The management of the Athens water resource system: Methodological issues

Demetris Koutsoyiannis

Department of Water Resources, Faculty of Civil Engineering,
National Technical University of Athens

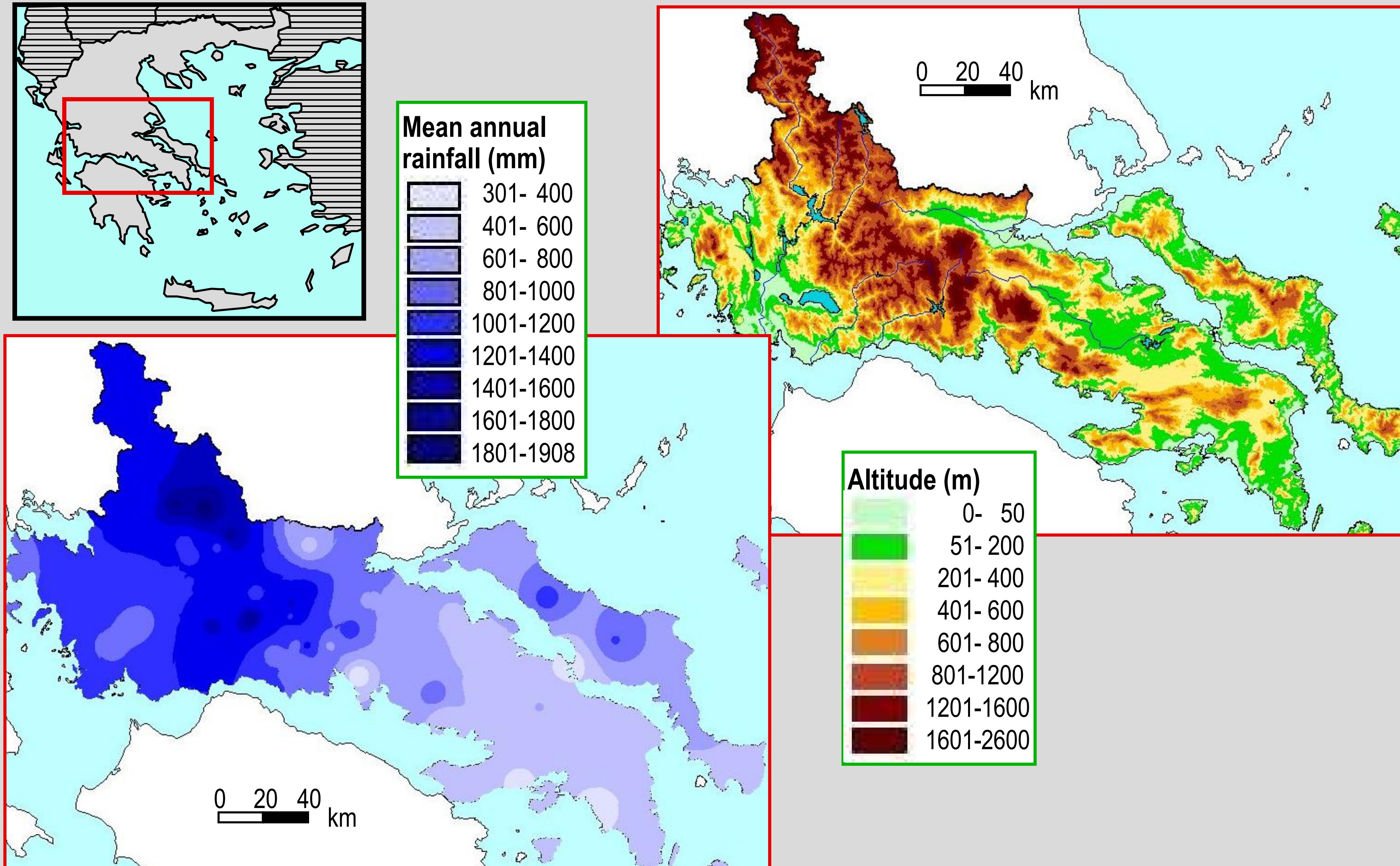
A seminar given at the Hydrologic Research Center

San Diego, 22 November 2005

Acknowledgments

Ὕσον, ὗσον Ζεῦ κατὰ τῆς ἀρούρης τῶν Ἀθηναίων

Do rain, do rain Zeus against the earth of Athenians (Ancient Greek prayer)



Parts of the presentation

1. The Athens water resource system

   History – Components – Technical characteristics

2. Hydrologic issues

   Diagnosis – Explanation – Operational synthesis

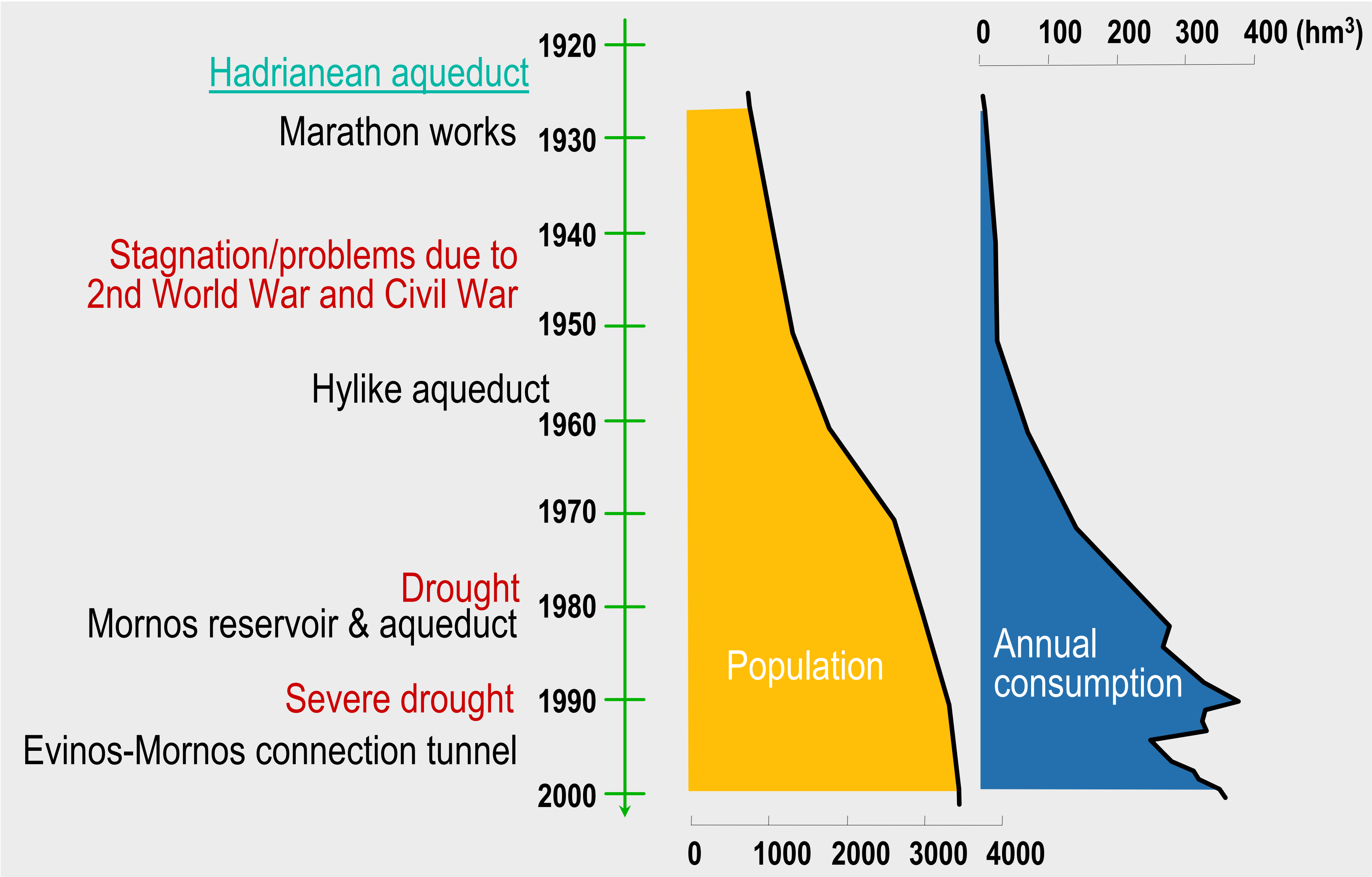
3. Hydrosystem operation issues

   Parameterization – Simulation – Optimization

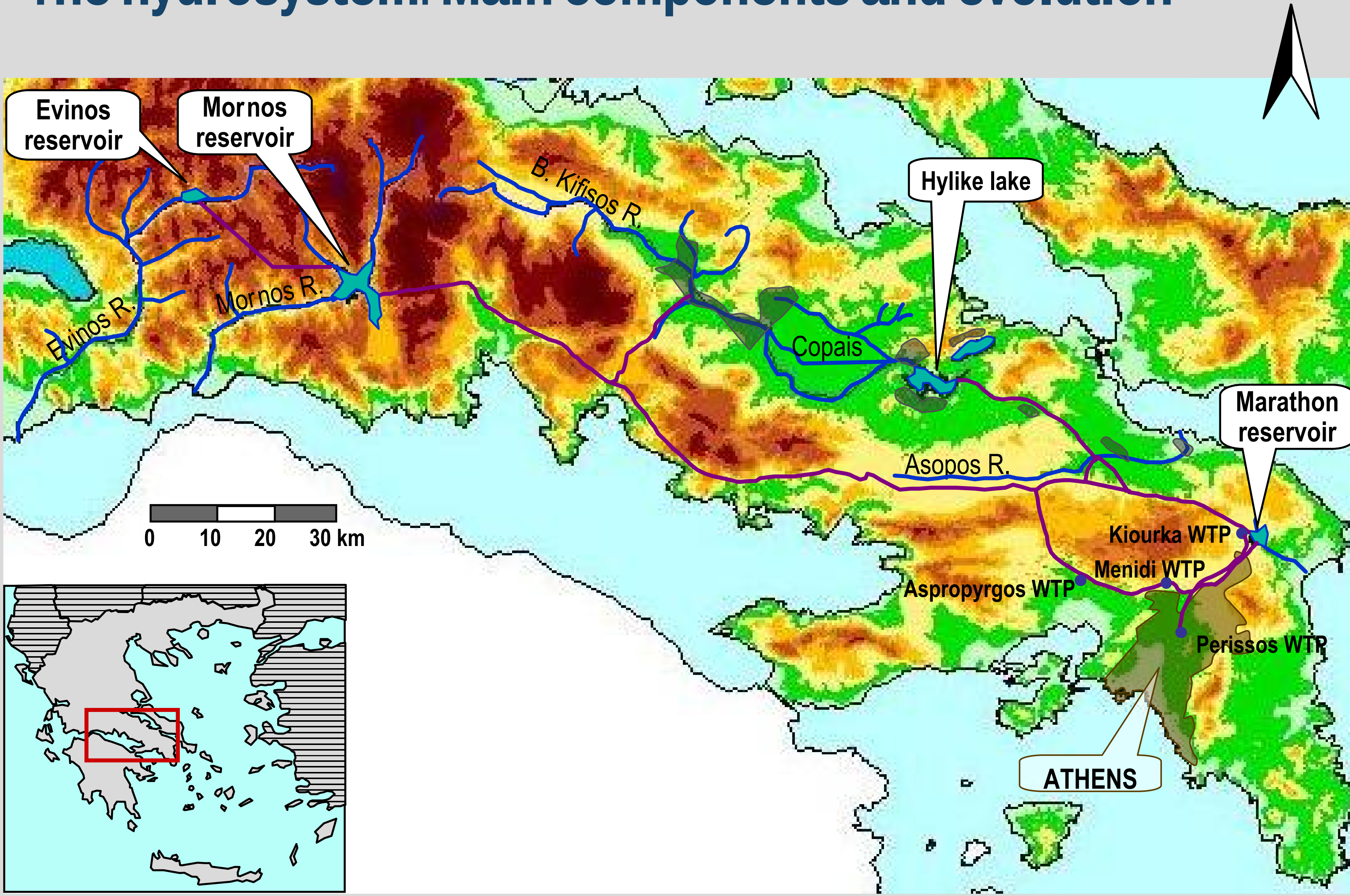
4. Decision support tool integration

   Data acquisition – Software systems – Management plans

Evolution of water consumption – Milestones



The hydrosystem: Main components and evolution

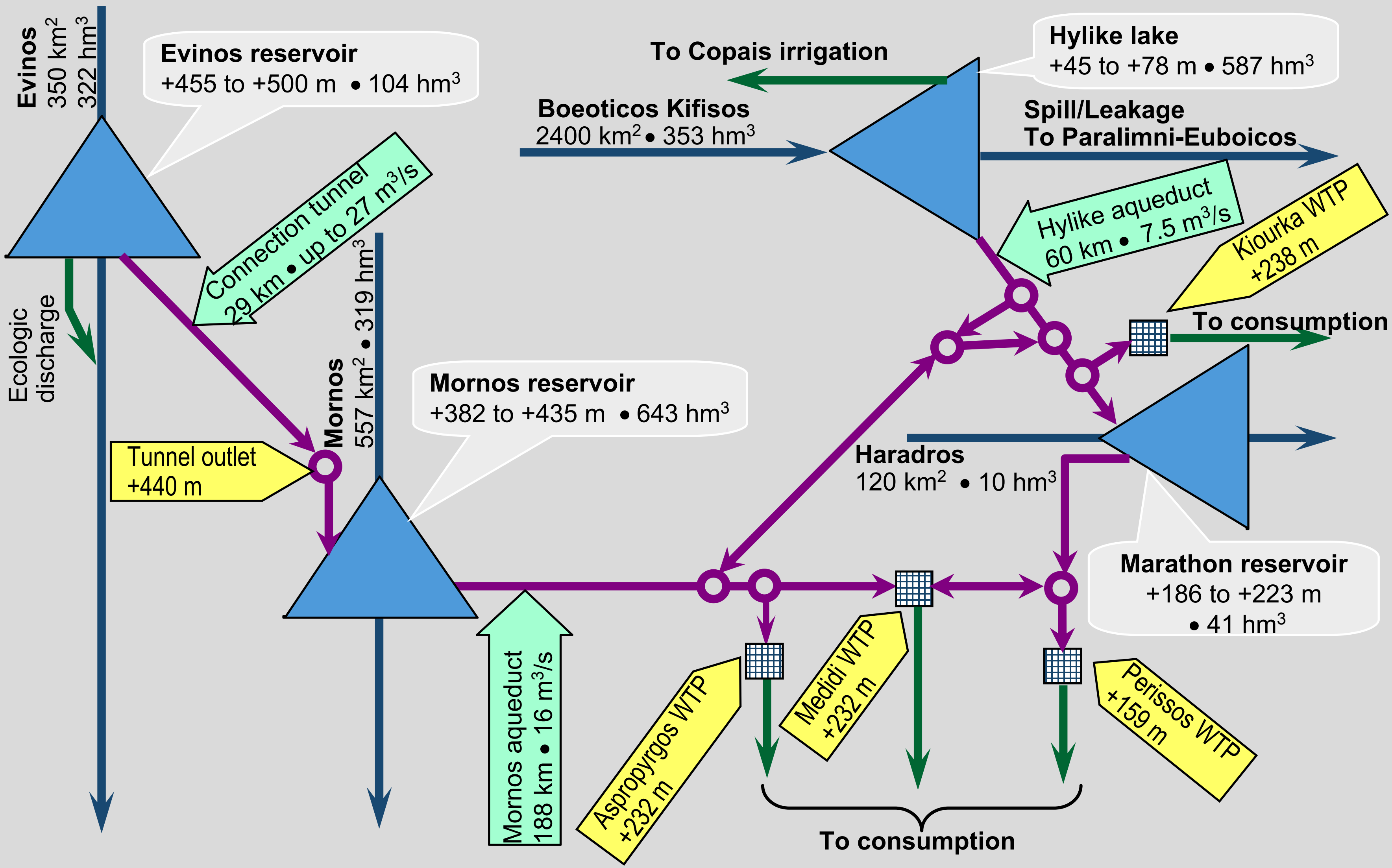


Classification of water resources

	SURFACE WATER		GROUNDWATER
Basin	Primary (Reservoirs)	Secondary (Reservoirs)	Backup (Boreholes)
Evinos 350 km²	Evinos 322 hm³/y		
Mornos 557 km²	Mornos 319 hm³/y		
Boeoticos Kifisos – Yliki 2400 km²		Yliki 353 hm³/y	B. Kifisos, middle course 136 hm³/y Yliki region 85 hm³/y
Haradros 120 km²		Marathon 10 hm³/y	
North Parnetha			Viliza 26 hm³/y Mavrosouvala 36 hm³/y

Area	Inflow	Pumping capacity	High spill	High leakage	Pumping

Hydrosystem: Current structure



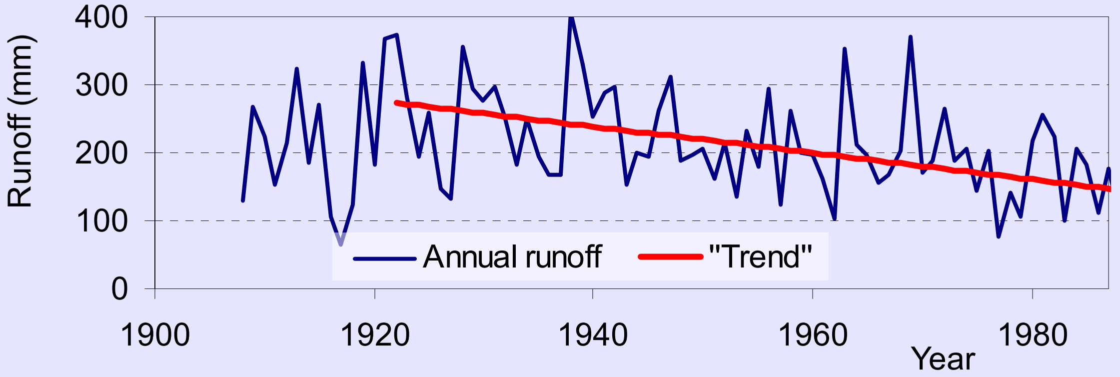
+ Boreholes (with connecting pipes) + Pumping stations + Small hydroelectric power plants

2. Hydrologic issues
2A. Diagnosis

Back in 1990s – Initial empirical observations

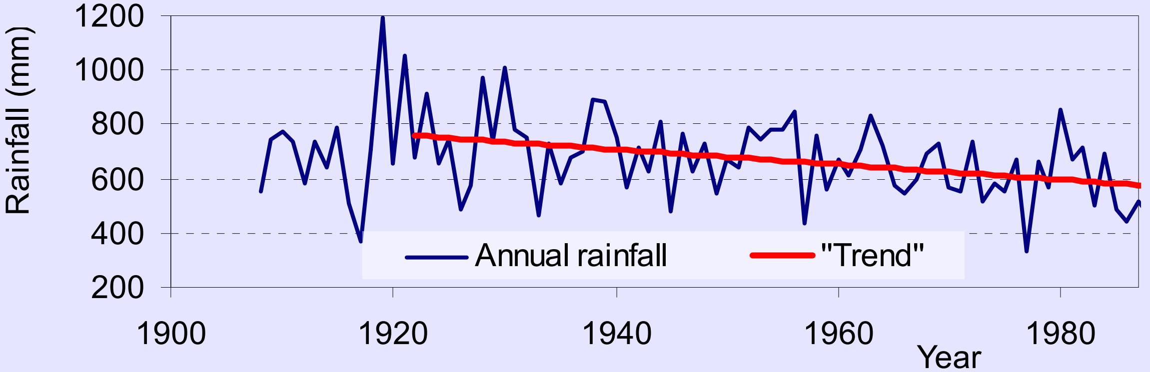
The historic time series of Boeoticos Kephisos runoff (Hydrologic years 1907/08-1986/87)

