

Information Analysis of Catchment Hydrological Patterns across Temporal Scales

Abstract

Catchment hydrological cycle takes on different patterns across temporal scales. The constitutive functions describing hydrological processes in specific models are usually employed at limited but fuzzy temporal scales. The interim between a daily runoff generation event and the long term water-energy correlation pattern requires further examination to justify a self-consistent understanding. In this research, we employ quantized entropy of clustered runoff observations to represent the prior uncertainty in determining the catchment's hydrological patterns at specific temporal scales. Mutual information between runoff observation and the catchment's water energy provisions, which are represented by precipitation and potential evapotranspiration, is employed to represent 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 items construct the framework of epistemic and aleatory uncertainty in evaluating the observation and simulation systems. We implement this information analysis with daily hydrometeorological data clustered at temporal scales from 10 days to 1 year in 24 catchments from MOPEX data set to detect the catchments' water-heat correlation patterns and simulation abilities of two existed 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 items across temporal scales are related with the seasonality type of catchments. 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.

Contents

1	Introduction	4
2	Methodology	8
2.1	Bits in Hydrological Simulation Context	8
2.2	Quantization Schemes for Runoff Differential Entropy	11
2.3	High Dimensional Mutual Information Estimator	12
3	Data	15
4	Results & Discussion	16
4.1	Aleatory Uncertainty	16
4.1.1	Result Anatomy	19
4.1.2	Cause Attribution	23
4.2	Epistemic Uncertainty	24
5	Conclusion	28

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 (Blöschl and Sivapalan, 1995). 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 (Freeze and Harlan, 1969), 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 models 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(Hofstadter, 2000). Given this fact, many conceptual hydrological models have been brought forward to provide coarser but valuable simulation without requiring detailed inputs or mass computation capacity. The mathematical analysis of the simplified forms offer an insight into the catchment hydrological mechanism that is blotted when aggregating the mass outputs produced by the distributed models(Gerrits et al., 2009; Xu et al., 2014). 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(Budyko, 1961). 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(Wang and Alimohammadi, 2012). 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 II Theorem(Fu, 1981; Choudhury, 1999; Yang et al., 2008). 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(Zhang and Dawes, 2001; Yang et al., 2007).

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(Tekleab et al., 2011). Former critics basically blame the deviation for excluding the impact of the changing soil moisture(Sankarasubramanian and Vogel, 2002; Sankarasubramanian and Vogel, 2003). By including the soil moisture storage item, some seasonal and monthly water balance models were developed(Thomas, 1981; Xiong and Guo, 1999; Zhang et al., 2008), which serve as temporal scale gap-filters of the long term water-heat correlation pattern and single precipitation-runoff phenomenon focused hydrological models. As have been declared, the introduction 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. The rationale of these functions are gaining hydrologists' concern due to a similar Darwinian ideological origin with the Budyko Curve(Wang and Tang, 2014). However, their specific dominant temporal scales remain ambiguous.

Given the pros and cons of the two groups of 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; (2), to what accuracy the data support the patterns; (3), to what extent the existing models capture this patterns.

This research tries to give primary response to these questions within the information theoretical framework.

The term *information* got mathematicized by Claude E. Shannon in 1948(Shannon, 1948). The notion that information is the combination of

bits and context(Bryant and Richard, 2003) sets the theoretical foundation of the digit revolution and broadens to find applications in many other areas, including hydrology and water resources(Singh, 1997; Singh, 2000; Singh, 2013).

Specific to hydrological simulation, information theory has been applied for model evaluation and uncertainty analysis as far back as the 1970s (Amoroch and Espildora, 1973; Chapman, 1986; Abebe and Price, 2003; Pokhrel and Gupta, 2010; Weijs et al., 2010; Weijs and Giesen, 2011) . Gong developed a comprehensive model evaluation framework based on *entropy* and *mutual information* (Gong et al., 2013) . 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.

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

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

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 (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.

Though the definition provides a seemingly crystalline framework to evaluate the observation and simulation systems , the hydrological context in which these bits 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(Weijs et al., 2013). It should also be noted that the intuitive significance of discrete entropy could not blindly generalize to differential entropy. We will address these issues in the following sections.

Besides, to calculate the high dimensional information terms is never an easy task. The strategy Gong adapts is to transform the high dimensional term into independent vectors using Independent Component Analysis Algorithm (ICA)(Hyvärinen et al., 2004). 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 may not be independent when the original term is highly non-linearly correlated. Thus, the method would overrate the entropy of the original data for neglecting the inner relevance among different dimension terms. Besides, the indirect calculation of mutual information through entropies could not avoid the problem of error accumulation.

In this research, 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-clustered 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 temporal scale hydrological terms to quantify the information flows within the hydrological cycle at that scale. Given the drawbacks of the existed high dimensional information estimators, we adapt a k-nearest neighbour distance method(Kraskov et al., 2004), 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(Cortes and Vapnik, 1995), we apply it to depict the *distances*. The theoretical clarification and algorithm are in the second section. Information terms to be calculated with this algorithm are followed. Finally, we discuss the interpretations of these terms 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 $-\log p$. Thus, the average information content of random variable X is:

$$H(X) = -\sum p(x)\log p(x) \quad (3)$$

$$h(X) = -\int f(x)\log f(x)dx \quad (4)$$

$H(X)$ and $h(X)$ are denoted as discrete and continuous Shannon Entropy, measured in bits for logarithm base 2. The unit bit is widely used in computer science because an ideally efficient encoding system is an exact physical implementation of information theoretical principles.

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 $-\log p \rightarrow \infty$ as $p \rightarrow 0$, 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:

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

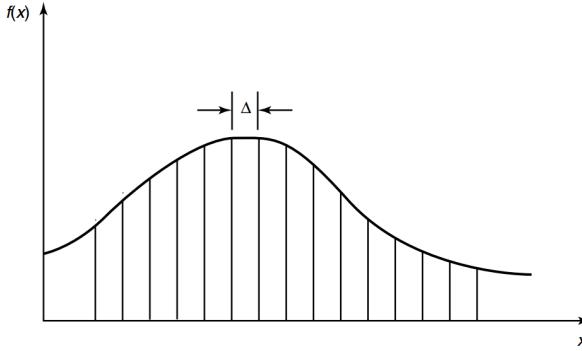


Figure 1: Quantization of a Continuous Random Variable

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(Cover and Thomas, 2012). 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 item we apply here is mutual information. Its discrete and continuous forms are as follows:

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

$$I(X; Y) = \int f(x,y) \log \frac{f(x,y)}{f(x)f(y)} dx dy \quad (7)$$

As can be derived:

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

E denotes expectation. The latter item in the middle and left part of equation (8) is denoted as 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 (Cover and Thomas, 2012).

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.

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 observation noise is denoted as *Aleatory Uncertainty*. The model serves as an information distiller or decoder that transfers the mass input observation data into simple simulations. The information loss during decoding is denoted as *Epistemic Uncertainty*.

Hydrological series encoded in different context can take up different amounts of bits. In this research, we restrict our attention to hydrological observations sampled discretely along the time domain base. The sample space is built on the clustered coordinates at various temporal scales 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(Li and Paul, 2009). 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 these 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(Gong et al., 2014) while incorporating significant temporal scales in detecting long term catchment hydrological behaviours.

Table 1: Estimated Information Terms

Classification	Estimated Terms
Observation Focused	$h(R_t)$
	$I(R_t; P_t) \dots I(R_t; P_t, P_{t-1} \dots P_{t-n})$
	$I(R_t; P_t, PE_t), I(R_t; P_t, P_{t-1}, PE_t, PE_{t-1}), \dots$
	$I(R_t; P_t, P_{t-1}, \dots, P_{t-n}, PE_t, PE_{t-1}, \dots PE_{t-n})$
	$I(R_t; P_t, P_{t-1}, PE_t, PE_{t-1}, R_t - 1), \dots$
	$I(R_t; P_t, P_{t-1}, \dots, P_{t-6}, PE_t, PE_{t-1}, \dots PE_{t-6}, R_{t-1}, \dots R_{t-n})$
Model Focused	TPWB: $I(R_t; Rs_t), I(R_t; P_t, PE_t, S_t)$
	Budyko: $I(R_t; Rs_t)$

P_t and PE_t denotes precipitation and potential evapotranspiration random variables at time step t . R_t and Rs_t denotes observed and simulated runoff random variables at time step t . $h(R_t)$ provides the base to estimate information content of runoff at different quantization schemes. By gradually introducing different hydrological terms with different previous input steps into the estimation of their mutual information with the runoff data, we can make a specific analysis of their information contributions .

