The hunt for deterministic structures in noisy hydrological data

Holger LANGE

Forest and Landscape Institute, Ås, Norway

skog+landskap

NORWEGIAN FOREST AND LANDSCAPE INSTITUTE

- Properties of river runoff time series
- Properties of *k* noise
- Microintroduction to order statistics
- Definition of entropy and complexity for time series
- The Complexity-Entropy Causality Plane (CECP)
- Results: H_S and MPR-C_{JS}
- Conclusions and Outlook

Properties of river runoff time series

- Available from many stations at daily resolution (from HCDN, GRDC, and other sources)
- Are usually non-stationary
- Are noisy, highly fluctuating (intermittent, volatile, ...)
- Contain trends and cycles, in particular the seasonal cycle
- Exhibit **long-range correlations or persistence**
  - expressed by the Hurst exponent , $\frac{1}{2} < H < 1$
  - Autocorrelation decays slowly as a power law, correlation length diverges
  - long («infinite») memory
- ...are complicated!
- Are they also complex? How complex?
- How much information do they contain?
- Is it possible to model them (statistically)?

Properties of k noise

skog+ landskap
NORWEGIAN FOREST AND LANDSCAPE INSTITUTE

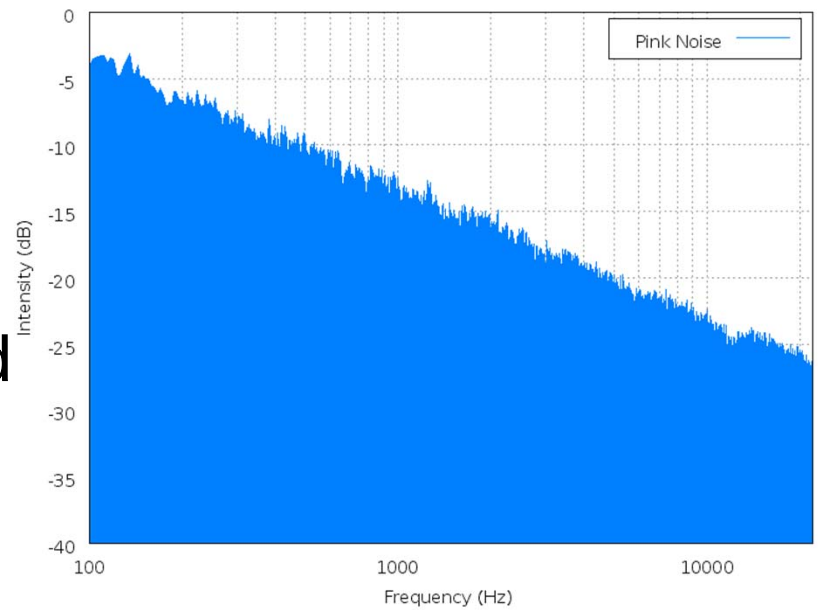
Generated from Gaussian noise by tunable distortion of the power spectrum: $P(f) \sim f^{-k} \; (f \to 0), k \geq 0$

Autocorrelation:

$$C(\tau) \sim \tau^{k-3} \; (\tau \gg 1, k < 3)$$

Hurst exponent:

$$H = \frac{k-1}{2} \; (1 < k < 3)$$



- Power law, long-range correlated
- Diverging correlation length
- Infinite memory

Is this a good description of runoff (deseasonalized time series)?

Approach: quantify information and complexity using order statistics

skog+landskap
NORWEGIAN FOREST AND LANDSCAPE INSTITUTE

Order Statistics

Given a sequence / time series $\{x_i\}, i = 1, \ldots, N$

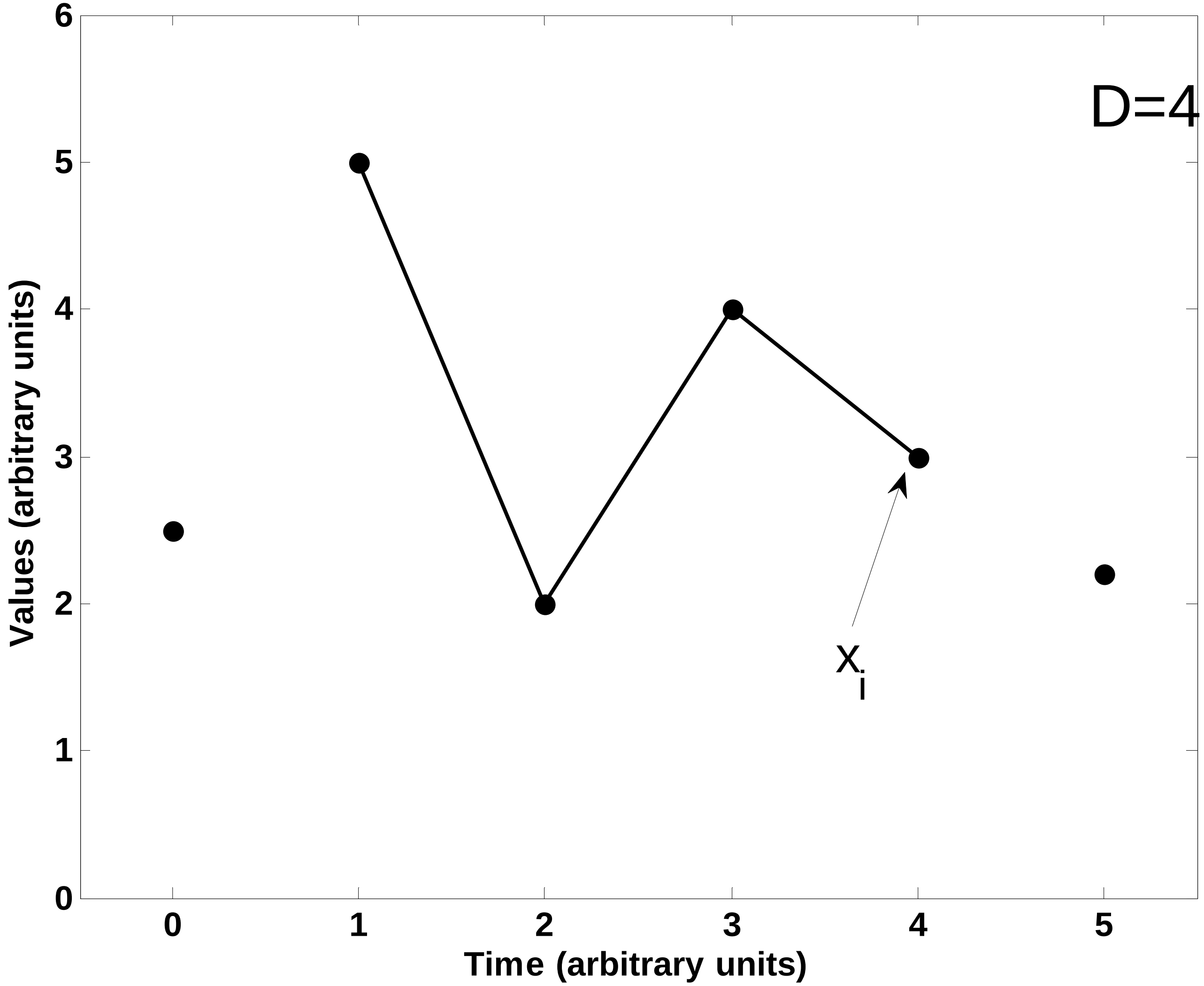
Define a subsequence („word") length $D << N$