A multi-year «trend» is observed



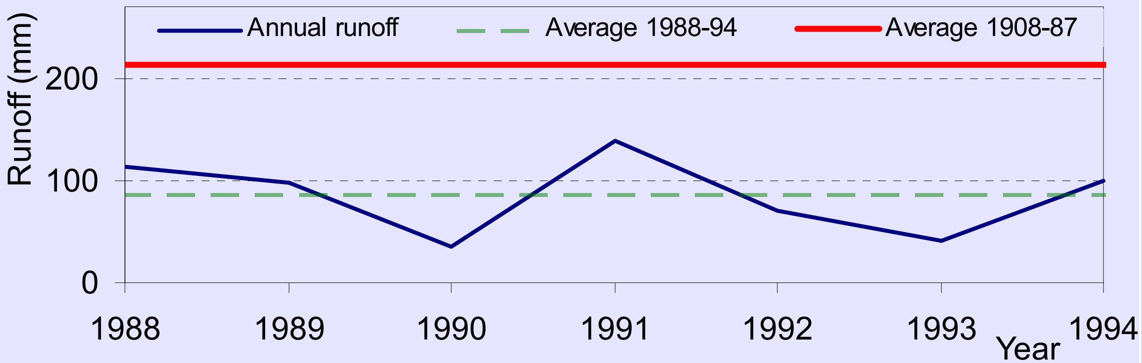
A similar «trend» in the rainfall time series

Explains the «trend» in runoff



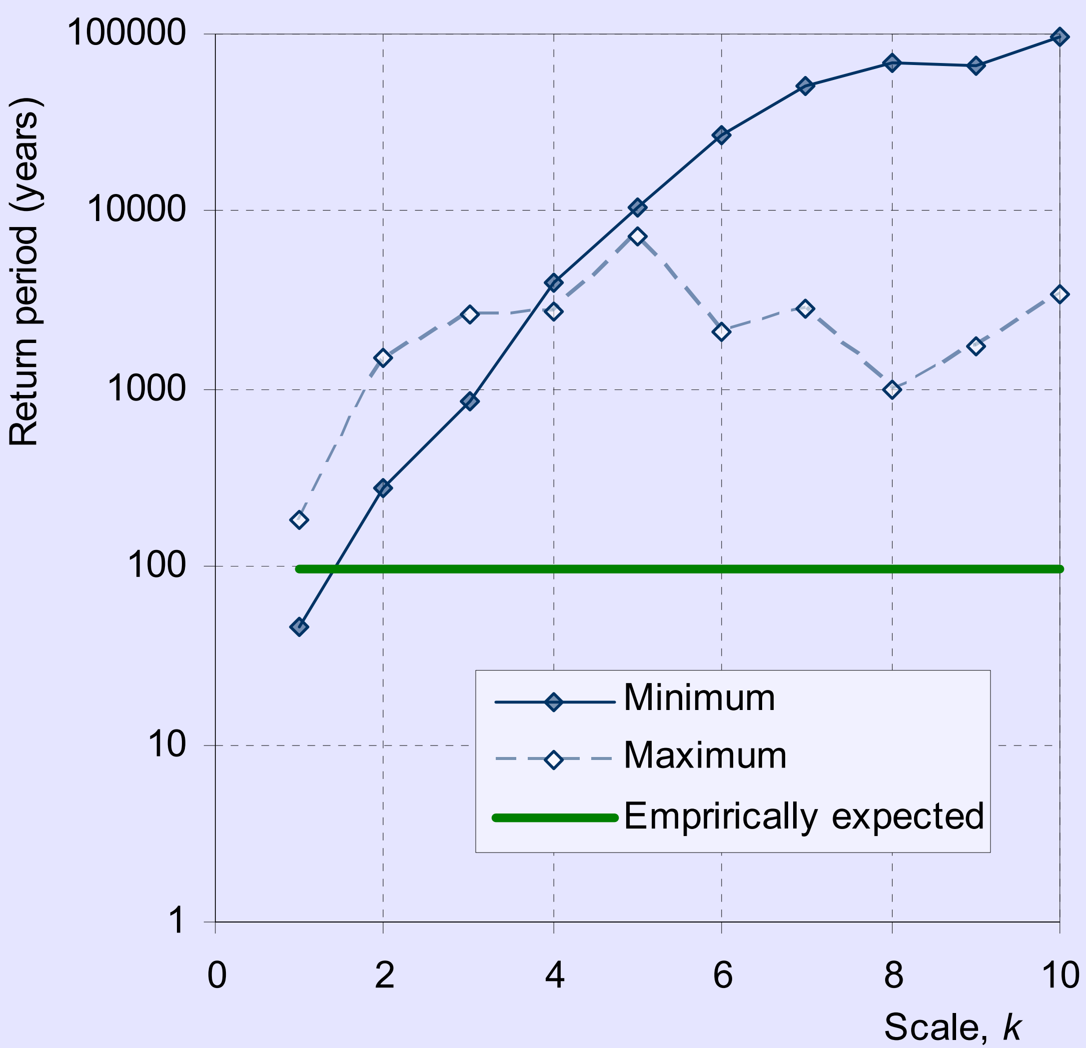
The next years were dry

Intense and persistent drought: Mean flow half that of the historic average, duration 7 years



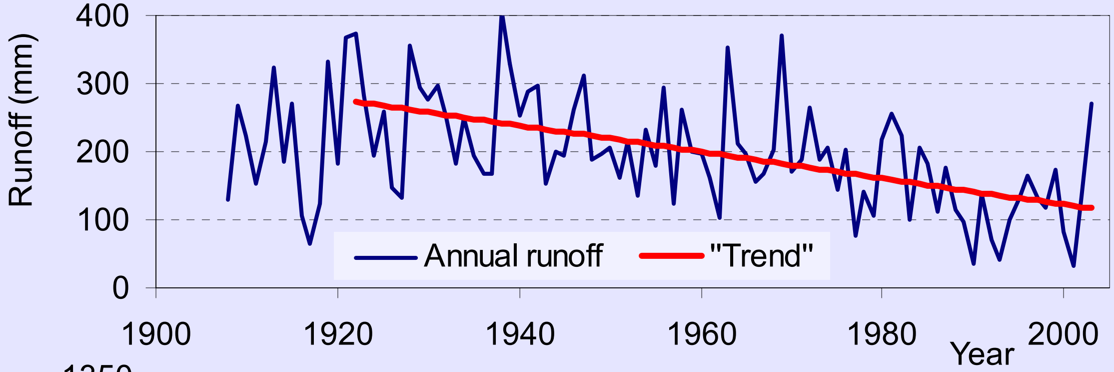
Return period of the persistent drought

- Assessment was done using classic hydrologic statistics
- At the annual scale, the drought was a record minimum but with typical magnitude
- Aggregated at larger scales, it appeared something extraordinary
- Similar behavior was observed for maxima on aggregate scales



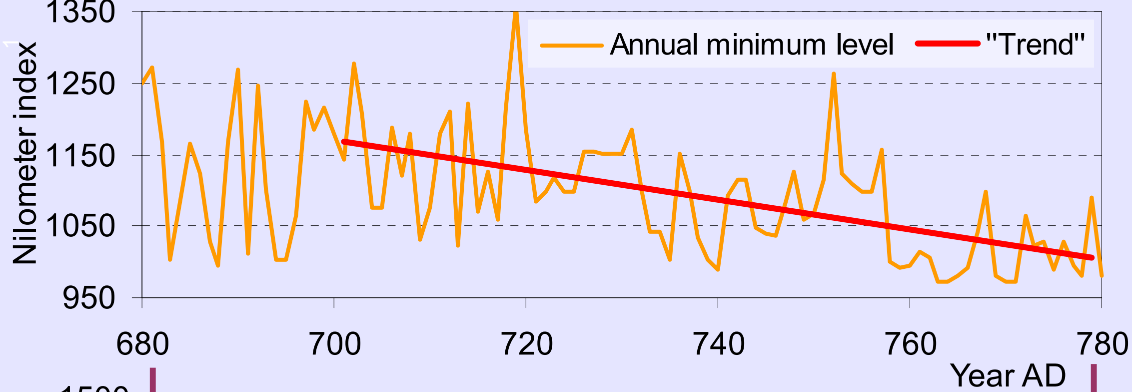
Comparisons with even longer series

The complete historic time series of Boeoticos Kephisos runoff



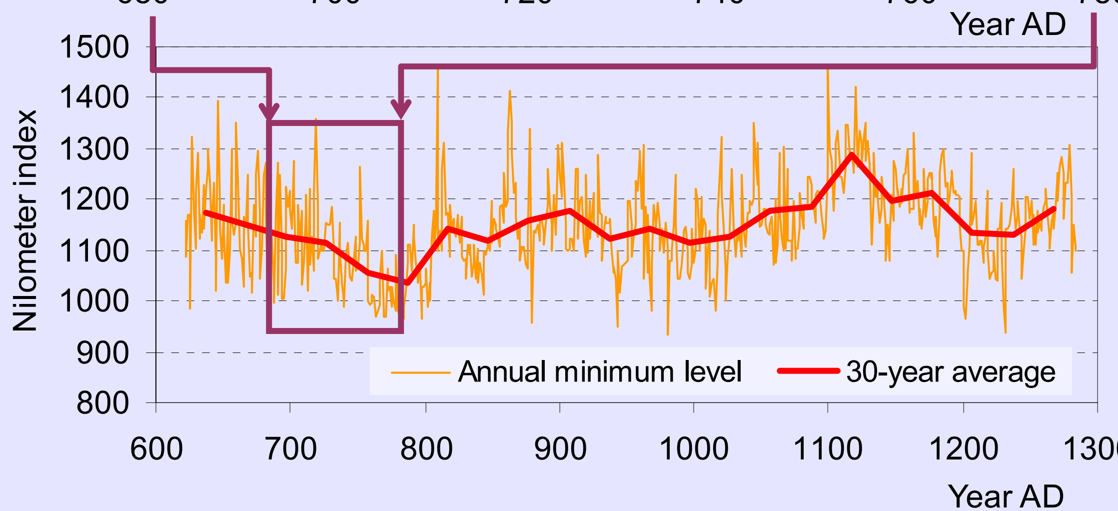
A part of the Nilometer series (an index of the minimum annual level in the Nile River*

A similar «trend»



The complete Nilometer series (622-1284 AD, 663 years)

Upward and downward fluctuations on all scales



* J. Beran (1994), *Statistics for Long-Memory Processes*, Chapman & Hall, New York, USA

The fluctuations on many scales and the "Hurst phenomenon"

- The "weird" (as compared to purely random processes) behavior of hydrologic and other geophysical processes was discovered by the English engineer E. H. Hurst* (1950) in the framework of the design of the High Aswan Dam in Nile ⇒ **Hurst phenomenon**
- The Polish-French mathematician and engineer B. Mandelbrot (1965-1971) related it to the biblical story of the seven fat and the seven thin cows ⇒ **Joseph effect**
- The behavior has been characterized with several other names ⇒ **long-term persistence, long-term memory, long-range dependence**
- Most of these names, even though correct, may be misleading for the conceptualization and understanding of the natural behavior and the causing mechanisms. Probably a better name ⇒ **multi-scale fluctuation**
- The behavior was verified to be omnipresent, not only in geophysical processes (hydrologic, climatic), but also in biological (e.g. tree rings), technological (e.g. computer networks), social and economical (e.g. stock market)
- In water resources design and management, it has unfavorable effects (increase of uncertainty)

* H. E. Hurst (1950), Long-Term Storage Capacity of Reservoirs, *Proc. American Society of Civil Engineers*, 76(11)

Easy detection and main effect of Hurst phenomenon

- Fundamental law of classic statistics $\mathrm{StD}[\bar{X}_n] = \dfrac{\sigma}{\sqrt{n}}$

  $\bar{X}_n$ = average of n variables

  σ = standard deviation of each variable

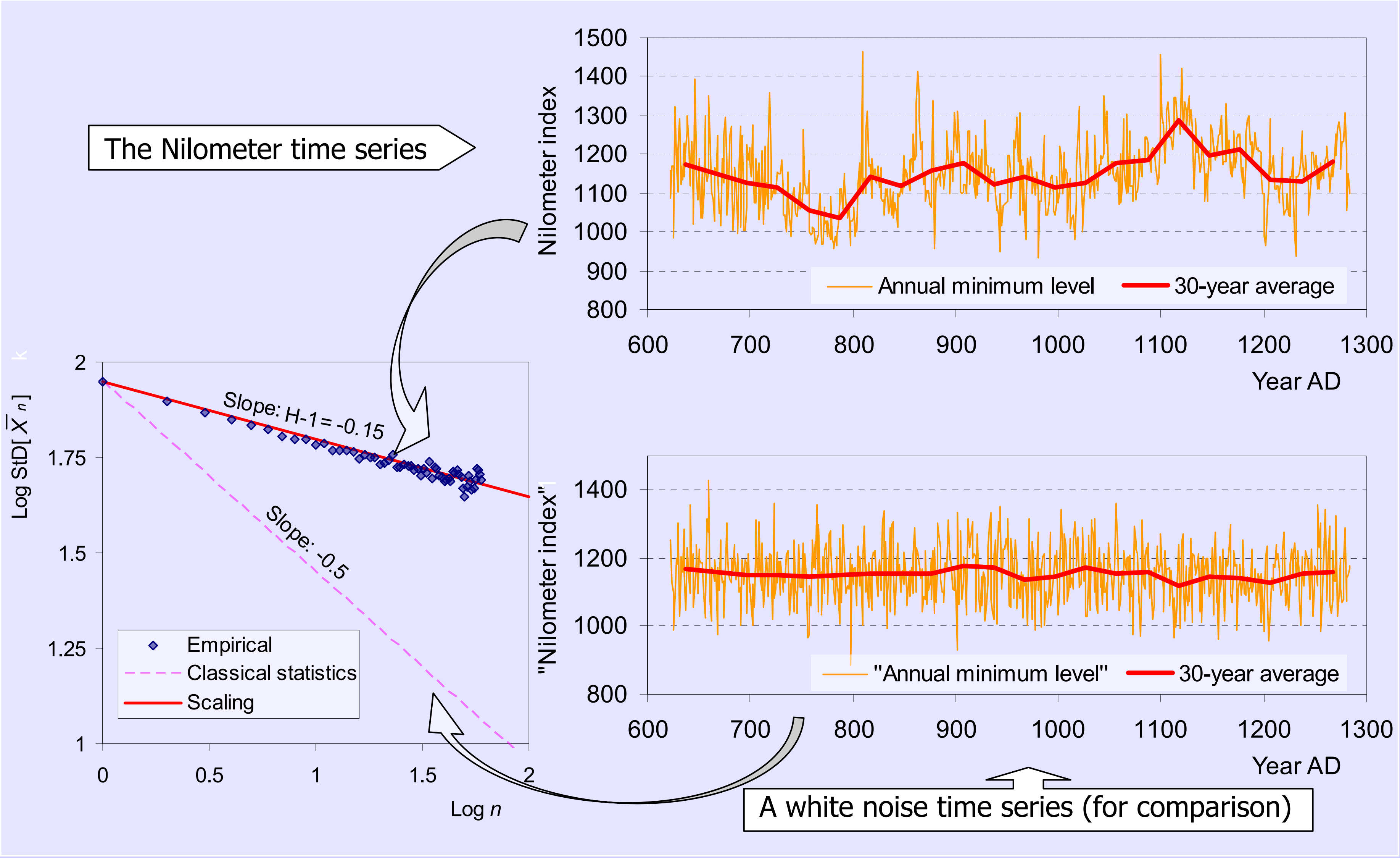
  n = aggregation scale (or sample size)

- Modified law traced natural processes $\mathrm{StD}[\bar{X}_n] = \dfrac{\sigma}{n^{1-H}}$, $H > 0.5$
- Example

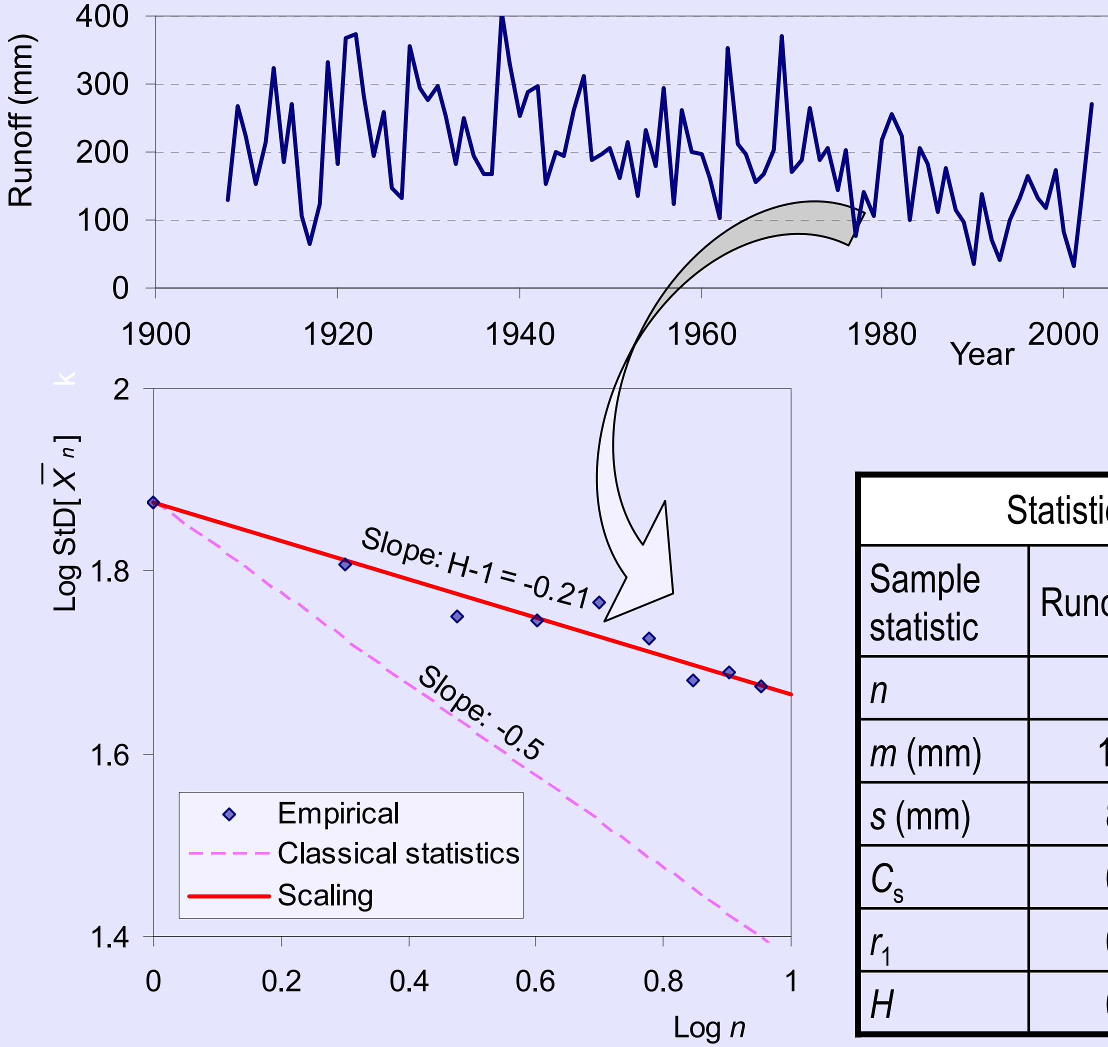
  To have $\mathrm{StD}[\bar{X}_n] / \sigma = 10\%$
  - n = 30 in classic statistics
  - n = 5 000 for the modified law with H = 0.8