TPWB and Budyko Model are two typical hydrological models selected in this research. TPWB is a two parameter water balance model(Xiong and Guo, 1999). The model adapts an adjusted Ol'dektop equation(Jobson, 1982) to depict the runoff generation and evapotranspiration at a monthly temporal scale and achieved satisfying performance. S_t is the state variable at time step t . It is employed to represent the influence of former hydrological influences. The Budyko Model is the combination of Budyko Curve and water balance equation as described above.

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 cluster the daily runoff observation data, the entropy difference in depicting them with quantization schemes introduced above is:

$$H(R_m) - H(R_n) = \begin{cases} h(R_m) - h(R_n) & ; \text{Quantization Scheme 1} \\ h(R_m) - h(R_n) - \log \frac{m}{n} & ; \text{Quantization Scheme 2} \end{cases} \quad (9)$$

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 (Kraskov et al., 2004):

$$I(X, Y) = \psi(k) - N^{-1} \sum_{i=1}^N [\psi(n_x(i) + 1) + \psi(n_y(i) + 1)] + \psi(N) \quad (10)$$

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 understanding of the equation 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(Dibike et al., 2001; Lin et al., 2006; Asefa et al., 2006; Behzad et al., 2009; Gong, 2012).

We use the following function to depict the distance between two model input samples x_1 and x_2 :

$$SVM_Metric(x_1, x_2) = |f(x_1) - f(x_2)| \quad (11)$$

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(Chang and Lin, 2011). 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(Shi and Eberhart, 1998). To avoid overfitting, we apply 3 cross validation in the support vector regression parameter estimation.

The calculating steps are as follows:

- (1) Re-cluster 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 11.
- (3) Estimate mutual information with equation 10.

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 clustered daily hydrologic records (including precipitation, potential evapotranspiration and runoff) from the MOPEX data set(Duan et al., 2006). 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 a 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.

Table 2: Catchment Information

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

4 Results & Discussion

This part is shown in the following manner: we list the typical estimations implemented at catchments from the four seasonality groups. The similarities and differences of estimations within and across groups are discussed following.

4.1 Aleatory Uncertainty

As is defined in equation(1), *Aleatory Uncertainty* equals to the difference between quantized runoff entropy and mutual information between runoff and hydrometeorological input observations. The values are determined by three factors besides the catchment's hydrological characteristics and observation accuracy. The first is pre-required accuracy of runoff estimation, which is determined by its quantization scheme. The second is the species of hydrometeorological inputs, since the incorporation of new input items is expected to decrease simulation uncertainty. The last factor is the inclusion of hydrological variables from former calculating steps, as previous hydrological behaviour may exert effects on current hydrological response. Given this analysis, we list the categorized estimations of *Aleatory Uncertainty* in table 3 and table 4.

Table 3: Aleatory Uncertainty of Absolute Constant Resolution

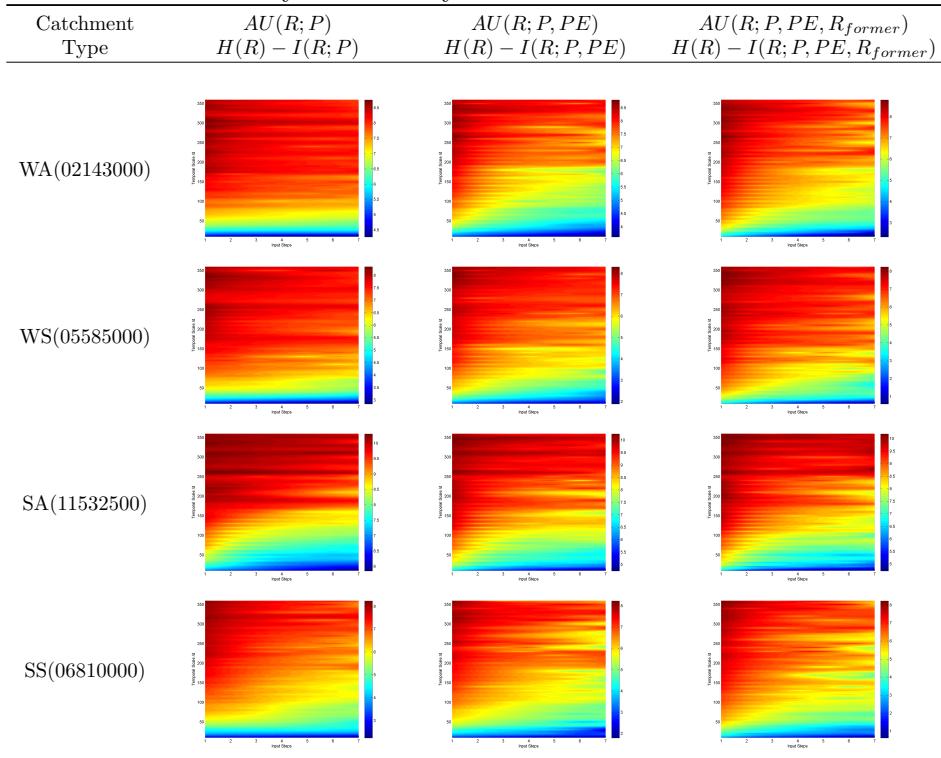
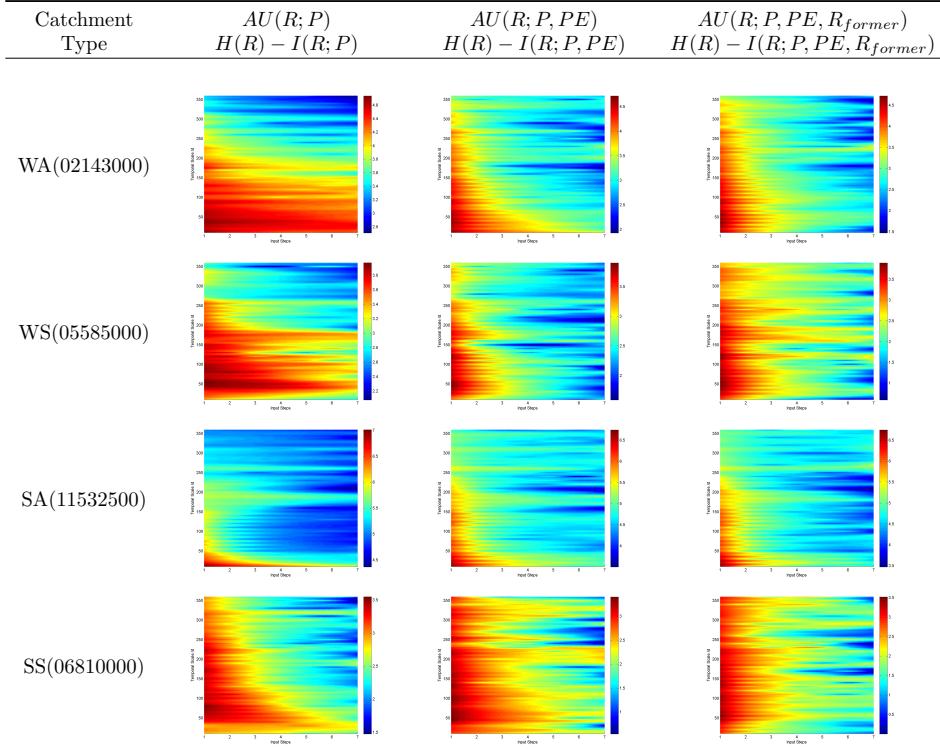


Table 4: Aleatory Uncertainty of Relative Constant Resolution

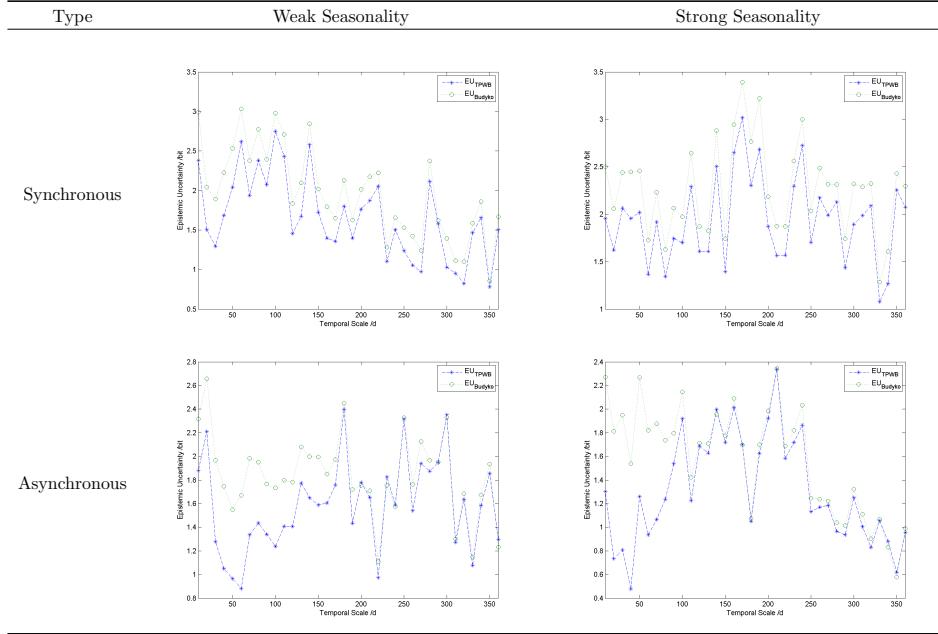


In each graph from the tables above, the abscissa represents the input steps, for example, 1 input step means that the *Aleatory Uncertainty* is estimated with inputs from current calculating step; 2 input steps means that the value is estimated with inputs from current and the previous calculating steps. The ordinate represents the estimating temporal scale, which varies from 10 days to a year.

