European Geosciences Union General Assembly 2006
Vienna, Austria, 2 - 7 April 2006
Session HS5: Low flow and drought – estimation, propagation in the hydrologic cycle and forecasting

Multiyear behaviour and monthly simulation and forecasting of the Nile River flow

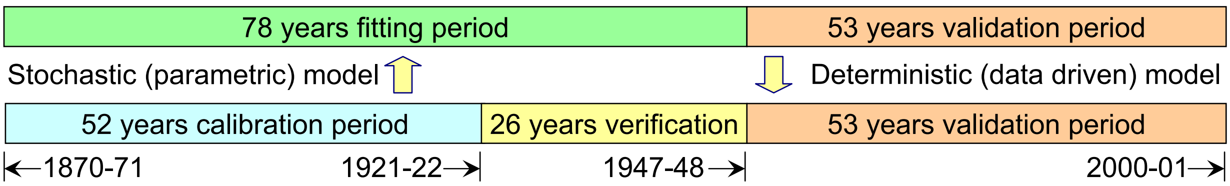
D. Koutsoyiannis
Department of Water Resources, National Technical University of Athens
Huaming Yao, and Aris Georgakakos
Georgia Water Resources Institute, School of Civil and Environmental Engineering, Georgia Institute of Technology

1. Abstract

Multiyear persistence of droughts is a typical natural behaviour that cannot be modelled by typical stochastic or deterministic approaches. As this persistence is closely related to the Hurst (or scaling) behaviour, a stochastic approach to represent multiyear persistence of droughts should also reproduce the Hurst phenomenon. An advanced, yet simple, stochastic methodology, is proposed based on the concept of maximum entropy that is able to represent multiyear persistence. The approach can be used to generate long-term simulations or shorter-term forecasts, and is demonstrated for the Nile River, the persistence behaviour of which motivated the discovery of the Hurst phenomenon. The analysis and demonstrations use the Nile flow record, the longest available flow record worldwide. The stochastic methodology is also compared with an analogue (local nonlinear chaotic) model and a connectionist (artificial neural network) model developed using the same flow record.

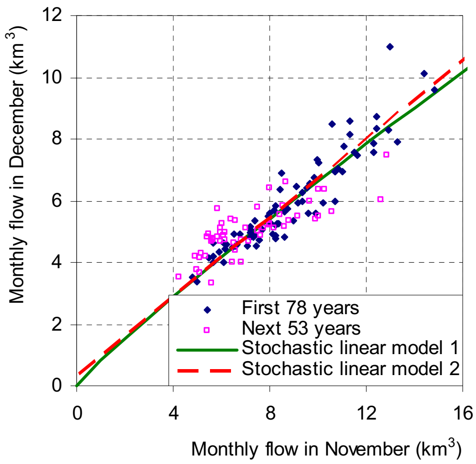
2. Background and data

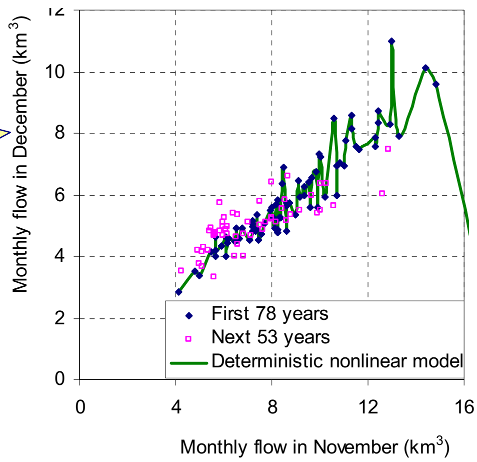
- Nile is the longest river of the world (6521 km)
- Due to large length, the travel time is of the order of a month
- This induces strong dependence on the monthly scale and makes monthly forecast possible
- The modern flow record at Aswan is one of the longest worldwide (131 years) and makes analysis and modelling more reliable
- In addition, there exist older instrumental records of annual maximum and minimum water level at the Roda Nilometer for more than 800 years
- All flow records as well as additional historical and archaeological data (Said, 1993) affirm the long-range dependence of the Nile flows and raise the question whether or not this dependence should be incorporated in the monthly forecast model
- Another important question is whether stochastic or deterministic models have better forecast skills; this is studied by comparing the performance of a stochastic model and two deterministic models (analogous, connectionist) on a 53-year validation period whose data were not used into model fitting



