer codes used in this work can be obtained from the first author upon request.

Appendix A: Accuracy and Precision of the Proposed Estimator

A1. Estimating Multivariate Information Entropy

[71] We evaluated the precision of the proposed estimator on two high-dimensional data sets for which the distributions are known (jointly Gaussian and jointly uniform), and so theoretical values can be computed. For a jointly Gaussian random vector $\mathbf{X}$ having covariance matrix Σ, the multivariate entropy is given by

$$H(\mathbf{X}) = \frac{1}{2}\log(2\pi e)^m|\Sigma|. \tag{A1}$$

For a jointly uniform random vector $\mathbf{X}$ in which each component X_i is uniform on (a_i, b_i), so that $r_i = a_i - b_i$, the multivariate entropy is given by

$$H(\mathbf{X}) = \log\left(\prod_{i=1}^{m} r_i\right). \tag{A2}$$

For the Gaussian cases, we used a synthetic data set of dimension $m = 10$ and having diagonal covariance matrix $\Sigma = \mathrm{diag}\,(10, 9, 8, \ldots 1)$. For the uniform cases, we used a synthetic data set of dimension $m = 10$ and having component ranges $[r_1,r_2,r_3,\ldots,r_{10}] = [10, 9, 8,\ldots,1]$. Since high-dimensional entropy does not vary under affine transformation, we applied randomly selected affine transformations to each data set. Further, several sample sizes were tested, and the test was replicated 10 times for each case. The results are shown below.

[72] (1) Jointly Gaussian distribution: In this case, the theoretical value of entropy is 19.17. The results clearly indicate that, as the sample size is increased, (a) the mean value of the estimate increases, (b) the variance of the estimate decreases, and (c) the absolute value of mean error decreases. The result indicates that the estimate of multivariate entropy is only slightly biased (0.44% for 1000 samples), and that making the sample size large enough can reduce the bias. Note however that average absolute bias is, in each case, smaller than two standard deviations (Table A1).

[73] (2) Jointly uniform distribution: In this case, the theoretical value of entropy is 15.10. The results are similar to those for the Gaussian distribution, but the bias tends to be positive, larger (2.78% for 1000 samples), and more significant (larger than two standard deviations) (Table A2).

Table A1. Result of Simulation Study—Joint Gaussian Distribution (Unit: nat)

Sample Size		1,000	10,000	100,000	1,000,000
Estimated entropy of 10 replicates	1	19.12	19.16	19.17	19.17
	2	19.10	19.15	19.16	19.18
	3	19.08	19.16	19.17	19.17
	4	19.21	19.13	19.16	19.17
	5	19.05	19.15	19.17	19.17
	6	19.04	19.18	19.17	19.17
	7	19.17	19.14	19.16	19.17
	8	19.13	19.15	19.17	19.17
	9	19.05	19.20	19.17	19.17
	10	18.90	19.19	19.17	19.17
Mean		19.09	19.16	19.17	19.17
Mean error		−8.5E-02	−8.5E-03	−5.5E-03	−1.2E-03
Standard deviation of error		8.46E-02	2.2E-02	4.01E-03	2.25E-03

Table A2. Result of Simulation Study—Joint Uniform Distribution (Unit: nat)

Sample Size		1,000	10,000	100,000	1,000,000
Estimated entropy of 10 replicates	1	15.49	15.31	15.20	15.15
	2	15.56	15.30	15.21	15.15
	3	15.54	15.29	15.20	15.15
	4	15.54	15.32	15.20	15.15
	5	15.45	15.32	15.20	15.15
	6	15.46	15.32	15.20	15.15
	7	15.57	15.30	15.20	15.15
	8	15.55	15.32	15.20	15.15
	9	15.54	15.31	15.19	15.15
	10	15.49	15.33	15.20	15.15
mean		15.52	15.31	15.20	15.15
mean error		42.0E-02	21.0E-02	10.0E-02	5.0E-02
Standard deviation of error		3.99E-02	1.15E-02	3.9E-03	8.25E-04

[74] (3) Estimating 2-D mutual information: We also evaluated the performance of the estimator for a 2-D joint normal distribution where random variable X,Y are jointly distributed as $N(0,\Sigma)$ with covariance matrix $\Sigma = \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix}$, so that the MI between X and Y is determined by the correlation coefficient ρ.

$$I(X;Y) = -\frac{1}{2}\log\left(1-\rho^2\right). \tag{A3}$$

Figure A1 shows the MI given by equation (A3) and estimated by the PCA/ICA based estimator proposed in this paper, with ρ on the x axis and MI on the y axis. For each value of ρ, 1000 samples were generated. The results show that the proposed estimator can give an accurate and precise estimate of 2-D MI.