Bomb in the foundation of climatology

Incongruity of natural processes with typical random processes : (a) The Nilometer series



Incongruity of natural processes with typical random processes : (b) The Boeoticos Kephisos time series



The time series of the Boeoticos Kephisos runoff

Statistical characteristics of all processes			
Sample statistic	Runoff (mm)	Rainfall (mm)	Temperature (^{o}C)
n	96	96	96
m (mm)	197.6	658.4	17.0
s (mm)	87.6	158.9	0.72
C_s	0.36	0.44	0.34
r_1	0.34	0.10	0.31
H	0.79	0.64	0.72

Mathematical description of the Hurst phenomenon

- The mathematical description of the Hurst phenomenon is done on grounds of probability theory and particularly theory of stochastic process
- The simple relationship

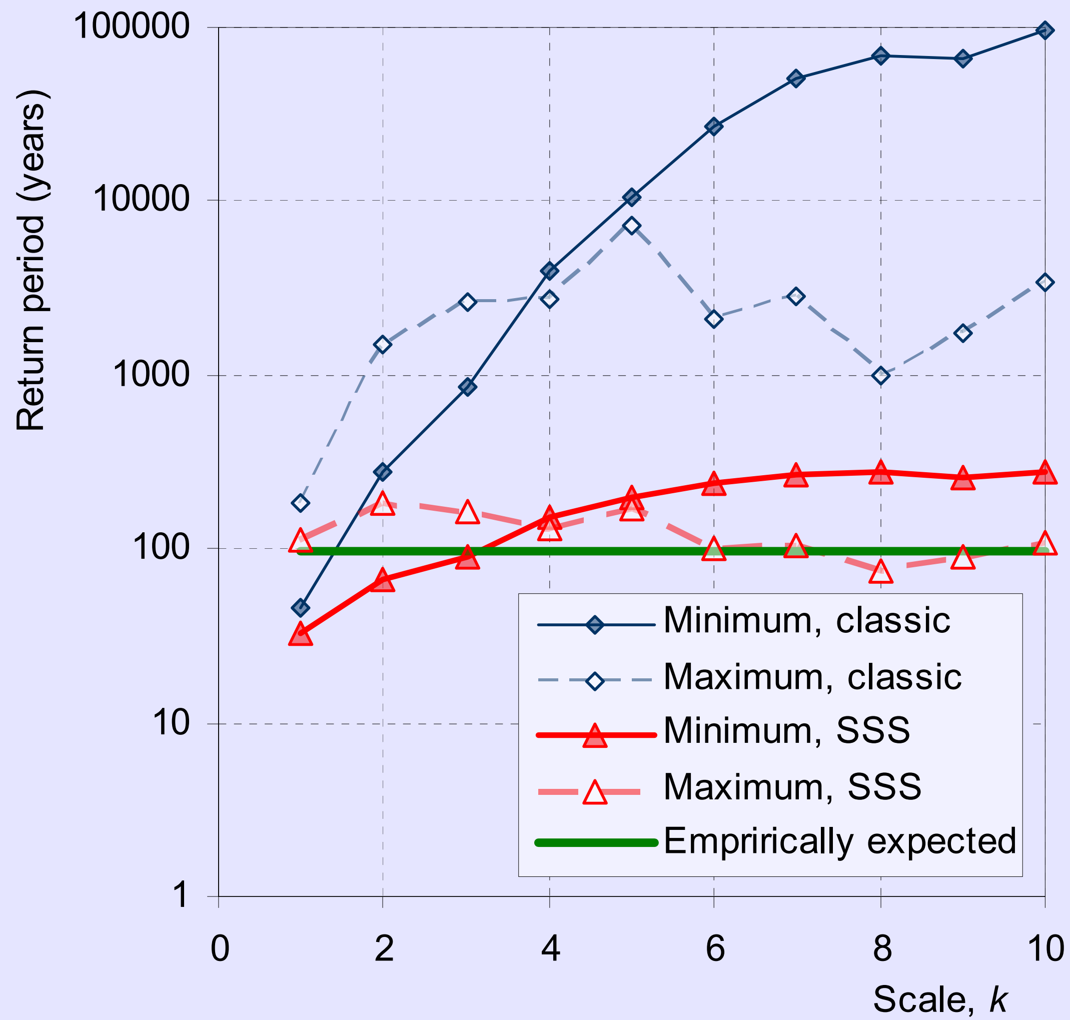
$$\mathrm{StD}[\bar{X}_n] = \frac{\sigma}{n^{1-H}}$$

  entails a definition (good for our purposes) of a model (stochastic process) reproducing the Hurst phenomenon; *n* is meant as a scale of aggregation (rather than sample size)
- (Hurst used a different formalism, in terms of the so called rescaled range, which is complicated and probably misleading)
- Today the stochastic process with the above property is called a **S**elf-**S**imilar process with **S**tationary intervals or a **S**imple **S**caling **S**tochastic process (abbreviated as an **SSS** process)
- The SSS process was introduced by the Russian mathematician A. Kolmogorov* (1940) who called it Wiener Spiral
- A significant contribution on the SSS process is due to the American mathematician J. Lamperti (1962) who called it a Semi-Stable Process
- The link of the SSS process with the Hurst phenomenon is due to B. Mandelbrot (1965), who called it Fractional Brownian Noise

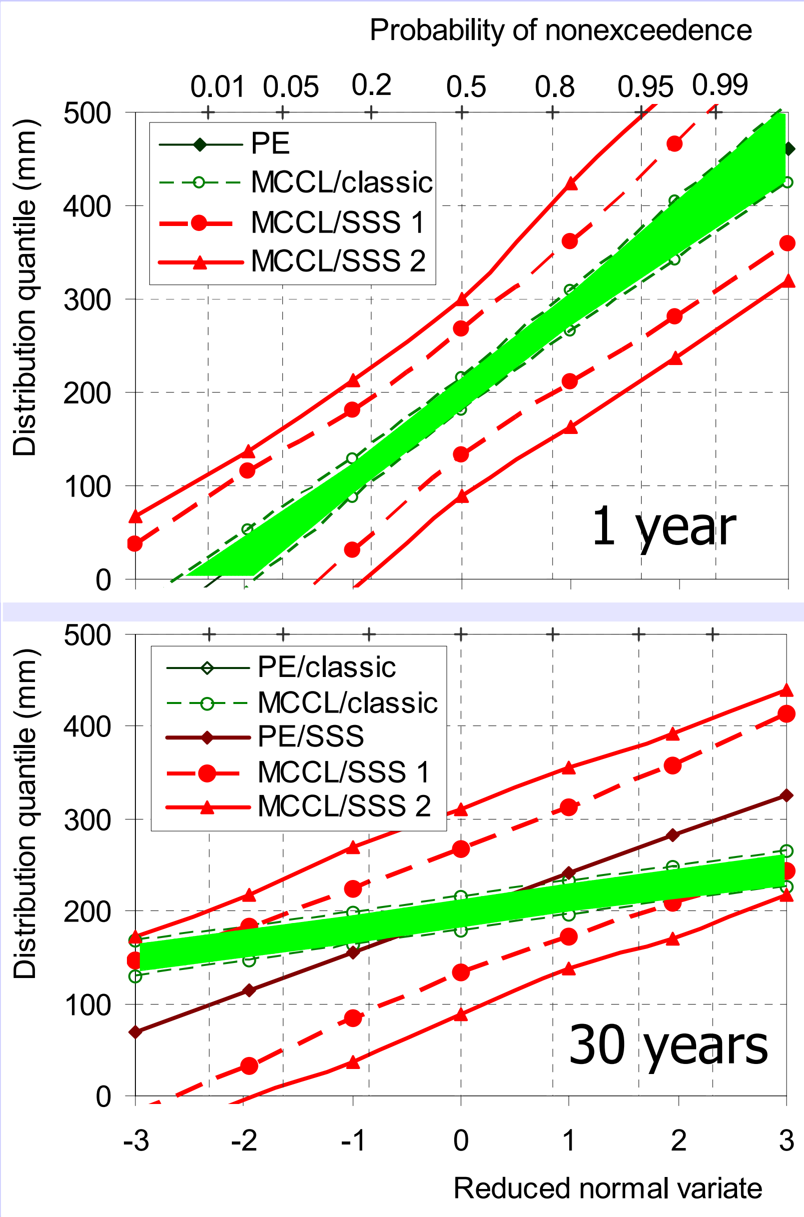
* A. N. Kolmogorov (1940), Wienersche Spiralen und einige andere interessante Kurven in Hilbertschen Raum, Comptes Rendus (Doklady) Acad. Sci. USSR (N.S.) 26, 115–118

Back to Boeoticos Kephisos – Adoption of the SSS process

- The trend is a natural and usual behavior
- The persistent drought is not extraordinary; it is a natural and expected behavior



Implications on uncertainty: Boeoticos Kephisos runoff



Statistical model	Total uncertainty in runoff (due to variability and parameter estimation) % of average	
	Annual scale	30-year scale
Classic	200	50
SSS	270	200

Classic model
Climate is what you expect
Weather is what you get

SSS model
Weather is what you get … immediately
Climate is what you get
… if you keep expecting a long time

2. Hydrologic issues
2B. Explanation

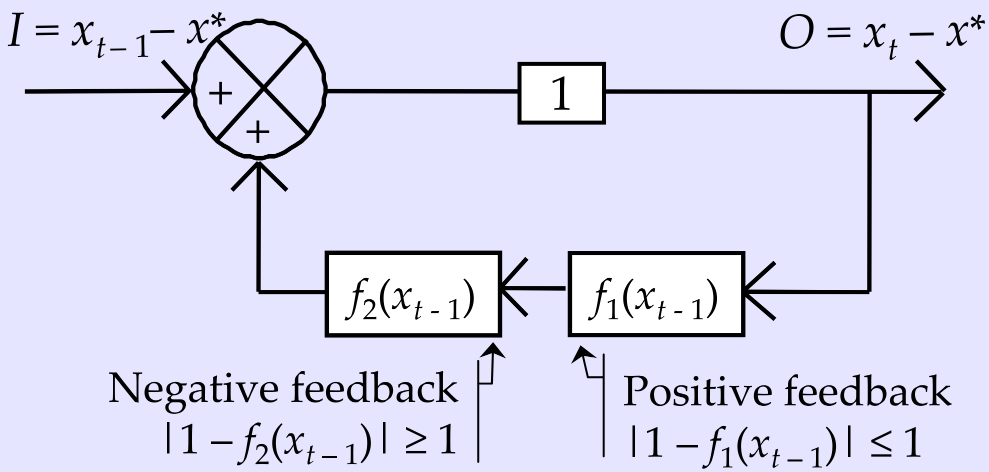
A climatic toy model: A simple system with nonlinear dynamics may produce the Hurst phenomenon

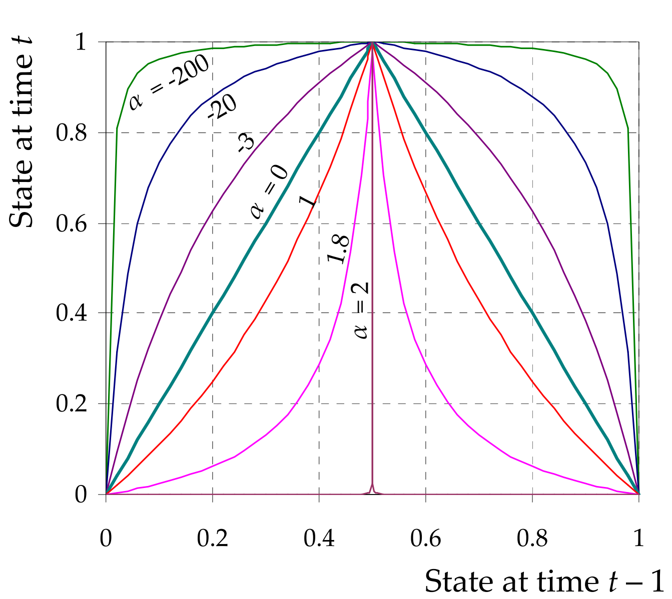
- A simplified climatic system is represented as a circuit with two feedback mechanisms, a positive (amplifying the departure from a stationary state x^*) and a negative (reducing this departure)
- The combined action of the two mechanisms could be represented by a generalized tent transform:

$$x_t = \frac{(2 - \alpha)\min(x_{t-1}, 1 - x_{t-1})}{1 - \alpha \min(x_{t-1}, 1 - x_{t-1})}$$

where $0 \le x_t \le 1$, $\alpha < 2$

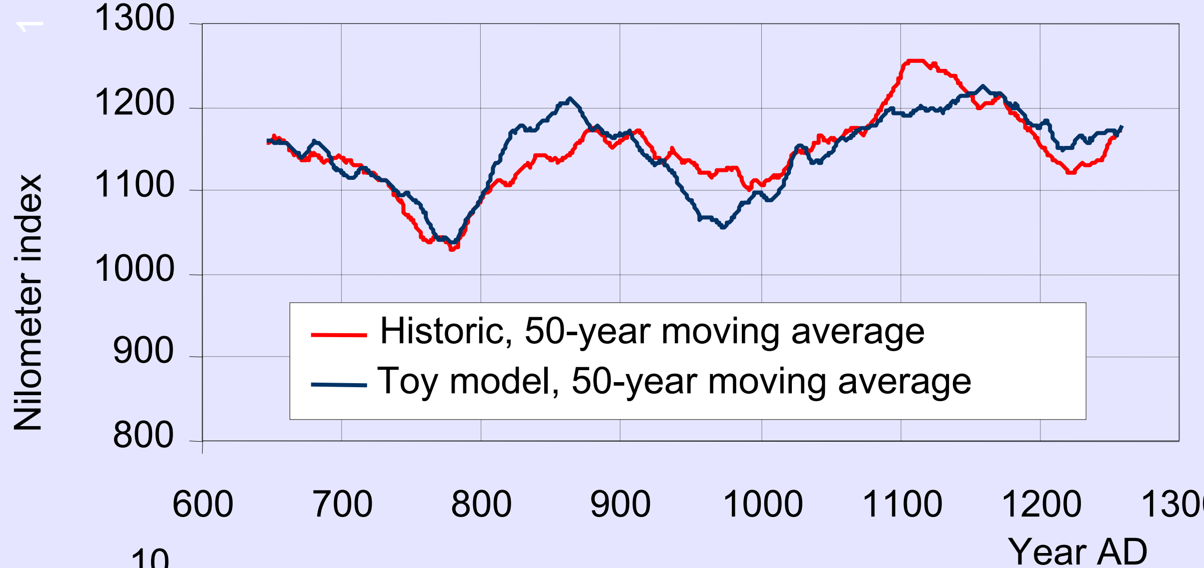
- The parameter α could be assumed to vary in time, following the same tent transform with a constant parameter β



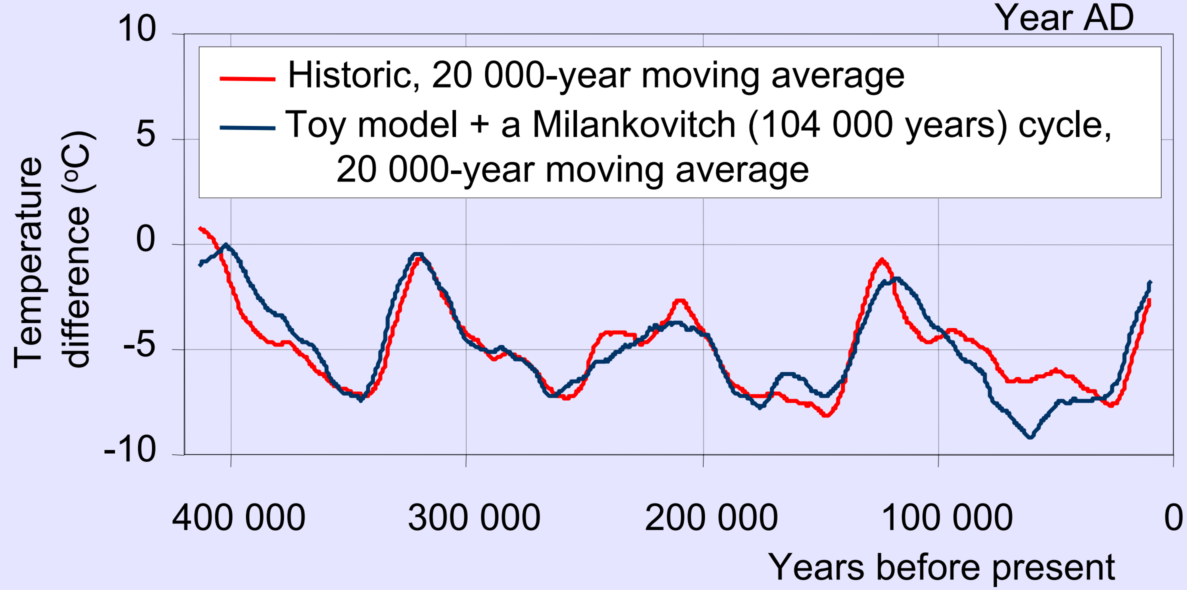


Demonstration continued: Toy model fitted to two long time series

The Nilometer time series

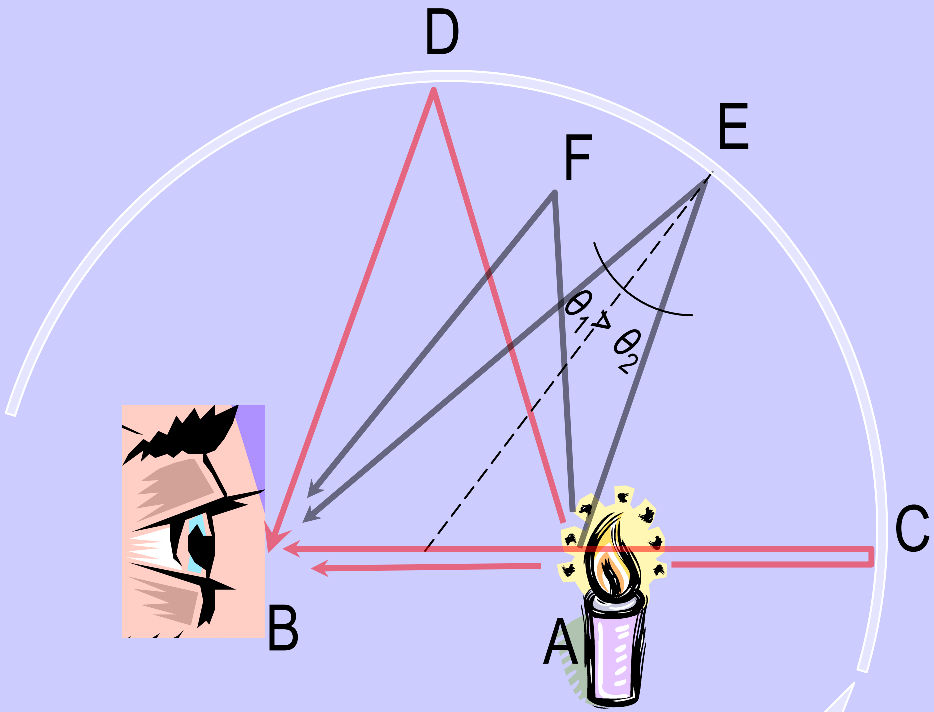


The Vostok (Antarctica) ice core deuterium data set going back to 422 766 years before present* Reconstructed temperature difference with reference to the mean recent time value



* Petit J.R., Jouzel J., Raynaud D., Barkov N.I., Barnola J.M., Basile I., Bender M., Chappellaz J., Davis J., Delaygue G., Delmotte M., Kotlyakov V.M., et al., Climate and atmospheric history of the past 420,000 years from the Vostok ice core, Antarctica, *Nature*, 399, 429-436, 1999.

