Information Analysis of Catchment Hydrological Patterns across Temporal Scales

Baoxiang Pan[a,*], Zhentao Cong[a]

[a] *Institute of Hydrology and Water Resources, Tsinghua University, Beijing*

Abstract

Catchment hydrological cycle takes on different patterns across temporal scales. The interim between daily hydrological processes and long range water-energy correlation pattern requires further examination to justify a self-consistent understanding. In this paper, we use quantized entropy of runoff observations to represent the prior uncertainty in determining catchment's hydrological patterns. Mutual information between runoff observation and catchment's water energy provisions, as represented by precipitation and potential evapotranspiration, is employed to denote the uncertainty decrease given the existed observations. Mutual information between runoff observation and simulation is employed to denote the uncertainty decrease given the models. The differences of these terms, as constrained by the functional transformation of Bayes' theorem, construct the framework of epistemic and aleatory uncertainty in evaluating the observation and simulation systems. We implement this information analysis with daily hydrometeorological data aggregated at temporal scales from 10 days to 1 year in 24 catchments from MOPEX data set to detect the catchments' hydrological patterns revealed by data and two existed water balance models. An improved approach combining K-nearest neighbour method and support vector regression is employed to tackle with high dimensional information item estimation. The estimations of information contents and flows of hydrological

*Corresponding author
Email address: panbaoxiang@hotmail.com (Zhentao Cong)
URL: https://github.com/morepenn (Baoxiang Pan)

terms across temporal scales are related with the catchments' seasonality type. It also shows that information distilled by the monthly and annual water balance models applied here does not correspond to that provided by observations around temporal scale from two months to half a year. This calls for a better understanding of seasonal hydrological mechanism.

Keywords: information theory, temporal scale, hydrological model, Budyko

1. Introduction

A major realm of hydrological community is to figure out the components of hydrological cycle. Each component should be determined either by observation or an independent governing equation to guarantee the solvability of the problem. The accuracy of observation and domain of governing functions usually change with scales. The term *scale* here refers to a characteristic time (or length) of a process, observation or model [1]. Besides the universal conservation equation that suits for any spatial and temporal scale that we care about, each process-oriented hydrological model seeks for the proper complementary constitutive functions that govern the water movement at scales it focuses on. There has long been two perspectives in reaching a temporal scale harmonious explanation of hydrological processes, specifically, bottom-up and top-down. We make a brief review of them before introducing the information theoretical framework to quantize the uncertainty in seeking for the interface of the two groups of models across temporal scales.

Since the blueprint brought forward by Freeze and Harlan [2], every advance in observation technique and calculation capacity would revitalize the seated reductionism intuition among hydrologists, which aims at reproducing the hydrological process in the greatest spatial and temporal detail, hoping that larger patterns are self-evident when "integrating" the calculating units along the spatial and temporal paths. However, we could not guarantee the universality of the phenomenological constitutive functions or the accuracy of the integrating spatial and temporal paths. The outputs of the distributed models could not

verify the vast assumptions or parameterization schemes to support the model as a scientific attempt, nor could they provide insights of hydrological patterns at larger scales.

Hydrological behaviour of some parts within a catchment tends to cancel out the behaviour of other parts, with the result that it does not matter too much what happens on the low level, because most anything will yield similar high-level behaviour[3]. Given this, many conceptual hydrological models have been brought forward to provide coarser but valuable simulation without requiring detailed inputs or strong computation capacity. The mathematical analysis of the simplified forms offers an insight into the catchment hydrological mechanism that is blotted when aggregating the mass outputs produced by the distributed models[4, 5]. On the other hand, the simplicity also crippled such models from making down-scaling analysis. Their structures must be extended in order to depict microscopic hydrological processes.

A paradigm of the declarations above is Budyko Curve[6]. The curve links climate to annual catchment evaporation and runoff by characterizing an empirical relationship between the ratio of mean annual actual evaporation to mean annual rainfall and mean annual dryness index of the catchment[7]. A series of specific forms of Budyko Curve are obtained by selecting special solutions of the partial differential equation set constrained by the extreme boundary conditions and Buckingham Π Theorem[8, 9, 10]. This constitutive equation together with the water conservation function where soil moisture storage change is neglected constitute a determined equation set that depicts the water-heat correlation pattern at annual mean temporal scale[11, 12].

The strong assumption of stable soil moisture storage has caused controversy and limited the application of the model at seasonal or monthly temporal scales. Even at annual scale, water balance analysis using Budyko-type curve reveals that the aridity index does not exert a first order control in most of the catchments[13]. Former critics basically blame the deviation for excluding the impact of the changing soil moisture[14, 15]. By including the soil moisture storage term, some seasonal and monthly water balance models were

developed[16, 17, 18], which serve as temporal scale gap-fillers of the long term water-heat correlation pattern and single precipitation-runoff phenomenon focused hydrological models. As have been declared, the inclusion of any new term brings an increase to the degrees of freedom of the problem, which should be complemented either by observation or an independent complementary function. The huge cost of the former forces us to accept a less convincing but workable new constitutive function that governs the soil moisture change during the iterative simulation of water movement and storage. The rationale of these functions are gaining hydrologists' concern due to a similar Darwinian ideological origin with the Budyko Curve[19]. However, their specific dominant temporal scales and accuracy remain ambiguous.

Given the pros and cons of the bottom-up and top-down models, we are faced with the following problems in reaching a temporal scale consistent hydrological simulation system: (1), how catchment hydrological patterns evolve as temporal scale expands, explicitly, how important hydrological items connect with each other at different temporal scales; (2), to what accuracy the data support the patterns; (3), to what extent the existing models capture these patterns.

This research tries to give primary responses to these questions within the discipline of information theory.

The term *information* got mathematicized by Claude E. Shannon in 1948[20]. The notion that information is the combination of bits and context[21] sets the theoretical foundation of the digit revolution and broadens to find applications in many other areas, including hydrology and water resources[22, 23, 24].

Specific to hydrological simulation, information theory has been applied for model evaluation and uncertainty analysis as far back as the 1970s [25, 26, 27, 28, 29, 30] . Gong developed a comprehensive model evaluation framework based on entropy and mutual information [31] . In this framework, the uncertainty caused by the insufficiency and inaccuracy of data is attributed to *Aleatory Uncertainty*, while that caused by imperfect data processing is attributed to *Epistemic Uncertainty*. The sum of the two terms depicts the whole uncertainty of hydrological simulation.

$$Aleatory\ Uncertainty = H(X_o) - I(X_o; X_i) \tag{1}$$

$$Epistemic\ Uncertainty = I(X_o; X_i) - I(X_o; X_s) \tag{2}$$

Here X_o, X_i, X_s represent random variables of the observed output, input terms and the simulated term of a specific model. H denotes entropy. The entropy of a discrete random variable represents the average information content (uncertainty) of it. I is mutual information, which represents the information that two stochastic variables share, or the uncertainty loss of one variable due to the knowledge of the other.

The definition provides a crystalline framework to evaluate the observation and simulation systems. However, it must blend into the existing uncertainty estimation knowledge systems before its wide acceptance. Besides, the hydrological context in which these terms make sense and the specific calculating techniques should be strictly examined.

Hydrological terms are usually taken as continuous random variables at temporal or frequency domains that are observable over quantized coordinate points. Hydrological series represented at different coordinates hold different entropy and mutual information. It is impossible to tell the aleatory and epistemic uncertainty without clarifying the specific context or prior beliefs[32]. It should also be noted that the intuitive significance of discrete entropy could not be blindly generalized to differential entropy. We will address these issues in the following sections.