[75] The proposed estimator can also be applied to non-Gaussian 2-D distributions, as long as X and Y satisfy the assumption that they can be decomposed into two

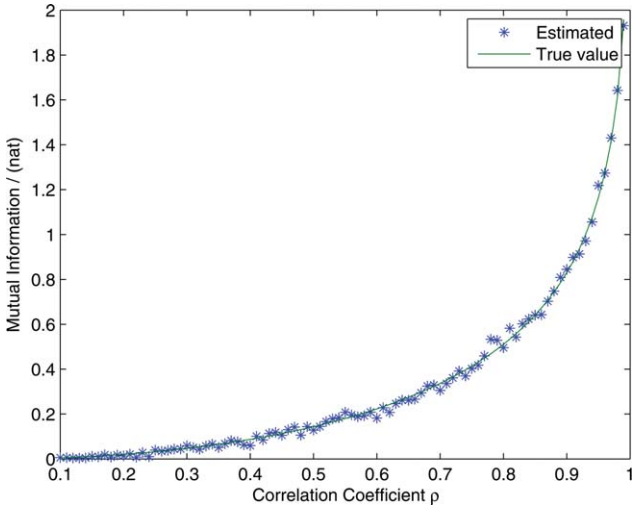

Figure A1. Evaluating the accuracy and precision of MI estimator.

Table B1. Mutual Information Between Each Independent Components Decomposed by FastICA[a]

	1	2	3	4	5	6	7	8	9	10
1	∞	0.0491	0.0636	0.0338	0.0440	0.0672	0.0351	0.0571	0.0393	0.0436
2	0.0491	∞	0.0640	0.0329	0.0385	0.0834	0.0345	0.0578	**0.1022**	**0.1998**
3	0.0636	0.0640	∞	0.0320	0.0351	0.0683	0.0317	**0.2158**	0.0504	0.0573
4	0.0338	0.0329	0.0320	∞	**0.1881**	0.0387	0.0660	0.0279	0.0284	0.0260
5	0.0440	0.0385	0.0351	**0.1881**	∞	0.0980	**0.1731**	0.0751	0.0240	0.0224
6	0.0672	0.0834	0.0683	0.0387	0.0980	∞	0.0419	0.0424	0.0431	0.0297
7	0.0351	0.0345	0.0317	0.0660	**0.1731**	0.0419	∞	0.0316	0.0294	0.0242
8	0.0571	0.0578	**0.2158**	0.0279	0.0751	0.0424	0.0316	∞	**0.1014**	**0.1549**
9	0.0393	**0.1022**	0.0504	0.0284	0.0240	0.0431	0.0294	**0.1014**	∞	**0.2708**
10	0.0436	**0.1998**	0.0573	0.0260	0.0224	0.0297	0.0242	**0.1549**	**0.2708**	∞

[a]Mutual information larger than 0.1 nat are highlighted with bold font. Note that this result is given by ASH method [*Fernando et al*., 2009].

independent signals via linear transformation. An intercomparison of 2-D MI estimators was given by *Khan et al*. [2007].

Appendix B: Confirming the Independence Between "Independent Components"

[76] As indicated in section 3, we compute the one-to-one MI between each "independent components" decomposed by FastICA (see Table B1). Here we use average-shifted histogram method (ASH method) [see *Fernando et al*., 2009] that has been shown to be sensitive to nonlinear correlations. The font of each cell indicates the MI between each pair of components; bold font means higher levels of MI (interdependence). Clearly, interdependence between most of the components is weak, confirming the independence of independent components decomposed by FastICA.

[77] However, in a few cases the interdependence is not negligible (although not very high). For example, the MI between components 9 and 10 is 0.2708. A scatterplot between components 9 and 10 (Figure B1) indicates the presence of a linearly correlated "dense region" that might be the cause of this. This interdependence will somewhat affect the accuracy of the proposed MI estimator. Work to further reduce this small degree of interdependence is ongoing; one possible approach is to divide the data set into different segments, separate the dense region from other parts, and apply the estimator, respectively.

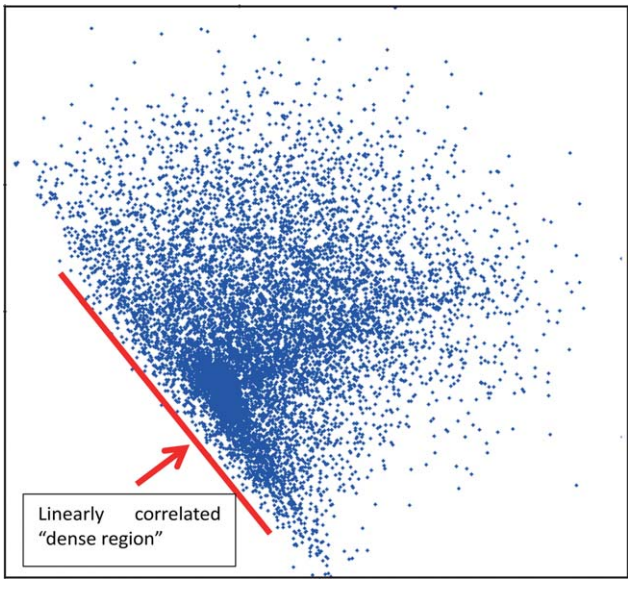

Figure B1. Scatterplots between components 9 and 10.