Towards a more general explanation: Nature loves extremes ...



A semi-cylindrical mirror

Why light follows the red paths from A to B (AB, ACB, ADB) and not other (the black) ones (e.g. AEB, AFB)?

- The red paths are those that (a) reach the mirror and (b) form an angle of incidence equal to the angle of reflection

  (True for most cases; not true for AB; not general or informative)

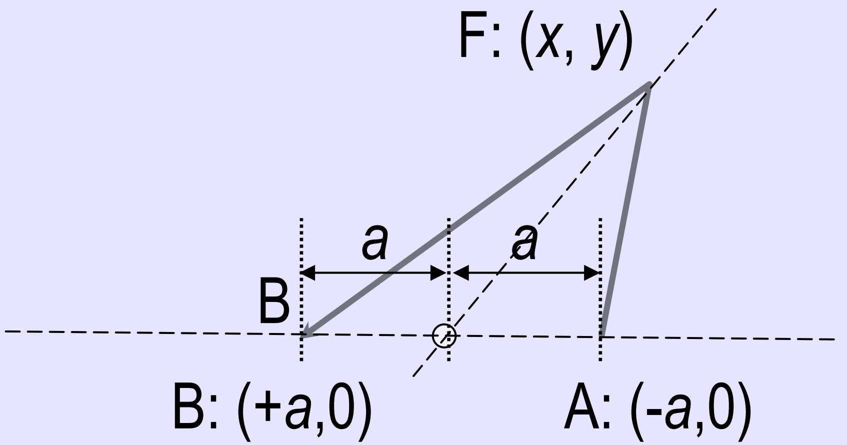
- The red paths have minimum travel time (or length)

  (Not true for ADB)

- The red paths have extreme (minimum or maximum) travel time (or length)

  (True)

The light example – no mirror

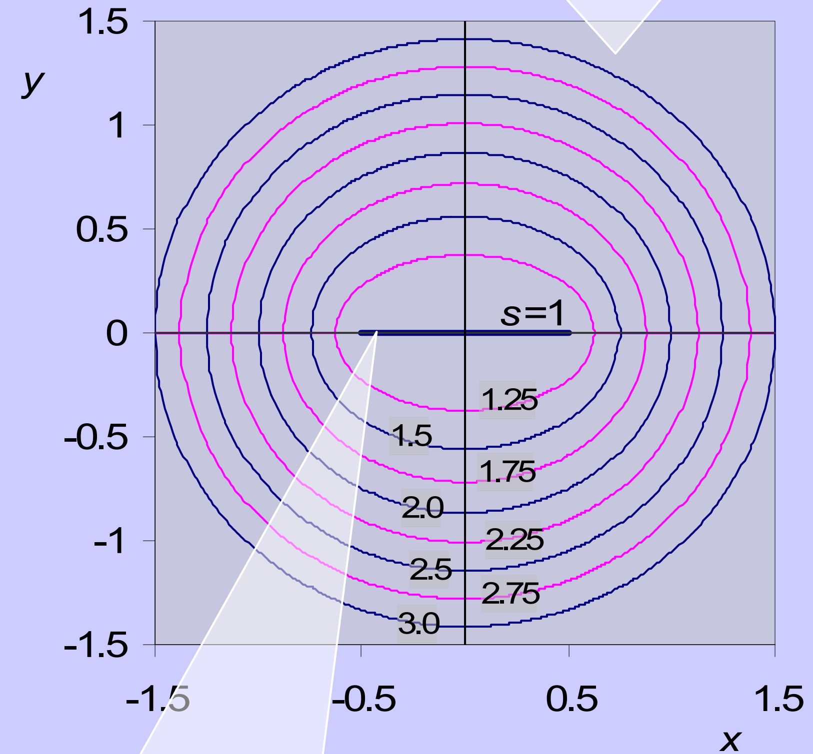


Assume that light can travel from A to B along a broken line with a break point F with coordinates (x, y).
(This is not restrictive: later we can add a second, third, … break points)
The travel distance is $s(x, y)$ = AF + FB where

$$AF = \sqrt{(x - a)^2 + y^2}$$

$$FB = \sqrt{(x + a)^2 + y^2}$$

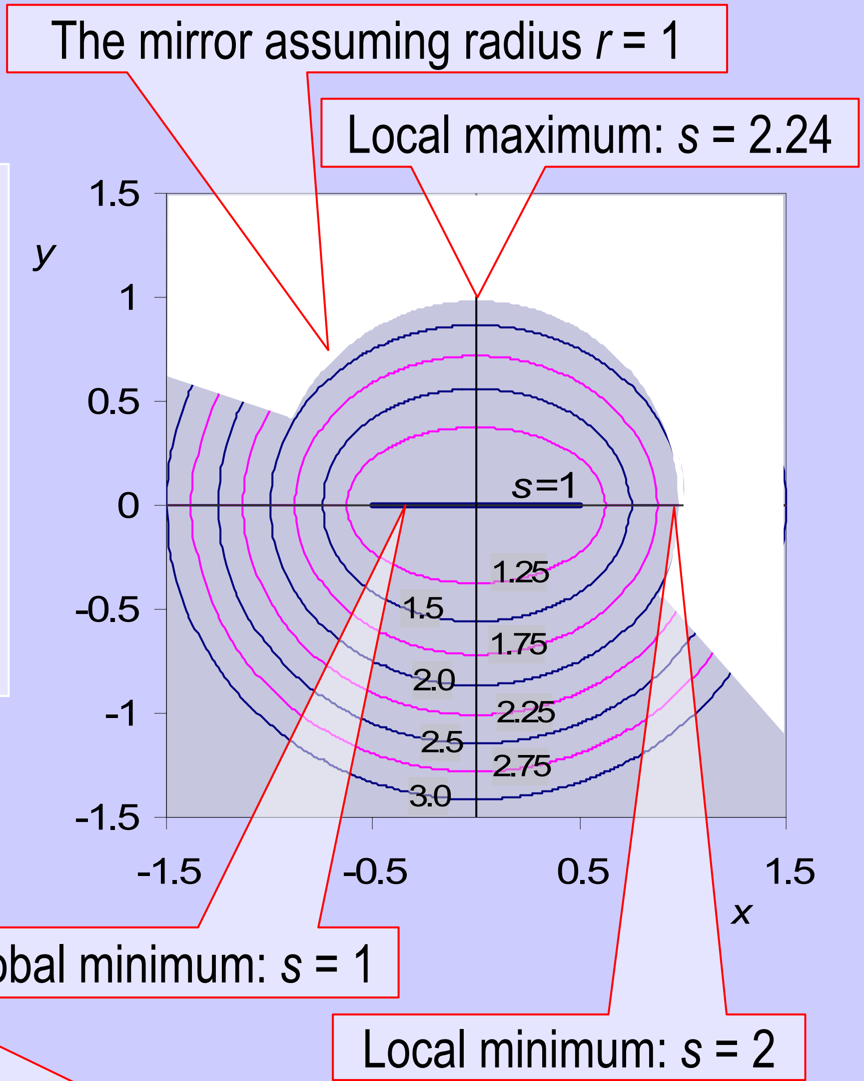
Contours of the distance $s(x, y)$ assuming $a = 0.5$

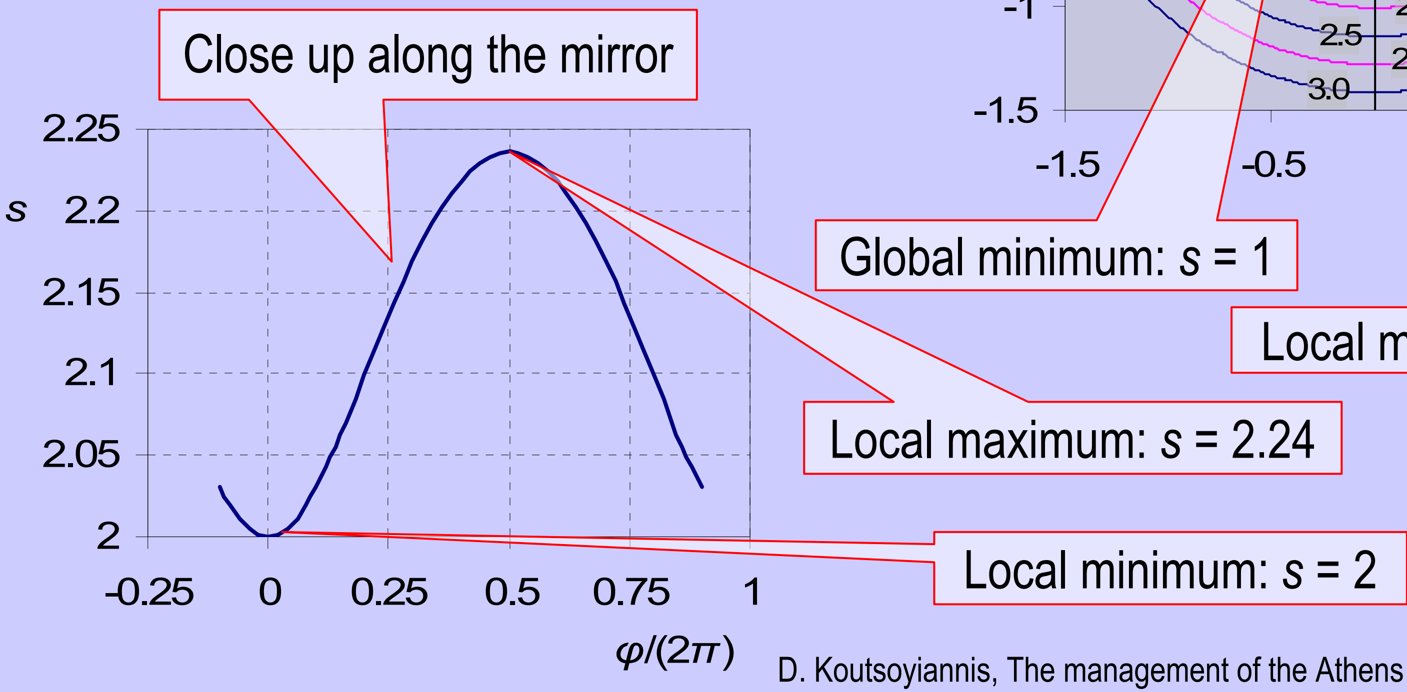


Line of minimum distance $s(x, y) = 1$
Infinite points F essentially describing the same path

The light example with mirror

- The mirror introduces an inequality constraint in the optimization: the point F should not be behind the mirror
- Two points of local optima emerge on the mirror surface (the curve where the constraint is binding)





How nature works? (a hypothesis …)

Property

- She preserves a few quantities (mass, momentum energy, ….)
- She optimizes a single quantity (Dependent on the specific system - Difficult to find what this quantity is)
- She disallows some states (Dependent on the specific system – Maybe difficult to find)

Mathematical formulation

- One equation per preserved quantity:

  $$g_i(\mathbf{s}) = c_i, \quad i = 1, \dots, k$$

  where c_i constants; **s** the size n vector of state variables ($n \geq k$, sometimes $n = \infty$)

- A single “optimation”:

  optimize $f(\mathbf{s})$

  [i.e. maximize/minimize $f(\mathbf{s})$] **This is equivalent to many equations** (as many as required to determine **s**)
  Conversely, many equations can be combined into an “optimation”

- Inequality constraints:

  $$h_j(\mathbf{s}) \geq 0, \quad j = 1, \dots, m$$

- In conclusion, we may find how nature works solving the problem:

  $$\begin{aligned} &\text{optimize } f(\mathbf{s}) \\ &\text{s.t. } g_i(\mathbf{s}) = c_i, \quad i = 1, \dots, k \\ &\phantom{\text{s.t. }} h_j(\mathbf{s}) \geq 0, \quad j = 1, \dots, m \end{aligned}$$

The typical “optimizable” quantity in complex systems …

- … is entropy – entropie – Entropie – entropia – entropía – entropi – entrópia – entroopia – entropija – энтропия – ентропія – 熵 – エントロピー – مقياس – אנטרופיה– εντροπία
- The word is ancient Greek (εντροπία, a feminine noun meaning: turning into; turning towards someone’s position; turning round and round)
- The scientific term is due to Clausius (1850)
- The entropy concept was fundamental to formulate the second law of thermodynamics
- Boltzmann (1877), then complemented by Gibbs (1948), gave it a statistical mechanical content, showing that entropy of a macroscopical stationary state is proportional to the logarithm of the number *w* of possible microscopical states that correspond to this macroscopical state
- Shannon (1948) generalized the mathematical form of entropy and also explored it further. At the same time, Kolmogorov (1957) founded the concept on more mathematical grounds on the basis of the measure theory

What is entropy?

- Entropy is defined on grounds of probability theory
- For a discrete random variable X taking values x_j with probability mass function $p_j \equiv p(x_j)$, j = 1,…,w, the Boltzmann-Gibbs-Shannon (or extensive) entropy is defined as

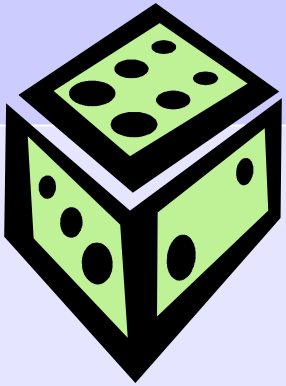
$$\varphi := E[-\ln p(X)] = -\sum_{j=1}^{w} p_j \ln p_j\,, \qquad \text{where} \sum_{j=1}^{w} p_j = 1$$

- For a continuous random variable X with probability density function $f(x)$, the entropy is defined as

$$\varphi := E[-\ln f(X)] = -\int_{-\infty}^{\infty} f(x) \ln f(x)\, dx, \quad \text{where} \int_{-\infty}^{\infty} f(x)\, dx = 1$$

- In both cases the entropy φ is a measure of **uncertainty** about X and equals the **information** gained when X is observed.
- In other disciplines (statistical mechanics, thermodynamics, dynamical systems, fluid mechanics), entropy is regarded as a measure of **order** or **disorder** and **complexity**.
- Generalizations of the entropy definition have been introduced more recently (Renyi, Tsallis)

Entropy maximization: The die example



- What is the probability that the outcome of a toss of a die will be i? (i = 1, …, 6)
- The entropy is:

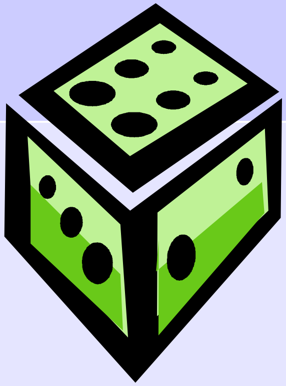
$$\varphi := E[-\ln p(X)] = -p_1 \ln p_1 - p_2 \ln p_2 - \dots - p_6 \ln p_6$$

- The equality constraint (mass preservation) is

$$p_1 + p_2 + \dots + p_6 = 1$$

- The inequality constraint is $p_i \geq 0$
- Solution of the optimization problem (e.g. by the Lagrange method) yields a single maximum: $p_1 = p_2 = \dots = p_6 = 1/6$
- This method, the application of the Maximum Entropy Principle (mathematically, an "optimation" form) is equivalent to the Principle of Insufficient Reason (Bernoulli-Laplace; mathematically, an "equation" form)

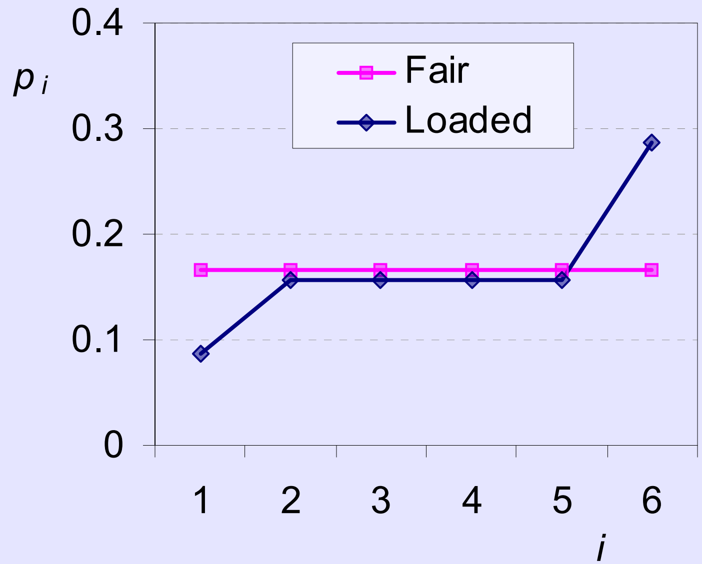
Entropy maximization: The loaded die example



- What is the probability that the outcome of a toss of a die will be i (i = 1, …, 6) if we know that it is loaded, so that $p_6 - p_1 = 0.2$?
- The IS principle does not work in this case
- The ME principle works. We simply pose an additional constraint:

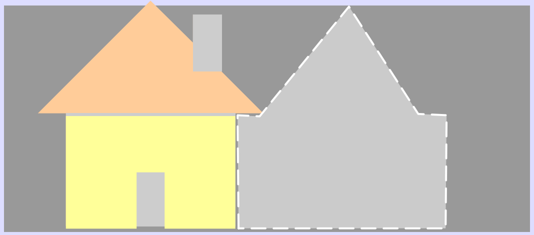
$$p_6 - p_1 = 0.2$$

- The solution of the optimization problem (e.g. by the Lagrange method) is a single maximum:



Entropy maximization: The temperature example

- What will be the temperature in my house (T_H), compared to that of the environment (T_E)? (Assume an open window and no heating equipment)
- Take a space of environment (E) in contact to the house (H) with volume equal to that of the house



- Partition the continuous range of kinetic energy of molecules into several classes i = 1 (coldest), 2, …, k (hottest)
- Denote p_i the probability that a molecule belongs to class i, and partition it to $p_{\mathrm{H}i}$ and $p_{\mathrm{E}i}$, if the molecule is in the house or the environment, respectively
- Form the entropy in terms of $p_{\mathrm{H}i}$ and $p_{\mathrm{E}i}$
- Maximize entropy conditional on $p_{\mathrm{H}i} + p_{\mathrm{E}i} = p_i$
- The result is $p_{\mathrm{H}i} = p_{\mathrm{E}i}$
- Equal number of molecules of each class are in the house and the environment, so $T_H = T_E$
- This could be obtained also from the IR principle