In addition to the theoretical considerations brought forward above, the technical challenge of high dimensional information term estimation is never an easy task. The strategy Gong adapts is to make a linear transformation of the original high dimensional term into independent vectors using Independent Component Analysis Algorithm (ICA)[33]. According to the *chain rule* of entropy, the sum of the entropies of the independent components differs from the entropy of the original term by $log|det(A)|$, where A is the ICA transform matrix. However, the ICA algorithm is no more than a linear transformation,

the vectors of the transformed matrix is very likely to be dependent when the original terms are highly non-linearly correlated. Thus, the method would overrate the entropy of the original data for neglecting the inner relevance among vectors in the transformed matrix. Besides, the indirect calculation of mutual information through entropies could introduce error accumulation.

In this research, we provide both intuitive and formal explanations of the *Aleatory Epistemic Uncertainty Evaluation Framework*. We will see that this framework is no more and no less than a functional transformation of the classical Bayes' theorem. In order to make senses of the "bits" estimated in this framework, we restrict the context to hydrological series (precipitation, potential evapotranspiration and runoff observations) laid at time domain sampling points. The original daily series are re-aggregated into series with temporal scales from ten days to a year (no moving cluster). The information contents of these terms at various temporal scales are represented with quantized entropy. The accuracy of the quantization scheme is determined by practical needs as is clarified in the following sections. We employ mutual information between different hydrological terms to quantify the information flows within the hydrological cycle at specific temporal scales. Given the drawbacks of the existed high dimensional information estimators, we adapt a k-nearest neighbour distance method[34], which uses the *distances* between samples to estimate high dimensional mutual information directly. Since the variable space is composed of different hydrological terms, we could not take it as an Euclidean Space and measure the sample points' distances with the popular *norms*. Considering the mathematical significance and strong information extraction ability of the support vector regression[35], we apply it to depict the *distances*. The theoretical clarification is in the second section. Information terms to be calculated are followed. Finally, we discuss the interpretations of the estimations and respond the questions we put forward above.

2. Methodology

2.1. Bits in Hydrological Simulation Context

It is intuitively believed that an infrequent sample of a random variable provides more surprisal, or information. The mathematical expression of this common sense is that information provided by an observation should be a decreasing function of its probability. If we further require the additive property of information between independent events, the form of information content attributed to a sample with probability p should be $-logp$. Thus, the average information content of random variable X is:

$$H(X) = -\Sigma p(x) log p(x) \tag{3}$$

$$h(X) = -\int f(x) log f(x) dx \tag{4}$$

$H(X)$ and $h(X)$ denote discrete and continuous Shannon Entropy, measured in bits for logarithm base 2.

While discrete entropy directly characterizes the average information content each observation brings to our knowledge, things become a little tricky for continuous situation. For continuous random variable, the probability of each value in the sample space is 0, since $-logp \to \infty$ as $p \to 0$, the information provided by each observation is infinite.

As is shown in Figure 1, let X^{Δ} be the discrete stochastic variable by scattering a continuous random variable X into bins with length of Δ in its probability density function image, we have:

$$H(X^{\Delta}) \to h(X) - log\Delta, \quad as\ \Delta \to 0 \tag{5}$$

This tells that differential entropy itself can not represent the average uncertainty of the information resource or the average information provided by each datum. However, if we only require an interval estimation, $h(X) - log\Delta$ would reveal the information content required to describe X to $-log\Delta$ bit accuracy[36].

Here $-log\Delta$ bit accuracy means X takes a same value in a bin-width of Δ in the p.d.f. curve.

The other term we apply here is mutual information. Its discrete and continuous forms are as follows:

$$I(X;Y) = \sum_{x,y} p(x,y) log \frac{p(x,y)}{p(x)p(y)} \tag{6}$$

$$I(X;Y) = \int \int f(x,y) log \frac{f(x,y)}{f(x)f(y)} dxdy \tag{7}$$

As can be derived:

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

E denotes expectation. The latter term in the middle and left part of equation 8 are called conditional entropy, which represents the residual uncertainty of a random variable given the knowledge of the other. Thus, $I(X;Y)$ denotes the uncertainty decrease of X given the knowledge of Y, and vice versa. It is always non-negative according to Jesen Inequality [36].

The continuous mutual information $I(X;Y)$ is the limit of the discrete mutual information of partitions of X and Y as these partitions become finer and finer. Thus it still represents the amount of discrete information that can be transmitted over a channel that admits a continuous space of values.

The definitions above are closely related with Bayesian Statistics. Bayes' theorem is stated mathematically as the following equation:

$$P(A|B) = P(A) \times \frac{P(B|A)}{P(B)}, \tag{9}$$

where A and B are events. $P(A)$ denotes prior distribution, $P(A|B)$ denotes posteriori distribution. $\frac{P(B|A)}{P(B)}$ is called standardised likelihood. Equation 9 quantizes how a subjective degree of belief should rationally change to account of evidence.

As have been declared, each probability distribution corresponds to its uncertainty or information content as is defined in Equation 3 or 4. We implement

the transformation on both sides of Equation 9. The specific steps are as follows:

(1) Make logarithmic transformation of Equation 9:

$$logP(A|B) = logP(A) + log\frac{P(AB)}{P(A)P(B)} \tag{10}$$

(2) Multiply each item by $-P(A,B)$:

$$-P(A,B)logP(A|B) = -P(A,B)logP(A) - P(A,B)log\frac{P(AB)}{P(A)P(B)} \tag{11}$$

(3) Sum or integrate each item in the sample space:

$$-\sum_A\sum_B P(A,B)logP(A|B) = -\sum_A\sum_B P(A,B)logP(A) - \sum_A\sum_B P(A,B)log\frac{P(AB)}{P(A)P(B)} \tag{12}$$

or

$$-\int\int P(A,B)logP(A|B)dAdB = -\int\int P(A,B)logP(A)dAdB - \int\int P(A,B)log\frac{P(AB)}{P(A)P(B)}dAdB \tag{13}$$

which simplifies to

$$H(A|B) = H(A) - I(A,B) \tag{14}$$

Given the correspondence of Equation 9 and 14, $H(A|B)$ represents posterior uncertainty, $H(A)$ represents prior uncertainty, $I(A,B)$ represents information connection between the two random variables.

In hydrological simulation, a general goal is to produce accurate runoff simulation with inputs from hydrometeorological series, underlying surface observations or other information sources. This is not only for the practical objective of efficient water resources utilization, but also for the scientific value that once the runoff process were characterized, each component into which the precipitation is partitioned gets determined.

The information theoretical paraphrase of this notion is that the information content of runoff observation depicts information required to figure out the catchment's hydrological compositions, which could be decreased due to the

information contribution of the input observations. The contribution is determined by the observation qualities and data processing procedures. Noise introduced due to observation inaccuracy and insufficiency is denoted as *Aleatory Uncertainty*, while that introduced by imperfect data processing is denoted as *Epistemic Uncertainty*

Hydrological series encoded in different context can take up different amounts of bits. For example, observations show that soil moisture dynamics in the Fourier domain require by far less coefficients to explain a specified variance level when compared to their time domain counterpart[37], thus the encoding of soil moisture requires far less bits in the frequency domain than in the time domain. In this research, we restrict our attention to hydrological observations sampled discretely along the time domain base. The sample space is built on the aggregated coordinates without considering seasonal fluctuation or any other temporal inconsistencies. This will increase the estimated information contents for neglecting the inner structures, but the endeavour to compress the data to their "true information contents" is endless for its logical paradox[38]. It will also impair the criterion's generality in evaluating the observation and simulation system.

With the sample spaces constructed, we apply the introduced terms to quantify the information contents and connections of catchment hydrological variables across temporal scales. The specific values to be estimated are listed in table 1. All the estimations are implemented at temporal scales from 10 days to a year. This range bypasses the difficulty of estimating discrete-continuous hybrid distributed daily precipitations[39] while incorporating significant temporal scales in detecting long term catchment hydrological behaviours.

Here P and EP denote precipitation and potential evapotranspiration random variables. R and Rs denote observed and simulated runoff random variables.