Appendix C: Calibrated Parameters

[78] Table C1-C6 show the calibrated parameters of HyMod, SAC-SMA, Xinanjiang model given by SCEM-UA. In each table, the prior range of parameters, posterior 95% confidence interval, best fit parameter, mean and median of posterior distribution, and the unit of parameters are presented. Some parameters, such as PCTIM of SAC-SMA, are not involved in calibration. The values of these parameters are directly presented in "Prior Range" and "Best Fit Parameter Set."

Table C1. Estimated Parameters of the HyMod Model (Leaf River)

Parameter	Prior Range	Posterior 95% Confidence Interval	Best Fit Parameter Set	Mean	Median	Units
C_{max}	(1,500)	(223.7,225.4)	224.4	224.5	224.7	mm
b_{exp}	(0.1,2)	(0.2602,0.265)	0.2614	0.2624	0.2623	
α	(0.1,0.99)	(0.8607,0.8627)	0.8614	0.8617	0.8607	
K_s	(0,0.1)	(0.0032,0.0034)	0.0033	0.0033	0.0034	
K_q	(0.1,0.99)	(0.4636,0.4656)	0.4646	0.4645	0.4645	

Table C2. Estimated Parameters of the HyMod Model (Chunky River)

Parameter	Prior Range	Posterior 95% Confidence Interval	Best Fit Parameter Set	Mean	Median	Units
C_{max}	(1,500)	(267.5,271.5)	269.5	269.6	267.1	mm
b_{exp}	(0.1,2)	(0.288,0.2985)	0.2936	0.2936	0.3024	
α	(0.1,0.99)	(0.8978,0.9043)	0.902	0.9014	0.9016	
K_s	(0,0.1)	(0.0104,0.0114)	0.0108	0.0109	0.011	
K_q	(0.1,0.99)	(0.5066,0.5106)	0.5085	0.5087	0.5073	

Table C3. Estimated Parameters of the HyMod Model (Chuzhou)

Parameter	Prior Range	Posterior 95% Confidence Interval	Best Fit Parameter Set	Mean	Median	Units
C_{max}	(1,500)	(132.3,152.8)	146.4	143.4	143.2	mm
b_{exp}	(0.1,2)	(1.6499,1.972)	1.8902	1.8239	1.856	
α	(0.1,0.99)	(0.5156,0.5268)	0.5207	0.5215	0.5196	
K_s	(0,0.1)	(0.0129,0.0136)	0.0132	0.0132	0.0134	
K_q	(0.1,0.99)	(0.6105,0.6181)	0.6155	0.6145	0.6137	

Table C4. Estimated Parameters of the SAC-SMA Model (Leaf River)

Parameter	Prior Range	Posterior 95% Confidence Interval	Best Fit Parameter Set	Mean	Median	Units
PCTIM	0.005		0.005			
ADIMP	0.4		0.4			
SARVA	0		0			
UZTWM	(10,300)	(21.55,56.80)	22.41	26.46	22.69	mm
LZTWM	(10,500)	(219.8,295.18)	296.7	275.6	283.2	mm
RSERV	0.3		0.3			
UZK	(0.1,0.75)	(0.3668,0.6401)	0.4126	0.4381	0.3872	
UZFWM	(5,150)	(21.94,44.18)	24.56	25.83	25.49	mm
ZPERC	(5,350)	(102.9,348.9)	348.1	292.0	341.6	
REXP	(1,5)	(3.4782,4.9519)	4.8689	4.4298	4.5764	
PFREE	(0,0.8)	(0.1037,0.2005)	0.1084	0.1291	0.1177	
LZFSM	(5,400)	(28.57,118.7)	124.7	58.43	52.70	mm
LZSK	(0.01,0.35)	(0.0262,0.1423)	0.0267	0.047	0.0358	
LZFPM	(10,1000)	(151.9,254.8)	158.8	203.0	204.4	mm
LZPK	(0.001,0.05)	(0.0061,0.0102)	0.0059	0.0084	0.0085	
SIDE	0		0			
SSOUT	1		1			
NASHN	(1,6)	(1.044,1.435)	1.072	1.177	1.142	
NASHK	(0.1,1)	(0.2967,0.3535)	0.2986	0.3106	0.3056	

Table C5. Estimated Parameters of the SAC-SMA Model (Chunky River)

Parameter	Prior Range	Posterior 95% Confidence Interval	Best Fit Parameter Set	Mean	Median	Units
PCTIM	0.005		0.005			
ADIMP	0.4		0.4			
SARVA	0		0			
UZTWM	(10,300)	(33.78,53.93)	33.47	46.50	53.41	mm
LZTWM	(10,500)	(257.1,279.3)	281.1	266.7	259.2	mm
RSERV	0.3		0.3			
UZK	(0.1,0.75)	(0.6942,0.7485)	0.7336	0.7289	0.7366	
UZFWM	(5,150)	(13.93,18.08)	16.73	15.97	15.91	mm
ZPERC	(5,350)	(37.35,59.24)	61.44	46.90	44.17	
REXP	(1,5)	(4.343,4.977)	4.910	4.573	4.460	
PFREE	(0,0.8)	(0.0697,0.1066)	0.0968	0.0841	0.0766	
LZFSM	(5,400)	(228.7,270.7)	274.7	246.3	228.4	mm
LZSK	(0.01,0.35)	(0.0325,0.0391)	0.0343	0.0359	0.0381	
LZFPM	(10,1000)	(49.39,94.14)	50.51	71.72	87.21	mm
LZPK	(0.001,0.05)	(0.0054,0.0073)	0.006	0.0061	0.0059	
SIDE	0		0			
SSOUT	1		1			
NASHN	(1,6)	(1.475,1.500)	1.500	1.493	1.495	
NASHK	(0.1,1)	(0.3962,0.4189)	0.4179	0.4081	0.4004	

