Information Analysis of Catchment Hydrologic Patterns across Temporal Scales

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

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

Abstract

Catchment hydrologic cycle takes on different patterns across temporal scales. The interim between event-scale hydrologic process and mean annual water-energy correlation pattern requires further examination to justify self-consistent understanding. In this paper, the temporal scale transition revealed by observation and simulation was evaluated in an information theoretical framework named *Aleatory Epistemic Uncertainty Estimation*. The Aleatory Uncertainty refers to posterior uncertainty of runoff given the input variables' observations. The Epistemic Uncertainty refers to the posterior uncertainty increase due to the imperfect observation *decoding* in models. Daily hydrometeorological observations in 24 catchments were aggregated from 10 days to 1 year before implementing the information analysis. Estimations of information contents and flows of hydrologic terms across temporal scales were related with the catchments' seasonality type. It also showed that information distilled by the monthly and annual water balance models applied here did 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 hydrologic mechanism.

Keywords: information theory, temporal scale, hydrologic model

*Corresponding author

Email addresses: `baoxianp@uci.edu` (Baoxiang Pan), `congzht@tsinghua.edu.cn` (Zhentao Cong)

1. Introduction

A major realm of hydrologic community is to figure out the components of hydrologic 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 equations usually change with scales. The term *scale* here refers to the characteristic time (or length) of a process, observation or model[1]. While large scale hydrologic patterns are expected to emerge by *integrating* detailed event-scale hydrologic control functions along the spatial and temporal paths, such reductionism approach often fails to distill the dominant factors that contribute to catchment's long range hydrologic behaviors. On the other hand, the holism perspective has been widely adopted to provide coarse explanation of catchment's mean annual water balance[2, 3]. A cut-through between the reductionism and holism paradigms is required for reaching a self-consistent understanding of hydrologic temporal scale transition[4].

One practical attempt toward this goal is to expand the existed mean annual models to fit for small temporal scale hydrologic simulation. Classical Budyko curve[5] which connects the ratio of catchment's actual evapotranspiration(E) to precipitation(P) and dryness index($\frac{PE}{P}$, where PE denotes potential evapotranspiration) can not exert first order control of water balance for excluding the impact of soil moisture change within the scales the model focuses on[6, 7, 8]. By including the soil moisture storage term(S), the expanded models could be applied for seasonal even monthly hydrologic simulation and prediction.

As is declared, the incorporation of new variable increases the system's freedom degree, which should be compensated by introducing new independent governing equation. In Budyko Curve, the water supply P is partitioned into actual evapotranspiration(E) and runoff(Q) with the evapotranspiration demand being PE. Accordingly, the adjusted models make a multi step precipitation partition given the water competition between catchment replenishment demand and evapotranspiration demand. Table 1 listed the analysis of some widely-accepted water balance models following this cognitive framework.

Due to the same constrains of extreme zero order and first order boundary conditions where water supply far surpasses or falls behind demand, the curves above take on similar shapes and achieve similar satisfactory performances in monthly scale simulation. Each model requires 2(in TPWB) to 4(in ABCD or DWBM) parameters to adjust curve concavity and position to fit for observations. The statistic characteristics of state variables and parameters differ as the modelling scale changes[12]. It is interesting to make a closer examination of data-revealed hydrologic pattern and model performance during scale transition given the wide temporal scale gap between event-scale hydrologic process and annual-mean water balance.

Variables in large temporal scale hydrologic models are the aggregation of themselves in small scale models. The goal of these models is to find out the control of the aggregated variables on that period's total water balance, which is determined by the inner-scale temporal distribution of hydrologic events and catchment's storage capacity[13]. For instance, given the same average water and energy supply, catchments with uniformly distributed rainfall and large storage capacity tend to generate less direct runoff and more evapotransipiration. These determinants are simplified as state varaible S in the water balance models. The motivation of this paper was to quantify data-revealed(potential) and model-revealed(achieved) control of hydrometeorological variables' mean values on catchments average water balance over different temporal scales. The estimations were implemented within an information theoretical framework named *Aleatory Epistemic Uncertainty Estimation(*AEUE*)*[14] for its mathematical elegance in assessing such insufficient information control problems.

The rest of the paper is structured as follows: in section 2, the definitions and properties of *AEUE* are briefly introduced before clarifying its logical extentions and technical adaptions in this work. Section 3 gives the data description. The results and their interpretations are in section 4 and 5. The last section draws conclusion and recommends directions for future work.