The order pattern of length D **of the series at time** i
is the permutation $\pi_D(i) = (r_0, r_1, \ldots, r_{D-1})$ **of** $(0, 1, \ldots, D-1)$ **satisfying**

$$x_{i-r_0} \geq x_{i-r_1} \geq \ldots \geq x_{i-r_{D-1}} \geq x_{i-r_{D-1}}$$

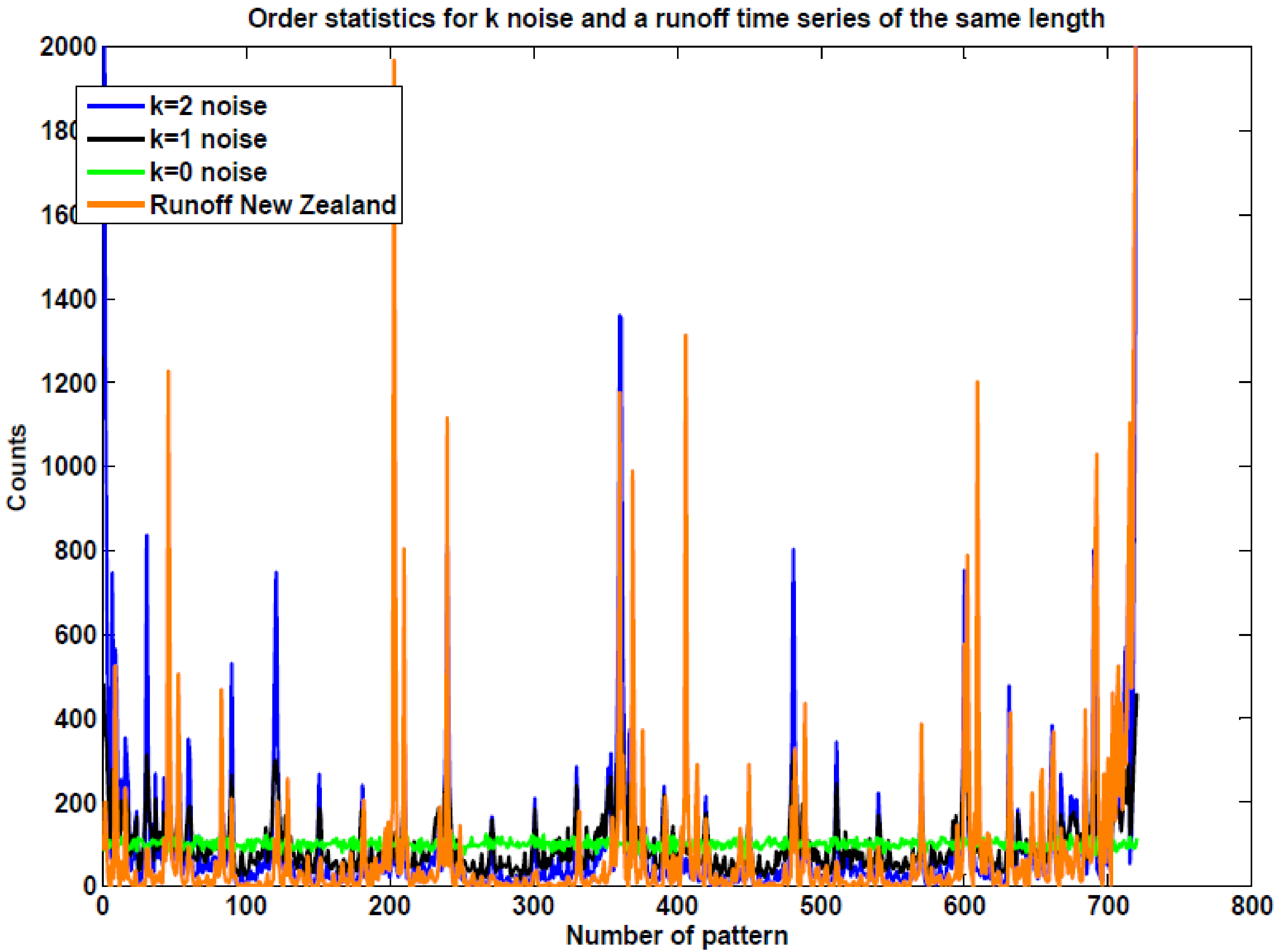
A clever coding of the order patterns (Keller 2008) reveals the order pattern frequency distribution

Order Statistics (continued)



Sequence of magnitudes:
1 – 4 – 2 -3

Order pattern: 3102
Keller number: 9



Relative frequency of pattern i: $p_i, i = 1, \ldots, D!$

(for reliable estimates, $N > 5D!$ recommended)

Information and Complexity of Time Series

skog+ landskap
NORWEGIAN FOREST AND LANDSCAPE INSTITUTE

1. **Information**

(first order in randomness)

Zero for constants, max for pure noise

Here: *Time-ordered* **Shannon entropy**

2. **Complexity**

(second order in randomness)

Zero for constants, zero for pure noise

Max for structured data

Here: **Jensen-Shannon MPR Complexity**:

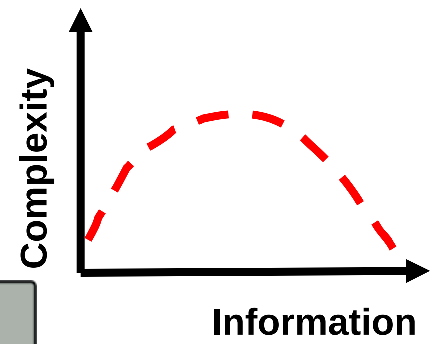
$$H_S = -\sum_{i=1}^{D!} p_i \ln p_i / \ln D!$$

$$C_{JS} = H_S Q_{JS}\left[P, P_u\right]$$

$$Q_{JS} = Q_0 \left\{ S\left[\left(P + P_u\right)/2\right] - S\left[P\right]/2 - S\left[P_u\right]/2 \right\}$$