3. Modelling approaches and underlying concepts

- According to the **stochastic approach** the flows are modelled as a stochastic process
- Maximization of Shannon entropy on a multivariate setting results in multivariate normal distribution (Papoulis, 1991)
- This entails linear dependence of lagged flows (stochastic linear model 2)
- Maximization of Tsallis (non-extensive) entropy (Tsallis, 2004) results in linear dependence of nonlinearly transformed flows using a normalizing transformation (stochastic linear model 1, only slightly different from 2; see panel 6)
- According to the **deterministic approach**, a deterministic relationship of lagged flows is assumed as in the hypothetical (caricature) case shown in the figure with a single time delay component, where the hypothetical relationship is a non-intersecting curve passing from all 78 points of the 'fitting' period (but the points of the 'validation' period lie outside the curve)





4. Synopsis of models

	Model type	Model specifications	Model abbreviation
Stochastic	Stochastic	Cyclostationary with short- and long-range dependence, using normalizing transformation of time series	S1
		As S1 but without normalizing transformation	S2
		PAR(2) without normalizing transformation	S3
Deterministic	Analogue (Local linear)	Single scale, 12 consecutive time delay items; 11 neighbors	A1
		Single scale, 13 consecutive time delay items; 24 neighbors	A2
		Two scales; 4 time delay items; 7 neighbors	A3
	Connectionist (Artificial neural network)	Single scale, 5 inputs, 2 layers, 2+2 hidden nodes	C1
		Single scale, 14 inputs, 2 layers, 11+11 hidden nodes	C2
		Two scales (delay times 1, 2, 12, 24), 2 layers 4+2 hidden nodes	C3

5. Statistical characteristics of the Aswan flow record

Month	μ (km^3)	σ (km^3)	C_s	C_k	τ_3	τ_4	H	ρ_{FGN1}	ρ_1	ρ_2	ρ_{12}
Aug	19.37	4.62	-0.09	-0.14	0.00	0.12	0.76	0.43	0.71	0.26	0.16
Sep	22.98	4.29	-0.12	-0.57	-0.02	0.07	0.74	0.39	0.80	0.51	0.17
Oct	16.33	3.65	0.41	0.31	0.08	0.14	0.76	0.44	0.88	0.70	0.24
Nov	8.79	2.34	0.42	-0.27	0.09	0.11	0.80	0.51	0.90	0.77	0.26
Dec	5.92	1.60	0.86	0.60	0.19	0.13	0.89	0.72	0.94	0.85	0.42
Jan	4.37	1.20	0.64	0.31	0.15	0.15	0.88	0.70	0.98	0.91	0.44
Feb	3.02	1.00	0.85	0.27	0.20	0.12	0.82	0.55	0.96	0.92	0.35
Mar	2.51	0.96	1.25	1.34	0.26	0.17	0.78	0.48	0.91	0.84	0.31
Apr	1.89	0.75	1.75	3.56	0.33	0.19	0.78	0.47	0.94	0.78	0.33
May	1.68	0.63	2.13	6.30	0.33	0.23	0.72	0.36	0.93	0.85	0.30
Jun	1.91	0.68	1.89	6.00	0.27	0.20	0.63	0.20	0.70	0.59	0.11
Jul	5.06	1.84	0.75	0.24	0.16	0.12	0.89	0.71	0.65	0.44	0.47
Average			0.90	1.50	0.17	0.14	0.79	0.50	0.86	0.70	0.30
Annual	93.85	20.16	0.35	-0.08	0.09	0.09	0.85	0.63	0.35	0.35	