2. Methodology

2.1. Aleatory and Epistemic Uncertainty in Hydrologic Simulation

It is intuitively believed that infrequent samples of a random variable bring more surprise, 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 $-logp$. The average information content of random variable X is:

$$H(X) = -\Sigma p(x) logp(x) \tag{1}$$

$$h(X) = -\int f(x) logf(x) dx \tag{2}$$

$H(X)$ and $h(X)$ are defined as discrete and continuous *Shannon Entropy*[15], both measured in *bits* for logarithm base 2. The two terms are connected with the following limitation relation[16]:

$$H(X^{\Delta}) \to h(X) - log\Delta, \quad as\ \Delta \to 0 \tag{3}$$

As is shown in Figure 1, X^{Δ} denotes the discrete random variable by scattering the continuous random variable X into bins with width of Δ in its probability density function(p.d.f). $h(X) - log\Delta$ represents the information content required to describe X to $-log\Delta$ bit accuracy[16]. Here $-log\Delta$ bit accuracy means X takes a same value in a Δ width bin in the p.d.f. curve.

Entropy maps the probability distribution function to a real number. Apply this functional map to both sides of Bayes' Theorem, we have:

$$-\sum_X \sum_Y P(X,Y) logP(X|Y) = -\sum_X \sum_Y P(X,Y) logP(X) - \sum_X \sum_Y P(X,Y) log\frac{P(XY)}{P(X)P(Y)} \tag{4}$$

or

$$-\int\int P(X,Y) logP(X|Y) dXdY = -\int\int P(X,Y) logP(X) dXdY - \int\int P(X,Y) log\frac{P(XY)}{P(X)P(Y)} dXdY \tag{5}$$

In the literature of information theory, they are denoted as:

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

or

$$h(X|Y) = h(X) - I(X,Y) \tag{7}$$

Each term in the equations above is named in correspondence to that in Bayes' Theorem. Explicitly, $H(X|Y)$ and $h(X|Y)$ are called *conditional entropy*, which represent the posterior uncertainty of X given the knowledge of Y. $H(X)$ and $h(x)$ represent the prior uncertainty. $I(X;Y)$ is called *mutual information*. It represents the information contribution one variable provides to the other.

The application of Bayes' Theorem in hydrologic simulation assessment[17, 18, 19] endows each of the information terms with its hydrologic significance. Concieve X as the hydrologic variable to be simulated and Y as the input variable observations, both taken as continuous random variables, equation 7 quantifies the residual uncertainty of the hydrologic system given the inaccurate and insufficient observation system. This uncertainty is named as *Aleatory Uncertainty(AU)*[14]:

$$Aleatory\ Uncertainty = H(X_o) - I(X_o;X_i) \tag{8}$$

Here X_o, X_i represent the observed output and input random variables of hydrologic models. It should be noted X_i is usually high dimensional, since information comes from different resources(both hydrometeorology and underlying surface observations) and lagging effects of former hydrologic behaviours. Models try to distill the largest information from X_i to construct their simulations X_s. Given that X_s is function of X_i, *Data Processing Inequality Theorem*[16] confirms that the potential maximum information models could distill(represented as $I(X_o, X_s)$) is no larger than that provided by the original data(represented as $I(X_o, X_i)$). A detailed proof is given in Appendix. The information loss due to imperfect input data processing is defined as *Epistemic Uncertainty(EU)*[14]:

$$Epistemic\ Uncertainty = I(X_o;X_i) - I(X_o;X_s) \tag{9}$$

Eqaution 8 and 9 construct the *Aleatory Epistemic Uncertainty Estimation(AEUE)* framework. The sum of AU and EU is the posterior uncertainty of the hydrologic system given the simulation system.

2.2. Extending AEUE for Temporal Scale Information Analysis

To implement $AEUE$ across temporal scales, daily hydrometeorologic observations were aggregated into different temporal scales. The aggregated data were used for estimation of each term in Equation 8 and 9. To achieve a more explicit information analysis, we adapted the strategy to gradually expand input variable species and lagging steps to detect the decreasing trajectory of AU. The decrease is attributed as the information contribution of the included variables. For example, $AU(Q; P, PE) - AU(Q; P)$(which simplified to $I(Q; P, PE) - I(Q; P)$ according to Equation 8) represents the information contribution of including energy supply constrains(PE) in simulation, while $AU(Q; P, P_{former}, PE, PE_{former}) - AU(Q; P, PE)$(which simplified to $I(Q; P, P_{former}, PE, PE_{former}) - I(Q; P, PE)$) represents the information contribution of former calculating steps' hydrologic lagging effects, in other words, it denotes the soil moisture memory significance.