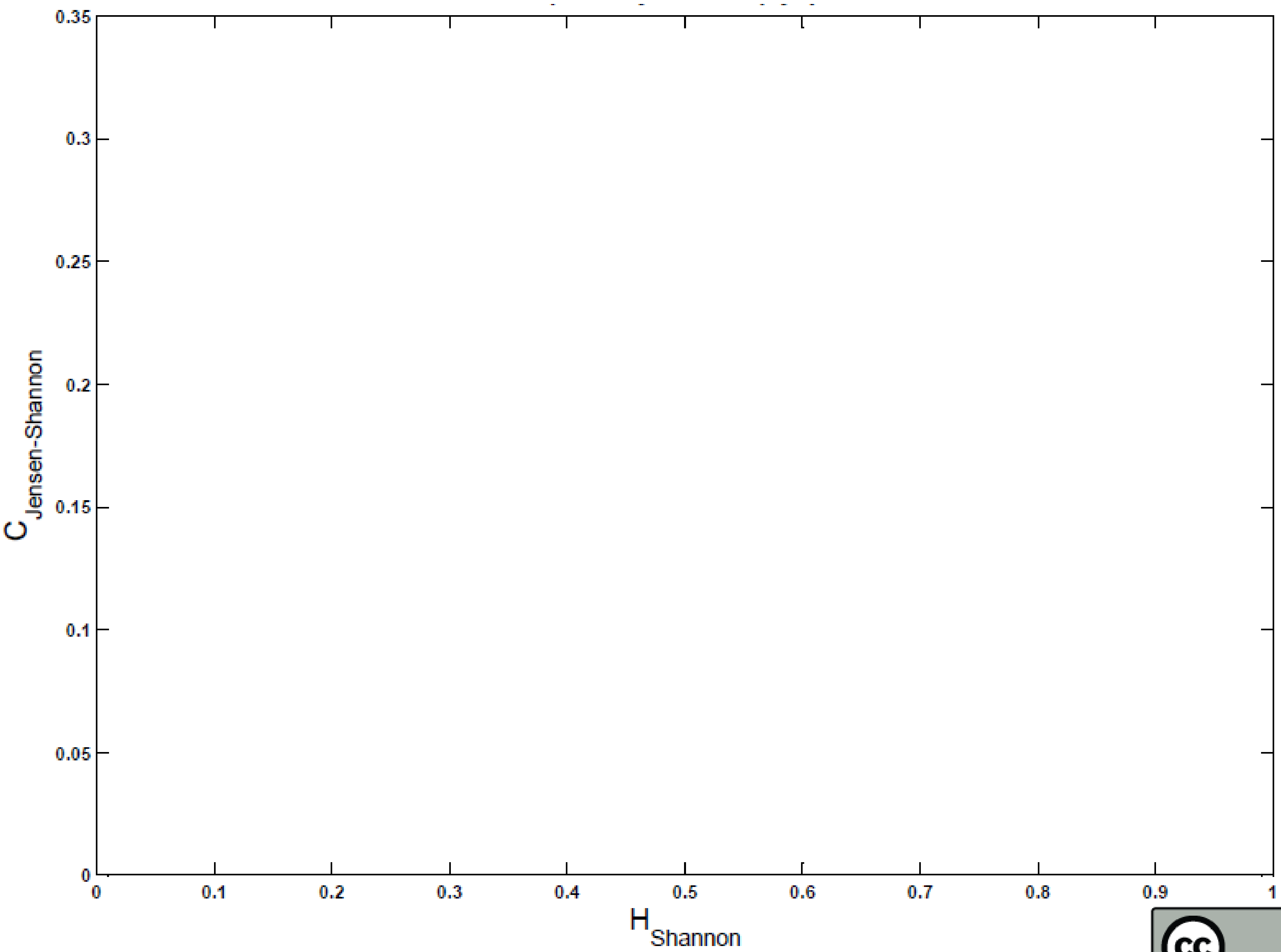
P_u is the uniform distribution and

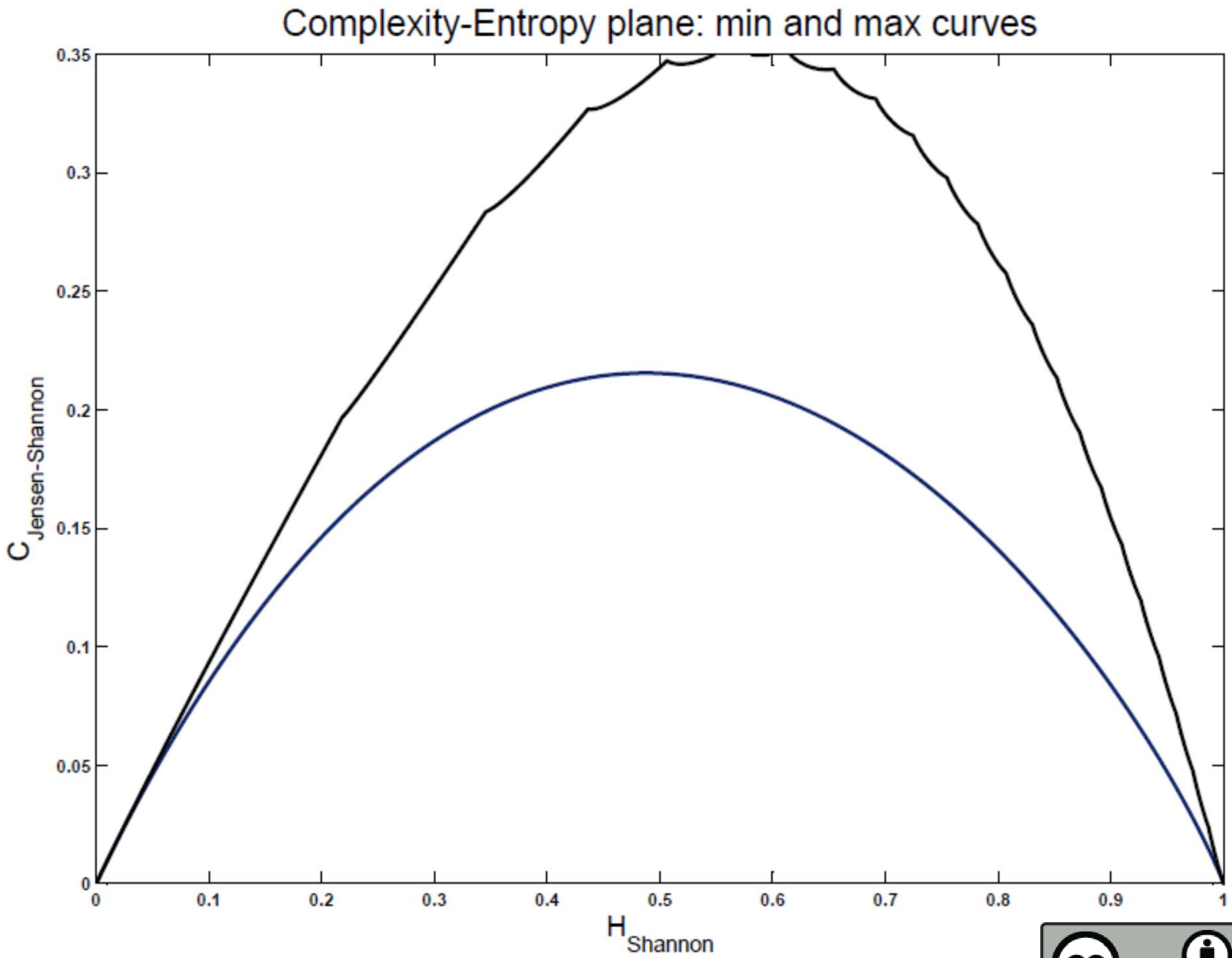
$$Q_0 = -2 / \left\{ \left(\frac{D!+1}{D!} \right) \ln\left(D!+1\right) - 2\ln\left(2D!\right) + \ln D! \right\}$$