As can be depicted from the estimations above, when we pre-require an absolute constant resolution of runoff estimation, *Aleatory Uncertainty* increases as the simulating temporal scale expands. However, for relative constant resolution, the value decreases or stays relatively stable as temporal scale expands. The changing rate varies with input species and steps. This phenomenon should be anatomized before digging into its causes.

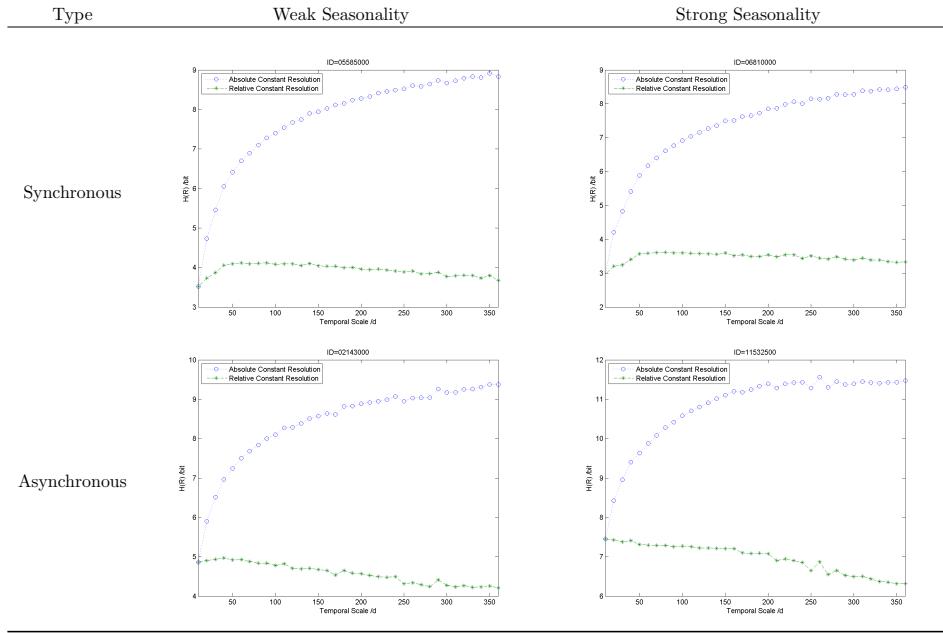
4.1.1 Result Anatomy

Table 5: Epistemic Uncertainty



The baseline of uncertainty estimation is constructed by quantized runoff entropy as shown in table 6. It depicts the uncertainty when no further prior assumption is incorporated given the estimating context, or in the terminology of Bayesian stochastic, it tells the prior uncertainty.

Table 6: Relative Magnitude of Quantized Runoff Entropy

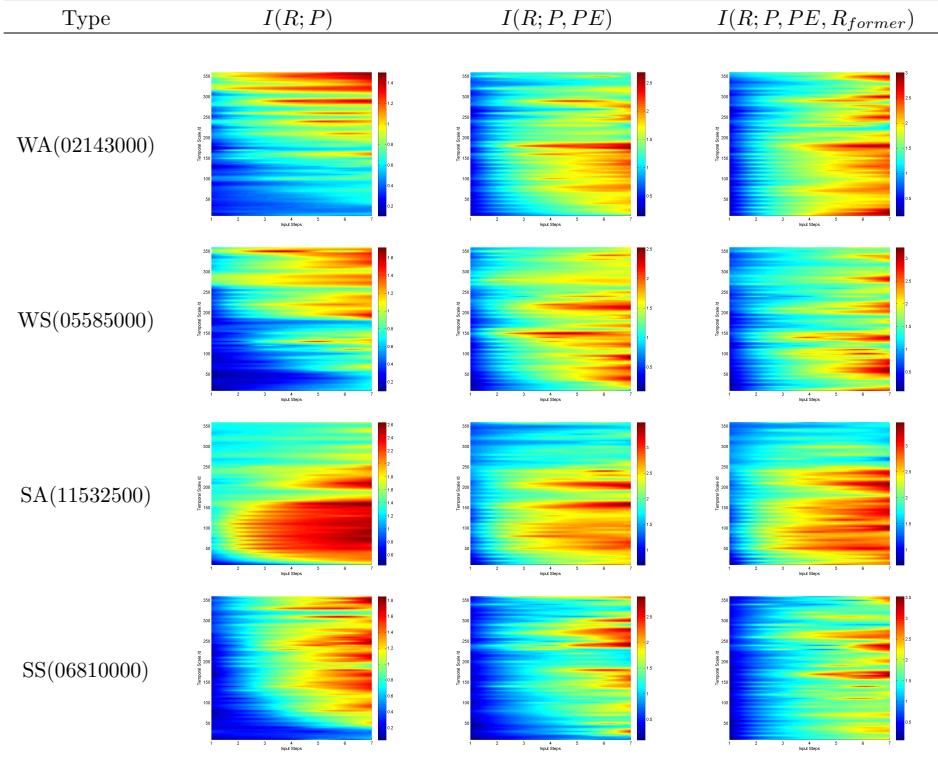


All the estimations are relative values on a same base of 0 bit accuracy at 10 days temporal scale. The runoff entropy of absolute constant resolution increases with temporal scales, while the increasing rate decreases as scale expands, making the curve take on a logarithm shape. This is the dominant factor that cause the increasing trend of *Aleatory Uncertainty* in table 3.

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.

Mutual information between runoff observation and hydrometeorological inputs are shown in table 7. They depict the uncertainty decrease given the input observations.

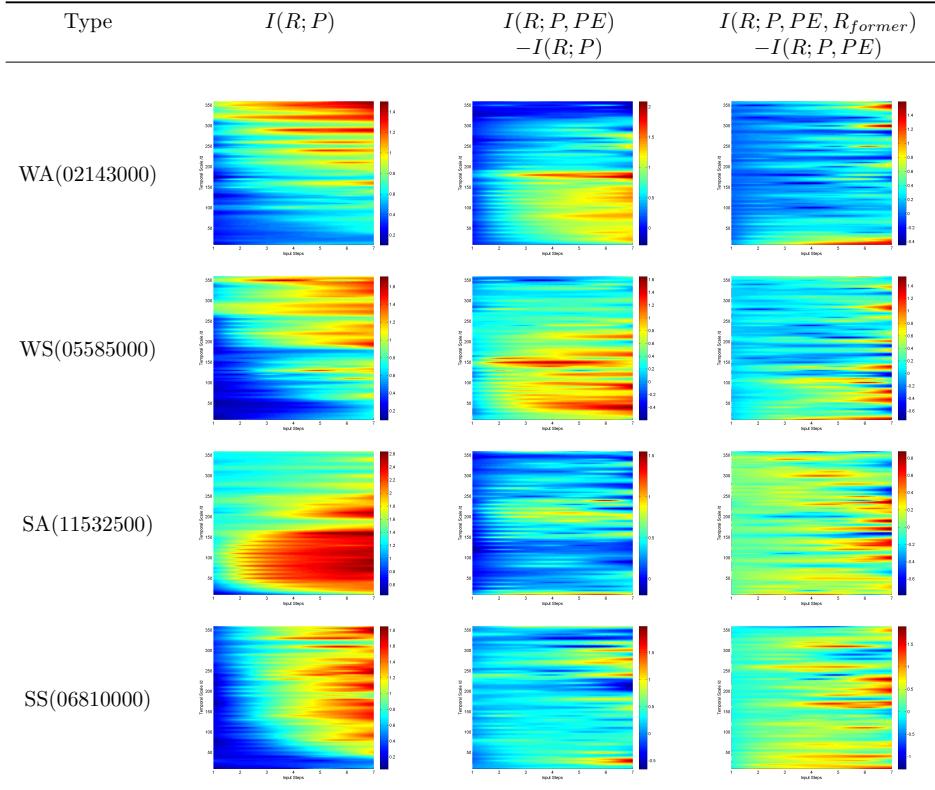
Table 7: Mutual Information Between Runoff and Input Data



The significances of the axis are the same as those in table 3 and table 4. As can be observed, mutual information between runoff observation and input observations increases as more items (PE and R_{former}) and former input data are incorporated. The increases differ between catchments across temporal scales.