EU of two typical water balance models were estimated. The Two Parameters Water Balance Model($TPWB$)[11] was preferred for its simplicity and satisfactory performance at monthly scale. The model structure was listed in Table 1. The other model was Budyko Model. It is the combination of Budyko Curve[20] and mass conservation function. The most significant distinction between the two models is that $TPWB$ adpats iterative structure. The performance of iterative models depends on its state variable's capacity to distill information of system's lagging effects and its constitutive functions' capacity to utilize the distilled information. These two factors were discerned through distinguishing $I(Q; P, P_{former}, PE, PE_{former})$, $I(Q; P, PE, S)$ and $I(Q; Qs)$, where S represents model's state variable and Q_s represents simulated runoff. The difference of the first two terms tells state variable's representativeness and difference between the last two terms tells constitutive function's data processing efficiency.

Given the analysis above, the explicit $AEUE$ framework estimates the following terms as listed in Table 2:

2.3. High Dimensional Mutual Information Estimator

The major obstacle $AEUE$ faces is the estimation of high dimensional mutual informaion terms in Table 2. Samples of finite length can not support accurate estimation of their probability distribution in high dimensional spaces. This phenomenon, known as dimensionality curse[21], has hindered the definition-based estimation of mutual information. Given this, some "nonplugin" methods have been developed. The basic idea is to construct relation between carefully designed sample statistics with its entropy[22] or to perform transform for a step-by-step entropy estimation[23]. Mutual information was estimated afterwards with the following equation:

$$I(X;Y) = H(X) + H(Y) - H(X,Y) \tag{10}$$

As can be detected, error may accumulate in these algorithms.

The estimator we adapted here belongs to the first mode, but makes direct estimation of mutual information. The formula is based on the concept of k nearest neighbour distances[24]:

$$I(X,Y) = \psi(k) - N^{-1}\sum_{i=1}^{N}[\psi(n_x(i)+1) + \psi(n_y(i)+1)] + \psi(N) \tag{11}$$

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. k takes 4 to balance the statistical error and systematic error according to Kraskov's suggestion[24].

An intuitive explanation of equation 11 is that it estimates mutual information with statistics that depicts the average concentrating density of each window opened around a sample point. Numerical experiments showed that even less than 30 sample size produces satisfying results. For a strict proof, please refer to Kraskov(2004).

The samples' distance function should be pre-defined to determine n_x and n_y in Equation 11. Since different dimensions hold specific hydrologic meanings and are not symmetric, the distance could not be efficiently depicted with

Euclidean Norm. Kernel trick was adapted to implicitly measure sample distances in their feature space. Kernels were chosen by optimizing their correspondent support vector regression(SVR)[25] performance implemented on the test set. Satisfactory kernel SVR performance suggests well-balanced compromise between minimizing variance and bias in the proper feature space. Results showed that kernel SVR was effective in performing high dimensional regression of hydrologic variables[26, 27, 28, 29, 30]. The following function was used to depict the distance between two input variable samples x_1 and x_2:

$$Kernel\_Distance(x_1, x_2) = |f(x_1) - f(x_2)| \tag{12}$$

$f(x)$ is the support vector regression function that fit the input to the output variable. n_x and n_y were determined after the calculation of distances between samples.

In practice, the support vector regression was implemented using the libsvm package[31]. Radial basic function kernel was adopted for its satisfying performance. The data were first normalized to $[-1, 1]$ to balance the impact of different dimensional terms. Results were sensitive to the penalty function parameter c and kernel parameter g, both of which were auto calibrated with particle swarm optimization algorithm[32]. .

3. Data

24 catchments with daily hydrologic records (including P,PE and Q) from MOPEX data set[33] were selected to implement temporal scale information analysis. Given their temporal water-energy distribution patterns, the selected catchments were 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 was based on the amplitude and phase of the average daily rainfall fitted with *sine* curve. If the amplitude

was less than 0.45, the catchment was taken as weak seasonality. If the phase of rainfall was inverse to that of potential evapotranspiration, it was taken as asynchronous rainfall energy climate type. The general conditions of the catchments were listed in Table 3. The vegetation, soil type, land use and other specifc catchment information are available from the following link: ftp://hydrology.nws.noaa.gov/pub/gcip/mopex/US_Data/

4. Aleatory and Epistemic Uncertainty Across Temporal Scales

4.1. Aleatory Uncertainty

Considering the implication of Equation 3, the *Aleatory Uncertainty* comparison across temporal scales requires a scheme to preset resolution for each temporal scale. Since the acceptable deviation of small temporal scale hydrologic models should be stricter than larger scale models, we pre-required relative constant resolution across temporal scales. Explicitly, the width of the bin into which the objective variable is clustered in the p.d.f. curve is proportional to its mean value. It was further assumed that the mean value to be proportional to its temporal scales. Thus, the quantization correction term in equation 3 is proportional to the logarithm of the temporal scale. For two scales m and n into which daily runoff observation data were aggregated, the entropy difference for depicting them with specific resolutions is:

$$H(R_m) - H(R_n) = h(R_m) - h(R_n) - log\frac{m}{n} \tag{13}$$

Given this baseline, the estimated *Aleatory Uncertainty* was shown in Figure 2:

In each subgraph above, the abscissa represents the input steps, for example, number n denotes that the current and $(n-1)$ lagging steps' input observations were used to decrease the uncertainty of runoff estimation. The ordinate represents the estimating temporal scale, which varied from 10 days to a year.