Table C6. Estimated Parameters of the Xinanjiang Model (Chuzhou)

Parameter	Prior Range	Posterior 95% Confidence Interval	Best Fit Parameter Set	Mean	Median	Units
KC	0.65		0.65			
UM	(10,20)	(18.07,20.00)	19.91	19.70	19.90	mm
LM	(60,90)	(79.59,89.95)	89.71	88.33	89.45	mm
C	(0.15,0.20)	(0.1535,0.1978)	0.1756	0.1781	0.1878	
WM	(120,200)	(194.0,200.0)	200.0	198.9	199.9	mm
B	0.3		0.3			
IM	0.01		0.01			
SM	(5,50)	(49.75,50.00)	49.98	49.95	49.99	mm
EX	1		1			
KG	(0.1,0.4)	(0.1461,0.1605)	0.1514	0.1517	0.1494	
KI	(0.1,0.4)	(0.1081,0.119)	0.1138	0.1134	0.1157	
CI	(0.5,0.99)	(0.8249,0.8434)	0.834	0.835	0.8376	
CG	(0.95,0.998)	(0.9828,0.9838)	0.9834	0.9833	0.9836	
UH	[0.3,0.6,0.1]		[0.3,0.6,0.1]			

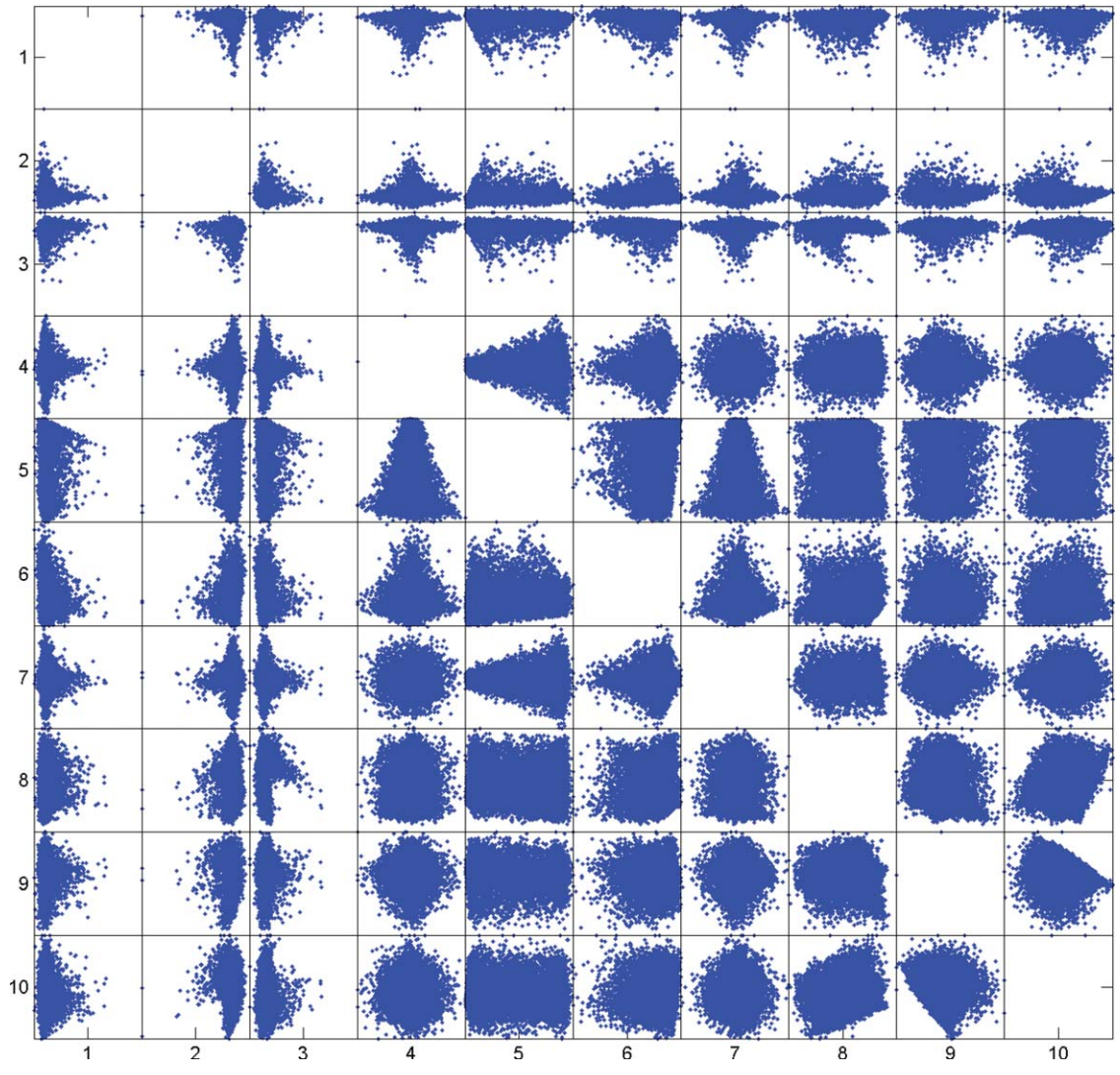

Figure C1. Scatterplots between independent components of the input data matrix constructed from the Leaf River data.