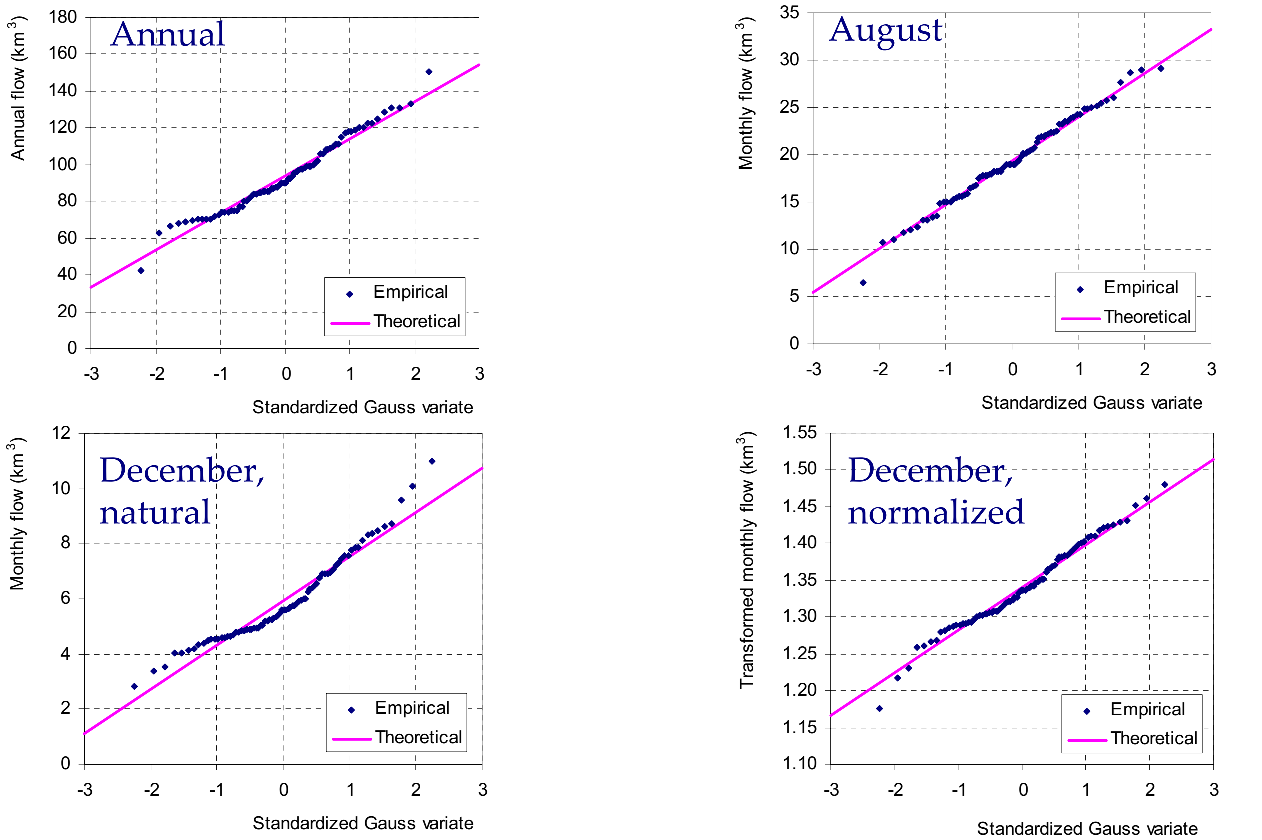
6. Marginal distributional properties

- During August-October, the Blue Nile flows dominate; these seem to be approximately normally distributed
- During November-July, other parts of the basin contribute more than Blue Nile, but with flows much lower than in August-October; these seem to be non-normally distributed with positive skewness and kurtosis
- On the annual scale the dominance of the high flows during August-October results in flows that are approximately normally distributed
- The normal distributions of August-October could be derived postulating Shannon entropy maximization; the non-normal distributions of November-July could be described by postulating Tsallis entropy maximization (Koutsoyiannis, 2005a)
- Non-normal Tsallis distributions (Tsallis et al., 1995) can be described by the normalizing transformation

$$z = g(x) = c + \operatorname{sgn}(x - c)\,\lambda \sqrt{\left(1 + \frac{1}{\kappa}\right) \ln\left[1 + \kappa \left(\frac{x - c}{\lambda}\right)^2\right]}$$

  where x and z are the natural and normalized flows, κ is a tail-determining dimensionless parameter, λ is a scale parameter with same units as x (which enables physical consistency) and c a translation parameter with same units as x; for $\kappa = 0$, z is identical to x
- The fitted parameters are $\kappa = 0$ (normal) for August-October, and $\kappa = 2.76$, $c = 0$ and $\lambda = 0.47$ km^3 for November-July

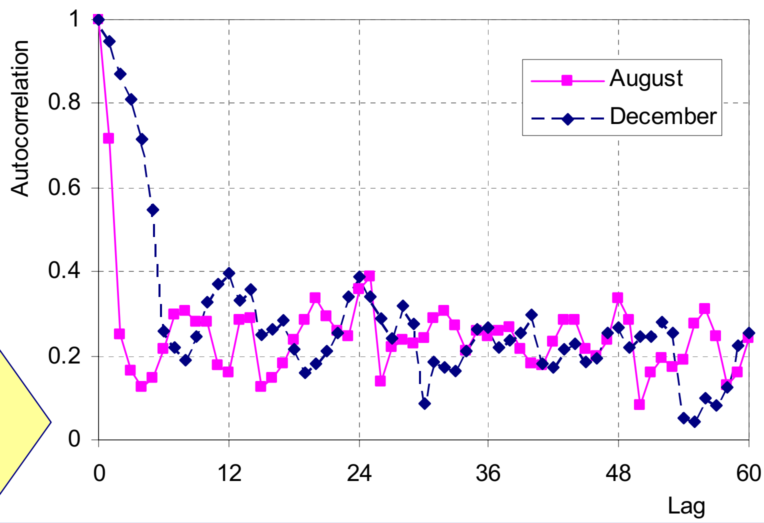
7. Normal probability plots: annual and monthly flows