The general pattern is that AU decreases as temporal scale expands or more lagging input observations were included. It could be depicted that AU was closely related to catchment's seasonality type, specifically:

$$AU_{SA} > AU_{WA} > AU_{WS} > AU_{SS} \tag{14}$$

The detailed analysis discerning each term's information contribution for different catchments were discussed in the next session.

4.2. Epistemic Uncertainty

The estimated *Epistemic Uncertainty* across temporal scales were shown in Figure 3:

For *TPWB* model, maximum *EU* appears around temporal scales from 2 months to half a year. This showed that at seasonal temporal scale, the model can not distill the information provided by the data effectively.

The *EU* difference between *TPWB* and Budyko Model was related to the catchment's seasonality. In 11 out of 14 asynchronous seasonality catchments, *EU* differs significantly at small temporal scales. The difference diminished as scale expands. In the rest 3 asynchronous catchments and 14 synchronous catchments, the difference stayed relatively constant across temporal scales.

5. Specific Information Analysis

5.1. Information Contribution of Included Input Terms

The including of new information sources could decrease simulation uncertainty. As was shown in Figure 4, the specific information contribution of including energy provision PE and observed previous runoff Q_p were obtained by subtracting the right column graphs from the left column ones. For instance, $AU(Q;P) - AU(Q;P,PE)$ denotes the information contributuion of considering PE in the simulation.

For all the 10 weak seasonality catchments and 5 out of 14 strong seasonality catchments, the information contribution of PE was more significant at temporal scales of less than half a year. It distributed more uniformly across temporal scales in the left 9 strong seasonality catchments.

The prominent information contribution of previous runoff at small scales in some catchments were attributed to runoff convergence influence.

5.2. Information Contribution of Soil Moisture Memory

Previous hydrologic behaviour exerts influence on current hydrologic response due to the storage capacity of soil moisture. Here this influence is defined as soil moisture memory and was represented by the difference between splines in each subgraph of Figure 2.

The second dissection scheme checks the information contribution of including lagged inputs into mutual information estimation. This is implemented by making differences between Aleatory Uncertainty estimated with different input steps, for instance, the nth spline in each graph from Figure 5 equals to the difference of the $(n+1)$th spline and nth spline in the corresponding graph from Figure 2.

It could be depicted that the first lagging steps' input variables provide most information contribution across all the temporal scales estimated here. As shown in the first column of Figure 5, this lagging effects were not significant when considering only the water provision. The consideration of energy provision is of key importance in estimating the soil moisture length.

5.3. Dissection of Model's Information Distilling Capacity

As was declared, the simulation capacity of iterative structure models depends on its capacity to distill lagging effects information and process such information. The information distilling and processing capacity were discerned in Figure 6.

The ordinate of each graph denotes mutual information. $I(Q; I_c, S)$ represents the mutual information between runoff and current input together with current state variable. It denotes model's capacity to distill lagging effects from previous hydrologic behaviours. The difference between $I(Q; I_c, S)$ and $I(Q; Q_s)$ denotes model's capacity to process the information it distilled.

It could be depicted that in synchronous climate catchments, the information distilled by $TPWB$ and Budyko Model increases as temporal scale expands, while in asynchronous climate catchments, the information distilling capcity of $TPWB$ does not change monotonously with temporal scales.

6. Discussion and Conclusion

The aggregation of event-scale hydrologic processes yields to the large temporal scale water-energy correlation pattern. The temporal scale transition was examined in the extended *Aleatory Epistemic Uncertainty Evaltuion* framework.

The *Aleatory Uncertainty* quantified the uncertainty caused by inaccurate and insufficient observation. For a large temporal scale, since the daily observations were aggregated, the *Large Number Law* guaranteed that the accumulated error tended to 0 when there being no systematic observation bias. Thus, AU was mainly attributed to the insufficience of data. The aggregated variables could exert certain control to the total water balance. The control significance was closely related to the seasonality type as quantified in the previous session.

The *Epistemic Uncertainty* of a monthly and mean annual water balance model were estimated. The performance of $TPWB$ was evaluated by quantifying its information distilling capacity and data processing efficiency. Results showed that information distilled by the models applied here did not correspond to the information provided by input observations around temporal scale from two months to half a year. This called for a better understanding of seasonal hydrologic mechanism. The information distillng capacity difference of $TPWB$ and Budyko Model were related to the inner-year distribution of water and energy. In asynchronous catchments, the difference converged to 0 at half year scale, which suggested close hydrologic cycle.