To clarify the information contribution of each item, we take two dissection schemes on graphs in table 7.

The first dissection scheme checks the information contribution of incorporating PE and R_{former} into mutual information estimation. This is implemented by making differences between columns in table 7.

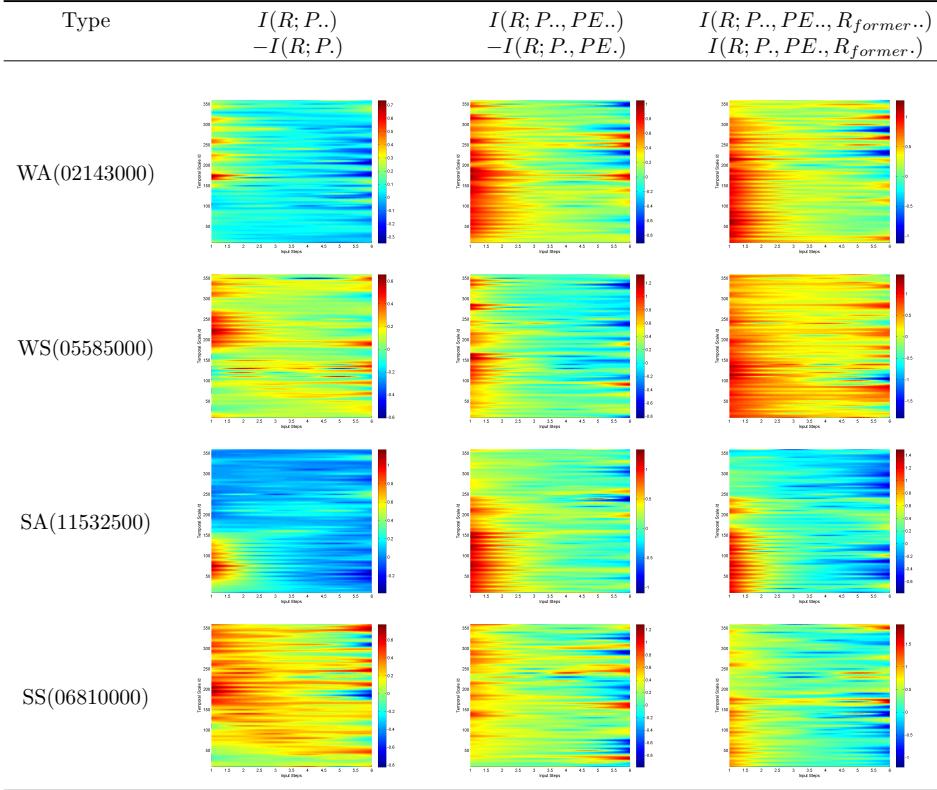
Table 8: Information Contribution of PE and R_{former}


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.

The second dissection scheme checks the information contribution of including former inputs into mutual information estimation. This is implemented by making differences between mutual information estimated with different input steps, for instance, the n th spline in each graph from table 9 equals to the difference of the $(n + 1)$ th spline and n th spline in the corresponding graph from table 7.

Table 9: Information Contribution of Former Inputs



Graphics in table 9 depict the information contribution rate ($\frac{\partial I}{\partial Input_Step}$) when including former 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.

4.1.2 Cause Attribution

As is shown in table 8 and table 9, the inclusion of new hydrometeorological items and data from previous calculation steps can improve this ability if these items are correlated. It will not decrease mutual information even no statistical connection exists. This is due to the data-processing inequality(Cover and Thomas, 2012).

The data-processing inequality states that if random variables X, Y, Z

form a Markov chain in that order (denoted by $X \rightarrow Y \rightarrow Z$), then:

$$I(X;Y) \geq I(X;Z) \quad (12)$$

Since:

$$R \rightarrow Input_{original}, Input_{new} \rightarrow Input_{original} \quad (13)$$

We have:

$$I(R; Input_{original}, Input_{new}) \geq I(R; Input_{original}) \quad (14)$$

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

Inequality 14 guarantees the non-negativity of items in table 8 and table 9 (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 posterior uncertainty given the information of input observations, which is denoted as *Aleatory Uncertainty*, is smaller than the prior uncertainty, as is represent by $H(R)$. Its origin 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 clustering them together.

The other origin is the inherent uncertainty caused by the coarse temporal scale. A simple clustering 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.

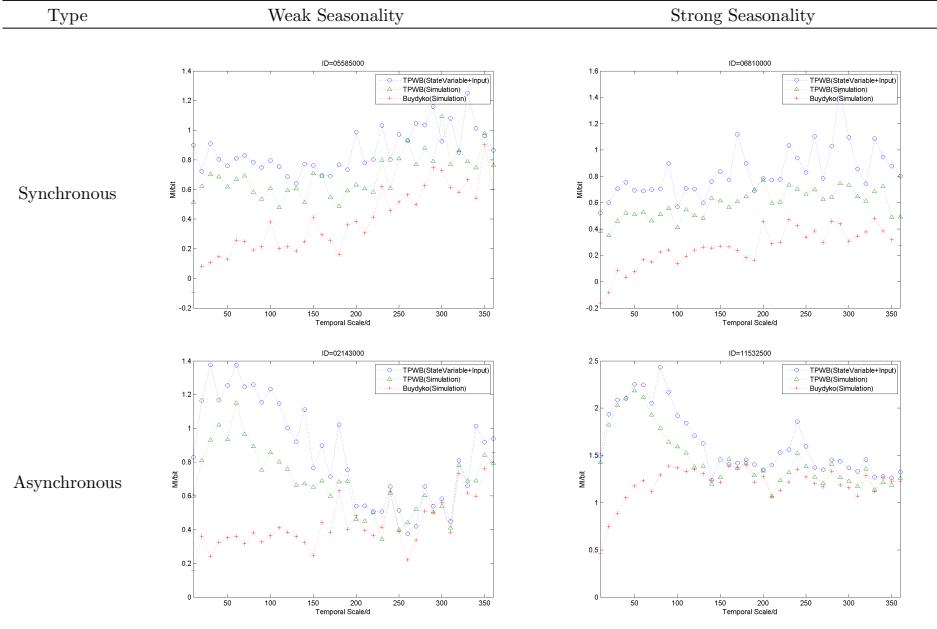
Given the reliability of the MOPEX dataset, we assume that the latter uncertainty source plays a dominant role. In other words, the *Aleatory Uncertainty* is mainly caused by data insufficiency rather than inaccuracy for large temporal scales.

4.2 Epistemic Uncertainty

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:

Table 10: Mutual Information Between Runoff and Simulation



As a monthly water balance model that takes iterative structure, TPWB applies state variable S to represent the influence of former hydrological processes. It could be observed that $I(R_t; P_t, PE_t, S_t)$ is always larger than $I(R, R_s)$, which means that the item S_t is not sufficient in representing former hydrological information. Both the two estimations increases with temporal scales in synchronous rainfall energy catchments(except Catchment 07019000). In asynchronous catchments, as temporal scale expands, they tend to increase and reach a maximum point around 1 to 2 months, after that, 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.

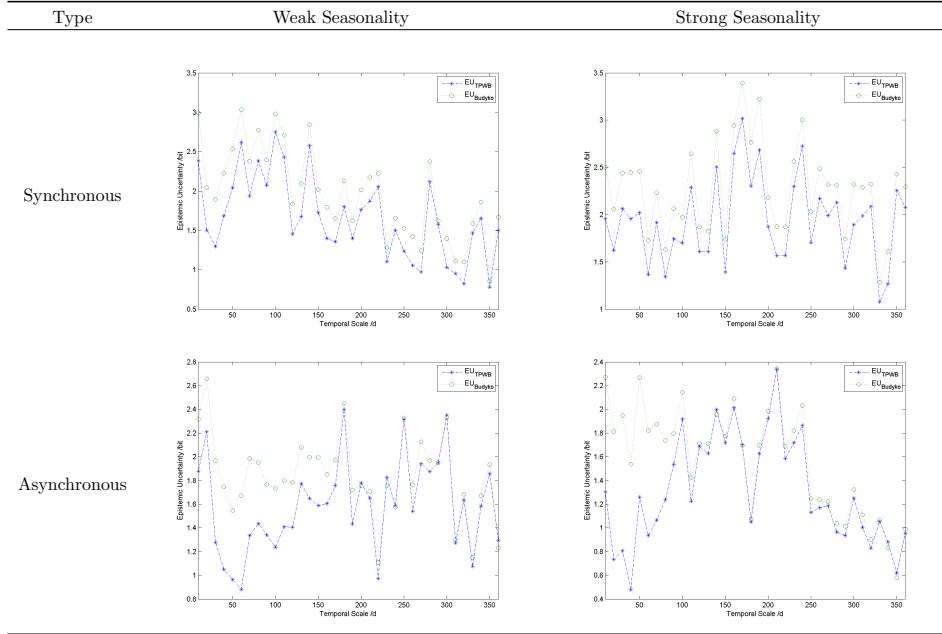
For Budyko Model, $I(R, R_s)$ increases with temporal scale while being smaller than that of TPWB (except Catchment 11025500 where the drought coefficient is extreme high).