Formalization of the principle of maximum entropy

- In a probabilistic context, the principle of ME was introduced by Janes (1957)
- In a probabilistic context, the principle of ME is used to infer unknown probabilities from known information
- In a physical context, it determines thermodynamical states
- The principle postulates that the entropy of a random variable should be at maximum, under some conditions, formulated as constraints, which incorporate the information that is given about this variable
- Typical constraints used in a probabilistic or physical context are:

Mass

Mean/Momentum

Non-negativity

$$\int_{-\infty}^{\infty} f(x)\, dx = 1, \quad E[X] = \int_{-\infty}^{\infty} x\, f(x)\, dx = \mu \qquad x \ge 0$$

Variance/Energy

Dependence/Stress

$$E[X^2] = \int_{-\infty}^{\infty} x^2 f(x)\, dx = \sigma^2 + \mu^2, \; E[X_i\, X_{i+1}] = \int_{-\infty}^{\infty} x_i\, x_{i+1}\, f(x_i, x_{i+1})\, dx_i\, dx_{i+1} = \rho\, \sigma^2 + \mu^2$$

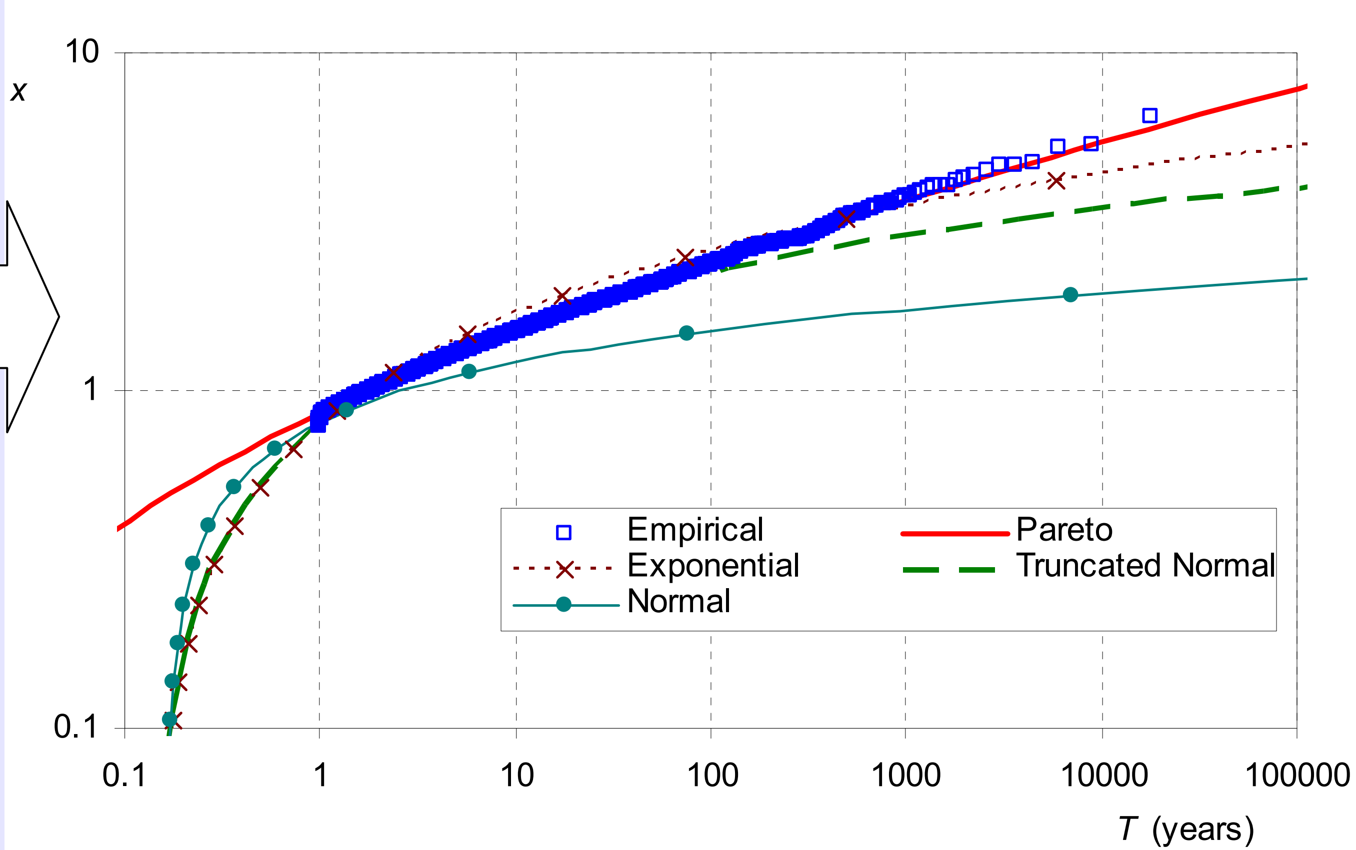
Some results of ME interesting to hydrology

- Assume that a hydrometeorological variable *X* (e.g. temperature, rainfall, runoff) is continuous and positive, has known mean μ and known variation σ/μ. Estimate the distribution function with only this information, applying the ME principle
- The results are:
  - Maximum entropy + Low variation → (Truncated) normal distribution
  - Maximum entropy + High variation → Power-type (Pareto) distribution
  - Maximum entropy + High variation + High return periods → State scaling
- The celebrated state scaling ($x_T \sim T^{\kappa}$, where T is the return period and x_T the corresponding quantile) is only:
  - a consequence of the ME principle,
  - an approximation, good for high return periods and for variables with high variation
- Real world time series (especially long ones) validate the applicability of the ME principle in hydrometeorological processes

ME application to extreme daily rainfall worldwide

Data set: Daily rainfall from 168 stations worldwide each having at least 100 years of measurements; series above threshold, standardized by mean and unified; period 1822-2002; 17922 station-years of data

$\mu = 0.28$
(mean minus threshold)
$\sigma/\mu = 1.19 > 1$
ME distribution:
Pareto
$\kappa = 0.15$
$\varphi_q = 1.16$



Conclusion:
Scaling
for $T >$ ~50 yr

Entropic quantities of a stochastic process

- The *order 1 entropy* (or simply *entropy* or *unconditional entropy*) refers to the marginal distribution of the process X_i:

$$\varphi := E[-\ln f(X_i)] = -\int_{-\infty}^{\infty} f(x) \ln f(x)\, dx, \quad \text{where} \int_{-\infty}^{\infty} f(x)\, dx = 1$$

- The *order n entropy* refers to the joint distribution of the vector of variables $\mathbf{X}_n = (X_1, \ldots, X_n)$ taking values $\mathbf{x}_n = (x_1, \ldots, x_n)$:

$$\varphi_n := E[-\ln f(\mathbf{X}_n)] = -\int_{D_n} f(\mathbf{x}_n) \ln f(\mathbf{x}_n)\, d\mathbf{x}_n$$

- The *order m conditional entropy* refers to the distribution of a future variable (for one time step ahead) conditional on known *m* past and present variables (Papoulis, 1991):

$$\varphi_{\mathrm{c},m} := E[-\ln f(X_1 | X_0, \ldots, X_{-m+1})] = \varphi_m - \varphi_{m-1}$$

- The *conditional entropy* refers to the case where the entire past is observed:

$$\varphi_{\mathrm{c}} := \lim_{m \to \infty} \varphi_{\mathrm{c},m}$$

- The *information gain* when present and past are observed is:

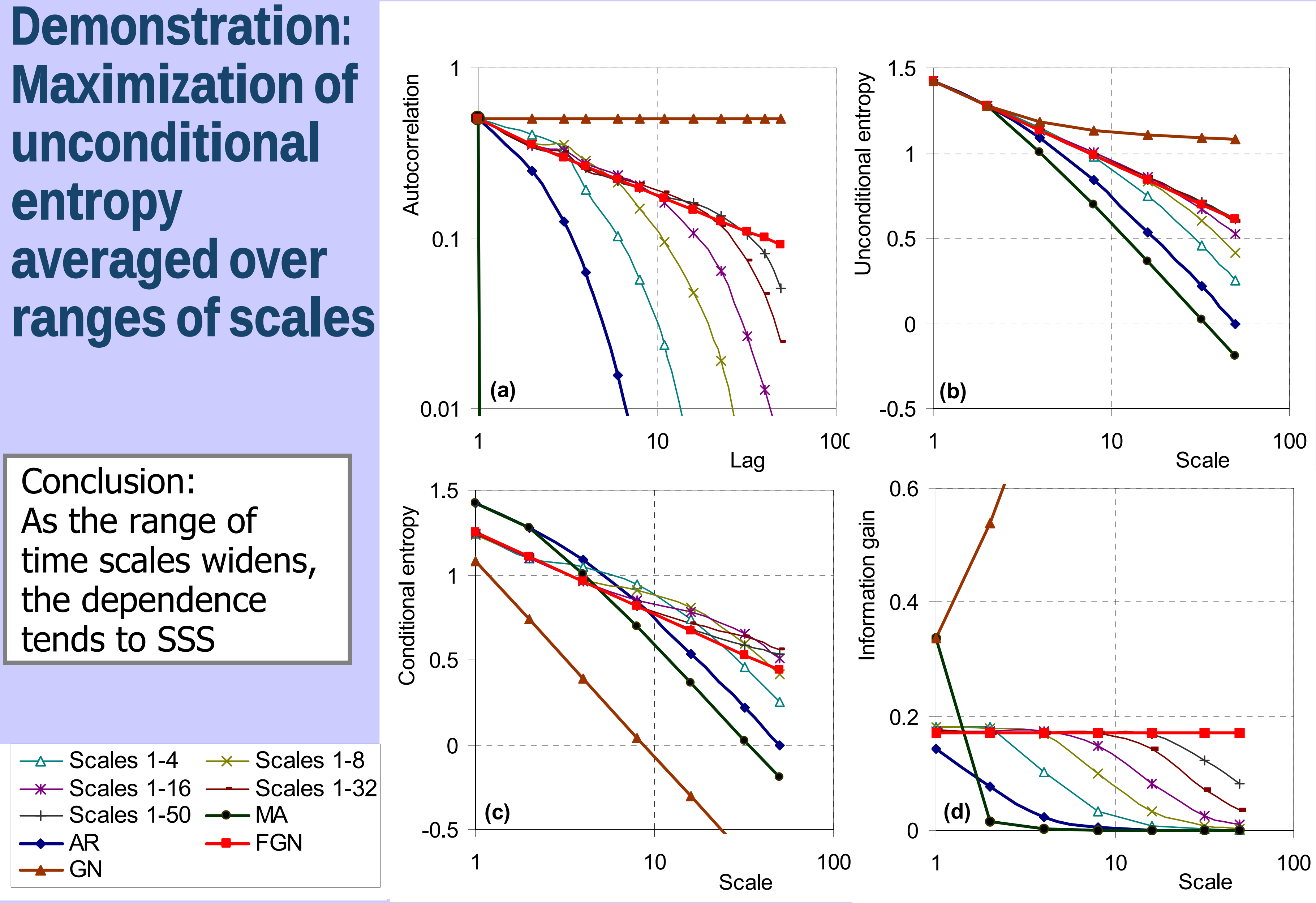
$$\psi := \varphi - \varphi_{\mathrm{c}}$$

Note: notation assumes stationarity

Entropy maximization for a stochastic process

- The purpose is to determine not only marginal probabilities but the dependence structure as well
- All five constrains are used (mass/mean/variance/dependence/non-negativity)
- The lag one autocorrelation (used in the dependence constraint) is determined for the basic (annual) scale but the entropy maximization is done on other scales as well
- The variation is low ($\sigma/\mu << 1$) and thus the process is virtually Gaussian (intermediate result). This is valid for annual and over-annual time scales
- For a Gaussian process the nth order entropy is given as $\varphi_n = \ln\sqrt{(2\pi e)^n \delta_n}$ where δ_n is the determinant of the autocovariance matrix $c_n := \mathrm{Cov}[\mathbf{X}_n, \mathbf{X}_n]$.
- The autocovariance function is assumed unknown to be determined by application of the ME principle. Additional constraints for this are:
  - Mathematical feasibility, i.e. positive definiteness of c_n (positive δ_n)
  - Physical feasibility, i.e. autocorrelation function (a) positive and (b) non increasing with lag and time scale
    (Note: periodicity that may result in negative autocorrelations is not considered here due to annual and over-annual time scales)

Demonstration: Maximization of unconditional entropy averaged over ranges of scales



Conclusion:
As the range of time scales widens, the dependence tends to SSS

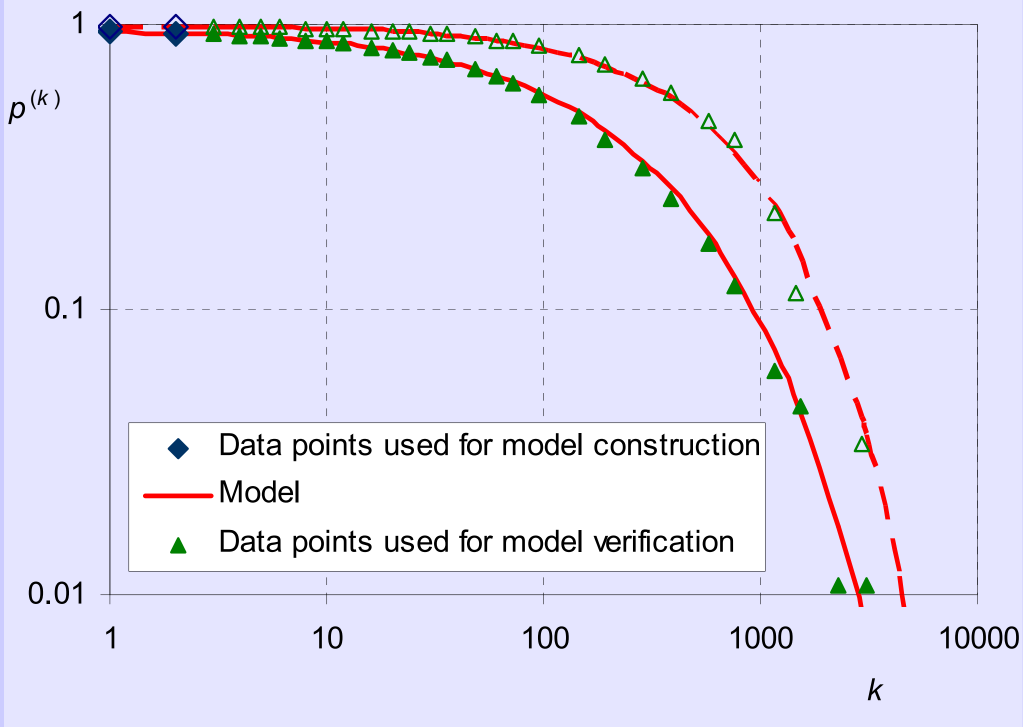
Results of the ME principle in stochastic processes

- Maximum entropy + Low variation + Dominance of a single time scale → Normal distribution + Time independence
- Maximum entropy + Low variation + Time dependence + Dominance of a single time scale → Normal distribution + Markovian (short-range) time dependence
- Maximum entropy + Low variation + Time dependence + Equal importance of time scales → Normal distribution + Time scaling (long-range dependence / Hurst phenomenon)
- The time scaling behavior is a result of the principle of maximum entropy
- The omnipresence of time scaling in numerous long hydrologic time series, validates the applicability of the ME principle

Another peculiar dependence explained by ME

- Rainfall at small scales is intermittent
- The dependence of the rainfall occurrence process is not Markovian neither scaling but in between; it has been known as clustering or overdispersion
- The models used for the rainfall occurrence process (point processes) are essentially those describing clustering of stars and galaxies
- The ME principle applied with the binary state rainfall process in more or less the same way as in the continuous state process explains this dependence

Probability $p^{(k)}$ that an interval of k hours is dry, as estimated from the Athens rainfall data set and predicted by the model of maximum entropy for the entire year (full triangles and full line) and the dry season (empty triangles and dashed line)



Interpretation of results

- The successful application of the ME principle in nature offers an explanation for of a plethora of phenomena (e.g. thermodynamic) and statistical behaviors including
  - the emergence of normal distribution, in many (but not all) cases
  - the scaling behavior in state, in other cases
  - the scaling behavior in time
  - the clustering behavior in rainfall occurrence
- This can be interpreted as dominance of uncertainty or ignorance in nature
- It harmonizes with the Socratic view: «Ἓν οἶδα, ὅτι οὐδέν οἶδα» (One I know, that I know nothing)
- This view was not a confession of modesty – Socrates regarded the knowledge of ignorance as a matter of supremacy
- In this respect, the knowledge of the dominance of uncertainty can assist to safer design and management of hydrosystems

2. Hydrologic issues
2C. Operational synthesis

Stochastic simulation/forecasting of hydrologic processes

- Question: Why simulated series?
- Answer:
  - Analytical solutions for a hydrosystem as complex as that of Athens are not feasible or would assume oversimplification of the system
  - Of numerical methods, Monte Carlo simulation (stochastic simulation) is the most convenient
  - Detailed inflow and other (rainfall, evaporation) hydrologic series are needed at many sites simultaneously and at several time scales for Monte Carlo simulation the hydrosystem
  - The acceptable failure probability level for Athens is of the order of 10^{-2}: one failure in 100 years on the average
  - For a reasonable estimation error in the failure probability we need 1000-10 000 years of data
  - Historic hydrologic records are too short

Requirements for stochastic simulation

1. Multivariate model
2. Multiple time scales of operation: annual to monthly or sub-monthly
3. Multiple time scales of preservation: multi-year (reproduction of the Hurst phenomenon) to sub-monthly (reproduction of sub-annual periodicity)
4. Preservation of essential marginal statistics up to third order (skewness)
5. Preservation of joint second order statistics
   - autocorrelations of any type and any lag
   - concurrent cross-correlations
6. Parsimony of parameters
7. Performance in simulation mode (steady state simulations) and in forecast mode, given the current and historic values (terminating simulations)

Models with such features did not exist (particularly, the ARMA type models were not useful)

Stochastic simulation strategy

- Stage 1: Generate annual time series
  - Use a parsimonious model yet capable of describing over-annual scaling
  - No need to describe sub-annual periodicity
- Stage 2: Disaggregate the annual into sub-annual time series
  - Use a parsimonious model structure such as PAR(1)
  - Couple it to the annual model
  - So, no need to describe over-annual scaling explicitly
- A one stage procedure to handle over-annual and sub-annual properties simultaneously has also been studied but not implemented operationally so far