The evaluations also revealed some counter-intuitive phenomenon that needs to be stressed and explained. The meaning of soil storage capacity from a large temporal scale perspective was not as physically clear as it is in event scale. The state variable S is influenced by the distribution of hydrologic processes and soil properties. The strict definition is required to explain the uncertainty differences in different seasonality catchments.

Appendix: Data Processing Inequality

The *Data Processing Inequality* states that if X, Y, Z form a Markov Chain(deonted as $X \to Y \to Z$), which means X and Z are conditionally independent given Y, then:

$$\mathrm{I(X;Z)} \leq I(X;Y)$$

The proof is given due to the non negative effects of information:

$$\mathrm{I(X;Z)}= \mathrm{H(X)}\text{-}\mathrm{H(X}|Z) \leq H(X) - H(X|Y,Z) = H(X) - H(X|Y) = I(X;Y)$$

Specifically, if $Z = f(Y)$, then $X \to Y \to f(Y)$, we have:

$$\mathrm{I[X;g(Y)]} \leq I[X;Y]$$

This confirms that information could not be produced through data processing.

For hydrologic simulation, due to the non negative effects of information, the inclusion of new input terms will not increase *AU* in our context. From a functional perspective, models acts as a function that transfers input variables into output variables, with the estimated state variables as intermediate products. Variables from models form the following Markov Chain:

$$\mathrm{Q} \to Input_{previous}, Input_{current} \to S, Input_{current} \to Q_s$$

This clarified the reason to perform model information diagnosis as presented in Figure 6.

7. Acknowledgements

We are grateful to the financial supports offered by the National Science Foundation of China(51479088, 51179083 and 91225302). For detailed algorithm 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] L. Zhang, W. R. Dawes, Response of mean annual evapotranspiration to vegetation changes at catchment scale, Water resources research 37 (3) (2001) 701–708.

[3] 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).

[4] 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.

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

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

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

[8] 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.

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

[10] 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.

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

[12] Q. Hu, Y. Wang, K. Liu, Z. Wang, Monthly runoff simulation based on improved two-parameter monthly water volume balance model, Journal of Hohai University(Natural Sciences) 6 (2007) 007.

[13] A. Porporato, E. Daly, I. Rodriguez-Iturbe, Soil water balance and ecosystem response to climate change, The American Naturalist 164 (5) (2004) 625–632.

[14] 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.

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

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

[17] J. Freer, K. Beven, B. Ambroise, Bayesian estimation of uncertainty in runoff prediction and the value of data: An application of the glue approach, Water Resources Research 32 (7) (1996) 2161–2173.

[18] G. Kuczera, E. Parent, Monte carlo assessment of parameter uncertainty in conceptual catchment models: the metropolis algorithm, Journal of Hydrology 211 (1) (1998) 69–85.

[19] K. Beven, J. Freer, Equifinality, data assimilation, and uncertainty estimation in mechanistic modelling of complex environmental systems using the glue methodology, Journal of hydrology 249 (1) (2001) 11–29.

[20] 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.

[21] R. Bellman, R. Kalaba, On adaptive control processes, Automatic Control, IRE Transactions on 4 (2) (1959) 1–9.

[22] N. Leonenko, L. Pronzato, V. Savani, et al., A class of rényi information estimators for multidimensional densities, The Annals of Statistics 36 (5) (2008) 2153–2182.

[23] 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.

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

[25] D. Basak, S. Pal, D. C. Patranabis, Support vector regression, Neural Information Processing-Letters and Reviews 11 (10) (2007) 203–224.

[26] 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.

[27] 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.

[28] 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.

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

[30] 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.

[31] 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.

[32] 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.

[33] 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.

Table 1: Structure Analysis of Water Balance Models[1]

Model	Replenishment Demand	Evapotranspiration Demand	Constitutive Functions
ABCD[9]	b	PE	$\Delta S = \frac{S+P}{2a} - \sqrt{(\frac{S+P+b}{2a})^2 - \frac{(S+P)b}{a}}$ $E = \Delta S \times (1 - e^{\frac{-PE}{b}})$
DWBM[10]	$S_{max} - S + PE$	PE	$\Delta S = S_{max} - S + PE + P - [(S_{max} - S + PE)^{\omega} + P^{\omega}]^{\frac{1}{\omega}}$ $E = \Delta S + S + PE - [(\Delta S + S)^{\omega_2} + PE^{\omega_2}]^{\frac{1}{\omega_2}}$
TPWB[11]	SC	PE	$\Delta S = (S + P - E) \times tanh(\frac{S+P-E}{SC}) - S$ $E = C \times PE \times tanh(\frac{P+S}{PE})$