$I(R, R_s)$ of the two models approach a same value as temporal scale expands except in catchments from the SS group. In SS group, $I(R, R_s)$ of

the two models increases parallel as temporal scale expands.

Given the estimations above, we present the *Epistemic Uncertainty* of two model employed here:

Table 11: Epistemic Uncertainty



For TPWB, peak value of *Epistemic Uncertainty* 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.

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. As temporal scale expands, the *Epistemic Uncertainty* of the two models approaches because of a less close hydrological connection between calculating steps.

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

We present an explanation of the estimations using data processing inequality. 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:

$$R \rightarrow Input_{previous}, Input_{current} \rightarrow S, Input_{current} \rightarrow R_s \quad (15)$$

which could be simplified as:

$$R \rightarrow Input \rightarrow S, Input_{current} \rightarrow R_s \quad (16)$$

given the data-processing inequality, we have:

$$I(R; Input) \geq I(R; S, Input_{current}) \geq I(R; R_s) \quad (17)$$

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

5 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.

Appendix

Mutual Information Between Runoff and Input Data

Table 12: WA Catchments

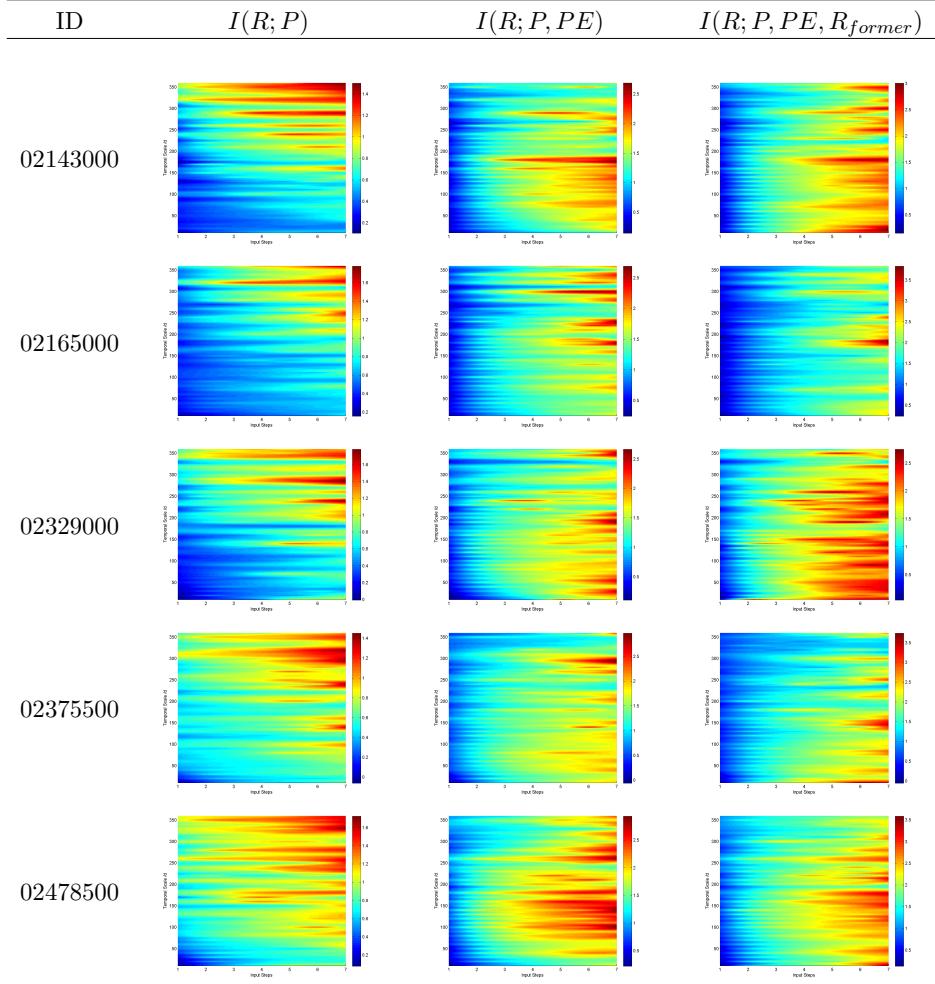


Table 13: WS Catchments

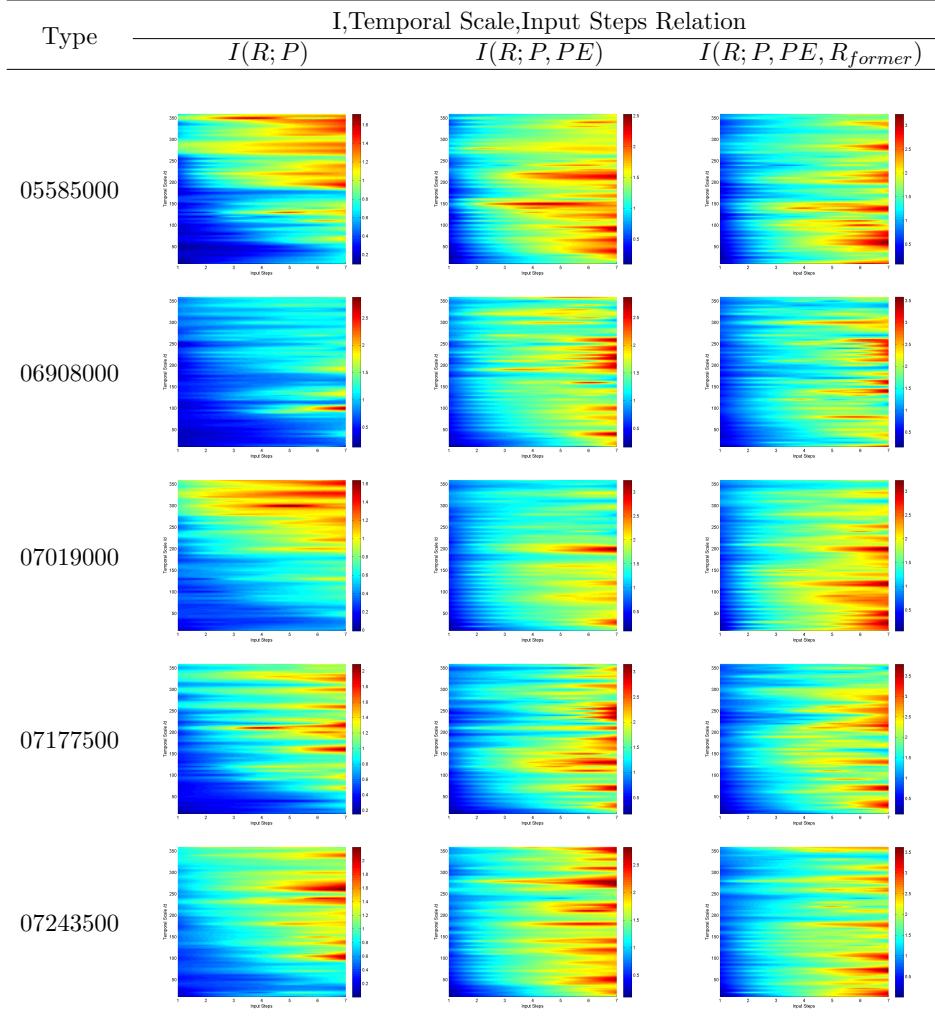


Table 14: SA

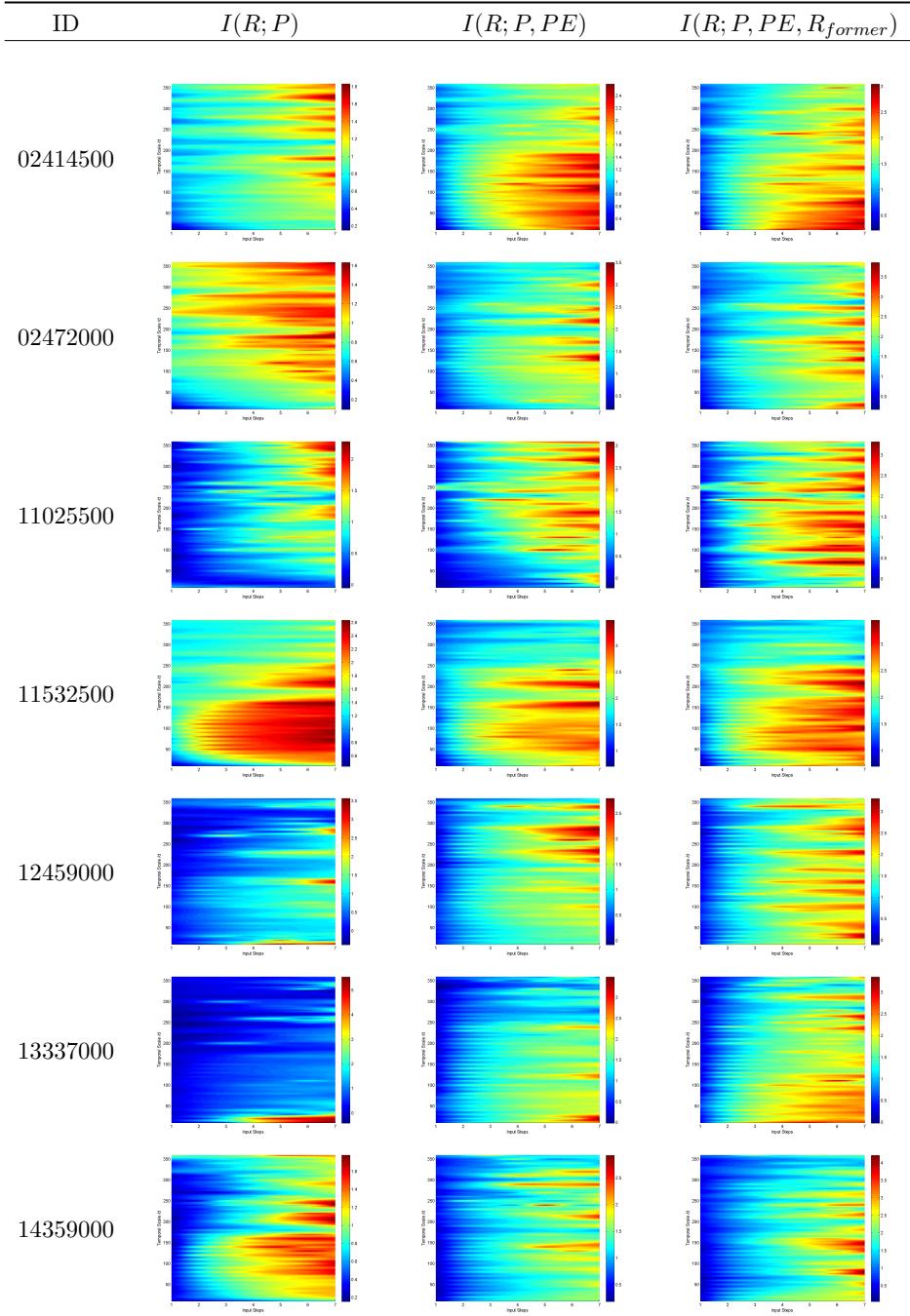
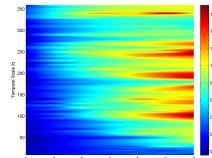
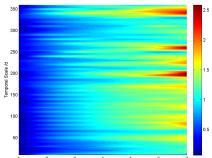
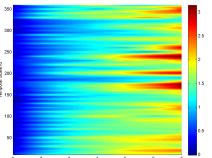
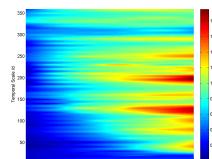
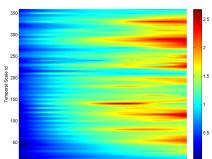
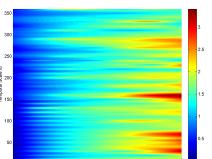
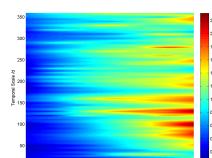
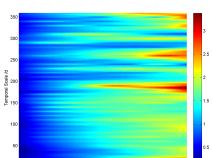
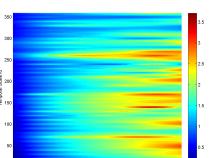
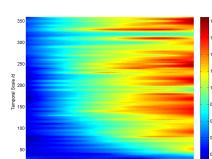
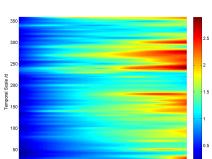
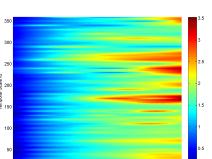
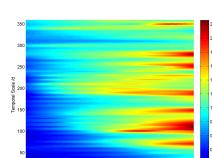
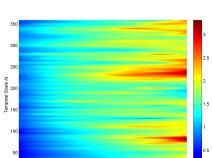
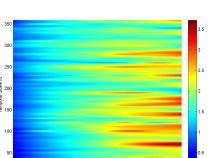
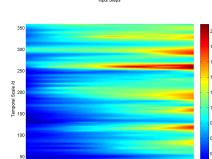
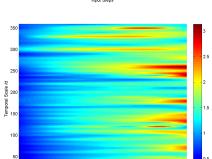
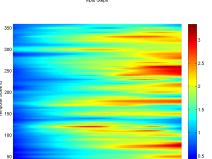
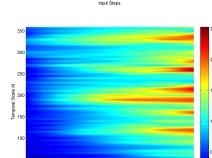
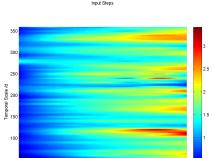
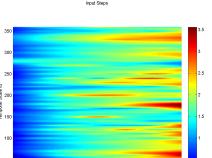


Table 15: SS Catchments

ID	$I(R; P)$	$I(R; P, PE)$	$I(R; P, PE, R_{former})$
05418500			
05454500			
05484500			
06810000			
06892000			
06914000			
07183000			