The estimations are classified into two groups. In the observation focused group, $h(R)$ sets the base for prior uncertainty estimation. Mutual information between runoff observation and different input observation terms are

estimated to analyse their respective information contributions in decreasing the prior uncertainty. Specifically, we used $I(R;P,PE)-I(R;P)$ to represent the information contribution of including PE in hydrological simulation, while $I(R;P,P_{former},PE,PE_{former})-I(R;P,PE)$ is used to represent the information contribution of including former calculating steps' hydrological conditions. For small temporal scale hydrological simulation, hydrological conditions in former calculating steps, in the form of lagged values (such as $P_{t-1},...,P_{t-n},PE_{t-1},...PE_{t-n},R_{t-1},...R_{t-n}$), could exert significant influence on current hydrological responses. We include a maximum of 6 lagged steps in the case study.

In the model focused group, two typical water balance models are applied to check their information processing capacities. The Two Parameters Water Balance Model(TPWB)[17] adapted an adjusted Ol'dektop equation[40] to depict the evapotranspiration and runoff generation at a monthly temporal scale and achieved satisfying performance. The constructive functions of TPWB are as follows:

$$E = C \times PE \times tanh(\frac{P}{PE}) \tag{15}$$

$$R = (S_{t-1} + P - E) \times tanh(\frac{S_{t-1} + P - E}{SC}) \tag{16}$$

C and SC are the two parameters. As is shown, TPWB uses an iterative structure. S_t is the state variable at time step t, which is used to represent the influence of former hydrological conditions. We compare $I(R;P,PE,S)$ with $I(R;P,P_{former},PE,PE_{former})$ to discern the model's capacity of distilling information from former inputs. We also compare $I(R;P,PE,S)$ with $I(R;Rs)$ to discern the model's capacity of digesting the distilled state variable.

The Budyko Model is the combination of Budyko Curve and water balance equation as described above. We adapt Choudhury-Yang Equation in this research.

2.2. Quantization Schemes for Runoff Differential Entropy

Since runoff observations are taken as continuous random variables in our hydrological simulation context, $h(R)$ can not characterize the average information content each runoff observation brings to our knowledge of the hydrological behaviour. Certain quantization schemes should be pre-setted to justify the significance of the estimation. We apply two quantization schemes here:

1. Absolute constant resolution across temporal scales.
2. Relative constant resolution across temporal scales.

As has been clarified, a $-log\Delta$ bit accuracy description of a continuous random variable X depicts it to the resolution that X takes a same value in a bin-width of Δ in the its p.d.f. curve.

For Quantization Scheme 1, the bin-width Δ into which we discretize the runoff observation data stays the same as the evaluating temporal scale expands.

For Quantization Scheme 2, the bin-width Δ into which we discretize the runoff observation data is proportional to the mean value of the runoff observation at the specific temporal scale. We further assume that the mean value of the runoff random variable to be proportional to its temporal scale. In this way, the discretization bin-width is proportional to the temporal scale. The quantization correction term is proportional to the logarithm of the temporal scale according to equation 5.

Thus, given two scales m and n into which we aggregate the daily runoff observation data, the entropy differences in depicting them with quantization schemes introduced above are:

Quantization Scheme 1

$$H(R_m) - H(R_n) = h(R_m) \quad - \quad h(R_n) \tag{17}$$

Quantization Scheme 2

$$H(R_m) - H(R_n) = h(R_m) - h(R_n) - log\frac{m}{n} \tag{18}$$

2.3. High Dimensional Mutual Information Estimator

Due to the curse of dimensionality, the high dimensional terms in table 1 could not be accurately estimated with primitive information estimators such as bin-counting or kernel density approaches. Besides, we want to make a direct estimation of mutual information to avoid error assumption. In this research, we adapt a widely accepted non-plug-in mutual information estimator and make some adjustments for its application in hydrological simulation context. The original method is derived from the k nearest neighbour entropy estimation approach [34]:

$$I(X,Y) = \psi(k) - N^{-1}\sum_{i=1}^{N}[\psi(n_x(i)+1) + \psi(n_y(i)+1)] + \psi(N) \tag{19}$$

Here $\psi(x)$ is the digamma function, $\psi(x) = \Gamma(x)^{-1}d\Gamma(x)/dx$. k is order of nearest neighbour, $n_x(i)$ and $n_y(i)$ are the numbers of samples that are within the k-th nearest criss-cross surrounding sample point i. For this research, k takes 4 in accordance with Hyvärinen's implementation.

An intuitive explanation of equation 19 is that it estimates mutual information with statistics that depict the average concentrating density of each window opened around a sample point. Numerical experiments show that even less than 30 sample size produces satisfying results. For a strict proof, please refer to Kraskov(2004).

We should notice that the widths of the windows are determined by the ordered *distance functions* we select to define the distances between samples. Since each dimension of a single sample represents different hydrological terms, the hydrological modelling space can not be taken as Euclidean. Thus, the Euclidean *norms* can not reflect the *geodesic distances* between points.

One approach to make a justifiable distance between samples is to map the points to their feature space through a certain transformation and calculate the *norm* in that space. The linear regression from the transformed points to the simulating variable forms an integrated model. This is in fact the idea of non-linear support vector regression(SVR). Non-linear SVR uses the kernel trick to

implicitly map its inputs into high-dimensional feature spaces. The method has been proven to be of great accuracy in runoff generation modelling[41, 42, 43, 44] and long range runoff simulation[45]. Thus, we apply the following function to depict the distance between two model input samples x_1 and x_2:

$$SVM\_Metric(x_1, x_2) = |f(x_1) - f(x_2)| \tag{20}$$

Here $f(x)$ is the support vector regression function that fit the input to the output of the sample.

In practice, the support vector regression is implemented using the libsvm package[46]. We select the radial basic function kernel to make the non-linear transformation in the support vector regression algorithm for its satisfying performance. The data are first scaled to $[-1, 1]$ to balance the impact of different dimensional terms. The result of SVR is sensitive to the penalty function parameter c and kernel parameter g, both of which are auto calibrated with particle swarm optimization algorithm[47]. To avoid overfitting, we apply 3 cross validation in the support vector regression parameter estimation.

The calculating steps are as follows:

(1) Re-aggregate the original hydrological data (daily precipitation, potential evapotranspiration and runoff) into different temporal scale terms.

(2) Calculate the model irrelevant information terms at these temporal scales.

(3) Implement hydrological simulation and calculate the model relevant mutual information terms.

The specific procedure of high dimensional mutual information estimating is as follows:

(1) Train support vector machine to find suitable mapping type (by choosing kernel function) and parameters.

(2) Use the trained support vector machine to estimate the distances between high dimensional inputs using equation 20.

(3) Estimate mutual information with equation 19.

All the codes are available at the github URL:
http://github.com/morepenn/matlab/tree/master

3. Data

We implement our simulation and estimation with aggregated daily hydrological records (including precipitation, potential evapotranspiration and runoff) from the MOPEX data set[48]. Given their annual water-energy distribution patterns, the selected 24 catchments are classified into 4 groups, explicitly, weak seasonality with synchronous rainfall energy distribution(WS), weak seasonality with asynchronous rainfall energy distribution(WA), strong seasonality with synchronous rainfall energy distribution(SS) and strong seasonality with asynchronous rainfall energy climate (SA). The classification standard is based on the amplitude and phase of the average daily rainfall fitted with sine curve. If the amplitude is less than 0.45, the catchment is taken as weak seasonality. If the phase of rainfall is inverse to that of potential evapotranspiration, it is taken as asynchronous rainfall energy climate type. The detailed information of the catchments are listed in table 2.