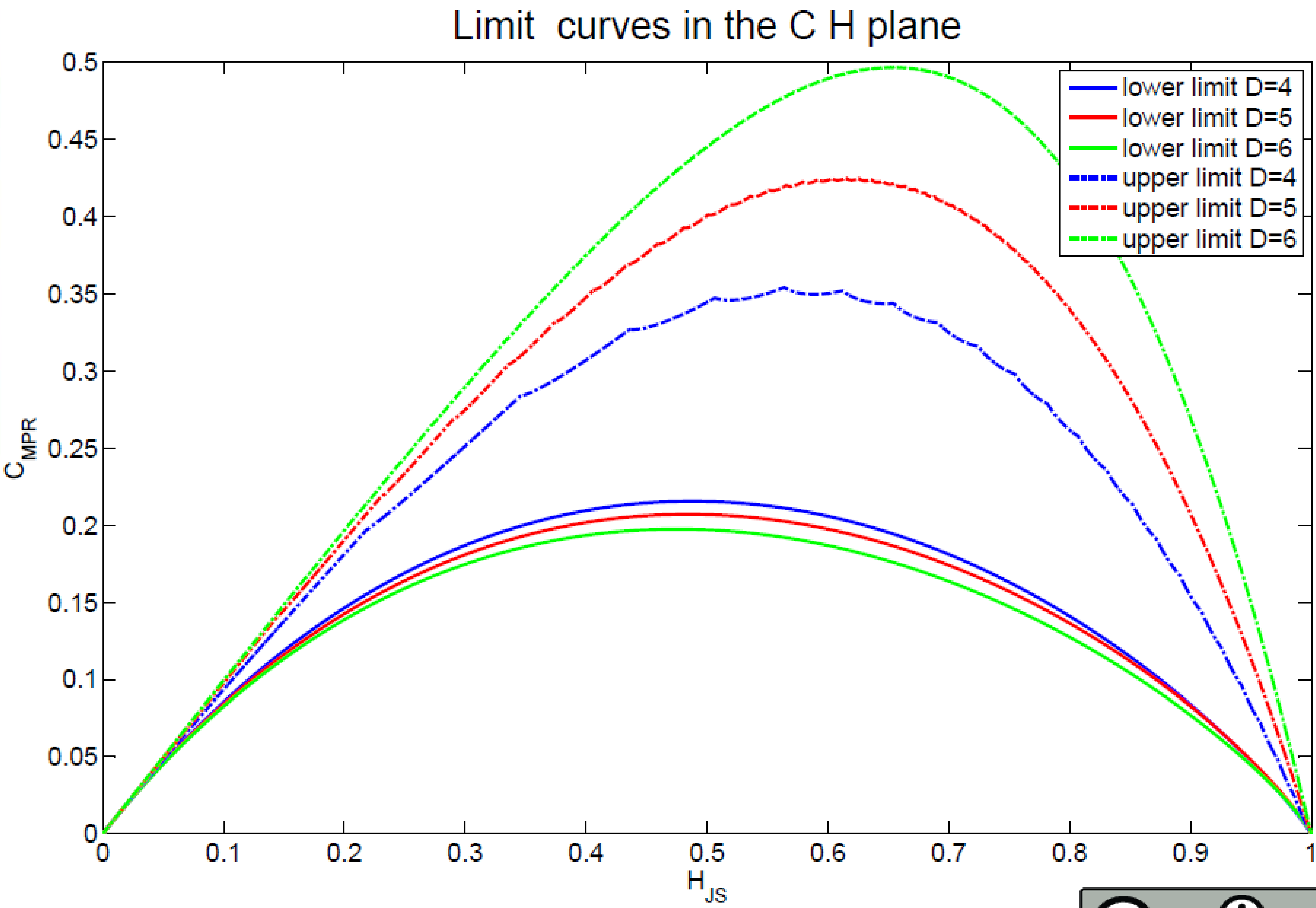


Martín, **P**lastino and **R**osso (2006)

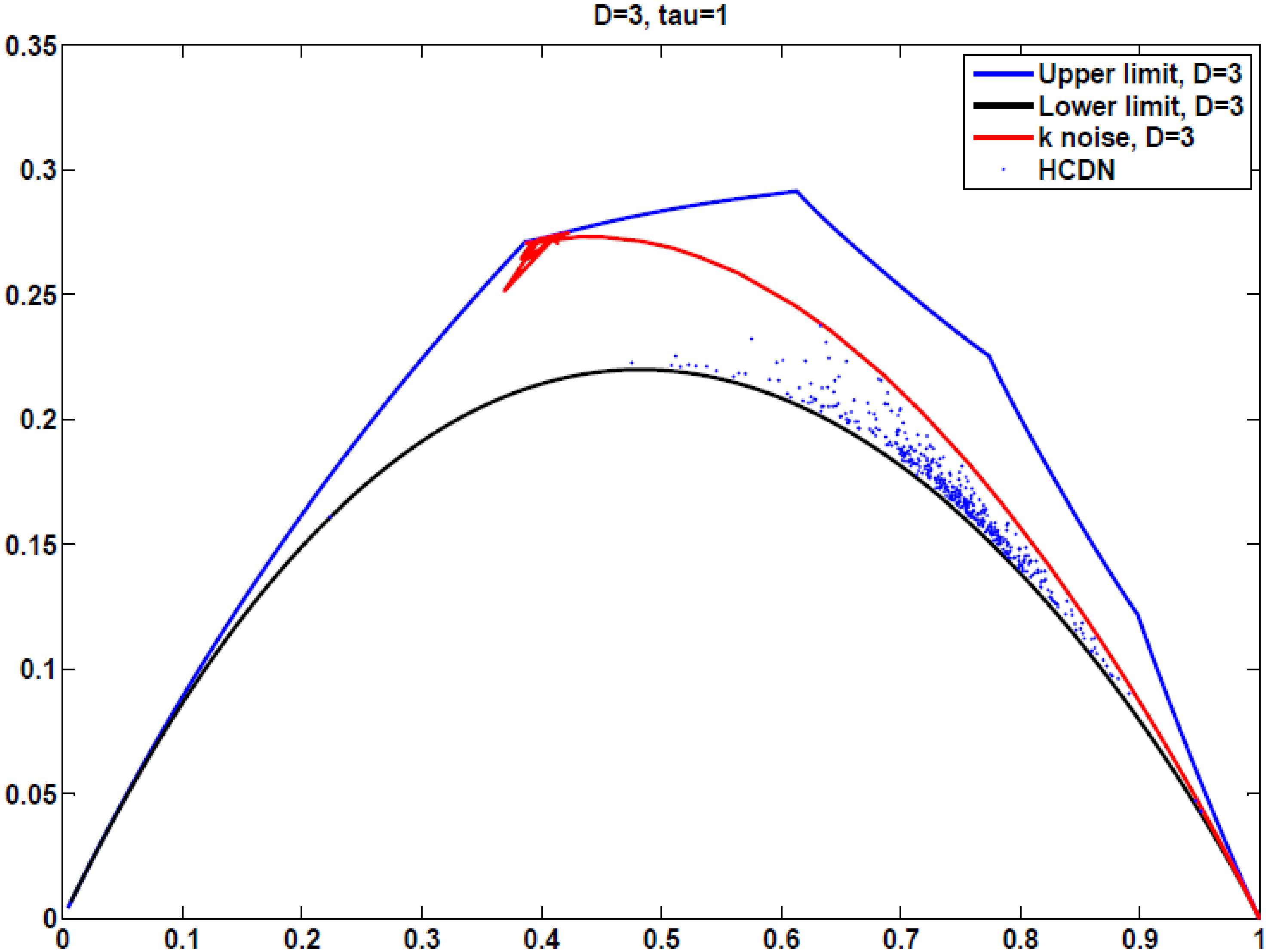
The Complexity-Entropy Causality Plane (CECP)