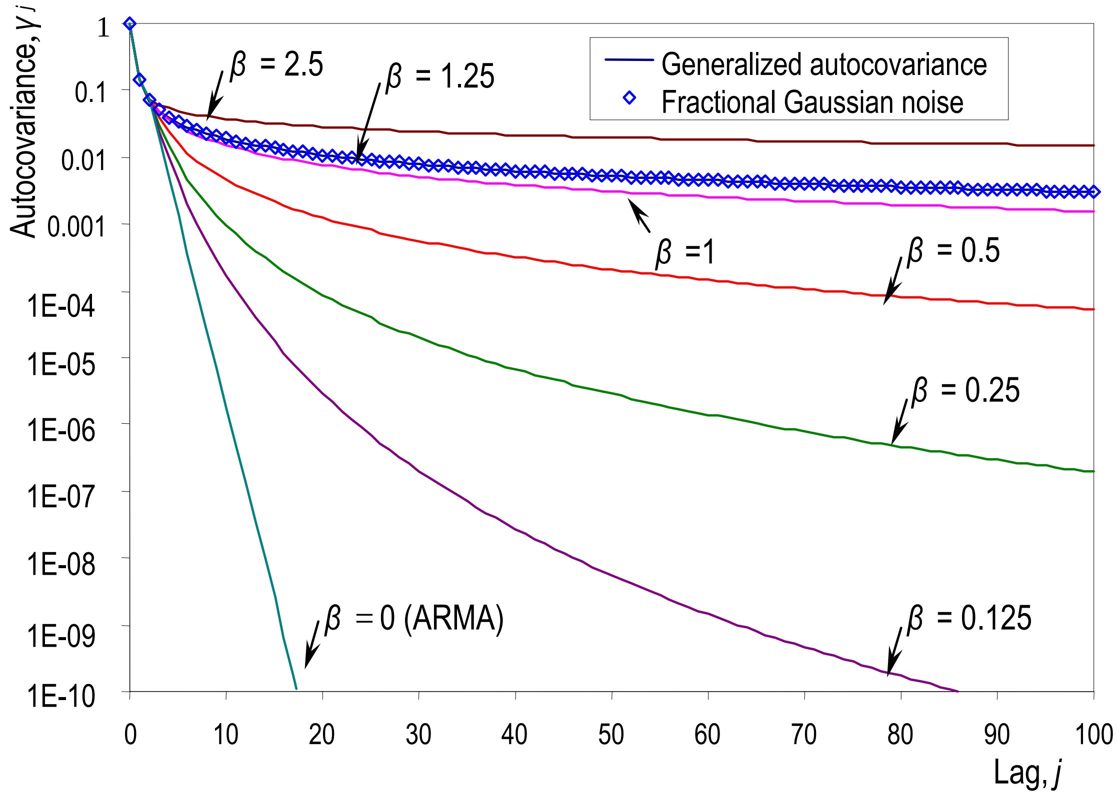
Annual model: The generalized autocovariance function (GAS)

- General GAS expression

  $$\gamma_j = \gamma_0\,(1 + \kappa\,\beta\,|j|^{\alpha})^{-1/\beta}$$

  where γ_j: lag j autocovariance; γ_0: variance; κ, α, β: parameters
- Fittings options
  - Optimize parameters to best fit historic autocorrelograms
  - Preserve explicitly γ_1, γ_2 and Hurst exponent
  - Explicit preservation of more γ_i is also possible
- GAS behavior
  - For $\beta = 0 \Rightarrow$ ARMA: $\gamma_j = \gamma_0 \exp(-\kappa\,|j|^{\alpha})$
  - For $\kappa = (1/\beta)\,(1 - 1/\beta)^{-\beta}\,(1 - 1/2\beta)^{-\beta}$ and $\alpha = 1 \Rightarrow$ FGN

Demonstration of GAS for α = 1 and several values of β



Annual model: Generalized generating scheme for any covariance structure

Typical (backward) moving average (**BMA**) scheme: $X_i = \ldots + a_1 V_{i-1} + a_0 V_i$ where V_i independent random variables and a_i numerical coefficients

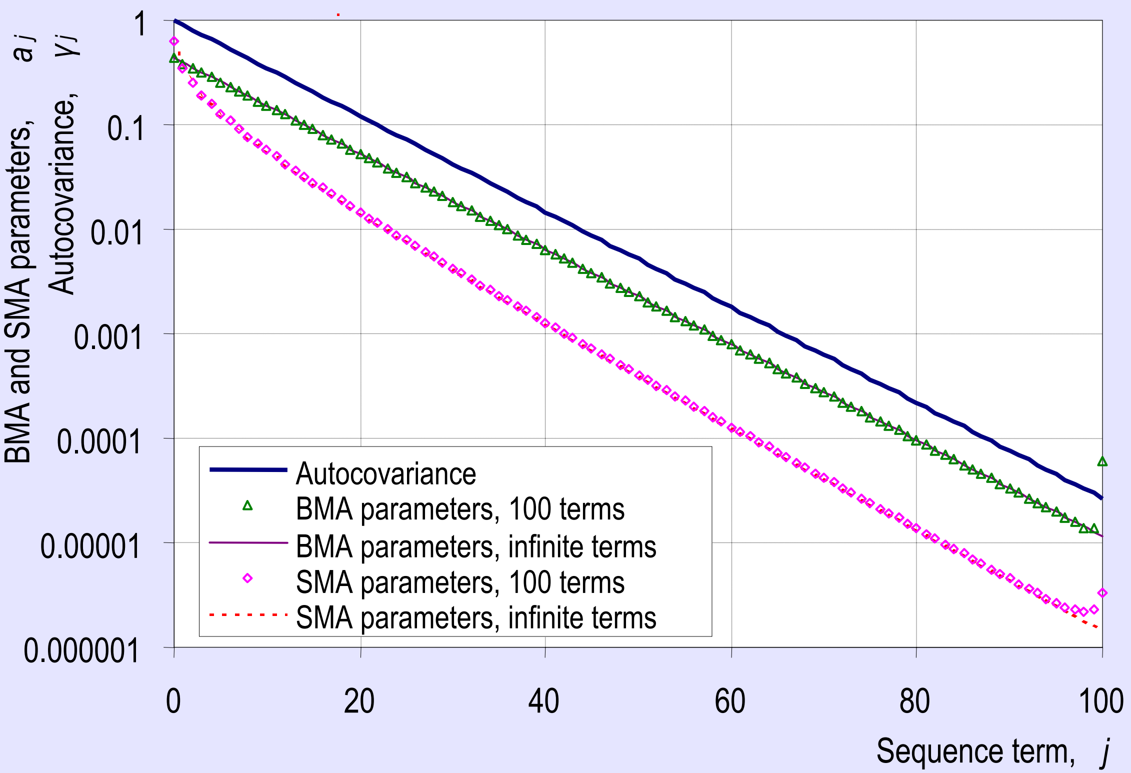
Symmetric moving average (**SMA**) scheme $X_i = \ldots + a_1 V_{i-1} + a_0 V_i + a_1 V_{i+1} + \ldots$

SMA has several advantages over BMA. Among them, it allows a closed solution for a_i:

$s_a(\omega) = [2\, s_\gamma(\omega)]^{1/2}$

where $s_a(\omega)$ and $s_\gamma(\omega)$ the Discrete Fourier Transforms of the series a_j and γ_j, respectively.

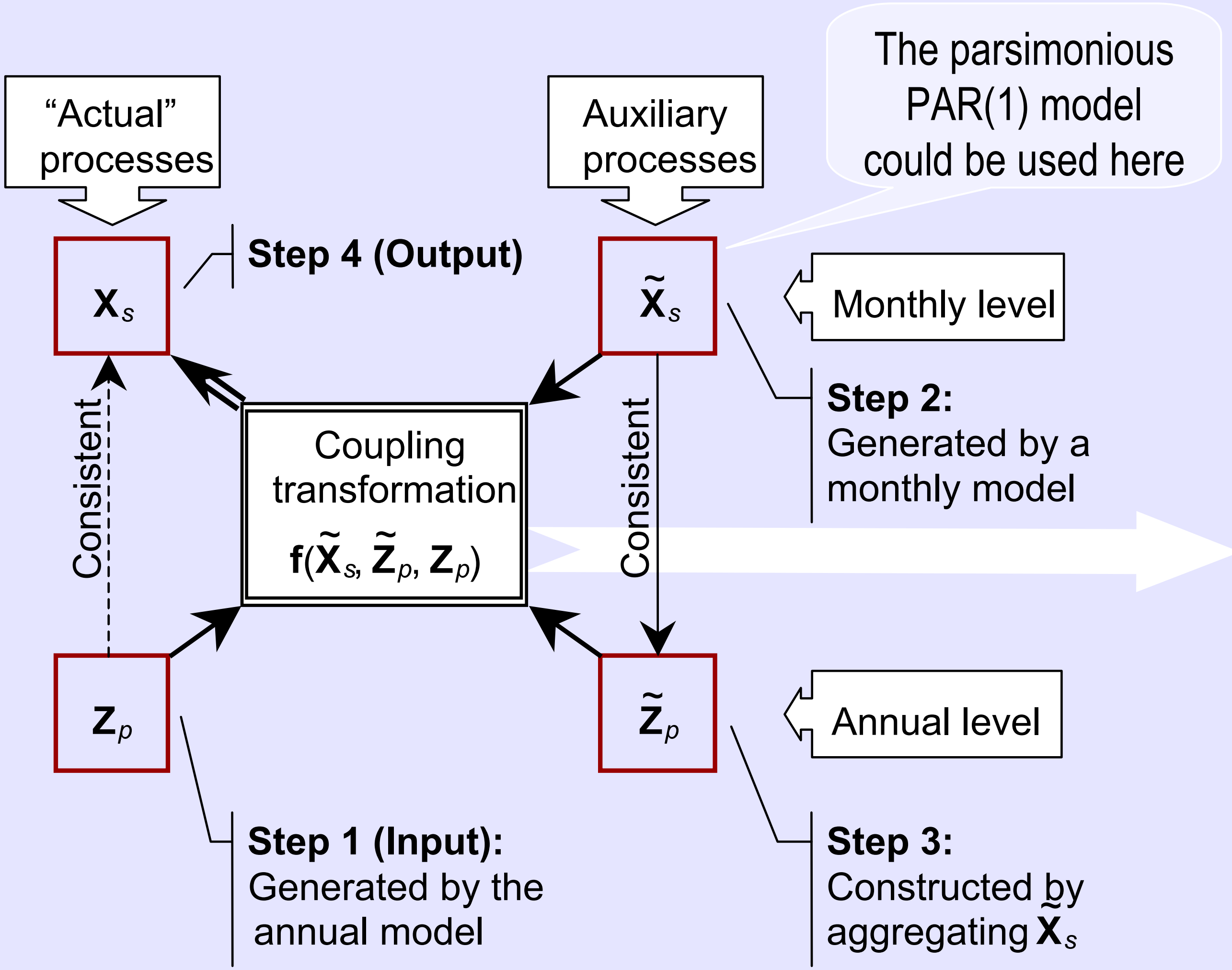
Both schemes are applicable for multivariate problems



Annual model: Stochastic simulation in forecast mode

- In forecast mode, the observed present and past values must condition the hydrologic time series of the future
- This is attainable using a two-step algorithm
  1. Generate future time series without reference to the known present and past values
  2. Adjust future time series using the known present and past values and a linear adjusting algorithm
- The linear adjusting algorithm:
  1. is expressed in terms of covariances among variables
  2. preserves exactly means, variances and covariances
  3. is easily implemented

Coupling stochastic models of different time scales



The linear transformation

$$\mathbf{X}_s = \tilde{\mathbf{X}}_s + \mathbf{h}\,(\mathbf{Z}_p - \tilde{\mathbf{Z}}_p)$$

where

$$\mathbf{h} = \mathrm{Cov}[\mathbf{X}_s, \mathbf{Z}_p] \cdot \{\mathrm{Cov}[\mathbf{Z}_p, \mathbf{Z}_p]\}^{-1}$$

preserves the vectors of means, the variance-covariance matrix and any linear relationship that holds among $\mathbf{X}_s$ and $\mathbf{Z}_p$.

Handling of skewness in multivariate problems: Optimized decomposition of covariance matrices

- Consider any linear multivariate stochastic model of the form

$$\mathbf{Y} = \mathbf{a}\,\mathbf{Z} + \mathbf{b}\,\mathbf{V}$$

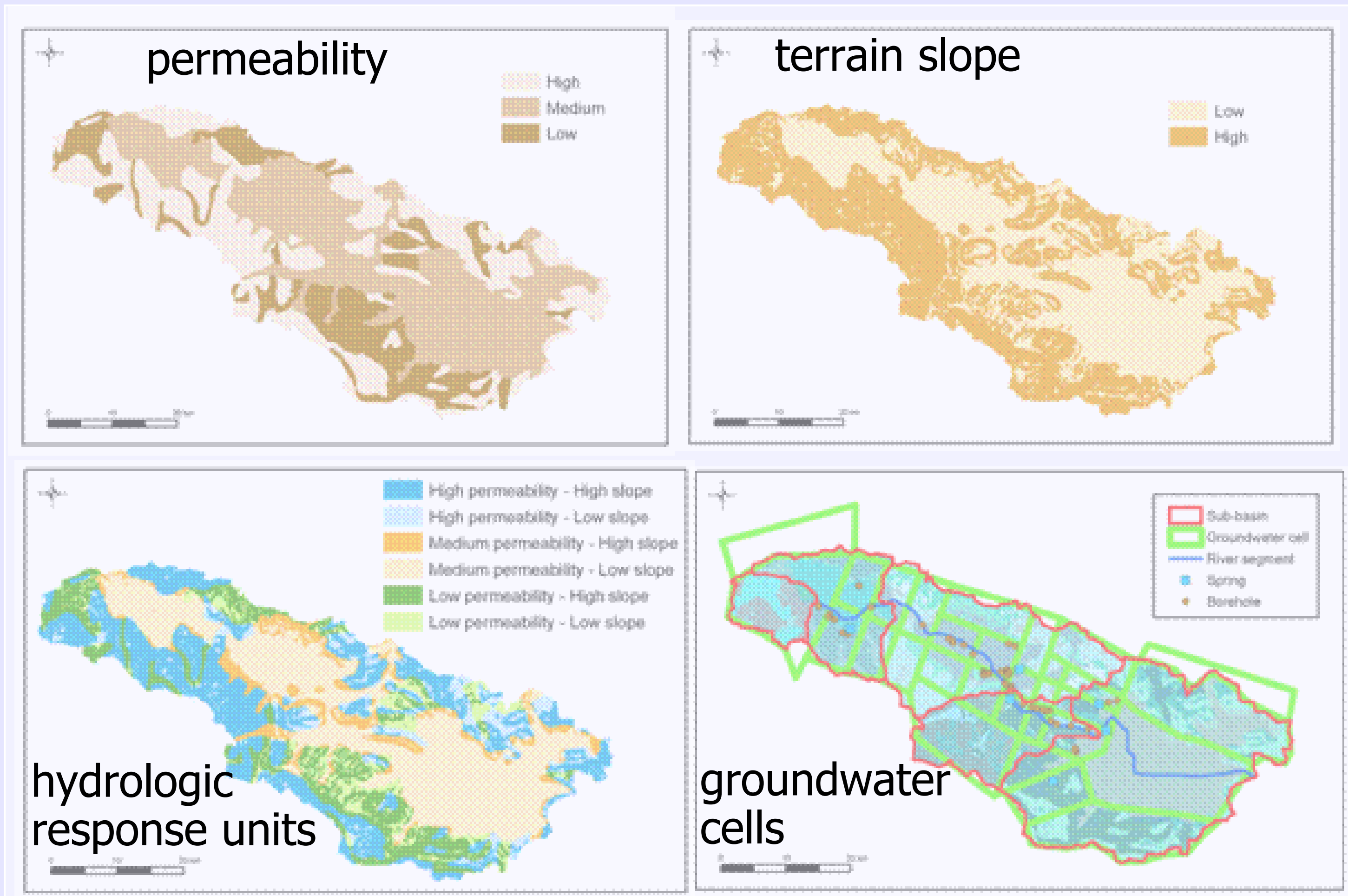
  where **Y**: vector of variables to be generated, **Z**: vector of variables with known values, **V**: vector of innovations, and **a** and **b**: matrices of parameters
- The parameter matrix **b** is related to a covariance matrix **c** by

$$\mathbf{b}\,\mathbf{b}^T = \mathbf{c}$$

- This equation may have infinite solutions or no solution (if **c** is not positive definite)
- The skewness coefficients **ξ** of innovations **V** depend on **b**
- The smaller the values of **ξ**, the more attainable the preservation of the skewness coefficients of the actual variables **Y**
- Therefore, the problem of determination of **b** can be seen as an optimization problem that combines
  - minimization of skewness **ξ**, and
  - minimization of the error $\|\mathbf{b}\,\mathbf{b}^T - \mathbf{c}\|$
- A fast optimisation algorithm has been developed for this problem
- The algorithm works even for **c** that are not positive definite

Models developed are not only stochastic ...

In the Boeoticos Kephisos River basin a hydrologic model of the entire hydrologic cycle had to be developed, which was demanding due to the extended karstic activity and the intensive withdrawals for irrigation



3. Hydrosystem operation issues
Parameterization – Simulation – Optimization

Typical problems to be answered

- Find the **maximum possible annual release** from the system:
  - for a certain (acceptable) **reliability level** (steady state conditions)
  - for a certain **combination of the system components** (e.g. primary resources)

  and determine the corresponding:
  - **optimal operation policy** (storage allocation; conveyance allocation; pumping operation)
  - **cost** (in terms of energy; economy; other impacts)
- Find the **minimum total cost**
  - for a given **water demand** (less than the maximum possible annual release)
  - for a certain (acceptable) **reliability level**

  and determine the corresponding:
  - **combination of the system components** to be enabled
  - **optimal operation policy** (storage allocation; conveyance allocation; pumping operation)
  - **alternative operation policies** (that can satisfy the demand but with higher cost)

Categories of problems

- Steady state problems for the current hydrosystem
  - (e.g., previous slide)
- Problems involving time
  - Availability of water resources in the months to come
  - Impact of a management practice to the future availability of water resources
  - Evolution of the operation policy for a temporally varying demand
- Investigation of scenarios
  - Hydrosystem structure: Impacts of new components (aqueducts, pumping stations etc.)
  - Demand: Feasibility of expansion of domain
  - Hydroclimatic inputs: Climate change
- Adequacy/safety under exceptional events – Required measures
  - Damages
  - Special demand occasions (e.g. 2004 Olympic Games)

The methodology: General aspects

Question 1: Simulation **or** optimization?

