

# 土壤含水量频域分析

April 7, 2015

- 背景知识
- 理论推导
-

- 不同时间尺度水文数据信息量（微分熵表示相对值）
- 不同时间尺度水文数据间互信息
- 模型提取信息的能力

- 理论分析
- 信息量计算方法改进

- 概率空间：不考虑非一致性，将日数据累加为不同尺度构成不同尺度下的概率空间
- 微分熵不表示绝对信息量，但是两微分熵之差表示同一精确度下信息量之差的绝对值

Table: Estimated Information Terms

Classification		Estimated Terms
Model		$H(Q_t)$
Irrelevant		$I(Q_t; P_t, EP_t), I(Q_t; P_t, P_{t-1}, EP_t, EP_{t-1}), \dots$ $I(Q_t; P_t, P_{t-1}, \dots, P_{t-n}, EP_t, EP_{t-1}, \dots EP_{t-n})$ $I(Q_t; P_t, P_{t-1}, EP_t, EP_{t-1}, Q_{t-1}), \dots$ $I(Q_t; P_t, P_{t-1}, \dots, P_{t-6}, EP_t, EP_{t-1}, \dots EP_{t-6}, Q_{t-1}, \dots Q_{t-n})$
Model	TPWB	$I(Q_t; Qs_t), I(Q_t; P_t, EP_t, S_t)$
Relevant	Budyko	$I(Q_t; Qs_t)$

- 固定时间尺度，不同前期输入项互信息表示了该尺度下相邻时间水文循环的依赖关系，前期信息贡献很小时为完整循环。
- 固定前期输入，不同尺度下互信息表示输入对输出的信息贡献，其差异由量方面决定
  - 数据造成，不同尺度数据信息量大小不同
  - 机制造成，不同尺度水文变量依赖程度不同

用信息熵之差表示数据造成的尺度间互信息的差别，进而可以得到由于机制造成的信息贡献的差别（对随机不确定性的重新解释——仅具有相对大小的意义）。

- ICA 方法缺陷：非线性相关数据计算得到的信息熵偏大；计算互信息造成误差累积。
- KNN+Support Vector Regression :

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

$$SVM\_Metric(x_1, x_2) = |f(x_1) - f(x_2)| \quad (1)$$

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



## MOPEX日水文数据（降水，潜在蒸散发，径流）

Table: Basin Conditions

Number	Location	Area( $\text{km}^2$ )	$P_{\text{mean}}(\text{mm})$	$EP_{\text{mean}}(\text{mm})$	$Q_{\text{mean}}(\text{mm})$
01048000	69.9392W, 44.7072N	1331	3.1848	1.9570	1.8973
02143000	81.4030W, 35.6840 N	215	3.5595	2.4159	1.5140
02165000	82.1764W, 34.4444 N	611	3.4290	2.6437	1.4767
02296750	81.8761W, 27.2219N	3541	3.5356	3.3299	0.6885
02329000	84.3842W, 30.5539N	2953	3.6193	3.0152	0.9050
02375500	87.2342W, 30.9650 N	9886	3.9772	2.9060	1.5039
02478500	88.5480W, 31.1480 N	6967	3.9457	2.8913	1.3403
07243500	96.0650W, 35.6750 N	5227	2.5619	3.5702	0.4392
08033500	94.3986W, 31.0247 N	9418	3.0982	3.5425	0.6016
08167500	98.3828W, 29.8606 N	3406	2.1148	4.1929	0.2851
08171000	98.0886W, 29.9942 N	919	2.2709	4.0492	0.3973
08172000	97.6497W, 29.6650N	2170	2.3097	3.9740	0.4521
08205500	99.1444W, 28.7364N	8881	1.8627	4.2143	0.0385
11025500	116.8653W, 33.1069N	290	1.4289	3.8556	0.0938
11080500	117.8050W, 34.2360 N	220	2.0235	4.0137	0.7134
11532500	124.0539W, 41.7894N	1577	7.5278	2.0572	6.0607

- 日数据按从1天到年际尺度组合
- 计算模型无关信息量
- 跑模型，计算模型相关信息量

# 结果 case 11532500

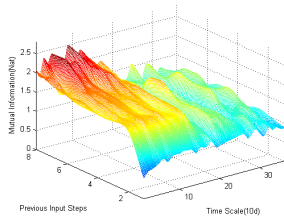
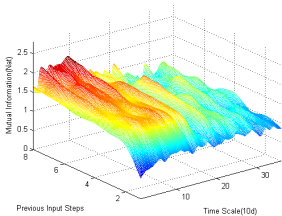
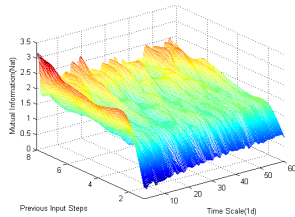
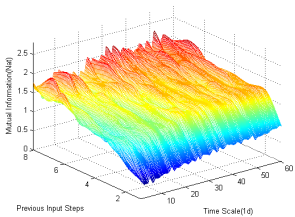


Figure: I-N-T

# 投影到互信息-时间尺度平面:

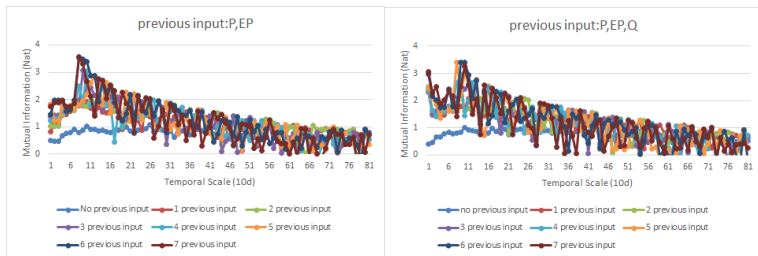


Figure: Compressed  $I(X_o; X_i)$ -T Slice

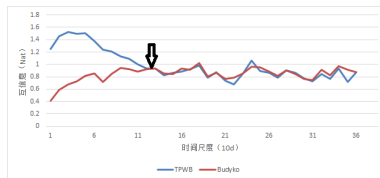
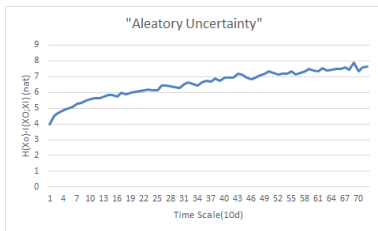


Figure:  $I(X_o; X_s)$ -T of TPWB and Budyko

# 结果



# 结果

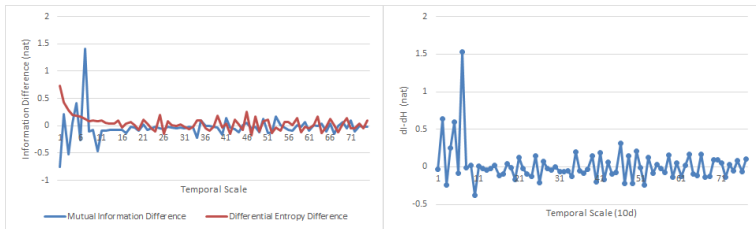
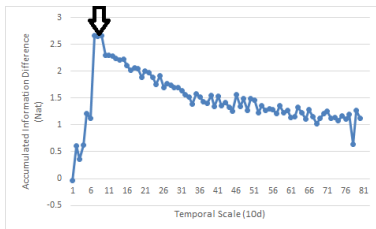


Figure: Relative Aleatory Uncertainty



There is a best performance temporal scale when clustering hydrological data into different scales!