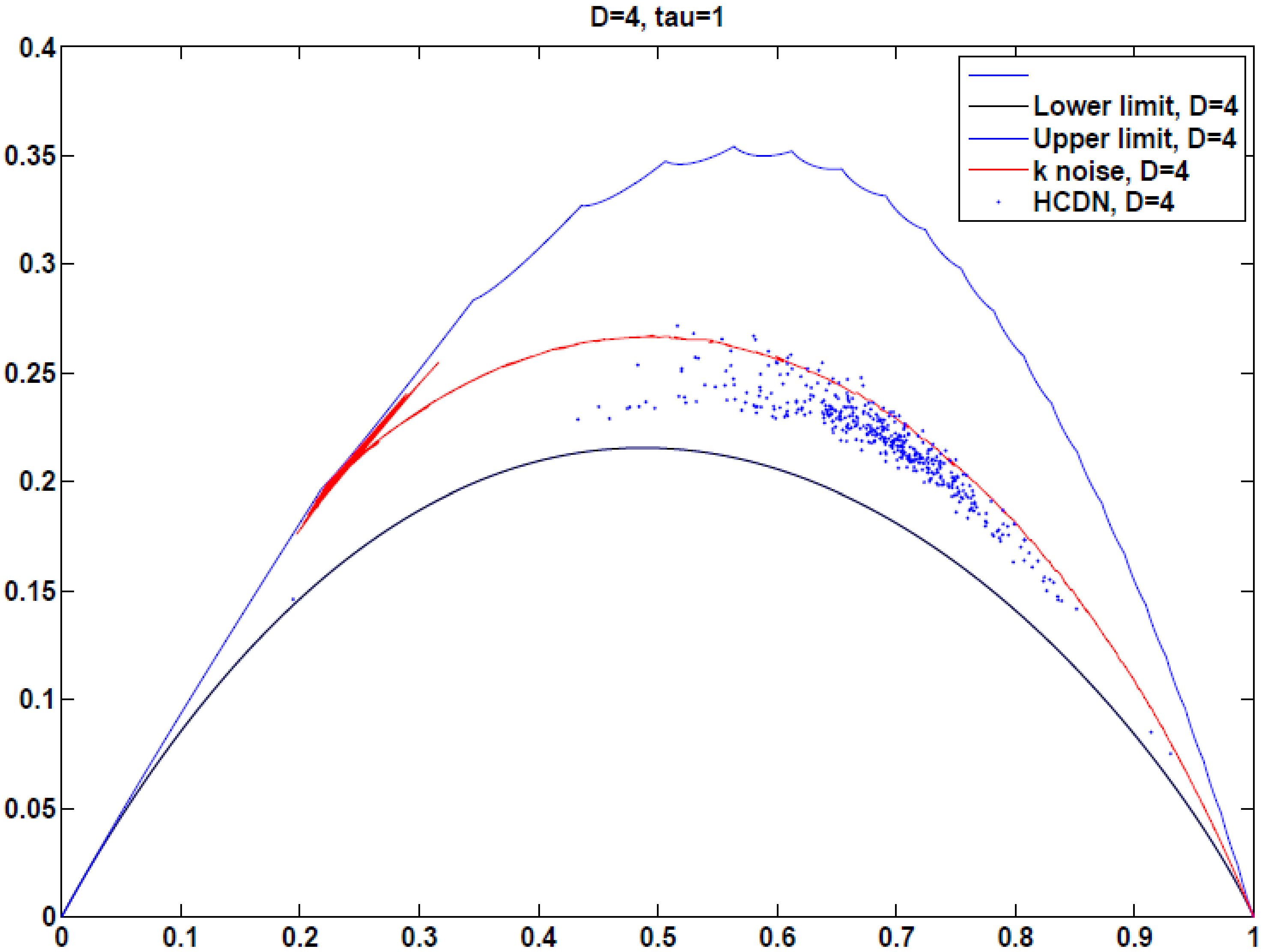

Complexity-Entropy plane: min and max curves
0.35
0.3
0.25
0.2
0.15
0.1
0.05
0
$C_{Jensen-Shannon}$
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1
$H_{Shannon}$


Limit curves in the C H plane
lower limit D=4
lower limit D=5
lower limit D=6
upper limit D=4
upper limit D=5
upper limit D=6
C_{MPR}
H_{JS}
0
0.05
0.1
0.15
0.2
0.25
0.3
0.35
0.4
0.45
0.5
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1