Appendix D: Outliers in the Data Matrix Causing Instability of the Proposed Estimator

[79] Figure C1 provides an illustration of the one-to-one scatter plots obtained between each independent signal of the input data matrix in the real-data study for the Leaf River. The plots reveal both non-Gaussianity and the existence of a large number of outliers that cluster far from the main body of the data, which causes instability in the estimator (negative values are obtained) unless these outliers are removed.

[80] **Acknowledgments.** Support for this study was provided by the National Science Foundation of China (contracts 51025931 and 50939004). Support for the second author was also provided by Andy Pitman and the Australian Research Council through the Centre of Excellence for Climate System Science (CE110001028). We thank the many research groups that kindly made available the ICA source codes online. In particular, we are grateful to Jinsong Chen of the Berkeley Lab Earth Sciences Division for clarifying some statistical concepts, to Hanbo Yang of Tsinghua University for conducting a MK trend test of the Leaf River discharge, precipitation, and pan-evapotranspiration data, and to Qi Wang of Exeter University for providing access to some important references. Finally, we are grateful to Jasper Vrugt, Grey Nearing, Steven Weijs, and several other reviewers of this manuscript, whose comments and suggestions resulted in significant improvements to the presentation and analysis.

References

Abebe, A. J., and R. K. Price (2003), Managing uncertainty in hydrological models using complementary models, *Hydrol. Sci. J.*, *48*(5), 679–692, doi:10.1623/hysj.48.5.679.51450.

Ajami, N. K., Q. Y. Duan, X. G. Gao, and S. Sorooshian (2006), Multimodel combination techniques for analysis of hydrological simulations: application to distributed model intercomparison project results, *J Hydrometeorol*, *7*(4), 755–768, doi:10.1175/JHM519.1.

Ajami, N. K., Q. Y. Duan, and S. Sorooshian (2007), An integrated hydrologic Bayesian multimodel combination framework: Confronting input, parameter, and model structural uncertainty in hydrologic prediction, *Water Resour. Res.*, *43*, W01403, doi:10.1029/2005WR004745.

Amorocho, J., and B. Espildor (1973), Entropy in assessment of uncertainty in hydrologic systems and models, *Water Resour. Res.*, *9*(6), 1511–1522, doi:10.1029/WR009i006p01511.

Bates, B. C., and E. P. Campbell (2001), A Markov chain Monte Carlo scheme for parameter estimation and inference in conceptual rainfall-runoff modeling, *Water Resour. Res*, *37*(4), 937–947, doi:10.1029/2000WR900363.

Bellman, R. E., and R. Corporation (1957), *Dynamic Programming*, Princeton University Press, Princeton NJ.

Beven, K. (1989), Changing ideas in hydrology—The case of physically-based models, *J. Hydrol.*, *105*(1-2), 157–172, doi:10.1016/0022-1694(89)90101-7.

Beven, K., and A. Binley (1992), The future of distributed models—Model calibration and uncertainty prediction, *Hydrol. Process.*, *6*(3), 279–298, doi:10.1002/hyp.3360060305.

Beven, K., P. J. Smith, and A. Wood (2011), On the colour and spin of epistemic error (and what we might do about it), *Hydrol. Earth Syst. Sci. Discuss.*, *8*, 5355–5386, doi:10.5194/hessd-8–5355-2011.

Box, G. E. P., and G. M. Jenkins (1976), *Time Series Analysis: Forecasting and Control*, revised ed., Holden-Day, San Francisco, Calif.

Boyle, D. P. (2000), Multicriteria calibration of hydrological models, Ph.D. thesis, Dep. of Hydrology and Water Resources, Univ. of Arizona, Tucson, Ariz.

Brazil, L. E. (1988), Multilevel calibration strategy for complex hydrologic simulation models, Ph.D. thesis, 217 pp., Colorado State Univ., Fort Collins, Colo.

Bulygina, N., and H. Gupta (2009), Estimating the uncertain mathematical structure of a water balance model via Bayesian data assimilation, *Water Resour. Res.*, *45*, W00B13, doi:10.1029/2007WR006749.

Bulygina, N., and H. Gupta (2010), How Bayesian data assimilation can be used to estimate the mathematical structure of a model, *Stochastic Environ Res. Risk Assess.*, *24*(6SI), 925–937, doi:10.1007/s00477-010-0387-y.