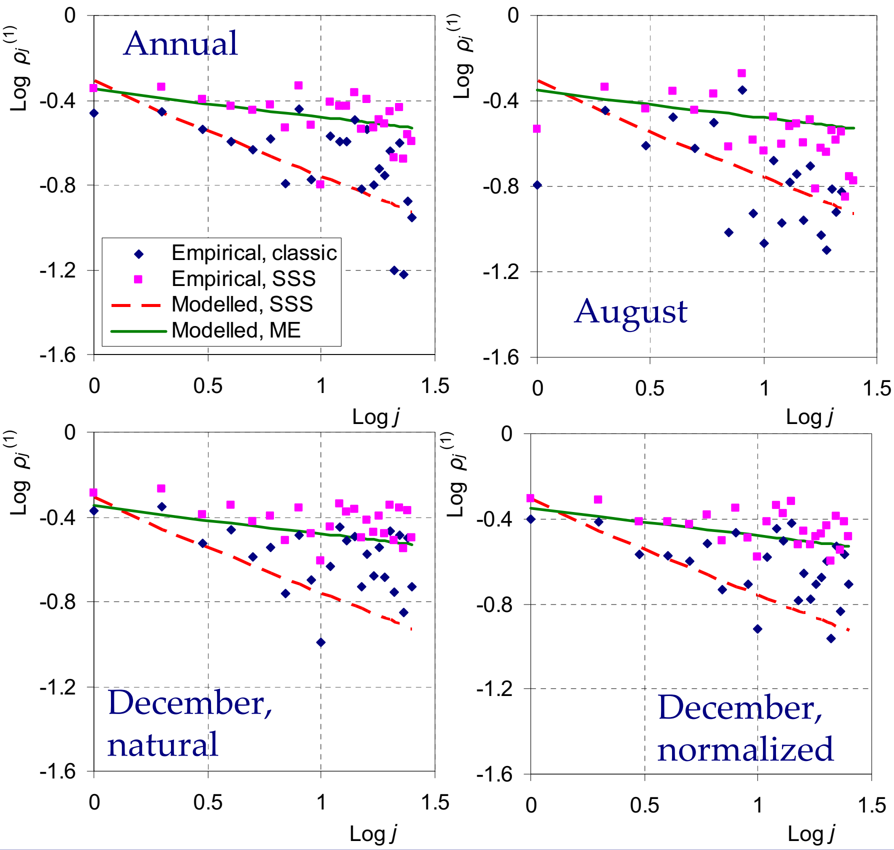
8. Dependence properties

- Monthly autocorrelations differ significantly from month to month for small lags (periodicity) but become very similar for large lags
- Clearly, the monthly autocorrelation function for large lags suggests long-range dependence (see also panel 9)
- At the annual scale as well as at the monthly scale with lags that are multiples of 12, the autocorrelation functions suggest a nearly power-law (Hurst) decay but not a simple scaling stochastic process (SSS or fractional gaussian noise)
- Entropy maximization at a multiple time scale setting as in Koutsoyiannis (2005b) but with two autocorrelation constraints (annual scale, lags 1 and 4) results in asymptotic scaling with almost power-law autocorrelation decay as in panel 9

Correlation coefficients of the standardized monthly flows, i.e. Corr(X_i, X_{i+j}) for $i = 1$ (August) and 5 (December)



9. Long range dependence: detection and modelling



Explanation

Autocorrelations of **annual** flows: Corr(Y_i, Y_{i+j}) vs. j

Autocorrelations of **monthly** flows: Corr(X_i, X_{i+12j})) vs. j for $i = 1$ (August) and 5 (December)

Empirical classic: the classical statistical estimates of autocorrelations

Empirical SSS: Modified for SSS processes estimates of autocorrelations (Koutsoyiannis, 2003)

Modelled SSS: the typical SSS dependence

Modelled, ME: Dependence derived by the principle of maximum entropy applied on multi-scale setting with two autocorrelation constraints (annual scale, lags 1 and 4); it is an asymptotic scaling dependence that tends to simple scaling for multi-year scales; the annual model is also applied to the monthly flows with same parameters

10. Stochastic model formalism

- The prediction W of the monthly flow one month ahead, conditional on a number s of other variables with known values that compose the vector $\mathbf{Z}$, is based on the linear model:

$$W = \mathbf{a}^T \mathbf{Z} + V$$

where $\mathbf{a}$ is a vector of parameters (the superscript T denotes the transpose of a vector or matrix) and V is the prediction error, assumed independent of $\mathbf{Z}$; for simplicity, Z is assumed standardized with zero mean and unit variance