- Simulation **versus** optimization (water resources literature and practice)
- Simulation methods **for** optimization (more mathematical literature)

Answer: Optimization coupled with simulation

Main advantages
- Determination of optimal policies
- Incorporation of mathematical optimization techniques

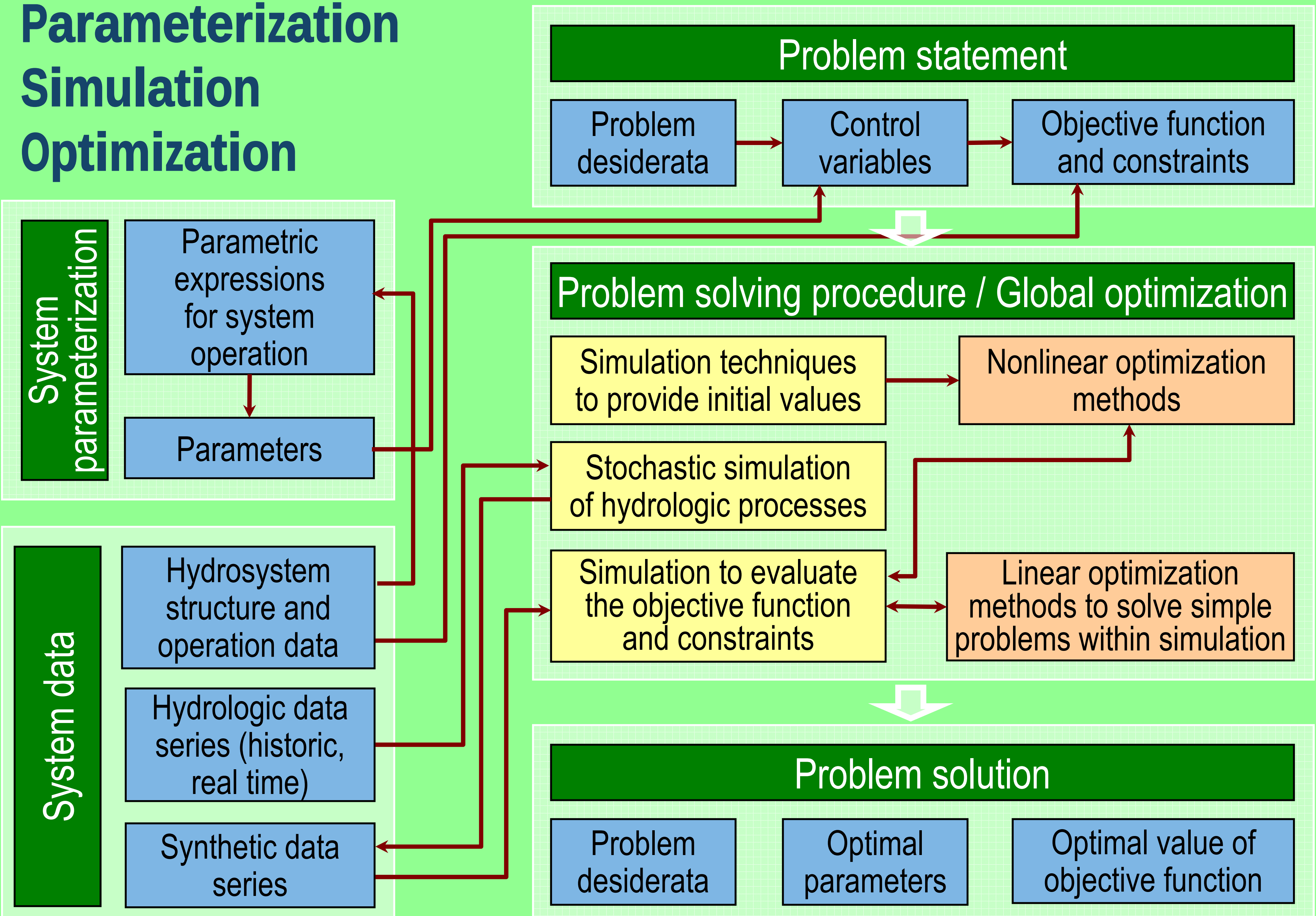
Main advantages
- Detailed and faithful system representation
- Better understanding of the system operation
- Incorporation of stochastic models

Question 2: Which are the control (decision) variables?

- Typically: Releases from system components in each time step

Answer: Introduction of **parametric control rules** with few **parameters** as control variables

Parameterization
Simulation
Optimization



{12,13,15,17}

Introduction to the parametric reservoir operation rule – Some analytical solutions

Maximize release from a simple reservoir system with single water use

- Case a: no conveyance restrictions; no leakages
  - Solution: Probability of spill equal at all reservoirs (**New York Rule;** Clark, 1950)
  - Under certain (rather common) conditions about the distribution of inflows:
- Case b: no conveyance restrictions; significant leakages; insignificant spills
  - Solution:
- Case c: restricted conveyance capacity; insignificant spills; no leakages
  - Solution:

Space rule
(Bower et al., 1962)

$$\frac{K_i - S_i}{E[CQ_i]} = \frac{\sum K - V}{\sum E[CQ]}$$

Leakage rule (Nalbantis & Koutsoyiannis, 1997)

$$S_i = \begin{cases} V & \text{for one reservoir} \\ 0 & \text{for all others} \end{cases}$$

Conveyance rule (Nalbantis & Koutsoyiannis, 1997)

$$\frac{S_i}{C_i} = \frac{V}{\sum C}$$

Notation: i = Reservoir index, K = Storage capacity, S = Storage, $V = \sum S$, CQ = Cumulative inflow, $E[\,]$ = expectation, C = Conveyance capacity

Formulation of the parametric reservoir operation rule

Initial linear parametric form

$S_i^* = a_i + b_i V$ (parameters a_i, b_i)

subject to $\Sigma a_i = 0$, $\Sigma b_i = 1$,

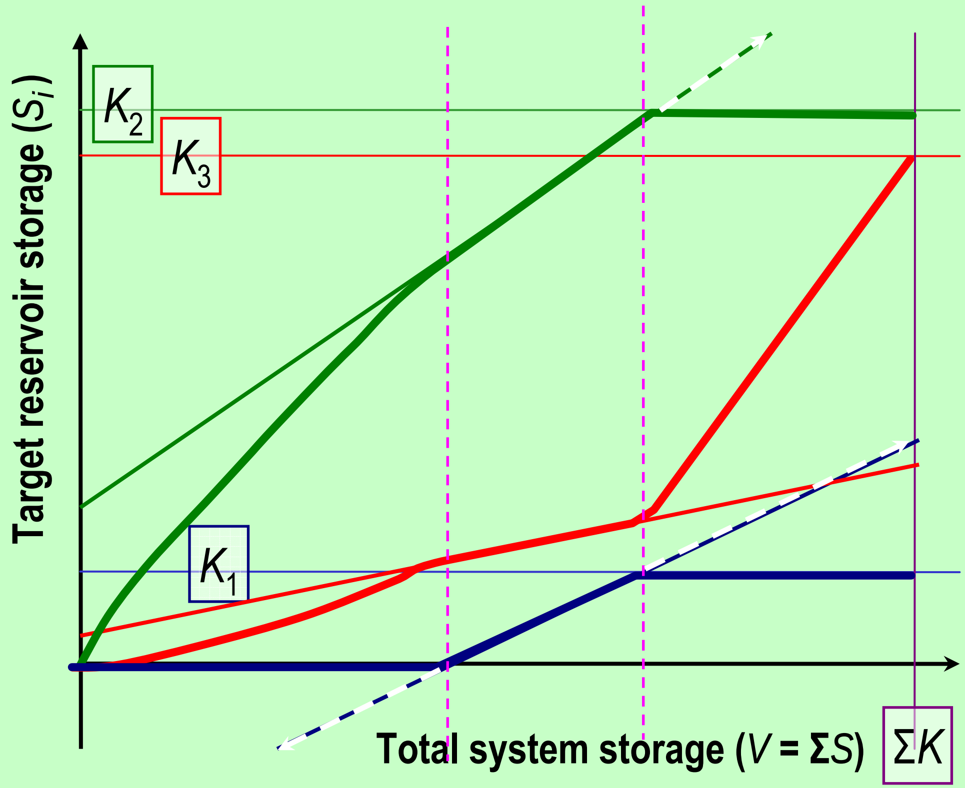
since $\Sigma S_i^* = V$

Corrected for physical constraints

$$S_i'^* = \begin{cases} 0 & a_i + b_i V < 0 \\ a_i + b_i V & 0 \le a_i + b_i V \le K_i \\ K_i & a_i + b_i V > K_i \end{cases}$$

Adjusted, nonlinear form

$$S_i''^* = S_i'^* + \frac{S_i'^*(1 - S_i'^*/K_i)}{\Sigma S_i'^*(1 - S_i'^*/K_i)}(V - \Sigma S_i'^*)$$



Two parameters per reservoir (a_i, b_i) = **Control variables**

Parameter values **determined by optimization** – depending on the objective function

Parameters may depend also on season (e.g., refilling-emptying period, or months)

2 × (reservoirs – 1) × seasons total parameters for the reservoir system

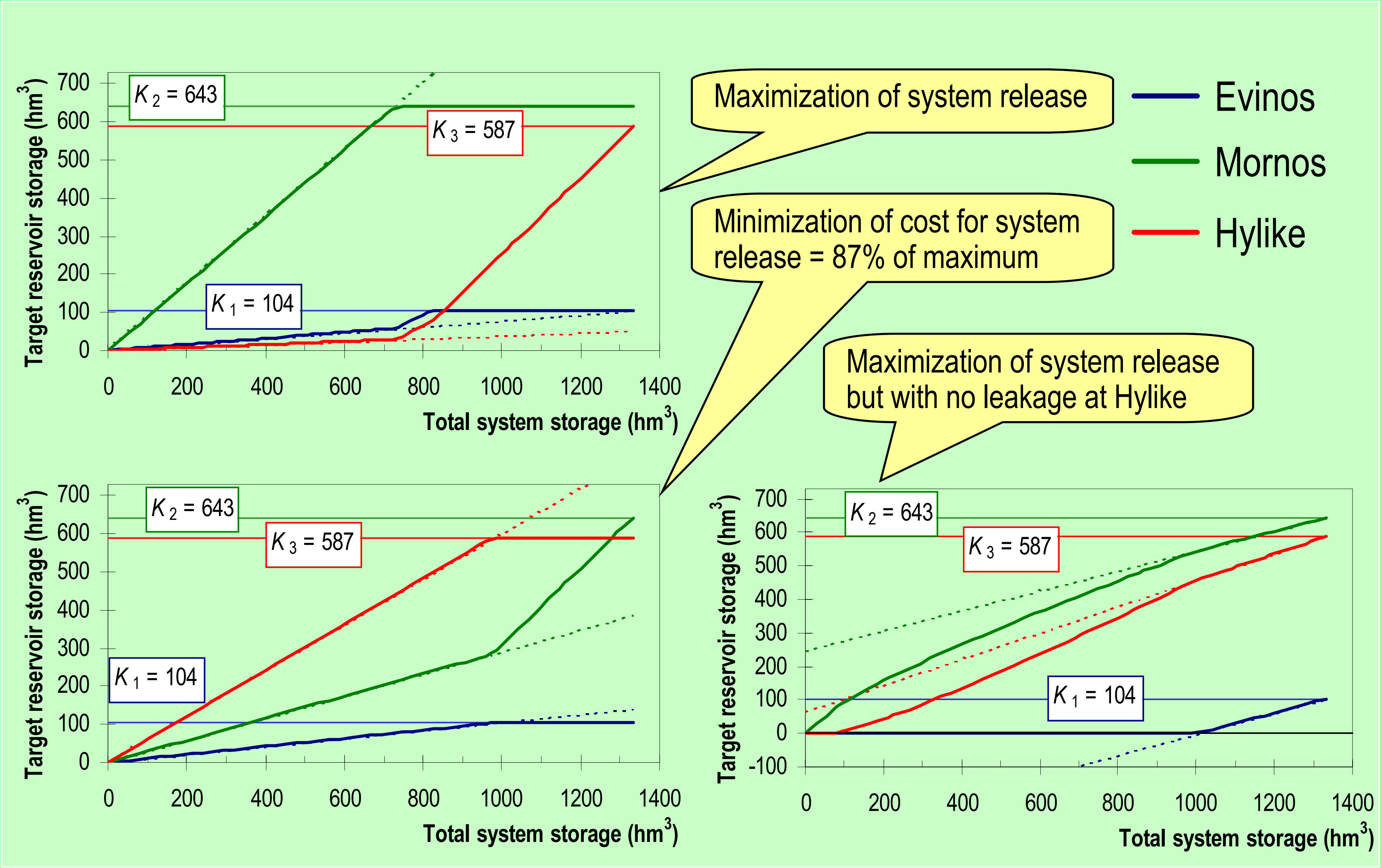
A comparison with non-parametric optimization

Problem: Find the maximum release that can be ensured by a system of **3 reservoirs** with **reliability 99%** (probability of failure 1%). Use **1000 years** of simulated data with **monthly time step**. Assume **steady state** conditions.

Non-parametric optimization	Parametric rule based optimization
Number of control variables: 1000 × 12 monthly releases × (3 – 1) reservoirs + 1 (problem target) = 24001	Number of control variables: 2 parameters/reservoir/ season × (3 – 1) reservoirs × 2 seasons + 1 (problem target) = 9 (as an order of magnitude)
Cannot be combined with simulation All physical constraints of the system must be entered as problem constraints	Can be combined with simulation Physical constraints of the system are handled by the simulation model
Control variables depend on inflow series Implicit assumption of known inflows (perfect foresight)	Control variables do not depend on inflow series but on their statistical properties No assumption of known inflows
The optimization model needs continuous runs with updated data	Once parameters are optimized, the system can be operated without running the model

{15,21}

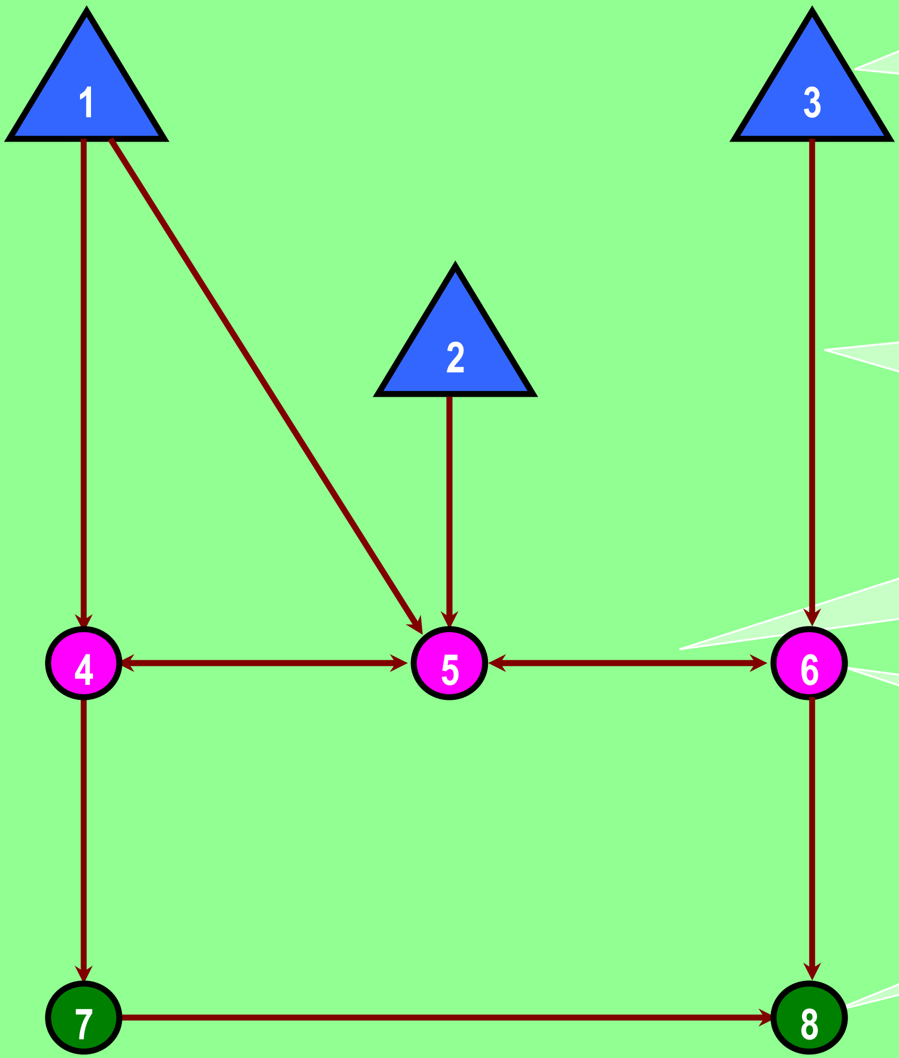
Application of the parametric rule – Optimal results



Considering the complete hydrosystem – Simulation

- Assuming that parameters a_i and b_i are known, the **target releases** from each reservoir will be also known in the beginning of each simulation time step
- The **actual releases** depend on several attributes of the hydrosystem (physical constraints)
- Their estimation is done using **simulation**
- Within simulation, an **internal optimization** procedure may be necessary (typically linear, nonparametric)
- Because parameters a_i and b_i are not known, but rather are to be optimized, simulation is driven by an **external optimization** procedure (nonlinear)

Hydrosystem components and attributes



Reservoir
Storage, *S*
Target release, *R*

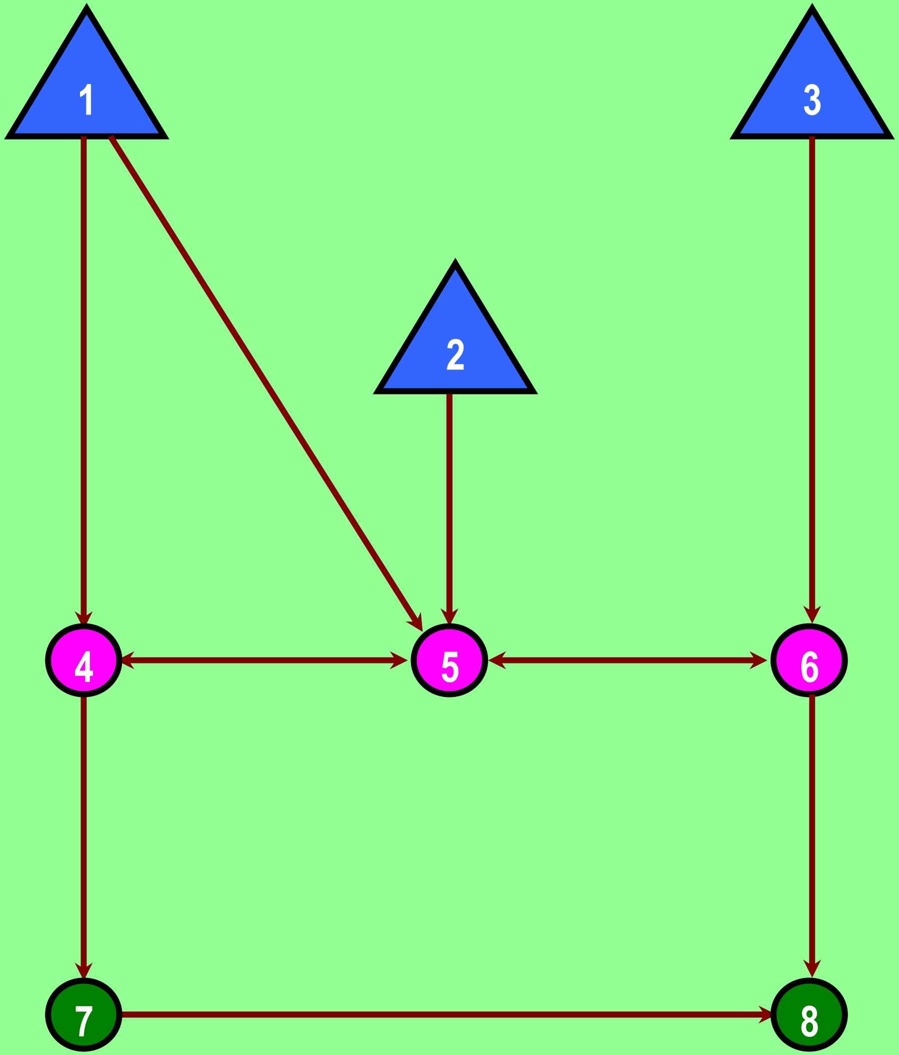
One direction aqueduct
Conveyance capacity, *C*
Unit conveyance cost, *u*

Two direction aqueduct
Conveyance capacities, C_{for}, C_{rev}
Unit conveyance costs, u_{for}, u_{rev}

Junction

Consumption point
Demand, *D*

Conveyance problem formulation



Given:

- Demands (D)
- Reservoir storages (S),
- Reservoir target releases ($R \leq S$; $\Sigma R = \Sigma D$; from parametric rule)