Bulygina, N., and H. Gupta (2011), Correcting the mathematical structure of a hydrological model via Bayesian data assimilation, *Water Resour. Res.*, *47*, W05514, doi:10.1029/2010WR009614.

Burnash, R. J. E., R. L. Ferral, and R. A. Mcguire (1973), *A Generalized Streamflow Simulation System*, 204 pp., Joint Fed.-State River Forecast Center, Sacramento, Calif.

Butts, M. B., J. T. Payne, M. Kristensen, and H. Madsen (2004), An evaluation of the impact of model structure on hydrological modelling uncertainty for streamflow simulation, *J. Hydrol.*, *298*(1-4), 242–266, doi:10.1016/j.hydrol.2004.03.042.

Chapman, T. G. (1986), Entropy as a measure of hydrologic data uncertainty and model performance, *J. Hydrol.*, *85*(1-2), 111–126, doi:10.1016/0022–1694(86)90079-X.

Clark, M. P., A. G. Slater, D. E. Rupp, R. A. Woods, J. A. Vrugt, H. V. Gupta, T. Wagener, and L. E. Hay (2008), Framework for understanding structural errors (fuse): A modular framework to diagnose differences between hydrological models, *Water Resour. Res.*, *44*, W00B02, doi:10.1029/2007WR006735.

Cover, T. M., and J. A. Thomas (2006), *Elements of Information Theory*, John Wiley, Hoboken, N. J.

Duan, Q., and J. Schaake (2002), Results from the 2nd International Workshop On Model Parameter Estimation Experiment (MOPEX), NWS, NOAA, Md.

Duan, Q. Y., N. K. Ajami, X. G. Gao, and S. Sorooshian (2007), Multi-model ensemble hydrologic prediction using Bayesian model averaging, *Adv. Water Resour.*, *30*(5), 1371–1386, doi:10.1016/j.advwatres.2006.11.014.

Ewen, J., G. O'Donnell, A. Burton, and E. O'Connell (2006), Errors and uncertainty in physically-based rainfall-runoff modelling of catchment change effects, *J. Hydrol.*, *330*(3-4), 641–650, doi:10.1016/j.jhydrot.2006.04.024.

Fenicia, F., H. Savenije, P. Matgen, and L. Pfister (2008), Understanding catchment behavior through stepwise model concept improvement, *Water Resour. Res.*, *44*, W01402, doi:10.1029/2006WR005563.

Fernando, T., H. R. Maier, and G. C. Dandy (2009), Selection of input variables for data driven models: An average shifted histogram partial mutual information estimator approach, *J. Hydrol.*, *367*(3-4), 165–176, doi:10.1016/j.jhydrol.2008.10.019.

Freer, J., K. Beven, and B. Ambroise (1996), Bayesian estimation of uncertainty in runoff prediction and the value of data: An application of the GLUE approach, *Water Resour. Res.*, *32*(7), 2161–2173, doi:10.1029/95WR03723.

Georgakakos, K. P., D. J. Seo, H. Gupta, J. Schaake, and M. B. Butts (2004), Towards the characterization of streamflow simulation uncertainty through multimodel ensembles, *J. Hydrol.*, *298*(1-4), 222–241, doi:10.1016/j.jhydrol.2004.03.037.

Gong, W. (2012), Watershed model uncertainty analysis based on information entropy and mutual information, Ph.D. thesis, Dep. of Hydraulic Engineering, Tsinghua Univ., Beijing, China.

Granger, C. W. J., and J. Lin (1994), Using the mutual information coefficient to identify lags in non-linear models, *J. Time Ser. Anal.*, *15*, 371–384.

Gupta, H. V., and H. Kling (2011), On typical range, sensitivity, and normalization of mean squared error and Nash-Sutcliffe efficiency type metrics, *Water Resour. Res.*, *47*, W10601, doi:10.1029/2011WR010962.

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

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

Gupta, H. V., M. P. Clark, J. A. Vrugt, G. Abramowitz, and M. Ye (2012), Towards a comprehensive assessment of model structural adequacy, *Water Resour. Res.*, *48*, W08301, doi:10.1029/2011WR011044.

Hero, A. O., B. Ma, O. Michel, and J. Gorman (2002), Applications of entropic spanning graphs, *IEEE Signal Process. Mag.*, *19*(5), 85–95, doi:10.1109/MSP.2002.1028355.

Hoeting, J. A., D. Madigan, A. E. Raftery, and C. T. Volinsky (1999), Bayesian model averaging: A tutorial, *Stat. Sci.*, *14*(4), 382–401.

Hsu, K. L., H. V. Gupta, X. G. Gao, S. Sorooshian, and B. Imam (2002), Self-organizing linear output map (Solo): An artificial neural network suitable for hydrologic modeling and analysis, *Water Resour. Res.*, *38*(12), 1302, doi:10.1029/2001WR000795.

Hyvarinen, A., and E. Oja (1997), A fast fixed-point algorithm for independent component analysis, *Neural Comput.*, *9*(7), 1483, doi:10.1162/neco.1997.9.7.1483.

Hyvarinen, A., and E. Oja (2000), Independent component analysis: Algorithms and applications, *Neural Netw.*, *13*(4-5), 411–430, doi:10.1016/S0893-6080(00)00026-5.