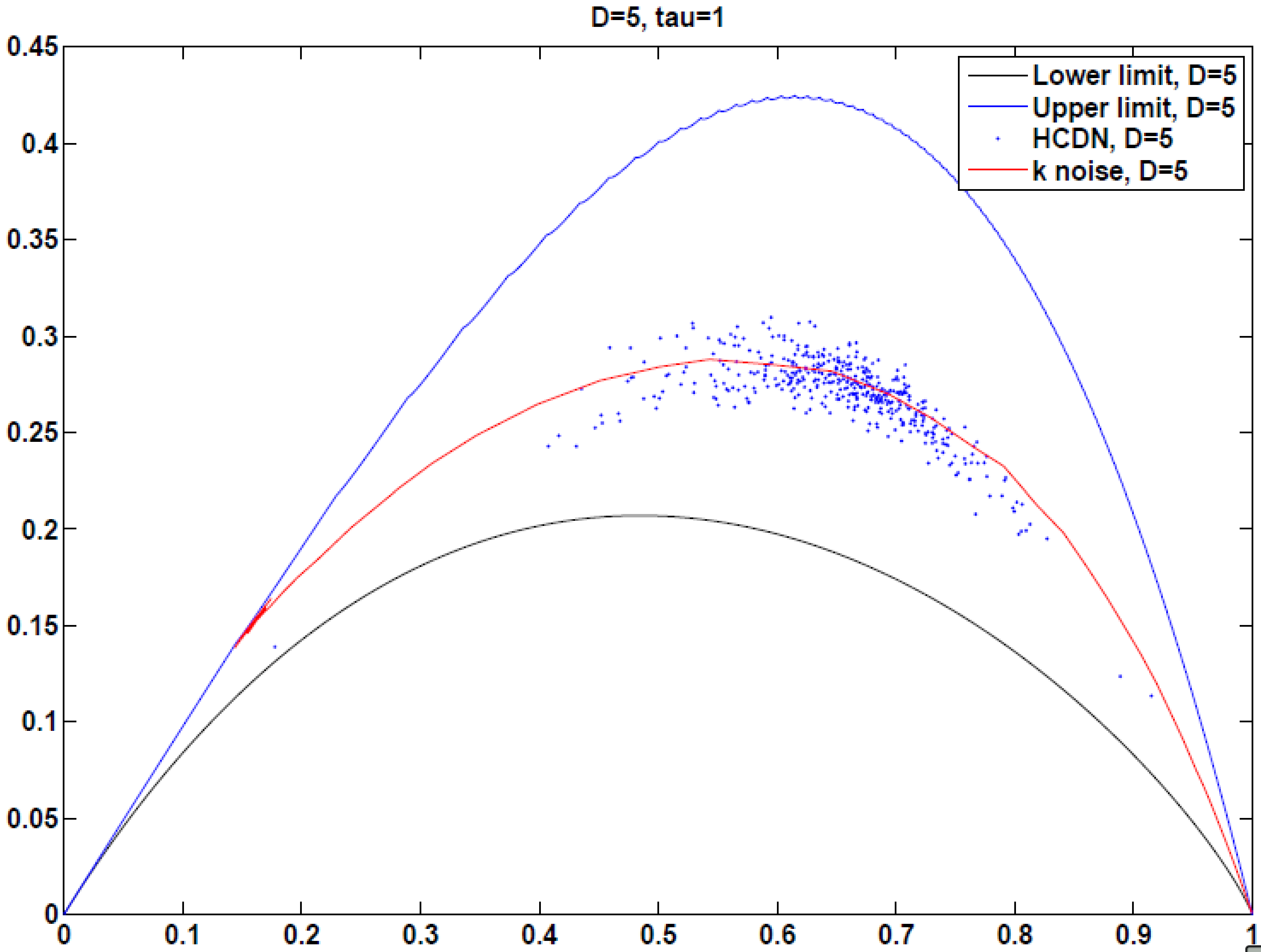
Runoff and *k* noise, D=3

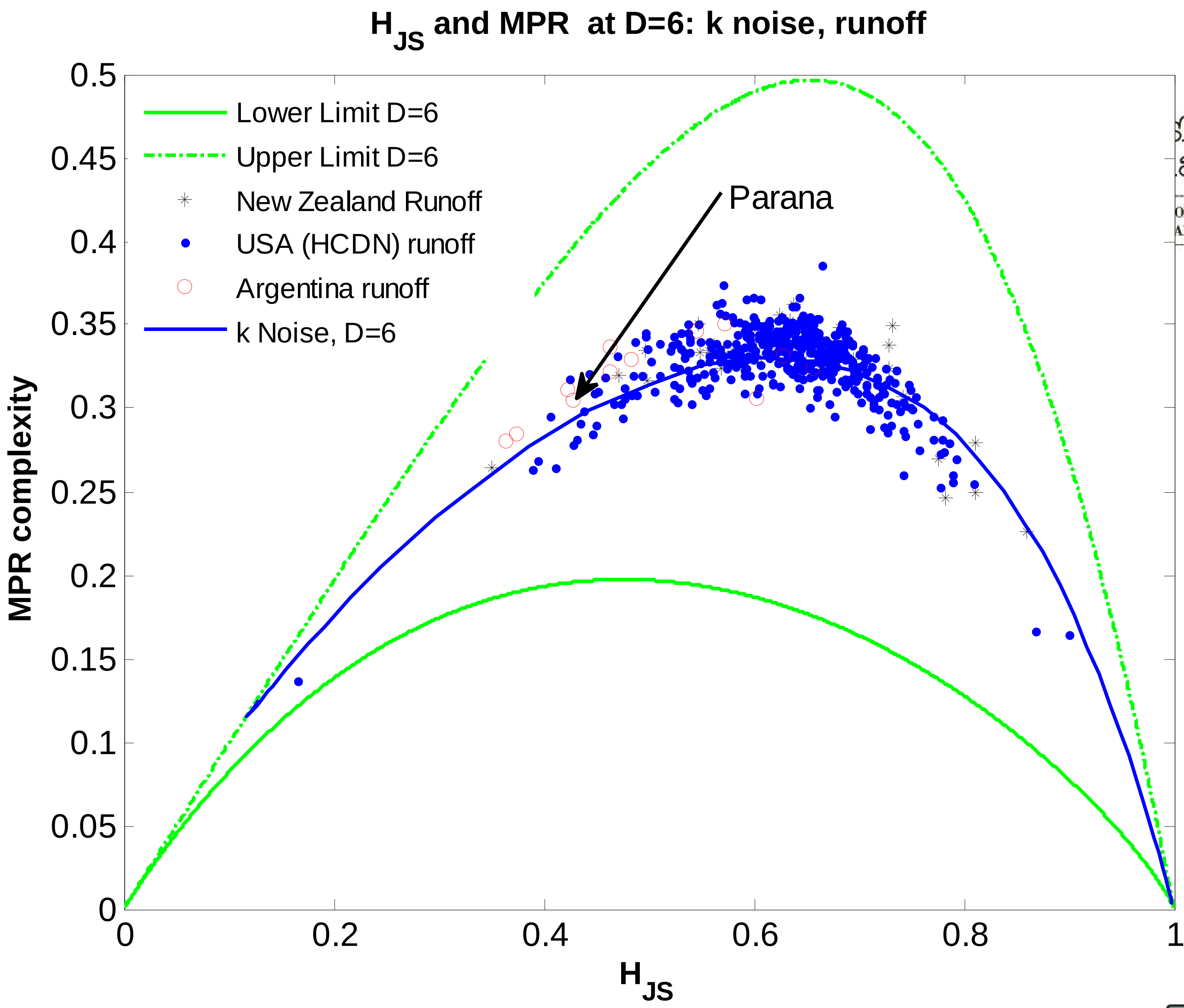


Runoff and *k* noise, D=4



Runoff and *k* noise, D=5