Information Analysis

Table 16: WA Catchments

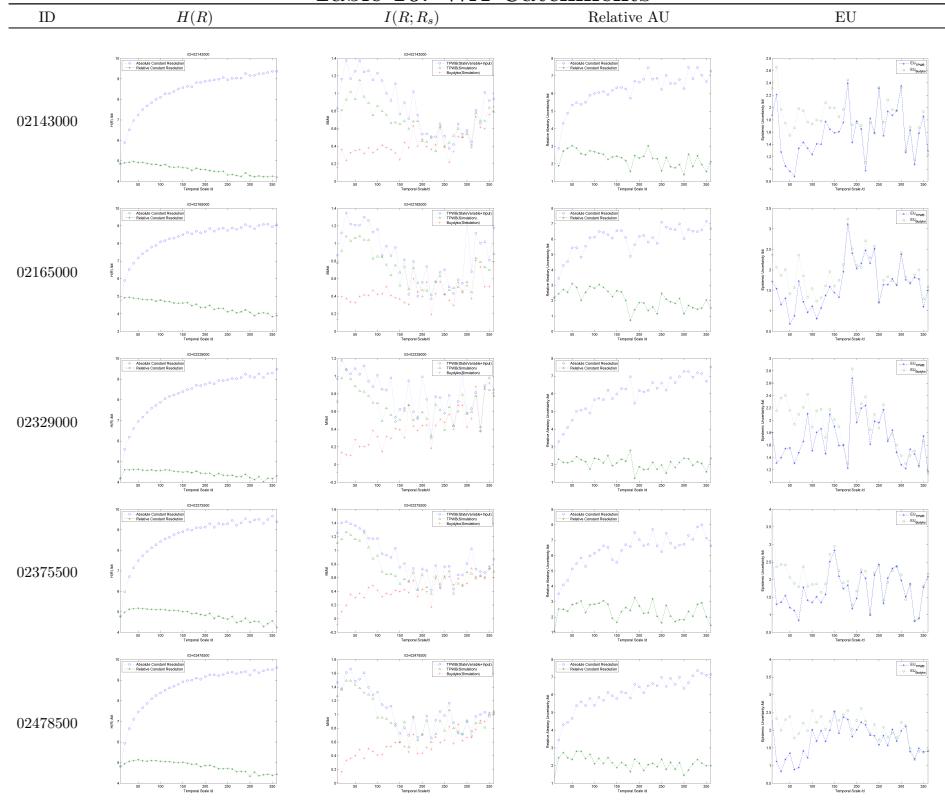


Table 17: WS Catchments

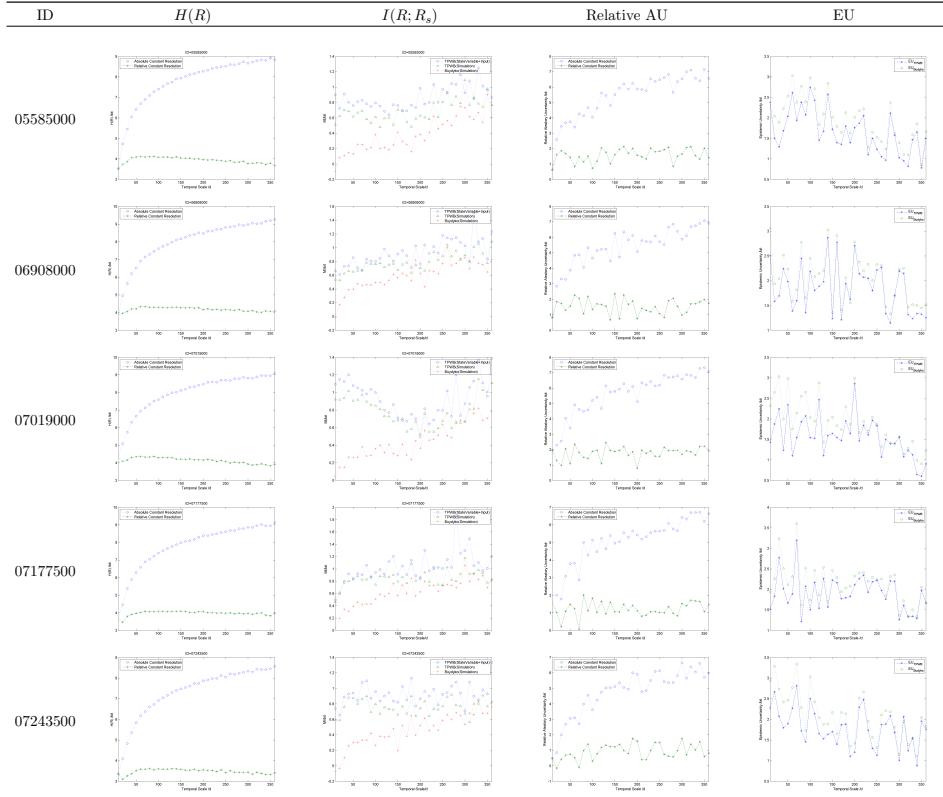


Table 18: SA

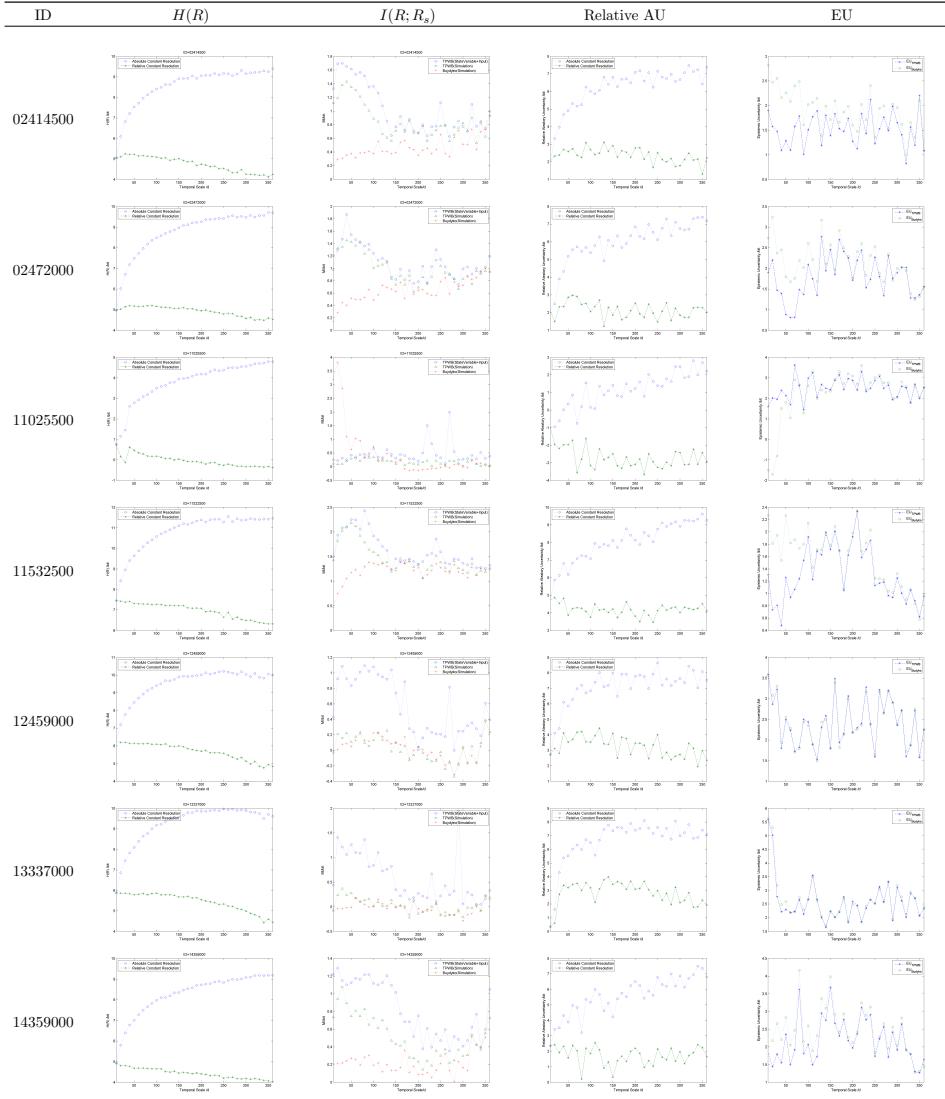
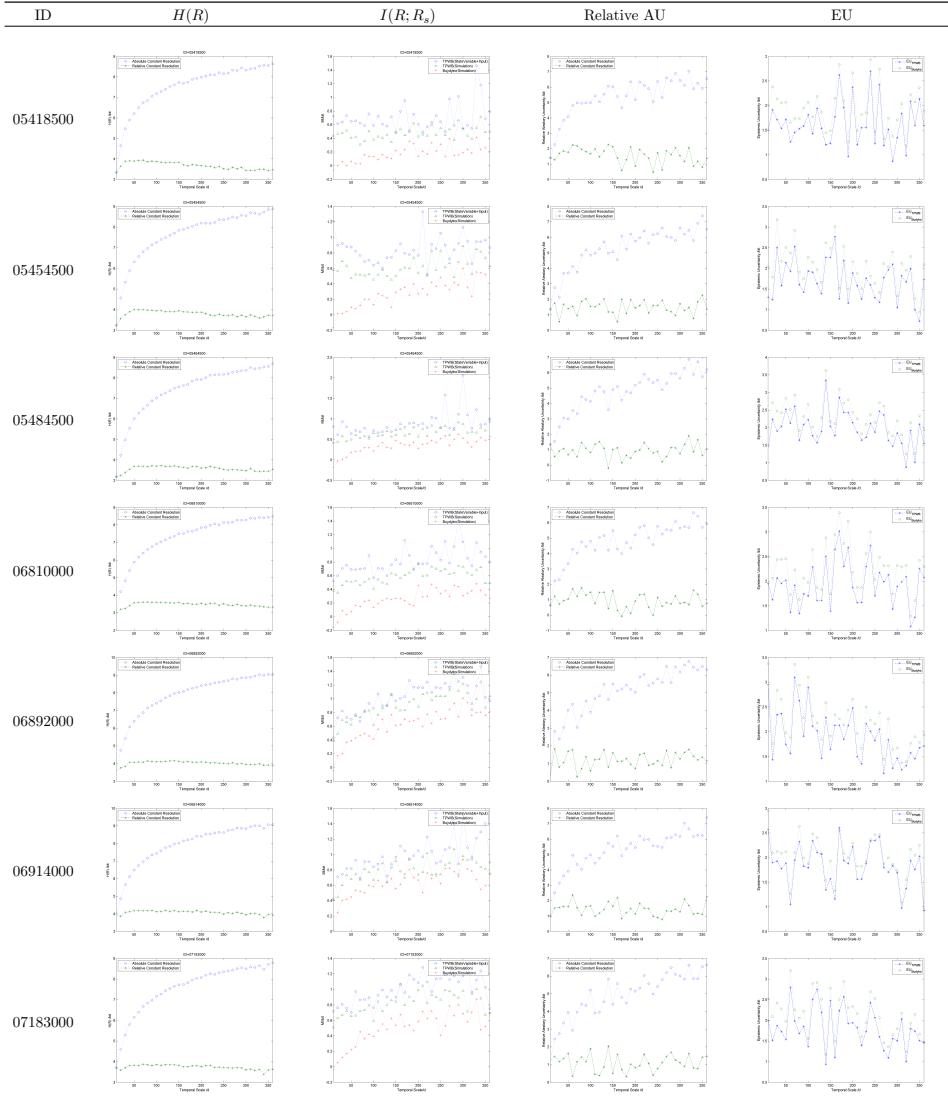