- After several trials, an optimal composition of $\mathbf{Z}$ was found to be the following
  - All available flow measurements of the same month on previous years; for simplicity the number of these elements is left unchanged, equal to the length of the fitting period (78 variables)
  - The flows of the two previous months of the same year (2 variables)

With this composition of $\mathbf{Z}$, the model takes account of both long-range and short-range dependence

- The model parameters are estimated from (Koutsoyiannis, 2000)

$$\mathbf{a}^T = \boldsymbol{\eta}^T \mathbf{h}^{-1}, \quad \mathrm{Var}[V] = 1 - \boldsymbol{\eta}^T \mathbf{h}^{-1} \boldsymbol{\eta} = 1 - \mathbf{a}^T \boldsymbol{\eta}$$

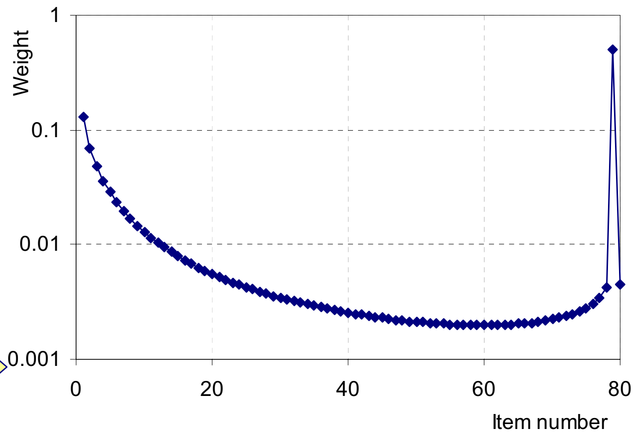
where $\boldsymbol{\eta} := \mathrm{Cov}[W, \mathbf{Z}]$ and $\mathbf{h} := \mathrm{Cov}[\mathbf{Z}, \mathbf{Z}]$

- In forecast mode, $V = 0$ (to obtain the expected value of W conditional on $\mathbf{Z} = \mathbf{z}$); in simulation mode V is generated from the normal distribution independently of $\mathbf{Z}$

11. Parameter estimation

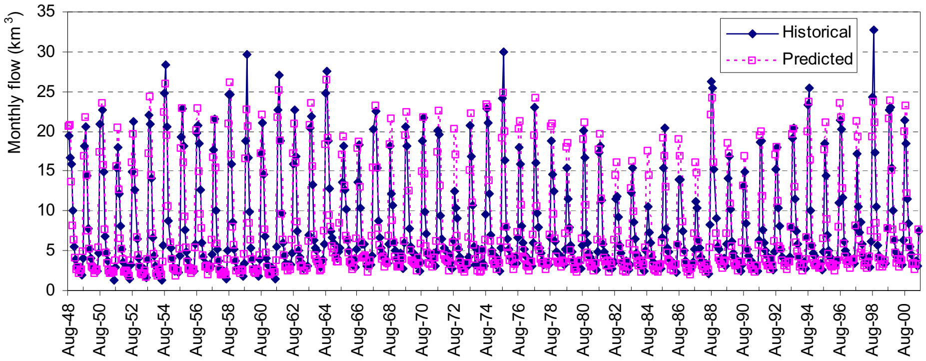
- Both $\mathbf{a}$ and Var[V] are estimated from the vector $\boldsymbol{\eta} := \mathrm{Cov}[W, \mathbf{Z}]$ and the matrix $\mathbf{h} := \mathrm{Cov}[\mathbf{Z}, \mathbf{Z}]$ that contain numerous items (in our case $80 + 80 \times 80 = 6480$ for each month; such a number of parameters cannot be estimated from 78 monthly data values)
- However, most covariances in $\boldsymbol{\eta}$ and $\mathbf{h}$ depend on:
  - 2-3 parameters (same for all months) expressing the long-range dependence, as estimated by application on the ME principle on a multi-time scale setting (a stationary component)
  - 2 parameters (per month) expressing the monthly autocovariances at the monthly scale (a cyclostationary component)
- All other covariances that cannot be derived from these parameters are left 'unestimated' (in terms of statistics) and are calculated by the ME principle, applied on a single scale
- The entropy maximization in this case has an easy analytical solution that can be formulated as a generalized Cholesky decomposition (assuming that $\mathbf{h} = \mathbf{b}\,\mathbf{b}^T$)

Graphical depiction of the vector of weights **a** estimated by the ME principle for the month of July