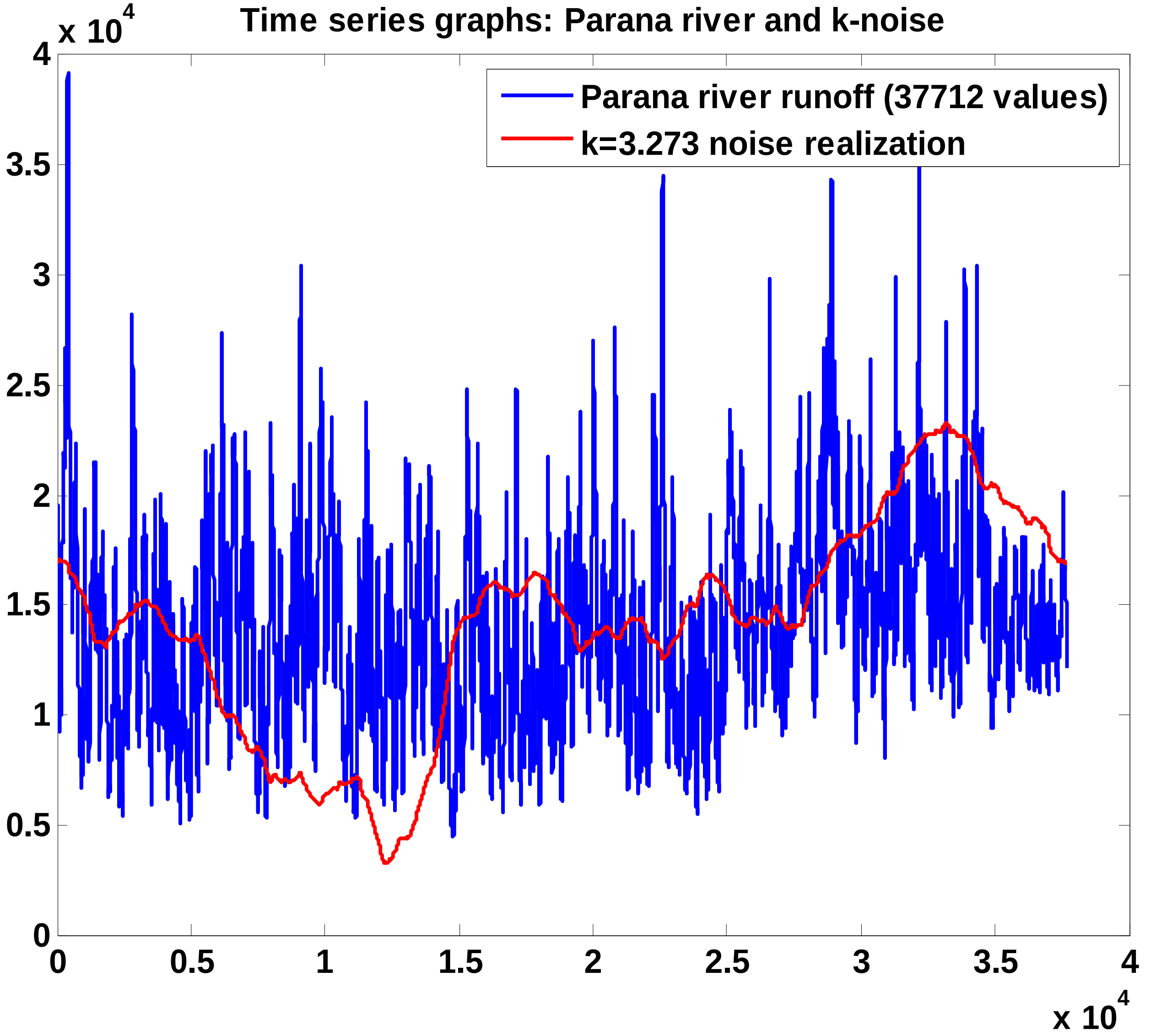




skog+landskap

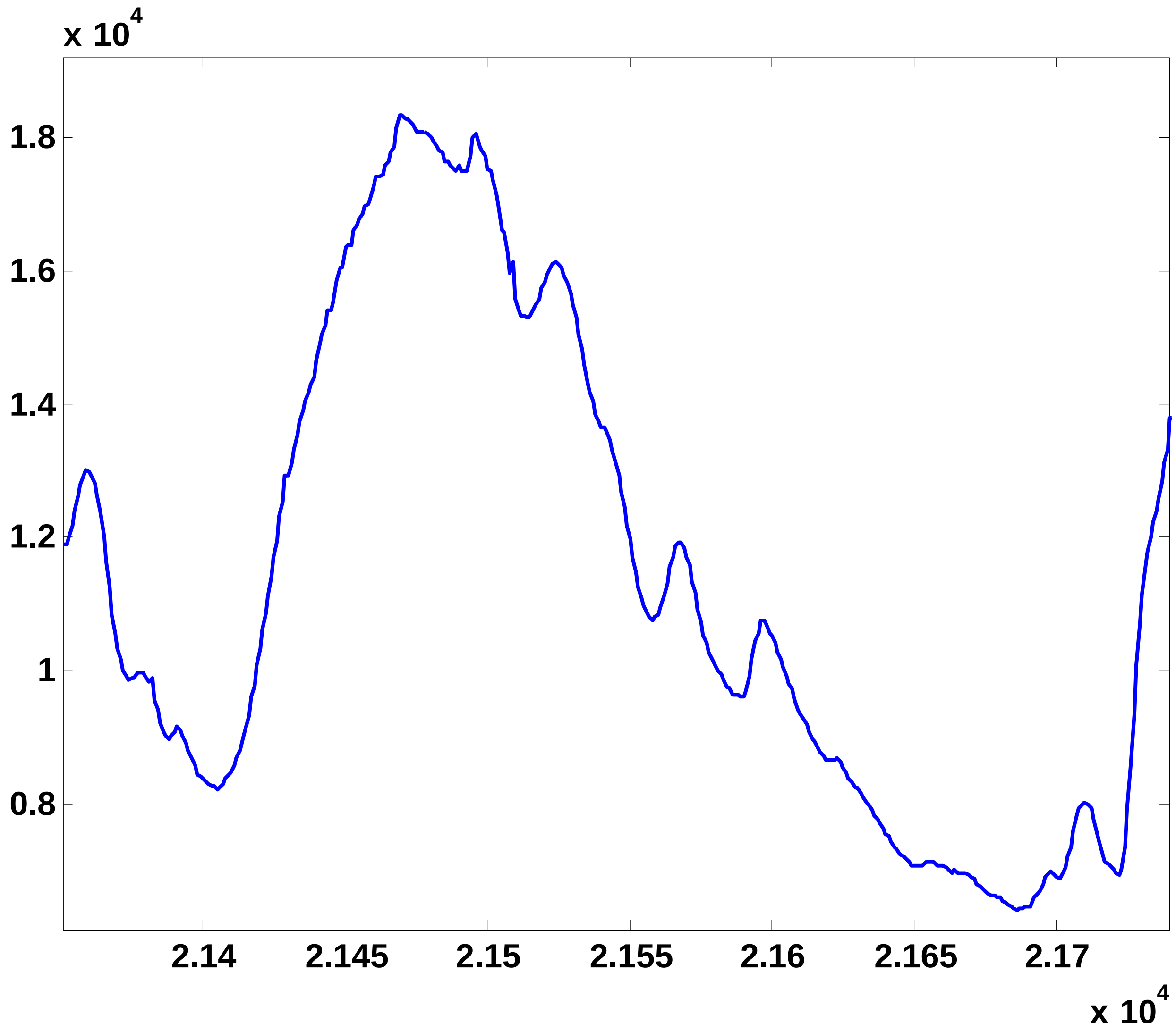
NORWEGIAN FOREST AND LANDSCAPE INSTITUTE

Runoff and tuned k noise (example)