Hyvarinen, A., J. Karhunen, and E. Oja (2001), *Independent Component Analysis*, John Wiley, New York.

Kaheil, Y. H., M. K. Gill, M. Mckee, and L. Bastidas (2006), A new Bayesian recursive technique for parameter estimation, *Water Resour. Res.*, *42*, W08423, doi:10.1029/2005WR004529.

Kavetski, D., G. Kuczera, and S. W. Franks (2006a), Bayesian analysis of input uncertainty in hydrological modeling: 1. Theory, *Water Resour. Res.*, *42*, W03407, doi:10.1029/2005WR004368.

Kavetski, D., G. Kuczera, and S. W. Franks (2006b), Bayesian analysis of input uncertainty in hydrological modeling: 2. Application, *Water Resour. Res.*, *42*, W03408, doi:10.1029/2005WR004376.

Khan, S., S. Bandyopadhyay, A. R. Ganguly, S. Saigal, D. J. I. Erickson, V. Protopopescu, and G. Ostrouchov (2007), Relative performance of mutual information estimation methods for quantifying the dependence among short and noisy data, *Phys. Rev. E.*, *76*(2), 26209.

Kuczera, G. (1982), On the relationship between the reliability of parameter estimates and hydrologic time-series data used in calibration, *Water Resour. Res.*, *18*(1), 146–154, doi:10.1029/WR018i001p00146.

Leiva-Murillo, J. M., and A. Artes-Rodriguez (2007), Maximization of mutual information for supervised linear feature extraction, *IEEE Trans. Neural Netw.*, *18*(5), 1433–1441, doi:10.1109/TNN.2007.891630.

Leonenko, N., L. Pronzat, and V. Savani (2008), A class of Renyi information estimators for multidimensional densities, *Ann. Stat.*, *36*(5), 2153–2182, doi:10.1214/07-AOS539.

Li, M., D. Yang, J. Chen, and S. S. Hubbard (2012), Calibration of a distributed flood forecasting model with input uncertainty using a Bayesian framework, *Water Resour. Res.*, *48*(8), W08510, doi:10.1029/2010WR010062.

Lin, Z., and M. B. Beck (2007), On the identification of model structure in hydrological and environmental systems, *Water Resour. Res.*, *43*, W02402, doi:10.1029/2005WR004796.

Marshall, L., D. Nott, and A. Sharma (2007), Towards dynamic catchment modelling: A Bayesian hierarchical mixtures of experts framework, *Hydrol. Process.*, *21*(7), 847–861, doi:10.1002/hyp.6294.

Misirli, F., H. V. Gupta, S. Sorooshian, and M. Thiemann (2003), Bayesian recursive estimation of parameter and output uncertainty for watershed models, in *Calibration of Watershed Models, Water Sci.Appl. Ser.*, vol. *6*, edited by Q. Duan, pp. 113–124, AGU, Washington, D. C.

Moore, R. J. (1985), The probability-distributed principle and runoff production at point and basin scales, *Hydrolog Sci. J.*, *30*(2), 273–297, doi:10.1080/02626668509490989.

Nadarajah, S. (2005), A generalized normal distribution, *J. Appl. Stat.*, *32*(7), 685–694, doi:10.1080/02664760500079464.

Nearing, G. S., H. V. Gupta, W. T. Crow, and W. Gong (2013), An Approach to Quantifying the Efficiency of a Bayesian Filter, *Water Resour. Res.*, doi:10.1002/wrcr.20177.

Neuman, S. P. (2003a), Maximum likelihood Bayesian averaging of uncertain model predictions, *Stochastic Environ Res. Risk Assess.*, *17*(5), 291–305, doi:10.1007/s00477-003-0151-7.

Neuman, S. P. (2003b), Accounting for conceptual model uncertainty via maximum likelihood Bayesian model averaging, in *Calibration and Reliability in Groundwater Modelling: A Few Steps Closer to Reality*, edited by K. Kovar and Z. Hrkal, pp. 303–313, Int. Assoc. Hydrol. Sci., Wallingford, Conn.

Pokhrel, P., and H. V. Gupta (2010), On the use of spatial regularization strategies to improve calibration of distributed watershed models, *Water Resour. Res.*, *46*, W01505, doi:10.1029/2009WR008066.

Reichert, P., and J. Mieleitner (2009), Analyzing input and structural uncertainty of nonlinear dynamic models with stochastic, time-dependent parameters, *Water Resour. Res.*, *45*, W10402, doi:10.1029/2009WR007814.

Ruddell, B. L., and P. Kumar (2009), Ecohydrologic process networks: 1. Identification, *Water Resour. Res.*, *45*, W03419, doi:10.1029/2008WR007279.

Scott, D. W. (2004), *Handbook of Computational Statistics—Concepts and Methods*, Springer, New York.

Seo, D. J., H. D. Herr, and J. C. Schaake (2006), A statistical post-processor for accounting of hydrologic uncertainty in short-range ensemble streamflow prediction, *Hydrol. Earth Syst. Sci. Discuss.*, *3*(4), 1987–2035.