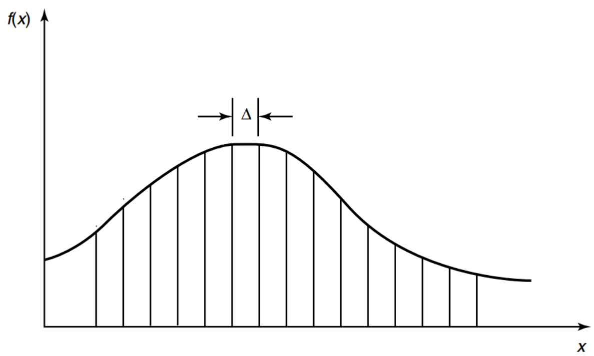


Figure 1: Quantization of Continuous Random Variable

Table 2: Information Terms to be Esitimated

Classification	Estimated Terms
Observation	$h(Q)$
Focused	$I(Q;P), I(Q;P,P_{former})$
	$I(Q;P,PE), I(Q;P,P_{former},PE,PE_{former})$
	$I(Q;P,P_{former},PE,PE_{former},Q_{former})$
Model	TPWB: $I(Q;Qs)$, $I(Q;P,PE,S)$
Focused	Budyko: $I(Q;Qs)$

Table 3: Catchment Condition

Climate Type	ID	Location	Area(km^2)	$P_{mean}(mm)$	$PE_{mean}(mm)$	$Q_{mean}(mm)$
	02143000	81°W,36°N	215	1299	882	553
	02165000	82°W,34°N	611	1252	965	539
WA	02329000	84°W,31°N	2953	1321	1101	330
	02375500	87°W,31°N	9886	1452	1061	549
	02478500	86°W,31°N	6967	1440	1055	489
	05585000	91°W,40°N	3349	922	993	232
	06908000	93°W,39°N	2901	1001	1066	261
WS	07019000	91°W,38°N	9811	1006	959	303
	07177500	96°W,36°N	2344	948	1259	221
	07243500	96°W,36°N	5227	935	1303	160
	02414500	86°W,33°N	4338	1371	976	542
	02472000	89°W,31°N	1924	1442	1059	509
	11025500	117°W,33°N	290	522	1407	34
SA	11532500	124°W,42°N	1577	2748	751	2212
	12459000	121°W,48°N	2590	1613	681	1105
	13337000	116°W,46°N	3056	1287	775	872
	14359000	123°W,42°N	5317	1052	851	510
	05418500	91°W,42°N	4022	854	1017	254
	05454500	91°W,41°N	8472	839	984	224
	05484500	94°W,41°N	8912	794	998	117
SS	06810000	96°W,40°N	7268	808	1027	173
	06892000	95°W,39°N	1052	941	1110	228
	06914000	95°W,38°N	865	950	1186	236
	07183000	96°W,37°N	9889	877	1250	187

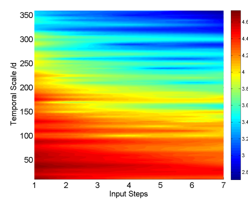
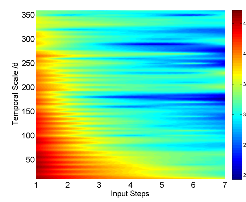
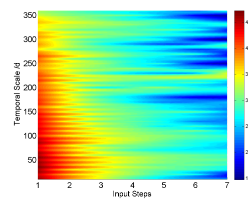
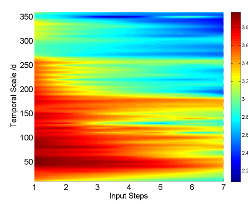
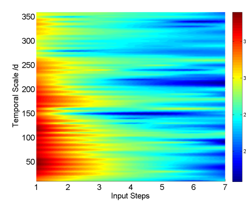
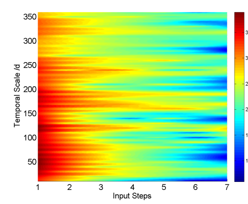
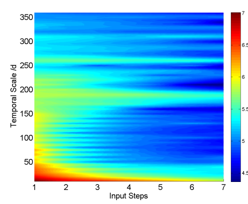
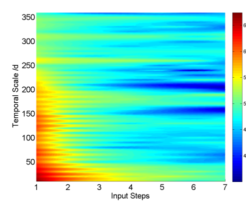
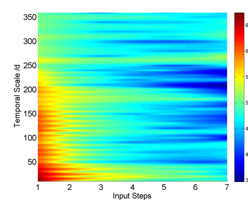
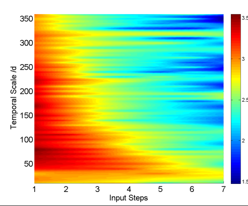
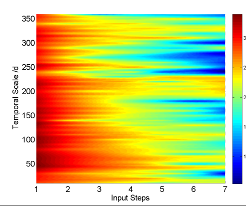
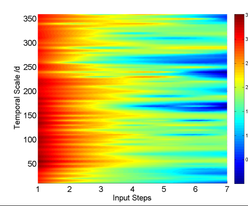
Catchment Type	$AU(Q;P)$	$AU(Q;P,PE)$	$AU(Q;P,PE,Q_{former})$
WA(02143000)			
WS(05585000)			
SA(11532500)			
SS(06810000)			