4. Results

4.1. Aleatory Uncertainty Across Temporal Scales

4.1.1. Aleatory Uncertainty of Absolute Constant Resolution

If we pre-require absolute constant resolution of runoff estimation, which means that the simulation deviation rate is constant across temporal scales as defined in Equation 18, the estimated *Aleatory Uncertainty* is shown as follows:

In each subgraph from Figure 2, the abscissa represents the input steps, for example, n input steps means that we use the current and $(n-1)$ lagging steps' input observations to decrease the uncertainty of runoff estimation. The ordinate represents the estimating temporal scale, which varies from 10 days to a year.

As can be depicted from the estimations above, *Aleatory Uncertainty* increases as the simulating temporal scale expands, decreases as more previous input observations are incorporated in the estimation.

4.1.2. Aleatory Uncertainty of Relative Constant Resolution

If we pre-require relative constant resolution of runoff estimation, which means that the simulation deviation rate is proportional to its mean value across temporal scales as defined in Equation 18, the estimated *Aleatory Uncertainty* is shown as follows: The significances of the coordinates in each subgraph are the same as those in Figure 2.

As can be depicted from the estimations above, *Aleatory Uncertainty* decreases as temporal scale expands. The incorporation of lagging input observations can decrease the value to different extent in different catchments with different input terms.

4.2. Epistemic Uncertainty Across Temporal Scales

The estimated *Epistemic Uncertainty* across temporal scales are shown as follows: For TPWB model, the peak value appears around temporal scales from 2 months to half a year. This calls for a more efficient information distiller, or put it in other words, a more efficient model to depict seasonal hydrological mechanism.

For the Budyko model, its *Epistemic Uncertainty* is much larger than that of TPWB model at temporal scales of less than half a year in 11 out of 14 asynchronous climate catchments. The difference is less significant for temporal scales of larger than half a year. In the rest 3 asynchronous climate catchments and 14 synchronous climate catchments, the *Epistemic Uncertainty* of Budyko model is smaller than that of TPWB model. However, the difference shows no abrupt change as the simulating temporal scale expands.

For both models, *Epistemic Uncertainty* is non-negative.

5. Discussion

5.1. Prior Uncertainty—Runoff Entropy

The baseline of uncertainty estimation is constructed by quantized runoff entropy as shown in the following figure. It depicts the uncertainty when no further prior assumption is incorporated given the estimating context, or in the terminology of Bayesian statistics, it tells the prior uncertainty.

All the estimations are relative values on a same base of 0 bit accuracy at 10 days temporal scale. The runoff entropy of absolute constant resolution increases with temporal scales. The increasing rate decreases as scale expands, making the curve take on a logarithm shape. This is the dominant factor that cause the increasing trend of *Aleatory Uncertainty* in Figure 2.

For relative constant resolution, most of the estimations reach their maximum points at temporal scales varying from 1 to 2 months, except for 5 out of 7 catchments from the asynchronous rainfall energy climate group, which take on a monotonically decreasing trend across the estimated temporal scales. The decreasing rates of entropy with temporal scales in catchments from synchronous climate groups are not as significant as those from asynchronous groups.

5.2. Mutual Information Between Runoff Observation and Input Data

Mutual information between runoff observation and hydrometeorological inputs are shown in the following figure. They depict the uncertainty decrease given the input observations. The significances of the coordinates are the same as those in Figure 2 and 3. As can be observed, mutual information between runoff observation and input observations increases as more observation terms (PE and R_{former}) and lagged input data are incorporated. To clarify the information contribution of each introduced term, we take two dissection schemes on graphs in Figure 6.

5.2.1. Mutual Information with Different Input Terms

The first dissection scheme checks the information contribution of incorporating PE and R_{former} into mutual information estimation. This is imple-

mented by making differences between columns in Figure 6:

For the estimations in all the 10 weak seasonality catchments and 5 out of 14 strong seasonality catchments, the inclusion of PE contributes more to increasing mutual information between runoff and input data at temporal scales of less than half a year. This contribution distributes more uniformly across temporal scales in the left 9 strong seasonality catchments.

The incorporation of former runoff contributes a lot to decrease uncertainty at small temporal scales in some of the catchments. This salient effect vanishes quickly as temporal scale expands. We attribute this mutation to the runoff convergence influence.

5.2.2. Mutual Information with Different Lagged Input Steps

The second dissection scheme checks the information contribution of including lagged inputs into mutual information estimation. This is implemented by making differences between mutual information estimated with different input steps, for instance, the nth spline in each graph from table 5.2.2 equals to the difference of the $(n+1)$th spline and nth spline in the corresponding graph from Figure 6. Subgraphics in Figure 8 depict the information contribution rate $(\frac{\partial I}{\partial Input\_Step})$ when including lagged observations. They represent the hydrological connections between temporally succeeded hydrological processes. The rate is larger than 0 because of the disclosure of the hydrological cycle. It decreases as more input steps are incorporated. The decreasing rate reflects the "memory length" of soil moisture.

5.3. Mutual Information Between Runoff Observation and Simulation

The mass content of hydrometeorological input observations can be distilled by models in the form of runoff simulation series. The noise introduced by imperfect data processing is denoted as *Epistemic Uncertainty*, which could be represented by the difference between mutual information provided by input data and the simulation. The former item has been estimated in the previous part. Mutual information between runoff observation and simulations generated

by TPWB model and Budyko Model at temporal scales from 10 days to a year are listed below:

As have been declared, in TPWB model, the state variable S is used to represent the influence of former hydrological conditions. Estimations show that $I(R; P, PE, S)$ is always larger than $I(R, R_s)$, which means that the model can not fully digest the distilled state variable to make accurate estimations. The relationship between these two items and temporal scales take on different patterns in catchments of different climate types. Both the two estimations increases with temporal scales in synchronous rainfall energy catchments(except Catchment 07019000). In asynchronous catchments, as temporal scale expands, the values decrease until the scale of half a year. Then, they increase in weak seasonality group or stay relatively stable in strong seasonality group. This explains the abrupt change of *Epistemic Uncertainty* differences between TPWB model and Budyko model around temporal scales of half a year.

We should point out that in two strong asynchronous seasonality catchments (12459000,13337000), although $I(R; P, PE, S)$ decreases as temporal scale expands from 10 days to half a year, $I(R, R_s)$ is much smaller than $I(R; P, PE, S)$. It stays low and relatively stable as temporal scale expands. The strong capacity of distilling information from former inputs does not guarantee a equal efficient digestion of the distilled item in these 2 catchments.

For Budyko Model, $I(R, R_s)$ increases with temporal scale while being smaller than that of TPWB (except Catchment 11025500 where the drought coefficient is extreme high).

5.4. Uncertainty Across Temporal Scales

The estimation of catchment hydrological processes is supported by the the observation and simulation system. There is no essential distinction between the uncertainties introduced in the two systems. Both of them are caused by insufficiency or inaccuracy of data, and are restricted by the data processing inequality[36].

The data-processing inequality states that if random variables X,Y,Z form a Markov chain in that order (denoted by $X \to Y \to Z$), then:

$$I(X;Y) \geq I(X;Z) \tag{21}$$

Since:

$$R \to Input_{original}, Input_{new} \to Input_{original} \tag{22}$$

We have:

$$I(R; Input_{original}, Input_{new}) \geq I(R; Input_{original}) \tag{23}$$

Here $Input_{original}$ denotes the original input observation items, $Input_{new}$ denotes the new introduced items.

Inequality 23 guarantees the non-negativity of items in table 5.2.1 and table 5.2.2 (the few negative points are attributed to estimation error). These values quantize the information contribution of hydrometeorological items from current and former calculating steps. As have been declared, the contributions also vary between catchments and temporal scales, though some common patterns exist in catchments of similar seasonality characteristics.