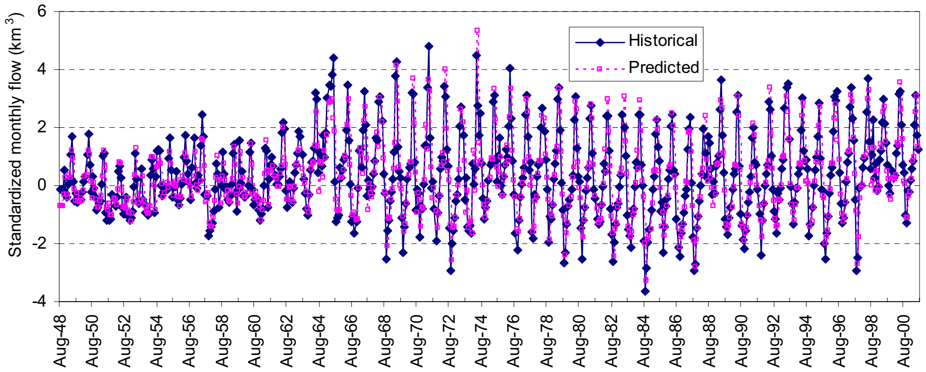
12. Results of stochastic model (validation period)

Natural values



The graphical depiction of monthly predictions (model S1), in comparison to historical values, indicates good performance of the model (see also panel 17)

Standardized values

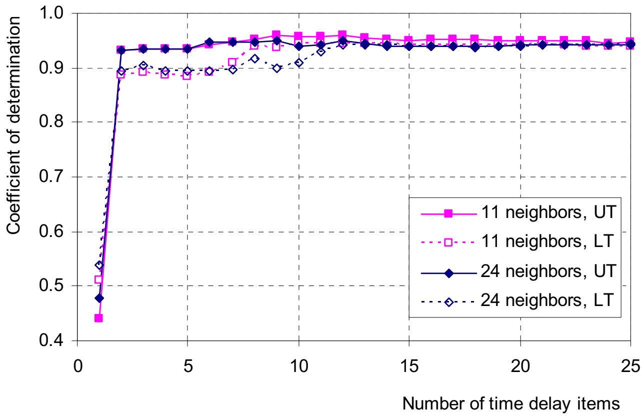


13. The analogue model

- This is a simple nonlinear prediction model: chaotic deterministic, data-driven and non-parametric
- The only adjustable parameters it uses are the embedding dimension m and the number of neighbours n
- The underlying assumptions are:
  - The system dynamics can be described by an attractor that can be embedded in an m-dimensional Euclidian space;
  - This attractor (and the state of the system) can be described in terms of time delay vectors $\mathbf{x}_i := [x_i, x_{i-\tau}, \dots, x_{i-(m-1)\tau}]^T$ where τ a positive integer (typically = 1)
  - Thus, the system dynamics is expressed as $\mathbf{x}_{i+1} = \mathbf{S}(\mathbf{x}_i)$ or $x_{i+1} = S_1(\mathbf{x}_i)$
  - The transformation $S_1(\mathbf{x}_i)$ is unknown but can be locally approximated from the data point nearest to $\mathbf{x}_i$ (an 'analogous' state) or else from n points nearest to $\mathbf{x}_i$
- The algorithm is very easy (Kantz & Schreiber, 1997; Georgakakos & Yao, 1995, 2001):
  - At the current time i, compose the state vector $\mathbf{x}_i$
  - In the calibration data set locate n vectors $\mathbf{y}_i^{(j)}$ ($j = 1, \dots, n$) nearest to $\mathbf{x}_i$
  - The prediction of x_{i+1} at time $i + 1$ is the average of $\mathbf{y}_{i+1}^{(j)}$ over j
- The model calibration is a trial-and-error procedure aiming at finding the optimal m and n that make the prediction error minimum at the verification period
- A two-scale modified version can be derived assuming $\mathbf{x}_i := [x_i, x_{i-1}, x_{i-12}, x_{i-24}]^T$

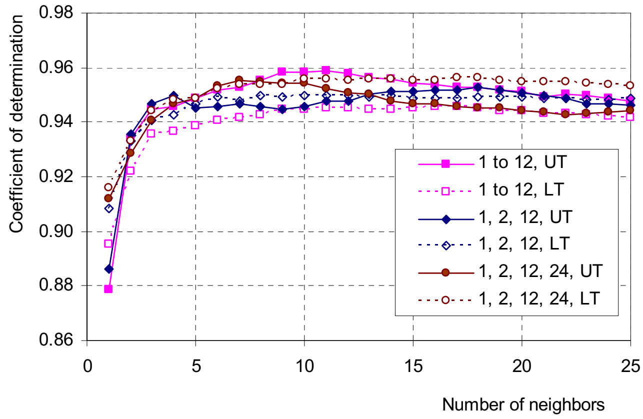
14. Fitting of the analogue model

Coefficients of efficiency (see definition in panel 17) attained by the analogue model for the verification period, as a function of: (up) number of delay items assuming fixed (11 or 24) number of neighbors; (down) the number of neighbors assuming the indicated delay items
UT: untransformed (natural) values;
LT: logarithmically transformed values