Figure 2. Aleatory Uncertainty

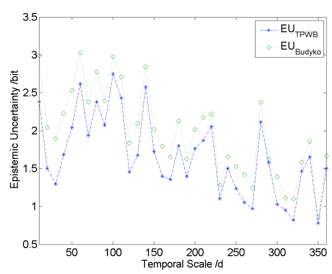
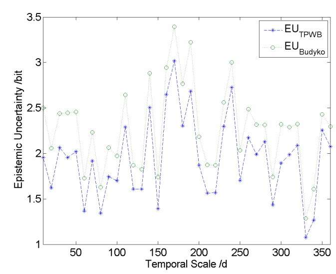
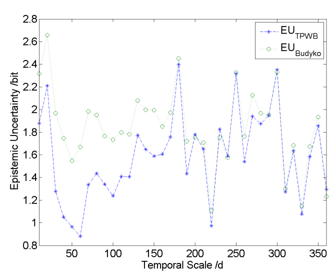
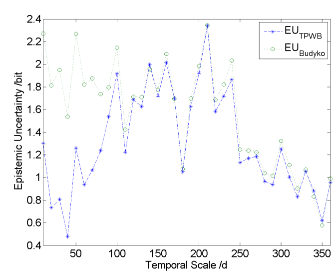
<table>
<tr><th>Type</th><th>Weak Seasonality</th><th>Strong Seasonality</th></tr>
<tr><td>Synchronous</td><td>
</td><td>
</td></tr>
<tr><td>Asynchronous</td><td>
</td><td>
</td></tr>
</table>

Figure 3. Epistemic Uncertainty

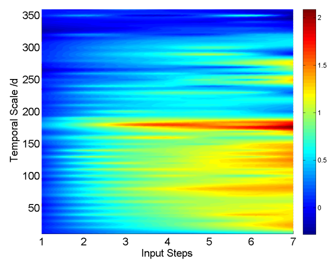
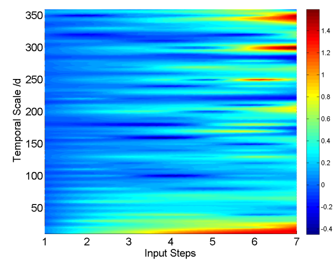
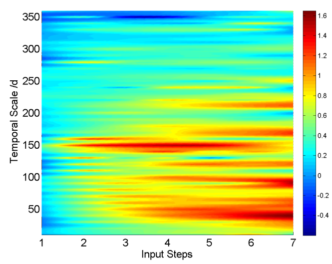
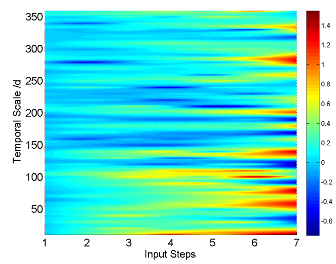
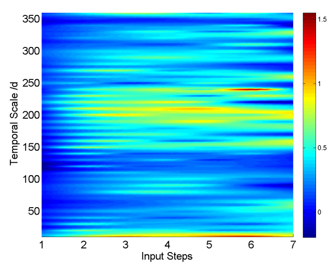
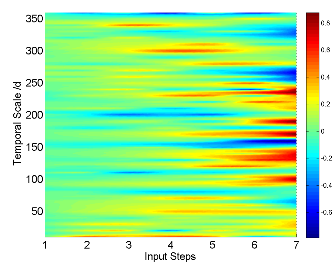
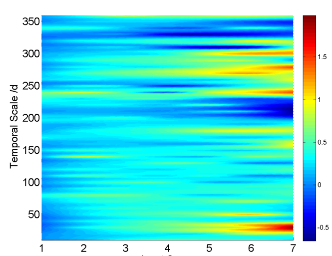
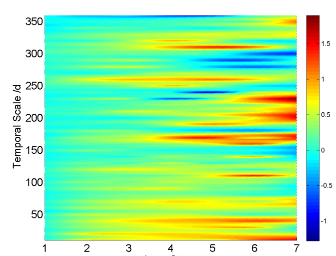
Type	$AU(Q;P) - AU(Q;P,PE)$	$AU(Q;P,PE) - AU(Q;P,PE,Q_p)$
WA(02143000)		
WS(05585000)		
SA(11532500)		
SS(06810000)		