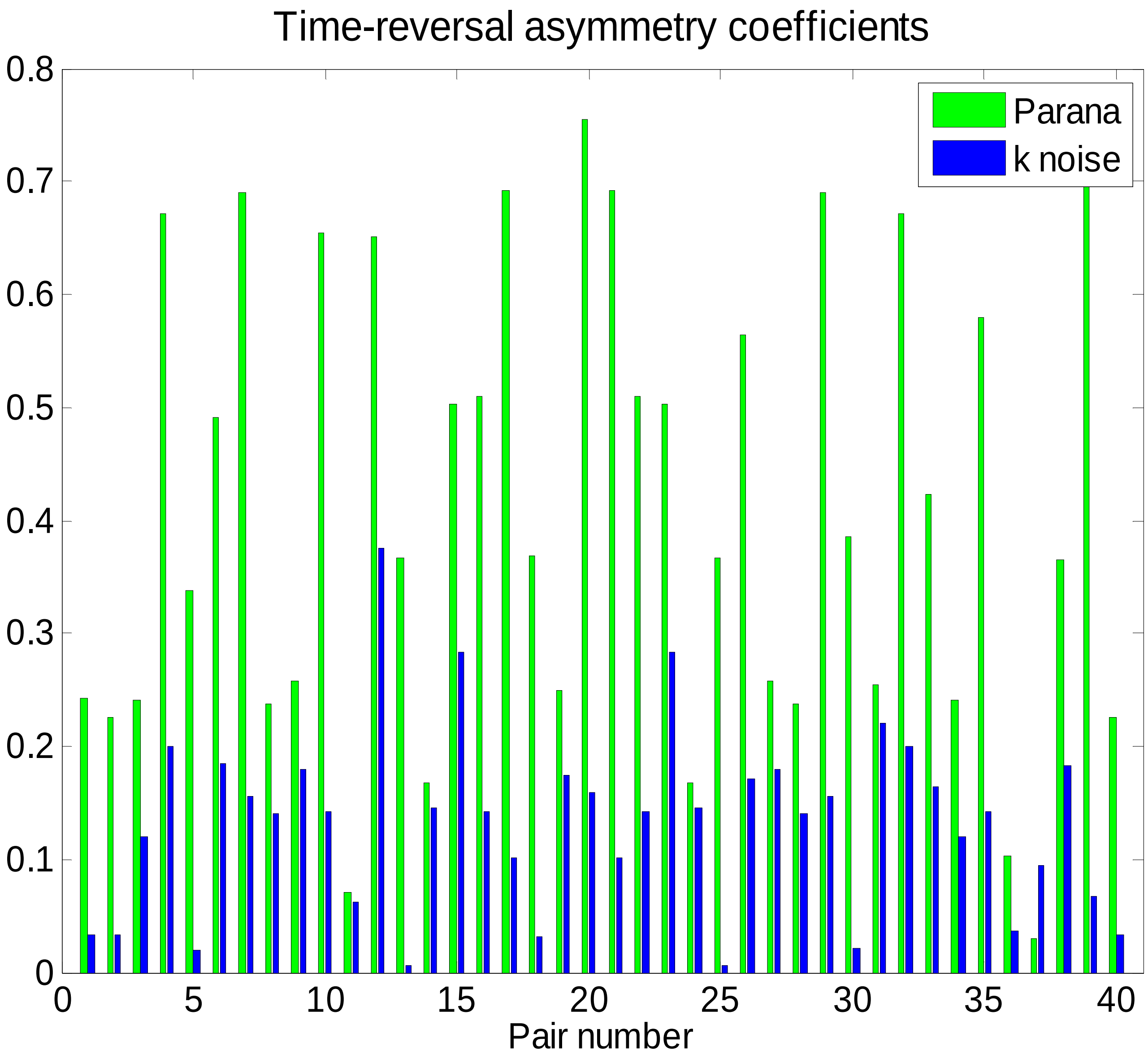
skog+ landskap

NORWEGIAN FOREST AND LANDSCAPE INSTITUTE



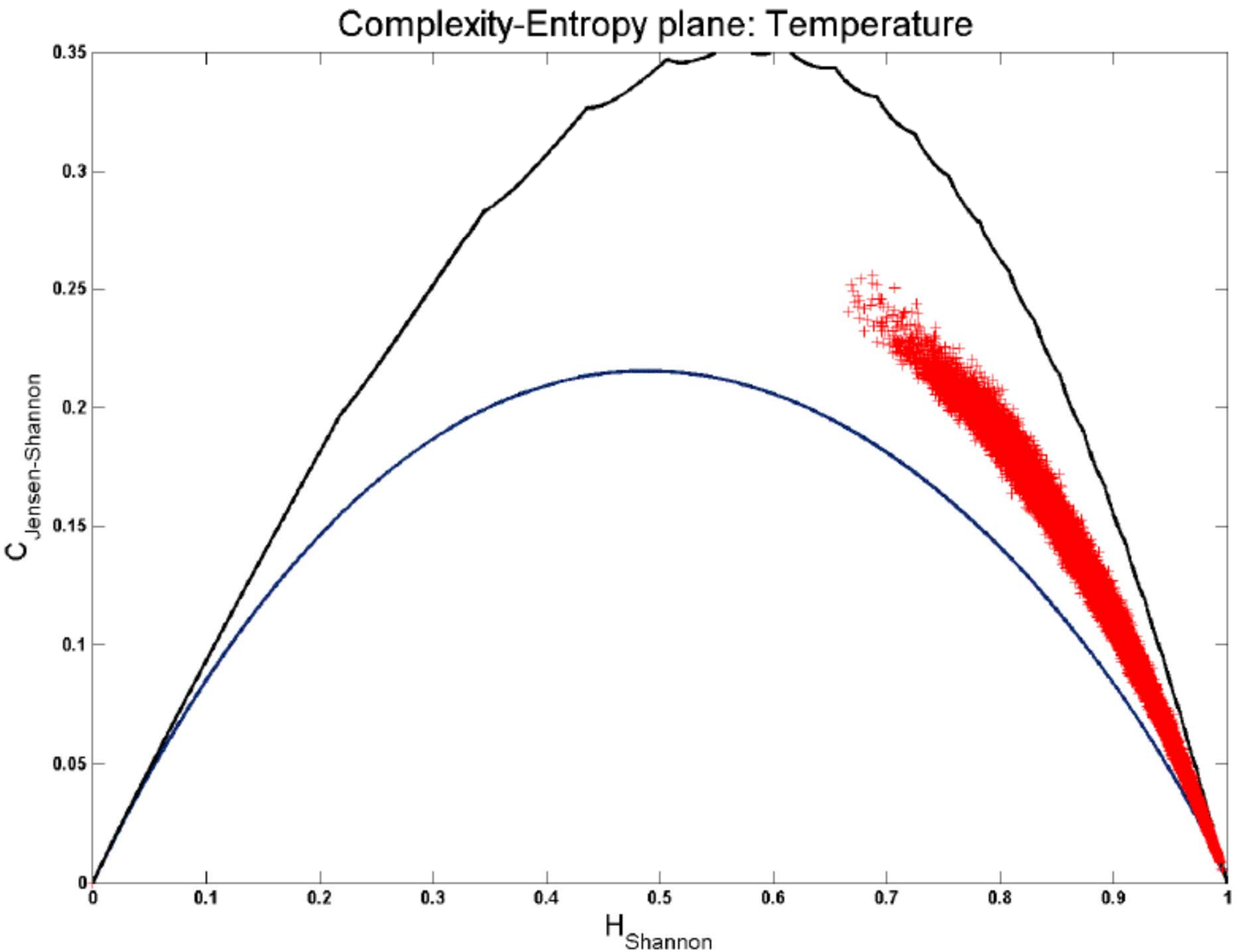
Runoff is time-reversal asymmetric,
k noise is not

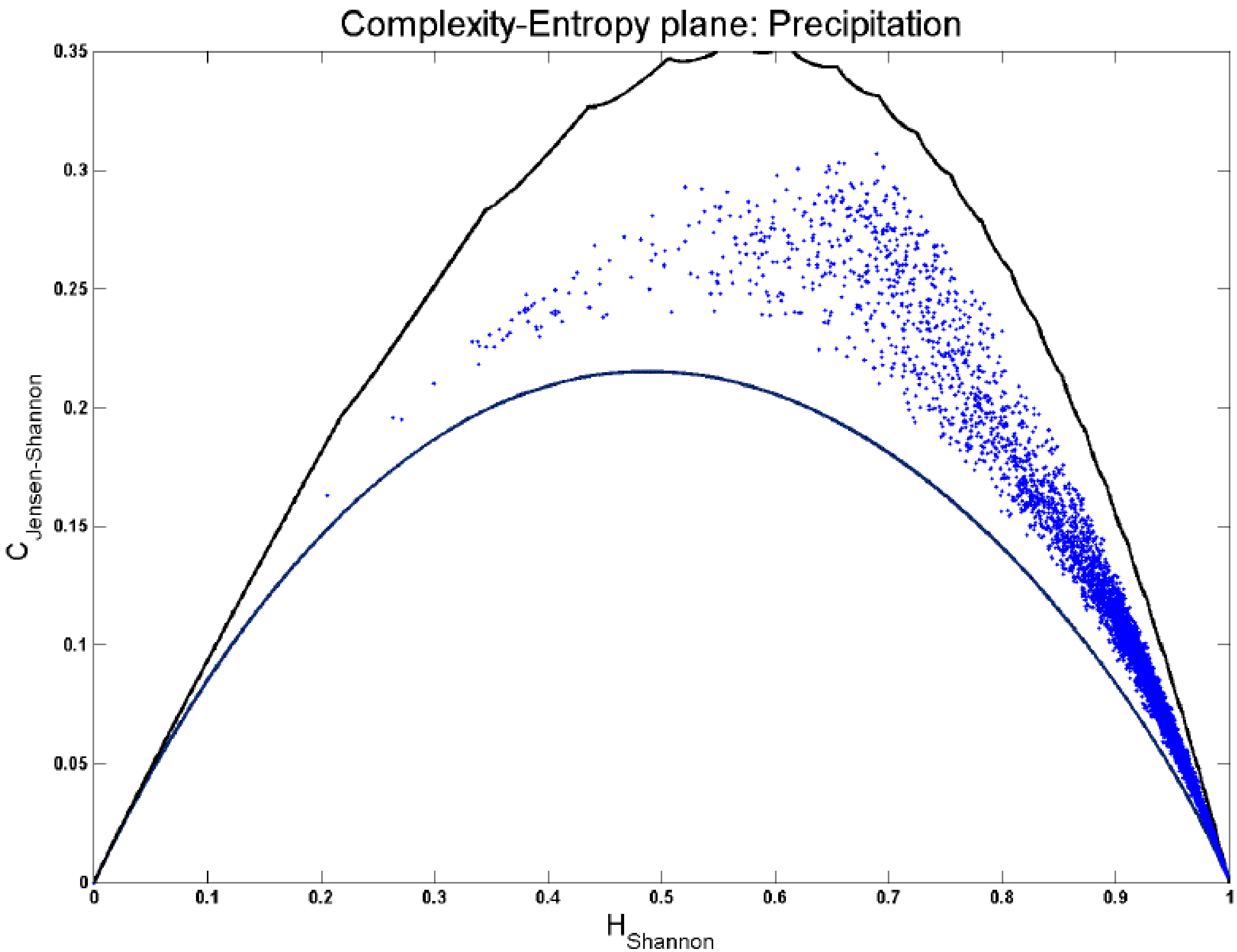
Time asymmetry of runoff and tuned k noise - quantified



skog+ landskap

NORWEGIAN FOREST AND LANDSCAPE INSTITUTE


Complexity-Entropy plane: Temperature
0.35
0.3
0.25
0.2
0.15
0.1
0.05
0
C Jensen-Shannon
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1
H Shannon



Complexity-Entropy plane: Precipitation
$C_{Jensen-Shannon}$
$H_{Shannon}$
0
0.05
0.1
0.15
0.2
0.25
0.3
0.35
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1

Conclusions and Outlook

- Runoff time series *are* complex (in the sense of the CECP)
  - The higher *D*, the more complex
- Runoff is *not* k noise
  - Time asymmetry is one part of that difference
- Correlation structure is intricate even at *small* lags
- Information and Complexity of runoff is qualitatively different from that of other variables
- Future plans:
  - Relate CECP position to auxiliary variables
  - Use CECP as sensible model evaluation tool
  - Construct a time series generator producing series with a prescribed H_S and C_{JS} – how do these compare to the observed runoff?

skog+ landskap

NORWEGIAN FOREST AND LANDSCAPE INSTITUTE