Shannon, C. E. (1948), A mathemetical theory of communication, *The Bell System Technical Journal*, *27*, 379–423, 623–656.

Sharma, A. (2000), Seasonal to interannual rainfall probabilistic forecasts for improved water supply management: Part 1—A strategy for system predictor identification, *J. Hydrol.*, *239*(1-4), 232–239, doi:10.1016/S0022-1694(00)00346-2.

Singh, V. P. (1997), The use of entropy in hydrology and water resources, *Hydrol. Process.*, *11*(6), 587–626, doi:10.1002/(SICI)1099-1085(199705)11:6<587::AID-HYP479>3.3.CO;2-G.

Singh, V. P. (2000), The entropy theory as a tool for modelling and decision-making in environmental and water resources, *Water SA*, *26*(1), 1–11.

Sivapalan, M., et al. (2003), IAHS decade on predictions in ungauged basins (Pub), 2003–2012: Shaping an exciting future for the hydrological sciences, *Hydrol. Sci. J.*, *48*(6), 857–880, doi:10.1623/hysj.48.6.857.51421.

Sorooshian, S., and J. A. Dracup (1980), Stochastic parameter-estimation procedures for hydrologic rainfall-runoff models—Correlated and heteroscedastic error cases, *Water Resour. Res.*, *16*(2), 430–442, doi:10.1029/WR016i002p00430.

Sorooshian, S., Q. Y. Duan, and V. K. Gupta (1993), Calibration of rainfall-runoff models—Application of global optimization to the Sacramento soil-moisture accounting model, *Water Resour. Res.*, *29*(4), 1185–1194, doi:10.1029/92WR02617.

Thiemann, M., M. Trosset, H. Gupta, and S. Sorooshian (2001), Bayesian recursive parameter estimation for hydrologic models, *Water Resour. Res.*, *37*(10), 2521–2535, doi:10.1029/2000WR900405.

Vrugt, J. A., H. V. Gupta, W. Bouten, and S. Sorooshian (2003), A shuffled complex evolution metropolis algorithm for optimization and uncertainty assessment of hydrologic model parameters, *Water Resour. Res.*, *39*(8), 1201, doi:10.1029/2002WR001642.

Vrugt, J. A., C. Diks, H. V. Gupta, W. Bouten, and J. M. Verstraten (2005), Improved treatment of uncertainty in hydrologic modeling: Combining the strengths of global optimization and data assimilation, *Water Resour. Res.*, *41*, W01017, doi:10.1029/2004WR003059.

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

Wagener, T., D. P. Boyle, M. J. Lees, H. S. Wheater, H. V. Gupta, and S. Sorooshian (2001), A framework for development and application of hydrological models, *Hydrol. Earth Syst. Sci.*, *5*(1), 13–26.

Wagener, T., N. Mcintyre, M. J. Lees, H. S. Wheater, and H. V. Gupta (2003), Towards reduced uncertainty in conceptual rainfall-runoff modelling: Dynamic identifiability analysis, *Hydrol. Process.*, *17*(2), 455–476, doi:10.1002/hyp.1135.

Weijs, S. V., and N. van de Giesen (2011), Accounting for observational uncertainty in forecast verification: An information-theoretical view on forecasts, observations, and truth B-5010-2008, *Mon. Weather Rev.*, *139*(7), 2156–2162, doi:10.1175/2011MWR3573.1.

Weijs, S. V., G. Schoups, and N. van de Giesen (2010a), Why hydrological predictions should be evaluated using information theory, *Hydrol. Earth Syst. Sci.*, *14*(12), 2545–2558, doi:10.5194/hess-14–2545-2010.

Weijs, S. V., R. van Nooijen, and N. van de Giesen (2010b), Kullback-Leibler divergence as a forecast skill score with classic reliability-resolution-uncertainty decomposition, *Mon. Weather Rev*, *138*(9), 3387–3399, doi:10.1175/2010MWR3229.1.

Xu, L., D. Yuan, and Z. Ye (2010), Parameter sensitivity analysis and multi-objective optimization on Hymod model, *Water Resour. Power*, *28*(11), 15–17.

Young, P. C., and M. Ratto (2009), A unified approach to environmental systems modeling, *Stochastic Environ. Res. Risk Assess.*, *23*(7SI), 1037–1057, doi:10.1007/s00477-008-0271-1.

Young, P. C., P. Mckenna, and J. Bruun (2001), Identification of non-linear stochastic systems by state dependent parameter estimation, *Int. J. Contr.*, *74*(18), 1837–1857, doi:10.1080/00207170110089824.

Zhang, X. S., F. M. Liang, R. Srinivasan, and M. Van Liew (2009), Estimating uncertainty of streamflow simulation using Bayesian neural networks, *Water Resour. Res.*, *45*, W02403, doi:10.1029/2008WR007030.

Zhao, R. J., Y. L. Zhuang, L. R. Fang, X. R. Liu, and Q. S. Zhang (1980), *The Xinanjiang model, paper presented at Hydrological Forecasting of the Oxford Symposium*, IAHS AISH Publ.