The data processing inequality can also be employed to explain patterns shown in Table 4.2 and Table 5.2. The state variable S in TPWB is the function of previous hydrological terms $Input_{previous}$. Its simulation R_s is the function of S and current hydrometeorology inputs $Input_{current}$. Thus:

$$R \to Input_{previous}, Input_{current} \to S, Input_{current} \to R_s \tag{24}$$

which could be simplified as:

$$R \to Input \to S, Input_{current} \to R_s \tag{25}$$

given the data-processing inequality, we have:

$$I(R; Input) \geq I(R; S, Input_{current}) \geq I(R; R_s) \tag{26}$$

The whole inequality explains the non-negativity of *Epistemic Uncertainty* in both models while the latter one explains why $I(R_t; P_t, PE_t, S_t)$ is no smaller than $I(R, R_s)$ in TPWB.

Though there is no mathematical difference between the uncertainty sources, it is helpful to distinct the uncertainties introduced by observation and simulation in order to make corresponding improvements[31].

Specific to the temporal scale analysis of hydrological patterns, the origin of observation uncertainty, defined as *Aleatory Uncertainty*, can be attributed to two sources. Th first one is observation bias. For consistent observations with no system error, this uncertainty source weakens as temporal scale expands due to the large number law. The daily observation errors tend to set off when aggregating them together.

The other origin is the inherent uncertainty caused by the coarse temporal scale. A simple aggregating of water quantity of different hydrological terms can not exert a strong control of the system. The variability of their temporal distribution takes effect in increasing the uncertainty.

Given the reliability of the MOPEX dataset, the latter uncertainty source is viewed as the dominant factor. In other words, the *Aleatory Uncertainty* is mainly caused by data insufficiency rather than inaccuracy for large temporal scales.

There are more than one models processing same observations. These models seek balance between structure complexity and accuracy. The structure complexities of the two models applied here are different. The TWPB model uses a state variable to represent former hydrological conditions. The Budyko model assumes a closed hydrological cycle in its calculating temporal scales. The ignorance of soil moisture profit and loss crippled its efficiency in monthly hydrological simulation. Graphics in Table 5.3 show their capacities in distilling information from the input observations. In models with iterative structures,

this capacity is divided into two parts, the first stresses the ability to extract lagged inputs' influences, the second stresses the ability to digest the distilled variable.

6. Conclusion

This research explores the hydrological patterns revealed by observations and models at temporal scales from 10 days to a year with an information theoretical approach. We apply the quantized differential entropy of runoff observations to represent the prior uncertainty in figuring out the catchment's hydrological compositions. Mutual information between hydrometeorological observations and runoff is applied to denote the best performance we could potentially reach given the existed observation system. The non-linear support vector regression processed data is taken as sufficient statistic in depicting high dimensional mutual information. The performances of two existed water balance models are represented by mutual information between runoff observations and their simulations. All the estimations are constrained by the data-processing inequality.

The estimations revealed the existence and flows of information in catchment across temporal scales, which could be used to explain hydrological patterns in the framework of aleatory and epistemic uncertainty. Results showed that these patterns are related to the seasonality type of the catchments, which calls for more case studies to figure out the mechanism under the phenomenon. It also shows that information distilled by the monthly and annual water balance models applied here does not correspond to the information provided by input observations around temporal scale from two months to half a year. This calls for a better understanding of seasonal hydrological mechanism.

7. Acknowledgements

We are grateful to the financial supports offered by the National Science Foundation of China(51479088, 51179083 and 91225302). In addition, the

MOPEX dataset applied here is available in the following link: ftp://hydrology.nws.noaa.gov/. The catchment serial numbers are the same as the original dataset. For detailed data description and estimation results, please refer to the following Github repository: http://github.com/morepenn/matlab/tree/master . Finally, we should express our special thanks to Hoshin V. Gupta from University of Arizona for his insights and kindness.

References

[1] G. Blöschl, M. Sivapalan, Scale issues in hydrological modelling: a review, Hydrological processes 9 (3-4) (1995) 251–290.

[2] R. A. Freeze, R. L. Harlan, Blueprint for a physically-based digitally-simulated hydrologic response model, Journal of Hydrology 9 (3) (1969) 237–258.

[3] D. R. Hofstadter, Gödel, Escher, Bach, An Eternal Golden Braid, Penguin, 2000.

[4] A. M. J. Gerrits, H. H. G. Savenije, E. J. M. Veling, et al, Analytical derivation of the budyko curve based on rainfall characteristics and a simple evaporation model, Water Resources Research 45 (4).

[5] X. Xu, D. Yang, H. Yang, et al, Attribution analysis based on the budyko hypothesis for detecting the dominant cause of runoff decline in haihe basin, Journal of Hydrology 510 (2014) 530–540.

[6] M. Budyko, The heat balance of the earth's surface, Soviet Geography 2 (4) (1961) 3–13.

[7] D. Wang, N. Alimohammadi, Responses of annual runoff, evaporation, and storage change to climate variability at the watershed scale, Water Resources Research 48 (5).

[8] B. Fu, Lansurface evaporation calculation, Meterology Science(China) 5 (1) (1981) 23–31.

[9] B. J. Choudhury, Evaluation of an empirical equation for annual evaporation using field observations and results from a biophysical model, Journal of Hydrology 216 (1) (1999) 99–110.

[10] H. Yang, D. Yang, Z. Lei, et al, New analytical derivation of the mean annual water-energy balance equation, Water Resources Research 44 (3).

[11] L. Zhang, W. R. Dawes, Response of mean annual evapotranspiration to vegetation changes at catchment scale, Water resources research 37 (3) (2001) 701–708.

[12] D. Yang, F. Sun, Z. Liu, et al, Analyzing spatial and temporal variability of annual water-energy balance in nonhumid regions of china using the budyko hypothesis, Water Resources Research 43 (4).

[13] S. Tekleab, S. Uhlenbrook, Y. Mohamed, et al, Water balance modeling of upper blue nile catchments using a top-down approach, Hydrology and Earth System Sciences 15 (7) (2011) 2179–2193.

[14] A. Sankarasubramanian, R. M. Vogel, Annual hydroclimatology of the united states, Water Resources Research 38 (6) (2002) 19–1.

[15] A. Sankarasubramanian, R. M. Vogel, Hydroclimatology of the continental united states, Geophysical Research Letters 30 (7).

[16] H. A. Thomas, Improved methods for national water assessment WR15249270[A].

[17] L. Xiong, S. Guo, A two-parameter monthly water balance model and its application, Journal of Hydrology 216 (1) (1999) 111–123.

[18] L. Zhang, P. N., H. K., et al, Water balance modeling over variable time scales based on the budyko framework–model development and testing, Journal of Hydrology 360 (1) (2008) 117–131.

[19] D. Wang, Y. A. Tang, A one-parameter budyko model for water balance captures emergent behavior in darwinian hydrologic models, Geophysical Research Letters 41 (13) (2014) 4569–4577.

[20] C. E. Shannon, A mathematical theory of communication, ACM SIGMOBILE Mobile Computing and Communications Review 5 (1) (1948) 3–55.

[21] R. Bryant, O. H. D. Richard, Computer systems: a programmer's perspective, Prentice Hall, 2003.

[22] V. P. Singh, The use of entropy in hydrology and water resources, Hydrological processes 11 (6) (1997) 587–626.

[23] V. P. Singh, The entropy theory as a tool for modelling and decision-making in environmental and water resources, WATER SA-PRETORIA- 26 (1) (2000) 1–12.

[24] V. P. Singh, Entropy theory and its application in environmental and water engineering, John Wiley & Sons, 2013.

[25] J. Amorocho, B. Espildora, Entropy in the assessment of uncertainty in hydrologic systems and models, Water Resources Research 9 (6) (1973) 1511–1522.

[26] T. G. Chapman, Entropy as a measure of hydrologic data uncertainty and model performance, Journal of Hydrology 85 (1) (1986) 111–126.

[27] A. J. Abebe, R. K. Price, Managing uncertainty in hydrological models using complementary models, Hydrological sciences journal 48 (5) (2003) 679–692.

