# Information Analysis of Watershed Hydrological Patterns Across Temporal Scales

Baoxiang Pan, Zhentao Cong and Dawen Yang
Institute of Hydrology and Water Resources, Tsinghua University

## Objectives

Explain the following issues in the context of Information Theory:

- The existence and transition of watershed hydrological patterns across temporal scales revealed by data.
- To what extent models capture these patterns.

## Introduction

Hydrological cycle takes on different patterns and calls for different models across temporal scales. The clustering of daily hydrological observations causes information loss of the time domain details, but, on the other hand, presents a water-heat correlation pattern as the temporal scale expands.
To quantify the two impacts during temporal upscaling, we employ two basic conceptions from *information theory*:

| | Entropy | Mutual Information |
|---|---|---|
| **Discrete** | $H(X) = -\Sigma p(x) log p(x)$ | $I(X;Y) = \sum_{x,y} p(x,y) log \frac{p(x,y)}{p(x)p(y)}$ |
| **Continuous** | $h(X) = -\int f(x) log f(x) dx$ | $I(X;Y) = \int\int f(x,y) log \frac{f(x,y)}{f(x)f(y)} dxdy$ |

Entropy is a measure of uncertainty of a random variable. Mutual information depicts the information decrease of a random variable given the knowledge of the other, and vice versa. Both of their dimensions are *nat* for logarithm base $e$.
Certain logical and methodological issues should be clarified before applying these terms to quantify the information existence and flow revealed by hydrological data and models.

## Logical Consideration

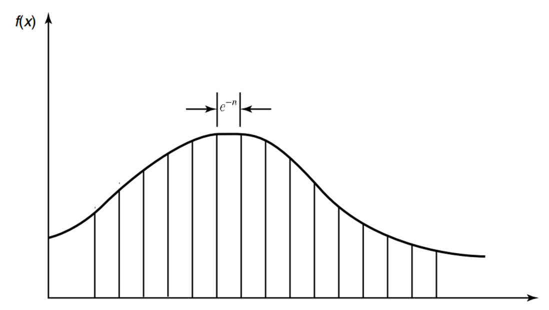
Table 1: Estimated Information Terms

| Classification | | Estimated Terms |
|---|---|---|
| Model Irrelevant | | $h(Q_t)$ |
| | | $I(Q_t; P_t, EP_t), I(Q_t; P_t, P_{t-1}, EP_t, EP_{t-1}), ...$ <br> $I(Q_t; P_t, P_{t-1}, ..., P_{t-n}, EP_t, EP_{t-1}, ...EP_{t-n})$ |
| | | $I(Q_t; P_t, P_{t-1}, EP_t, EP_{t-1}, Q_t - 1), ...$ <br> $I(Q_t; P_t, P_{t-1}, ..., P_{t-6}, EP_t, EP_{t-1}, ...EP_{t-6}, Q_{t-1}, ...Q_{t-n})$ |
| Model Relevant | TPWB | $I(Q_t; Qs_t),\ I(Q_t; P_t, EP_t, S_t)$ |
| | Budyko | $I(Q_t; Qs_t)$ |

**Symbol Explanation:** $h$ denotes differential entropy; $I$ denotes mutual information; $P_t$,$EP_t$,$Q_t$,$Qs_t$denotes precipitation, potential evapotranspiration runoff observation and runoff simulation at time step $t$. TPWB is a monthly iterative water balance model[1],$S_t$ is its state variable; Budyko is yearly water-heat correlation model.

Information content of continuous random variable is infinite. $h(X) + n$ is the number of nats on the average required to describe $X$ to $n$-nat [2].

The $n$-nat accuracy means $X$ takes a same value in a bin-width of $e^{-n}$ in the p.d.f curve:



We pre-require the relative bin-width stays the same during temporal upscaling:

$$\frac{e^{-p}}{m} = \frac{e^{-q}}{n}$$

Here $m$ and $n$ are two temporal scales at which we re-cluster the runoff data; $p$,$q$ are their accuracy requirements. Thus, the information content difference when quantizing runoff observations $Q$ to $p$ and $q$ nat accuracy approximates:

$$\begin{aligned} \Delta H &\approx h(Q_m) + p - h(Q_n) - q \\ &= h(Q_m) - logkm - h(Q_n) + logkn \\ &= h(Q_m) - h(Q_n) + log\frac{n}{m} \end{aligned}$$

Mutual information still represents the amount of discrete information that can be transmitted over a channel that admits a continuous space of values[2],thus:

- For a fixed temporal scale, the difference of mutual information with different previous input steps represents the correlation between temporal neighbouring hydrological cycles at this scale[3].
- For fixed previous input steps, mutual information estimated at different temporal scales represents the information contribution of the input observation to the output observation.