Figure 4. Information Contribution of PE and Q_p

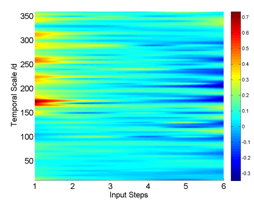
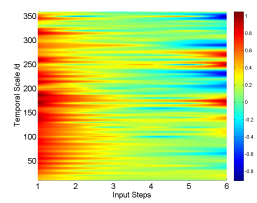
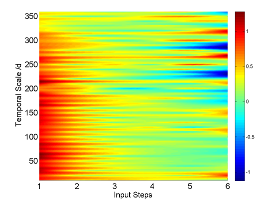
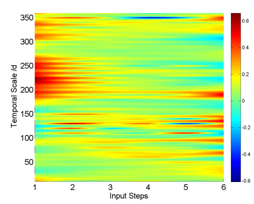
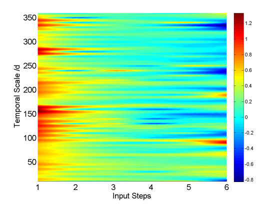
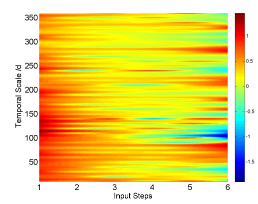
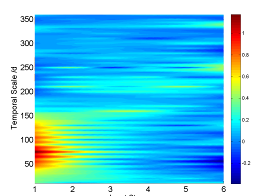
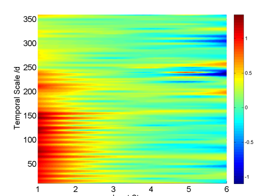
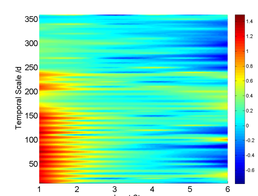
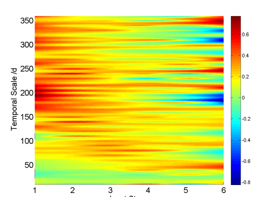
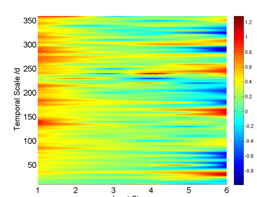
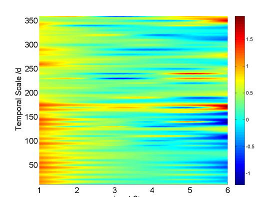
Type	$AU(R; P..)$ $-AU(R; P.)$	$AU(R; P.., PE..)$ $-AU(R; P., PE.)$	$AU(R; P.., PE.., R_{former}..)$ $AU(R; P., PE., R_{former}.)$
WA(02143000)			
WS(05585000)			
SA(11532500)			
SS(06810000)			

Figure 5. Information Contribution of Former Inputs

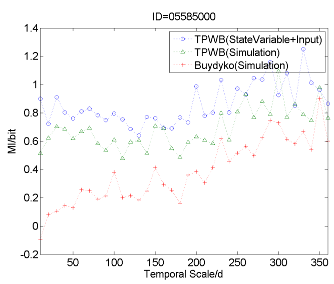
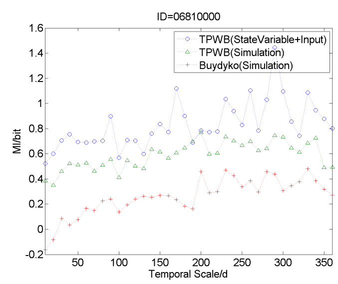
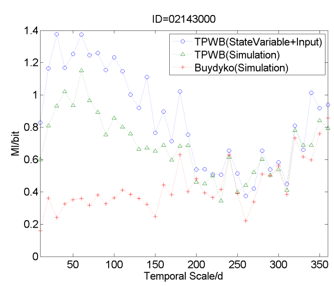
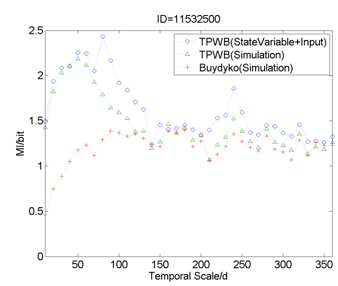
Type	Weak Seasonality	Strong Seasonality
Synchronous		
Asynchronous		

Figure 6. Mutual Information Between Runoff and Simulation