[28] P. Pokhrel, H. V. Gupta, On the use of spatial regularization strategies to improve calibration of distributed watershed models, Water resources research 46 (1).

[29] S. V. Weijs, G. Schoups, N. V. D. Giesen, Why hydrological predictions should be evaluated using information theory, Hydrology and Earth System Sciences 14 (12) (2010) 2545–2558.

[30] S. V. Weijs, N. V. D. Giesen, Accounting for observational uncertainty in forecast verification: An information-theoretical view on forecasts, observations, and truth, Monthly Weather Review 139 (7) (2011) 2156–2162.

[31] W. Gong, H. V. Gupta, D. Yang, et al, Estimating epistemic and aleatory uncertainties during hydrologic modeling: An information theoretic approach, Water Resources Research 49 (4) (2013) 2253–2273.

[32] S. V. Weijs, N. V. D. Giesen, M. B. Parlange, Data compression to define information content of hydrological time series, Hydrology. and Earth System Sciences 17 (8) (2013) 3171–3187.

[33] A. Hyvärinen, J. Karhunen, E. Oja, Independent component analysis, Vol. 46, John Wiley & Sons, 2004.

[34] A. Kraskov, H. Stögbauer, P. Grassberger, Estimating mutual information, Physical review E 69 (6) (2004) 066138.

[35] C. Cortes, V. Vapnik, Support-vector networks, Machine learning 20 (3) (1995) 273–297.

[36] T. M. Cover, J. A. Thomas, Elements of information theory, John Wiley & Sons, 2012.

[37] G. G. Katul, A. Porporato, E. Daly, A. C. Oishi, H.-S. Kim, P. C. Stoy, J.-Y. Juang, M. B. Siqueira, On the spectrum of soil moisture from hourly to interannual scales, Water Resources Research 43 (5).

[38] M. Li, V. Paul, An introduction to Kolmogorov complexity and its applications, Springer Science & Business Media, 2009.

[39] W. Gong, D. Yang, H. V. Gupta, G. Nearing, Estimating information entropy for hydrological data: One-dimensional case, Water Resources Research 50 (6) (2014) 5003–5018.

[40] H. E. Jobson, Evaporation into the atmosphere: Theory, history, and applications, Eos Transactions American Geophysical Union 63 (51) (1982) 1223–1224.

[41] Y. B. Dibike, S. Velickov, D. Solomatine, et al, Model induction with support vector machines: introduction and applications, Journal of Computing in Civil Engineering 15 (3) (2001) 208–216.

[42] T. Asefa, M. Kemblowski, M. McKee, A. Khalil, Multi-time scale stream flow predictions: The support vector machines approach, Journal of Hydrology 318 (1) (2006) 7–16.

[43] M. Behzad, K. Asghari, M. Eazi, M. Palhang, Generalization performance of support vector machines and neural networks in runoff modeling, Expert Systems with applications 36 (4) (2009) 7624–7629.

[44] W. Gong, Watershed model uncertainty analysis based on information entropy and mutual information, PhD thesis of Department of Hydraulic Engineering Tsinghua University, Beijing, China.

[45] J. Y. Lin, C. T. Cheng, K. W. Chau, Using support vector machines for long-term discharge prediction, Hydrological Sciences Journal 51 (4) (2006) 599–612.

[46] C. C. Chang, C. J. Lin, Libsvm: a library for support vector machines, ACM Transactions on Intelligent Systems and Technology (TIST) 2 (3) (2011) 27.

[47] Y. Shi, R. Eberhart, A modified particle swarm optimizer, in: Evolutionary Computation Proceedings,1998,IEEE World Congress on Computational Intelligence, The 1998 IEEE International Conference on, IEEE, 1998, pp. 69–73.

[48] Q. Duan, J. Schaake, V. Andreassian, et al, Model parameter estimation experiment (mopex): An overview of science strategy. and major results from the second and third workshops, Journal of Hydrology 320 (1) (2006) 3–17.

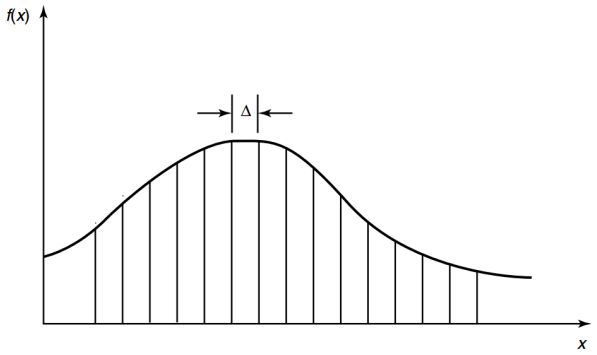


Figure 1: Quantization of Continuous Random Variable

Table 1: Information Terms to be Estimated

	Estimated Items
Observation	$h(R)$
Focused	$I(R;P), I(R;P,P_{former})$
	$I(R;P,PE), I(R;P,P_{former},PE,PE_{former})$
	$I(R;P,P_{former},PE,PE_{former},R_{former})$
Model	TPWB: $I(R;Rs)$, $I(R;P,PE,S)$
Focused	Budyko: $I(R;Rs)$

Table 2: Catchment Information

Climate Type	ID	Area(km^2)	$P_{mean}(mm)$	$PE_{mean}(mm)$	$R_{mean}(mm)$
	02143000	215	1299	882	553
	02165000	611	1252	965	539
WA	02329000	2953	1321	1101	330
	02375500	9886	1452	1061	549
	02478500	6967	1440	1055	489
	05585000	3349	922	993	232
	06908000	2901	1001	1066	261
WS	07019000	9811	1006	959	303
	07177500	2344	948	1259	221
	07243500	5227	935	1303	160
	02414500	4338	1371	976	542
	02472000	1924	1442	1059	509
	11025500	290	522	1407	34
SA	11532500	1577	2748	751	2212
	12459000	2590	1613	681	1105
	13337000	3056	1287	775	872
	14359000	5317	1052	851	510
	05418500	4022	854	1017	254
	05454500	8472	839	984	224
	05484500	8912	794	998	117
SS	06810000	7268	808	1027	173
	06892000	1052	941	1110	228
	06914000	865	950	1186	236
	07183000	9889	877	1250	187

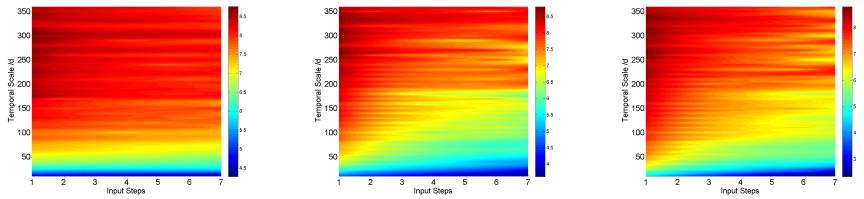
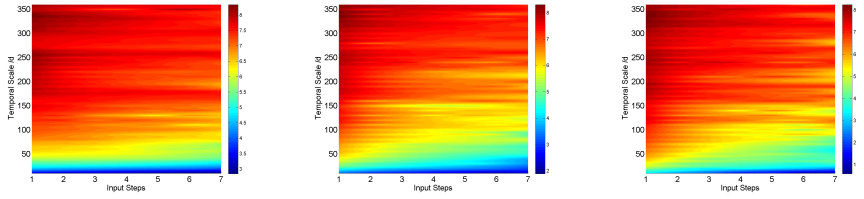
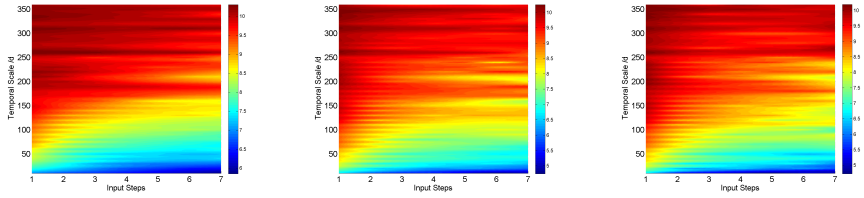
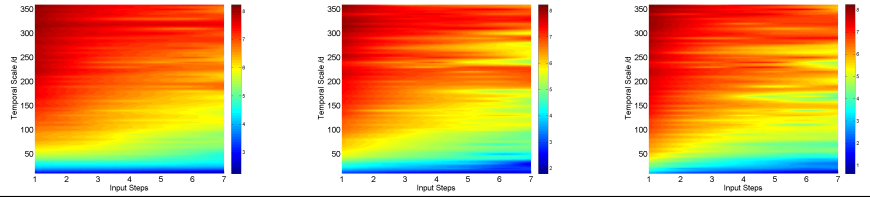
Catchment Type	$AU_a(R;P)$	$AU_a(R;P,PE)$	$AU_a(R;P,PE,R_{former})$
WA(02143000)			
WS(05585000)			
SA(11532500)			
SS(06810000)			