## Methodological Consideration

Due to the curse of dimensionality, the high dimensional mutual information terms in table 1 could not be accurately estimated. An improved approach combining K-nearest neighbour method and support vector regression is employed in this research.

$$\begin{cases} I(X,Y) = \psi(k) - N^{-1}\sum_{i=1}^{N}[\psi(n_x(i)+1) + \psi(n_y(i)+1)] + \psi(N) \\ SVM\_Metric(x_1, x_2) = |f(x_1) - f(x_2)| \end{cases}$$

The first equation estimated MI with statistics that depict the average concentrating density of each window opened around a sample point[4], the second equation applied the kernel trick in support vector regression to depict *distances* between high dimensional hydrological terms by implicitly mapping them into feature spaces[5]. Numerical experiments shows that even less than 30 sample size good results[4].

## Data & Method

- Re-cluster daily hydrological records (P,EP,Q) from MOPEX basins into temporal scales from 1 day to a year.
- Calculate the model irrelevant information terms.
- Implement hydrological simulation and calculate the model relevant mutual information terms.

## Result

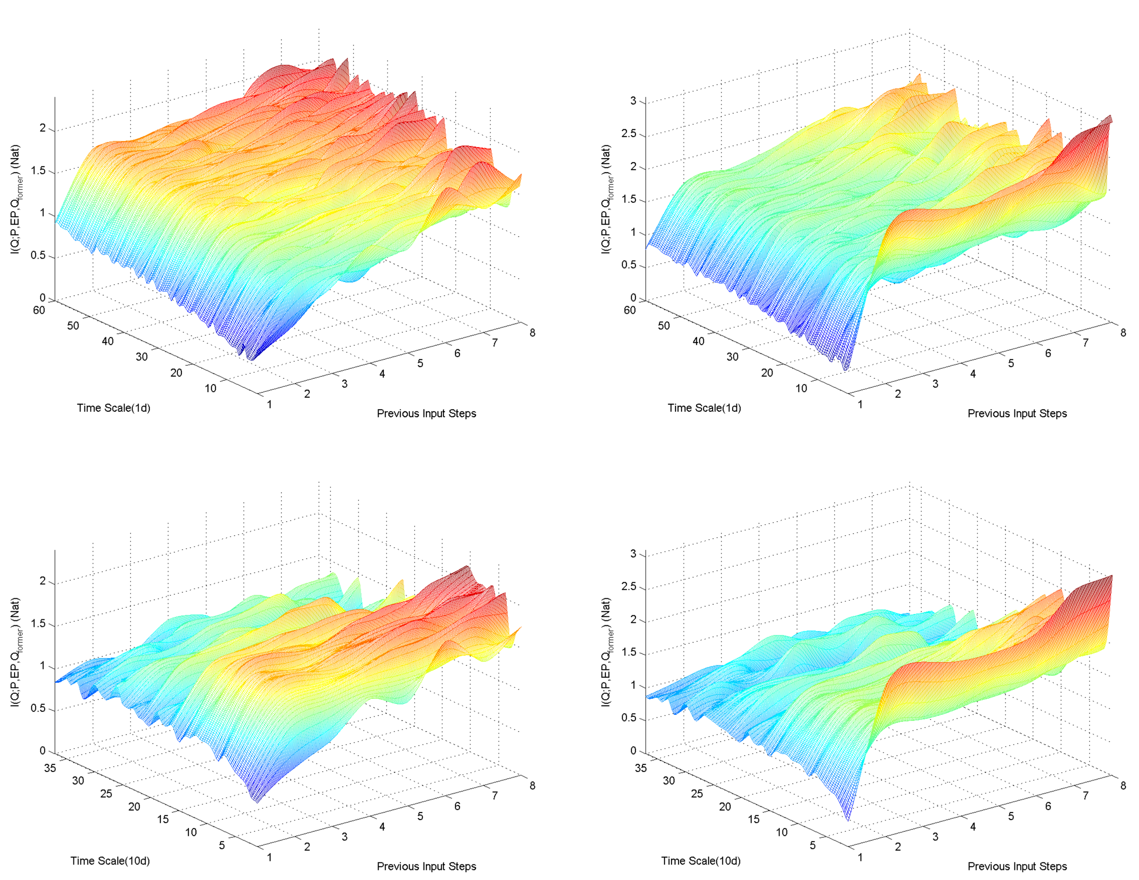


Figure 1: MI-Scale-Previous_Input Relationship

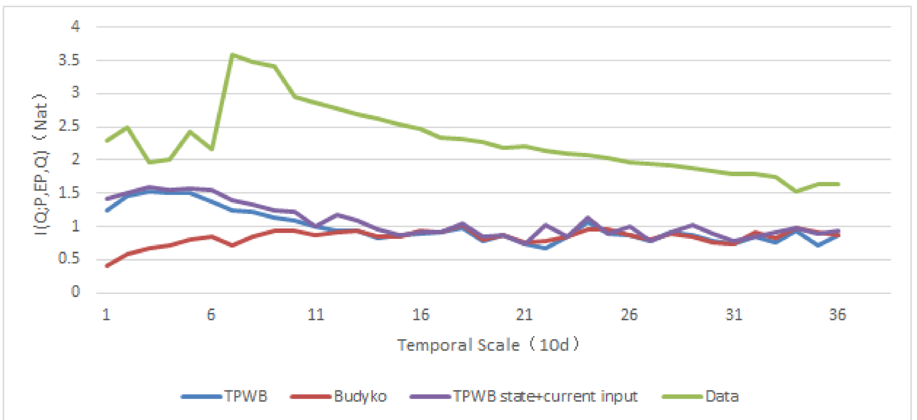


Figure 2: MI revealed by data and models

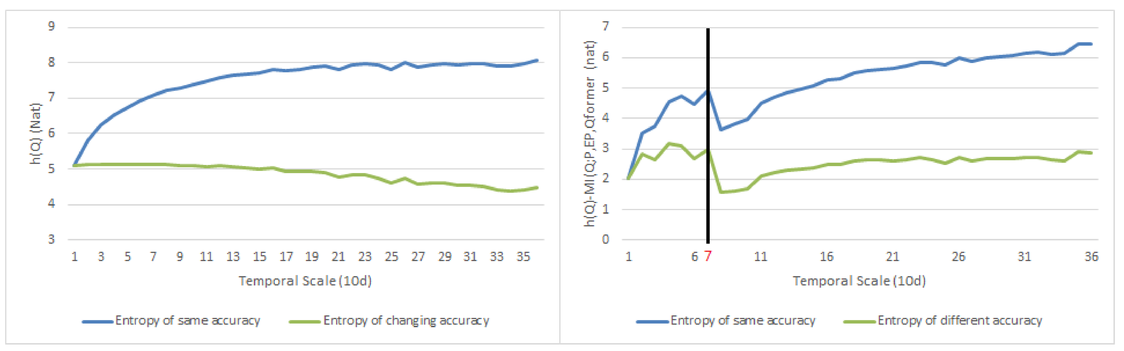