Table 19: SS



References

- Abebe, A. J. and Price, R. K. (2003). Managing uncertainty in hydrological models using complementary models. *Hydrological sciences journal*, 48(5):679–692.
- Amoroch, J. and Espildora, B. (1973). Entropy in the assessment of uncertainty in hydrologic systems and models. *Water Resources Research*, 9(6):1511–1522.
- Asefa, T., Kemblowski, M., McKee, M., and Khalil, A. (2006). Multi-time scale stream flow predictions: The support vector machines approach. *Journal of Hydrology*, 318(1):7–16.
- Behzad, M., Asghari, K., Eazi, M., and Palhang, M. (2009). Generalization performance of support vector machines and neural networks in runoff modeling. *Expert Systems with applications*, 36(4):7624–7629.
- Blöschl, G. and Sivapalan, M. (1995). Scale issues in hydrological modelling: a review. *Hydrological processes*, 9(3-4):251–290.
- Bryant, R. and Richard, O. H. D. (2003). *Computer systems: a programmer's perspective*. Prentice Hall.
- Budyko, M. (1961). The heat balance of the earth's surface. *Soviet Geography*, 2(4):3–13.
- Chang, C. C. and Lin, C. J. (2011). Libsvm: a library for support vector machines. *ACM Transactions on Intelligent Systems and Technology (TIST)*, 2(3):27.
- Chapman, T. G. (1986). Entropy as a measure of hydrologic data uncertainty and model performance. *Journal of Hydrology*, 85(1):111–126.
- Choudhury, B. J. (1999). Evaluation of an empirical equation for annual evaporation using field observations and results from a biophysical model. *Journal of Hydrology*, 216(1):99–110.
- Cortes, C. and Vapnik, V. (1995). Support-vector networks. *Machine learning*, 20(3):273–297.
- Cover, T. M. and Thomas, J. A. (2012). *Elements of information theory*. John Wiley & Sons.

- Dibike, Y. B., Velickov, S., Solomatine, D., and et al (2001). Model induction with support vector machines: introduction and applications. *Journal of Computing in Civil Engineering*, 15(3):208–216.
- Duan, Q., Schaake, J., Andreassian, V., and et al (2006). Model parameter estimation experiment (mopex): An overview of science strategy and major results from the second and third workshops. *Journal of Hydrology*, 320(1):3–17.
- Freeze, R. A. and Harlan, R. L. (1969). Blueprint for a physically-based digitally-simulated hydrologic response model. *Journal of Hydrology*, 9(3):237–258.
- Fu, B. (1981). Lansurface evaporation calculation. *Meteorology Science(China)*, 5(1):23–31.
- Gerrits, A. M. J., Savenije, H. H. G., Veling, E. J. M., and et al (2009). Analytical derivation of the budyko curve based on rainfall characteristics and a simple evaporation model. *Water Resources Research*, 45(4).
- Gong, W. (2012). Watershed model uncertainty analysis based on information entropy and mutual information. *PhD thesis of Department of Hydraulic Engineering Tsinghua University, Beijing, China*.
- Gong, W., Gupta, H. V., Yang, D., and et al (2013). Estimating epistemic and aleatory uncertainties during hydrologic modeling: An information theoretic approach. *Water Resources Research*, 49(4):2253–2273.
- Gong, W., Yang, D., Gupta, H. V., and Nearing, G. (2014). Estimating information entropy for hydrological data: One-dimensional case. *Water Resources Research*, 50(6):5003–5018.
- Hofstadter, D. R. (2000). *Gödel, Escher, Bach, An Eternal Golden Braid*. Penguin.
- Hyvärinen, A., Karhunen, J., and Oja, E. (2004). *Independent component analysis*, volume 46. John Wiley & Sons.
- Jobson, H. E. (1982). Evaporation into the atmosphere: Theory, history, and applications. *Eos Transactions American Geophysical Union*, 63(51):1223–1224.
- Kraskov, A., Stögbauer, H., and Grassberger, P. (2004). Estimating mutual information. *Physical review E*, 69(6):066138.

- Li, M. and Paul, V. (2009). *An introduction to Kolmogorov complexity and its applications*. Springer Science & Business Media.
- Lin, J. Y., Cheng, C. T., and Chau, K. W. (2006). Using support vector machines for long-term discharge prediction. *Hydrological Sciences Journal*, 51(4):599–612.
- Pokhrel, P. and Gupta, H. V. (2010). On the use of spatial regularization strategies to improve calibration of distributed watershed models. *Water resources research*, 46(1).
- Sankarasubramanian, A. and Vogel, R. M. (2002). Annual hydroclimatology of the united states. *Water Resources Research*, 38(6):19–1.
- Sankarasubramanian, A. and Vogel, R. M. (2003). Hydroclimatology of the continental united states. *Geophysical Research Letters*, 30(7).
- Shannon, C. E. (1948). A mathematical theory of communication. *ACM SIGMOBILE Mobile Computing and Communications Review*, 5(1):3–55.
- Shi, Y. and Eberhart, R. (1998). A modified particle swarm optimizer. In *Evolutionary Computation Proceedings, 1998, IEEE World Congress on Computational Intelligence, The 1998 IEEE International Conference on*, pages 69–73. IEEE.
- Singh, V. P. (1997). The use of entropy in hydrology and water resources. *Hydrological processes*, 11(6):587–626.
- Singh, V. P. (2000). The entropy theory as a tool for modelling and decision-making in environmental and water resources. *WATER SA-PRETORIA-*, 26(1):1–12.
- Singh, V. P. (2013). *Entropy theory and its application in environmental and water engineering*. John Wiley & Sons.
- Tekleab, S., Uhlenbrook, S., Mohamed, Y., and et al (2011). Water balance modeling of upper blue nile catchments using a top-down approach. *Hydrology and Earth System Sciences*, 15(7):2179–2193.
- Thomas, H. A. (1981). Improved methods for national water assessment. WR15249270[A].

- Wang, D. and Alimohammadi, N. (2012). Responses of annual runoff, evaporation, and storage change to climate variability at the watershed scale. *Water Resources Research*, 48(5).
- Wang, D. and Tang, Y. A. (2014). A one-parameter budyko model for water balance captures emergent behavior in darwinian hydrologic models. *Geophysical Research Letters*, 41(13):4569–4577.
- Weijs, S. V. and Giesen, N. V. D. (2011). Accounting for observational uncertainty in forecast verification: An information-theoretical view on forecasts, observations, and truth. *Monthly Weather Review*, 139(7):2156–2162.
- Weijs, S. V., Giesen, N. V. D., and Parlange, M. B. (2013). Data compression to define information content of hydrological time series. *Hydrology. and Earth System Sciences*, 17(8):3171–3187.
- Weijs, S. V., Schoups, G., and Giesen, N. V. D. (2010). Why hydrological predictions should be evaluated using information theory. *Hydrology and Earth System Sciences*, 14(12):2545–2558.
- Xiong, L. and Guo, S. (1999). A two-parameter monthly water balance model and its application. *Journal of Hydrology*, 216(1):111–123.
- Xu, X., Yang, D., Yang, H., and et al (2014). Attribution analysis based on the budyko hypothesis for detecting the dominant cause of runoff decline in haihe basin. *Journal of Hydrology*, 510:530–540.
- Yang, D., Sun, F., Liu, Z., and et al (2007). Analyzing spatial and temporal variability of annual water-energy balance in nonhumid regions of china using the budyko hypothesis. *Water Resources Research*, 43(4).
- Yang, H., Yang, D., Lei, Z., and et al (2008). New analytical derivation of the mean annual water-energy balance equation. *Water Resources Research*, 44(3).
- Zhang, L. and Dawes, W. R. (2001). Response of mean annual evapotranspiration to vegetation changes at catchment scale. *Water resources research*, 37(3):701–708.
- Zhang, L., N., P., K., H., and et al (2008). Water balance modeling over variable time scales based on the budyko framework–model development and testing. *Journal of Hydrology*, 360(1):117–131.