Figure 2. Aleatory Uncertainty of Absolute Constant Resolution

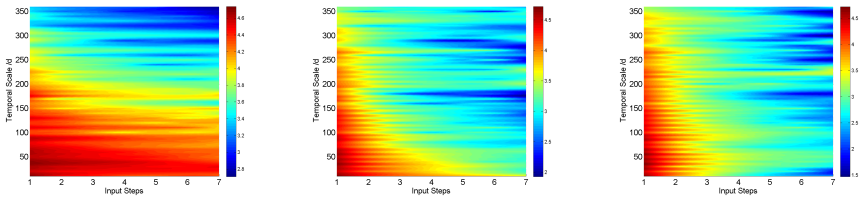
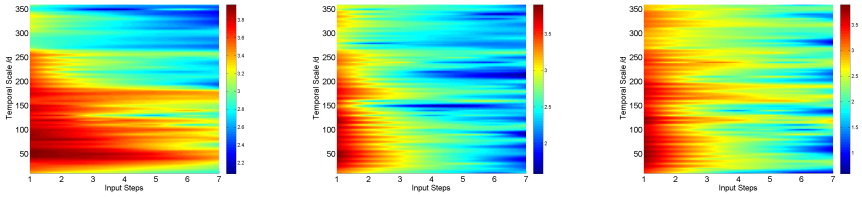
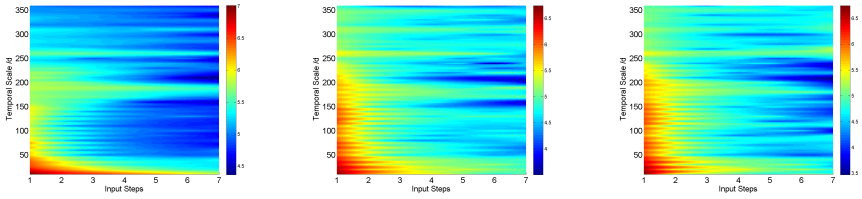
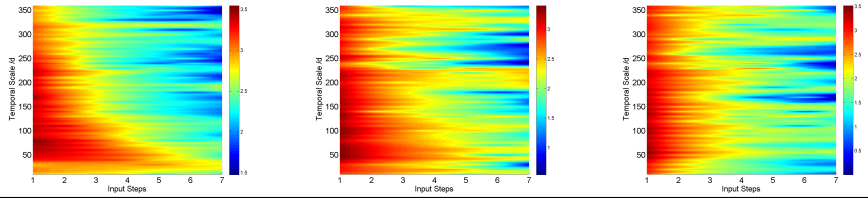
Catchment Type	$AU_r(R;P)$	$AU_r(R;P,PE)$	$AU_r(R;P,PE,R_{former})$
WA(02143000)			
WS(05585000)			
SA(11532500)			
SS(06810000)			

Figure 3. Aleatory Uncertainty of Relative Constant Resolution

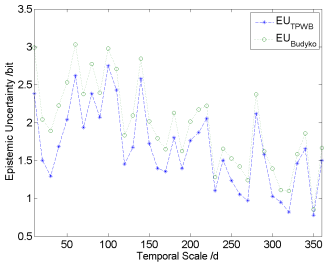
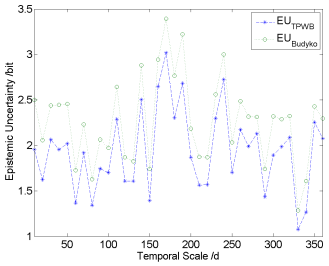
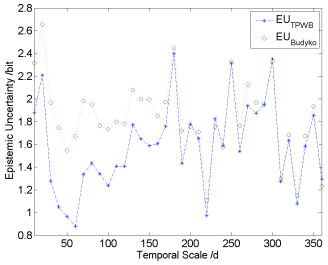
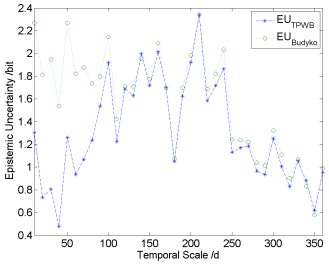
Type	Weak Seasonality	Strong Seasonality
Synchronous		
Asynchronous		

Figure 4. Epistemic Uncertainty

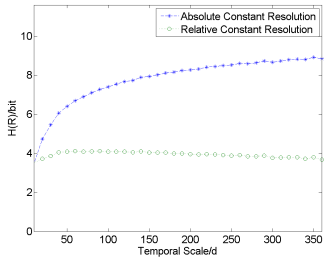
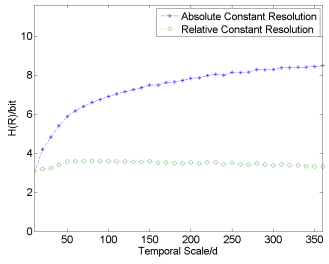
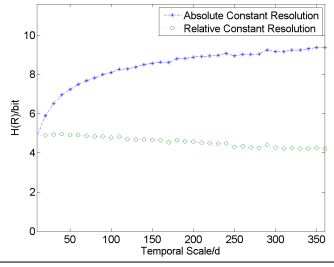
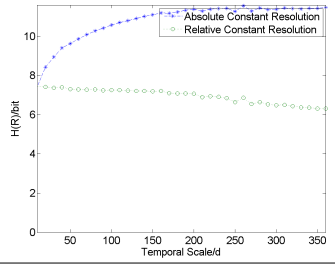
Type	Weak Seasonality	Strong Seasonality
Synchronous		
Asynchronous		

Figure 5. Relative Magnitude of Quantized Runoff Entropy

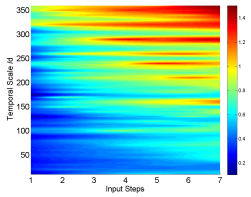
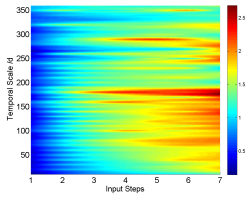
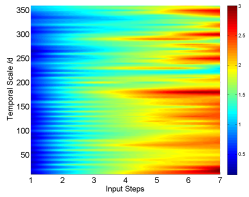
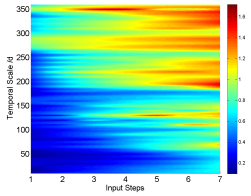
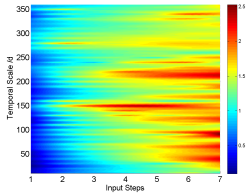
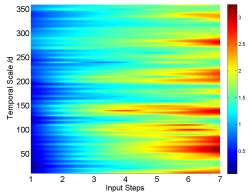
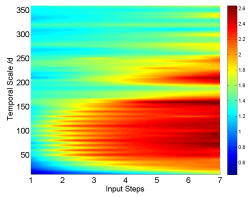
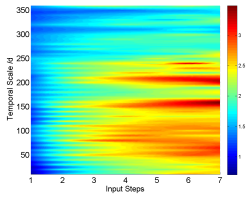
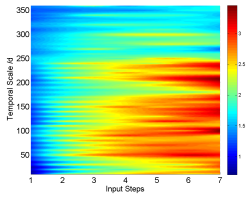
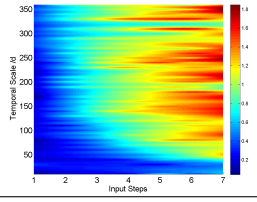
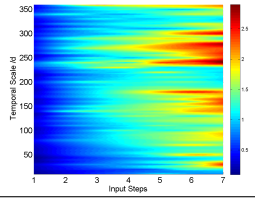
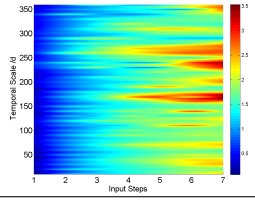
Type	$I(R;P)$	$I(R;P,PE)$	$I(R;P,PE,R_{former})$
WA(02143000)			
WS(05585000)			
SA(11532500)			
SS(06810000)			