Coefficients of efficiency in the verification period of the optimal configurations of the analogue model

Model	Untransformed values	Logarithmically transformed values
A1	0.959	0.945
A2	0.945	0.942
A3	0.955	0.954

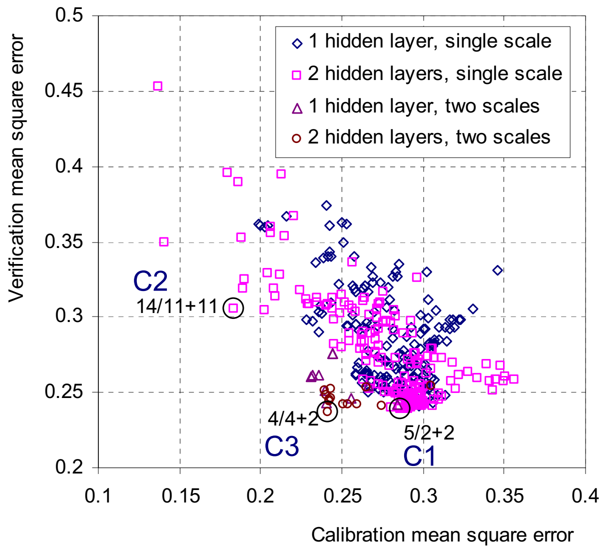


15. The connectionist model

- The connectionist model, also known as an artificial neural network model is a deterministic model based on the same assumptions as the analogue model
- The difference is that it expresses the transformation $x_{i+1} = S_1(\mathbf{x}_i)$ explicitly, as a weighted sum of linear or sigmoidal ($\phi(x) = 1/(1 + e^{b\,x-c})$ elementary functions; the elements of $\mathbf{x}_i$ represent the 'input nodes' on an 'input layer', the result x_{i+1} represents the 'output node' and the specific expression of $S_1(\mathbf{x}_i)$ corresponds to a geometric analogue of nodes and arcs forming a network, which has been called 'connectionist model' or metaphorically 'neural network model'
- The intermediate (between input and output) nodes are typically arranged in the so called 'hidden layers'; in our case, structures with one or two 'hidden layers' have been examined
- The model fitting, metaphorically known as 'training' or 'learning', is a nonlinear optimization procedure than minimizes fitting errors and is typically executed by the 'error backpropagation' method which is a version of a gradient descent method
- To avoid overfitting (i.e. use of too many components of elementary functions) two fitting measures should be used: the *calibration error* (in the calibration period) and the *verification error* (in the verification period; Georgakakos and Yao, 1995)
- The two errors typically display a conflicting behaviour; thus the solution of the optimization problem is the determination of a Pareto front rather than a single point
- As in the analogue model case, a two-scale modified version was also used

16. Fitting of the connectionist model

Plot of the attained verification error vs. the attained calibration error of a series of configurations of the connectionist model with 1 to 15 input nodes, 1 to 2 hidden layers, and 1 and 15, hidden nodes in each layer; the circled points in the Pareto front, for which the number of input and hidden nodes are marked, depict the solutions further explored



Performance indices for the calibration and verification period of the solutions C1-C3
Note: All indices refer to monthly standardized series whereas those in panel 14 refer to natural series

Model	Mean square error		Coefficient of efficiency	
	Calibration	Verification	Calibration	Verification
C1	0.289	0.241	0.749	0.435
C2	0.183	0.309	0.842	0.277
C3	0.241	0.240	0.794	0.438

17. Intercomparison of the prediction skill of models

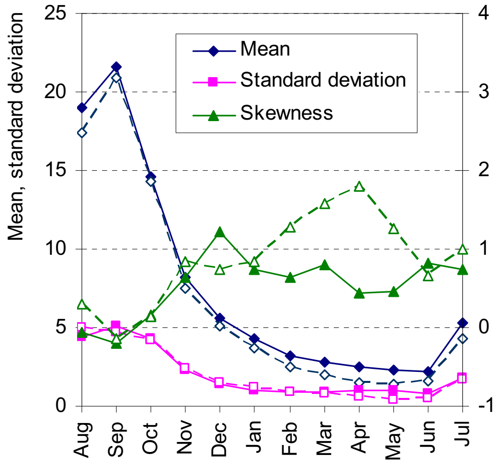
Performance index = coefficient of efficiency (CE = 1 – $E[(W – X)^2] / \mathrm{Var}[X]$)
for the validation period (53 years, all months simultaneously)