Figure 3: MI revealed by data and models

## Conclusion

- Th impact of previous runoff in daily hydrological simulation is huge, but vanishes quickly as temporal scale expands.
- The correlation between temporal neighbouring hydrological cycles weakens as scale expands.
- **There are spaces for models to improve their abilities to distil information from data.**
- **The re-clustered hydrological observations takes on an abnormal correlation at temporal scales of season.**

## References

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

[2] Thomas M Cover and Joy A Thomas.
*Elements of information theory.*
John Wiley & Sons, 2012.

[3] Wei Gong, Hoshin V Gupta, Dawen Yang, Kumar Sricharan, and Alfred O Hero.
Estimating epistemic and aleatory uncertainties during hydrologic modeling: An information theoretic approach.
*Water Resources Research*, 49(4):2253–2273, 2013.

[4] Alexander Kraskov, Harald Stögbauer, and Peter Grassberger.
Estimating mutual information.
*Physical review E*, 69(6):066138, 2004.

[5] Wei Gong.
Watershed model uncertainty analysis based on information entropy and mutual information.
*Ph.D. thesis, Dep. of Hydraulic Engineering, Tsinghua Univ., Beijing, China.*

## Contact Information

- Email: http://panbaoxiang@hotmail.com
- Github: https://github.com/morepenn
- Phone: +85 133 6672 0253