Figure 6. Mutual Information Between Runoff and Input Data

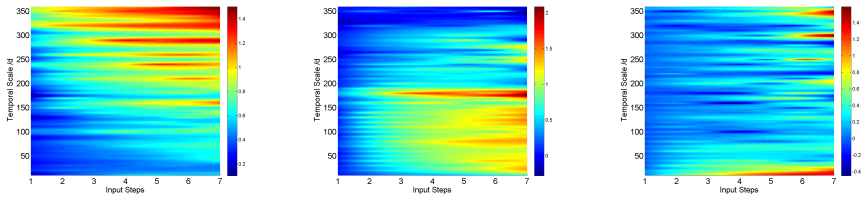
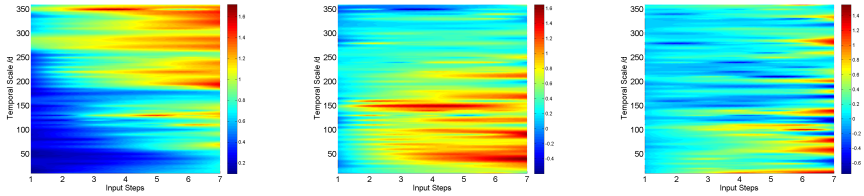
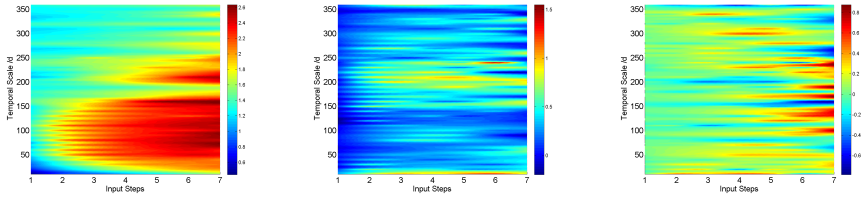
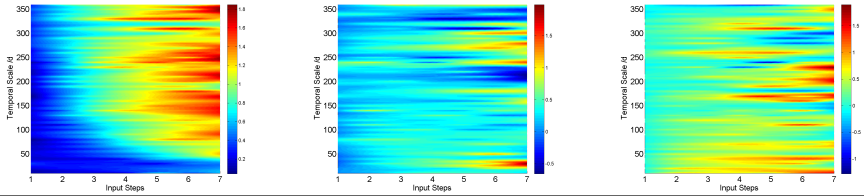
Type	$I(R;P)$	$I(R;P,PE)$ $-I(R;P)$	$I(R;P,PE,R_{former})$ $-I(R;P,PE)$
WA(02143000)			
WS(05585000)			
SA(11532500)			
SS(06810000)			

Figure 7. Information Contribution of PE and R_{former}

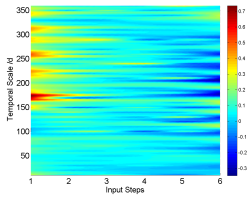
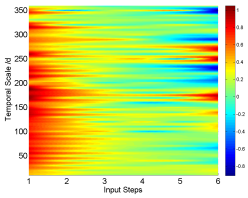
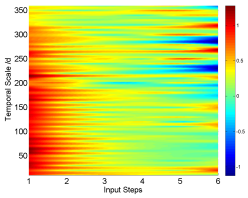
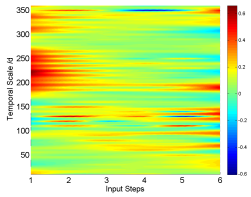
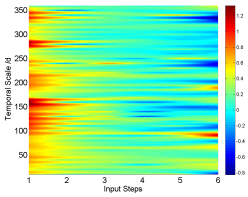
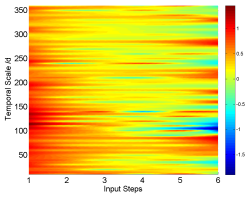
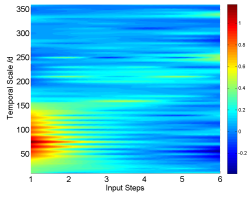
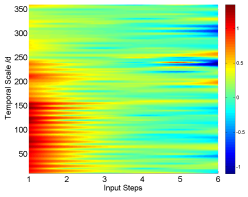
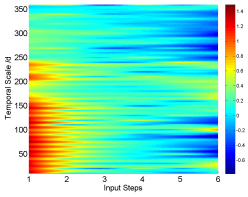
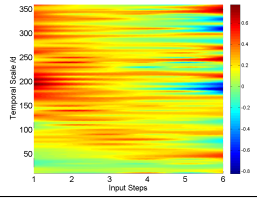
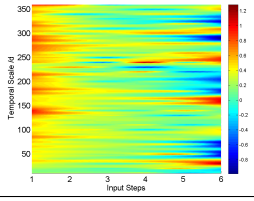
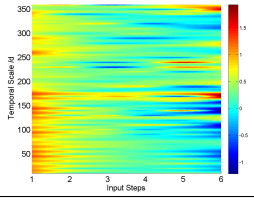
Type	$I(R;P..)$ $-I(R;P.)$	$I(R;P..,PE..)$ $-I(R;P.,PE.)$	$I(R;P..,PE..,R_{former}..)$ $I(R;P.,PE.,R_{former}.)$
WA(02143000)			
WS(05585000)			
SA(11532500)			
SS(06810000)			

Figure 8. Information Contribution of Former Inputs

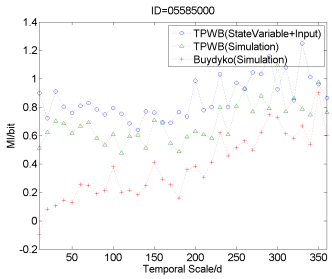
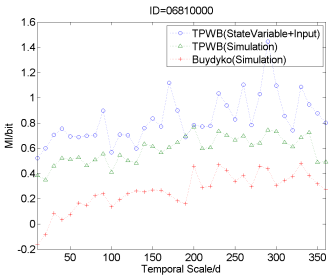
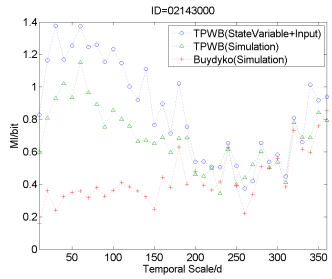
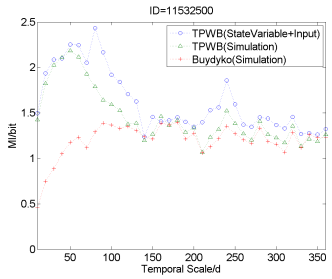
Type	Weak Seasonality	Strong Seasonality
Synchronous		
Asynchronous		

Figure 9. Mutual Information Between Runoff and Simulation