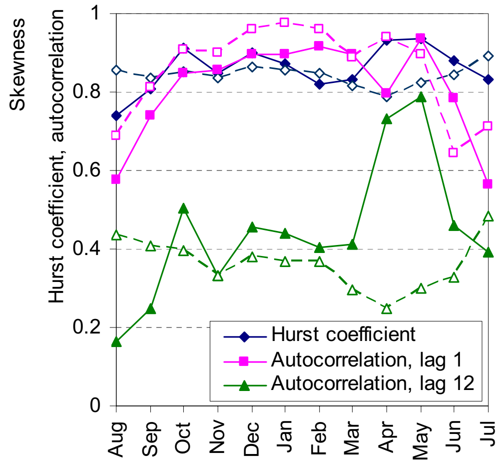
Model	Untransformed values	Logarithmically transformed values	Seasonally standardized untransformed values
S1	0.911	0.904	0.673
S2	0.907	0.899	0.675
S3	0.884	0.884	0.624
A1	0.840	0.613	-0.145
A2	0.847	0.623	-0.126
A3	0.879	0.851	0.490
C1	0.888	0.878	0.583
C2	0.775	0.791	0.280
C3	0.859	0.849	0.472

18. The behaviour of models in simulation mode

- The stochastic forecast models can be directly operate in simulation mode by generating the random component *V* (instead of equating it to zero)
- The analogue model cannot operate in simulation mode because soon it converges to an "attracting" periodic trajectory, same for all years
- The connectionist model, when the number of nodes is small, behaves like the analogue model resulting in an "attracting" periodic trajectory; otherwise (for more than 15-20 hidden nodes) it produces irregular trajectories, which however are statistically inconsistent with historical evolution of flows

Comparison of statistics of the historic 131-year record and a synthetic record of equal length generated by model S1





19. Conclusion: questions studied (and answers)

- Which of the models is based on the most consistent concept? (S)
- Which of the models is the simplest to construct? (A)
- Which of the models has the least number of parameters? (A)
- Which of the models has the best performance? (S)
- Which of the models can incorporate/reproduce long-range dependence? (Only S; but A and C can be altered in a two-scale setting thus enabling incorporation of a "medium-range" dependence)
- Does incorporation of long-range (or medium-range) dependence increase performance? (Yes: S1 > S3, S2 > S3; A3 > A1, A3 > A2)
- Which of the models can run in simulation mode, in addition to forecast mode? (Only S)
- How can the stochastic model, built on the hypothesis on maximum uncertainty (entropy), yield better forecasts than the deterministic models negating uncertainty? (Perhaps because it is closer to natural behaviour?)

20. References

- Georgakakos, A. and H. Yao (1995), Inflow forecasting models for the High Aswan Dam, Technical Project Report to the Egyptian Ministry of Public Works and Water Resources, School of Civil and Environmental Engineering, Georgia Tech.
- Kantz, H., and T. Schreiber (1997), *Nonlinear Time Series Analysis*, Cambridge University Press, Cambridge.
- Koutsoyiannis, D. (2000), A generalized mathematical framework for stochastic simulation and forecast of hydrologic time series, *Water Resources Research*, 36(6), 1519-1533.
- Koutsoyiannis, D. (2003), Climate change, the Hurst phenomenon, and hydrological statistics, *Hydrological Sciences Journal*, 48(1), 3-24.
- Koutsoyiannis, D. (2005a), Uncertainty, entropy, scaling and hydrological stochastics, 1, Marginal distributional properties of hydrological processes and state scaling, *Hydrological Sciences Journal*, 50(3), 381-404.
- Koutsoyiannis, D. (2005b), Uncertainty, entropy, scaling and hydrological stochastics, 2, Time dependence of hydrological processes and time scaling, *Hydrological Sciences Journal*, 50(3), 405-426.
- Papoulis, A. (1991), *Probability, Random Variables, and Stochastic Processes* (third edn.), McGraw-Hill, New York.
- Said, R. (1993), *The River Nile: Geology, Hydrology and Utilization*, Pergamon Press, Oxford.
- Tsallis, C. (2004), Nonextensive statistical mechanics: construction and physical interpretation, in *Nonextensive Entropy, Interdisciplinary Applications* (edited by M. Gell-Mann & C. Tsallis), Oxford University Press, New York, NY.
- Tsallis, C., S. V. F. Levy, A. M. C. Souza, and R. Maynard (1995), Statistical-Mechanical Foundation of the Ubiquity of Lévy Distributions in Nature, *Phys. Rev. Lett.*, 75, 3589–3593.
- Yao, H, and A. Georgakakos (2001), "Assessment of Folsom Lake Response to Historical and Potential Future Climate Scenarios," Journal of Hydrology, 249, 176-196.