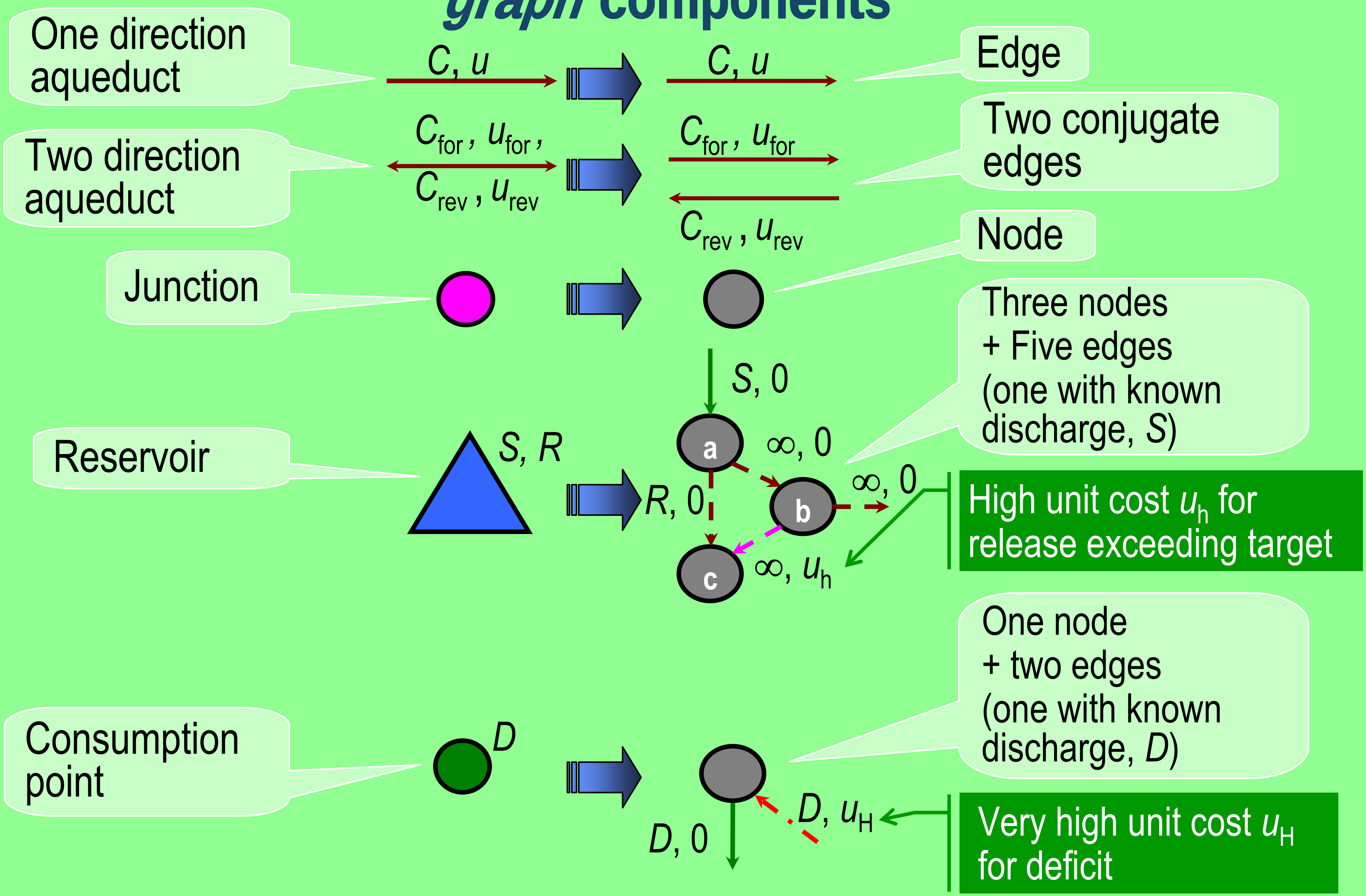
Required:

- Actual (feasible) consumptions (at consumption points)
- Actual (feasible) releases (from reservoirs)
- Aqueduct discharges
- Conveyance cost

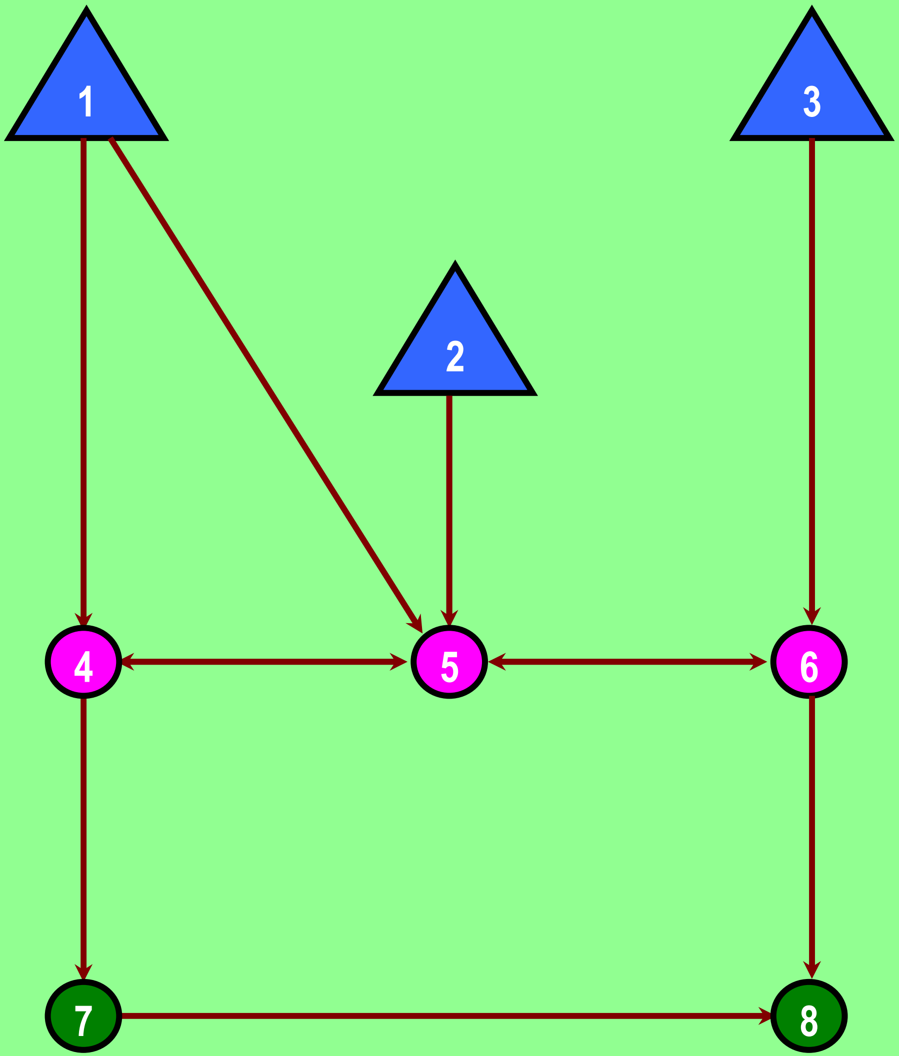
Conditions:

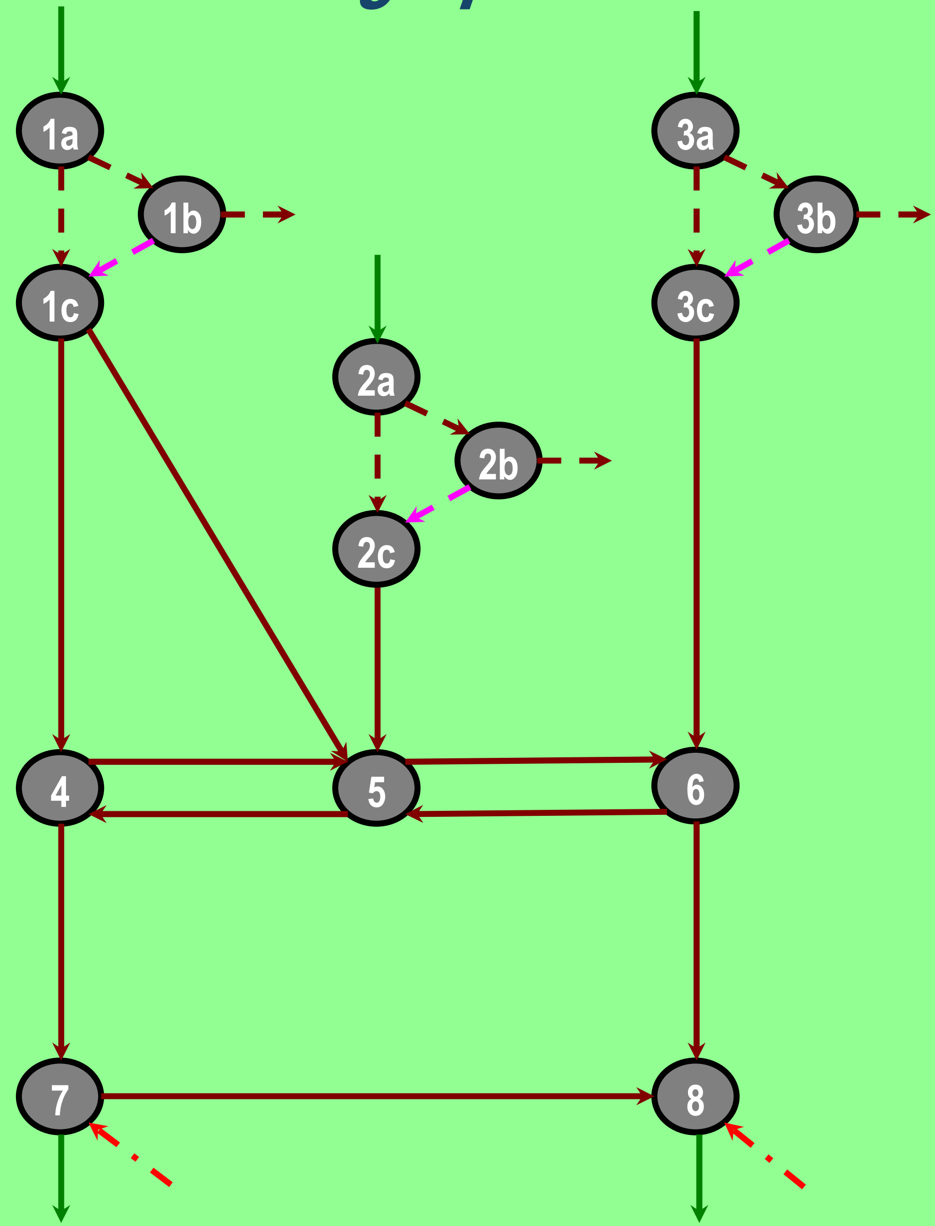
- If possible, no deficits at consumption points
- If possible, releases from reservoirs equal to target releases
- Minimum conveyance cost

Transformations of hydrosystem components to *graph* components

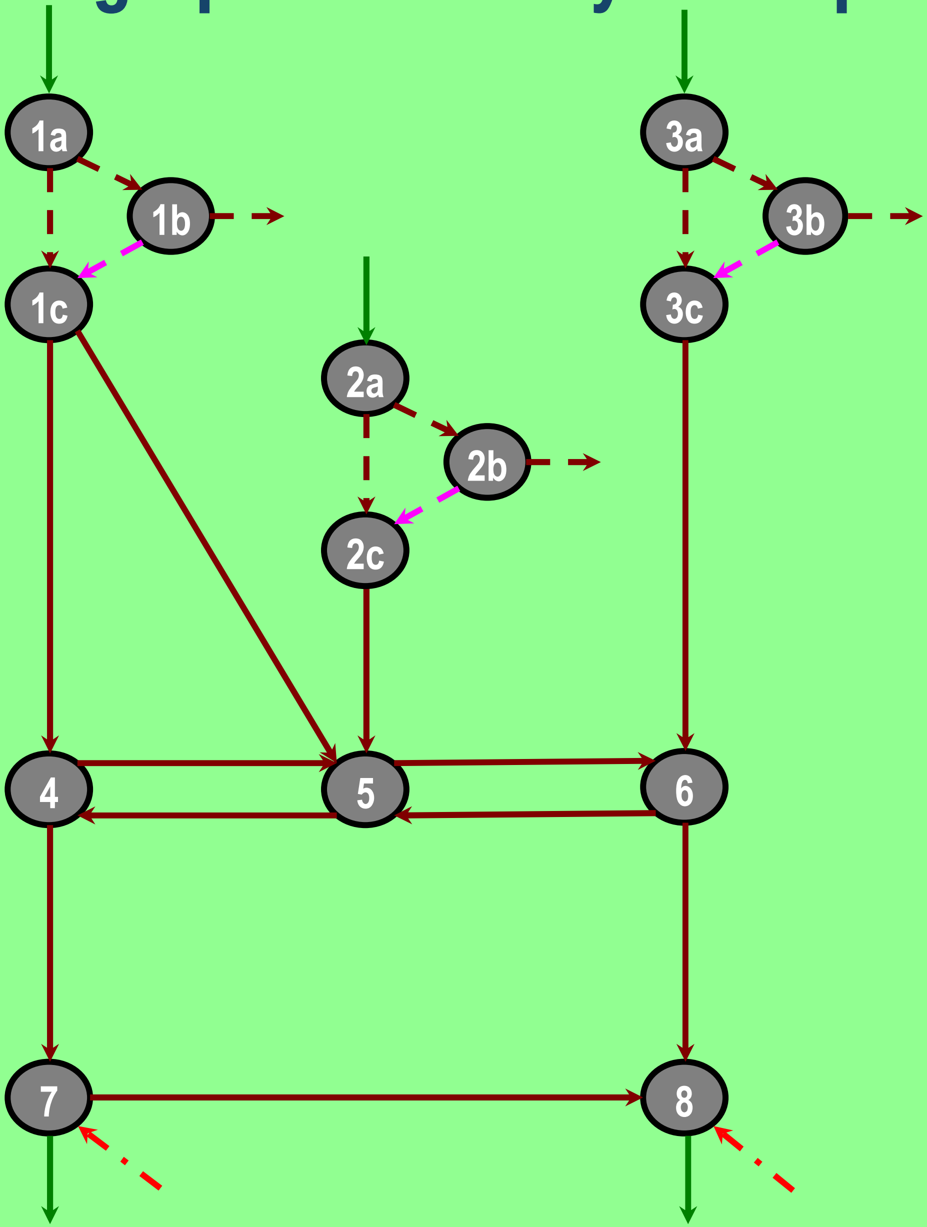


Hydrosystem and its transformation to *digraph*





Digraph solution by linear programming



Determine all unknown discharges Q_{ij} at edges *ij*, by **minimizing total cost**

$$TC = \Sigma_{ij}\, u_{ij}\, Q_{ij}$$

subject to equality constraints for each node *i*

$$\Sigma_j\, Q_{ij} - \Sigma_j\, Q_{ji} = 0$$

and to **inequality constraints** for each edge *ij*

$$0 \le Q_{ij} \le C_{ij}$$

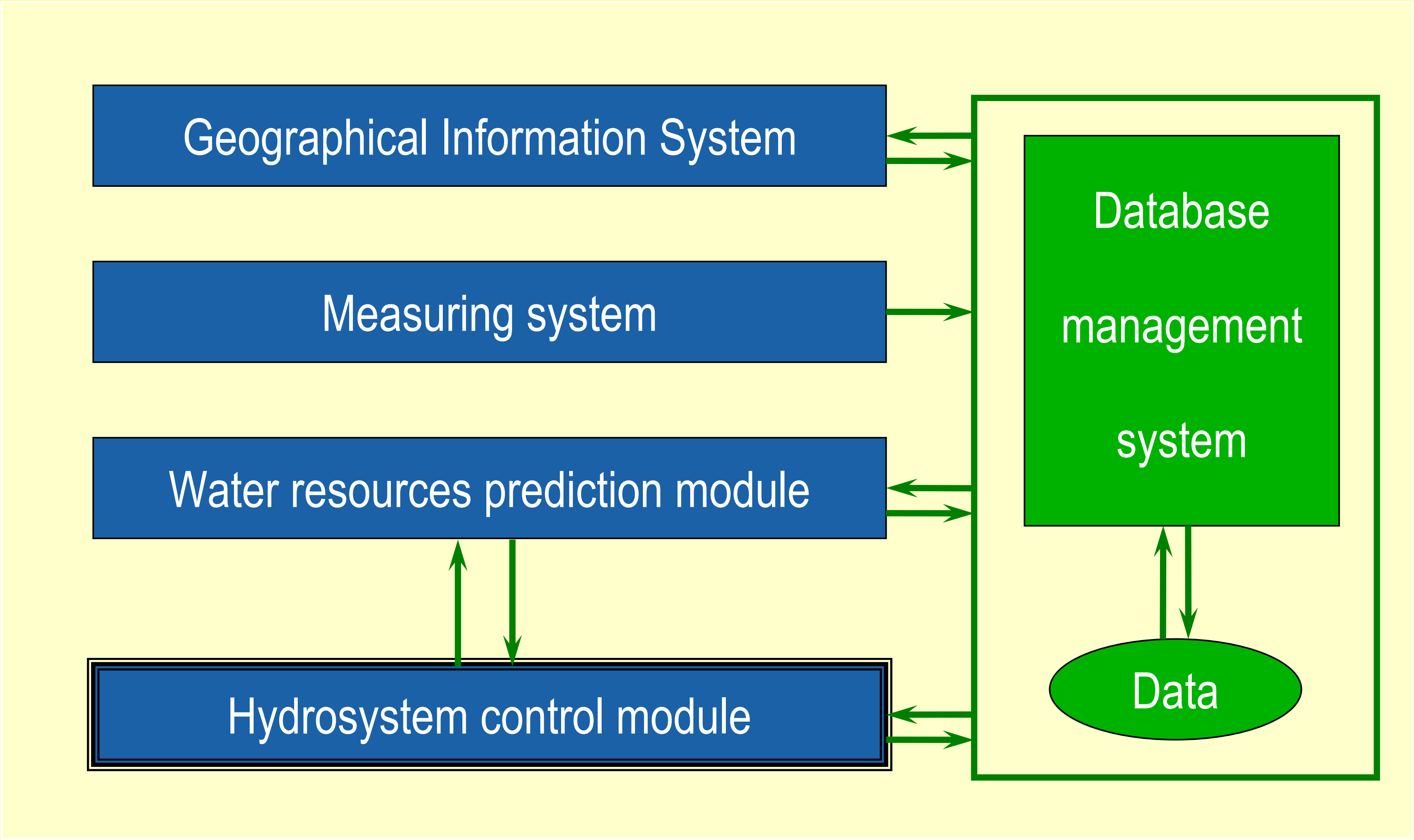
General evaluation and extensions of the parameterization-simulation-optimization method

- Is parametric rule underparametrized?
  - Nonlinear expressions with three parameters per reservoir did not outperform
  - Homogeneous linear expressions (one parameter per reservoir, $a_i = 0$) result in almost same optimal solutions
  - Considering seasonality (2 seasons) may improve results (slightly)
- How results of parametric rule based optimization compare to those of nonparametric optimization methods?
  - Generally, they are not inferior
  - In the non realistic case of *perfect foresight*, high dimensional methods may outperform parametric method *with no foresight* (slightly, by about 2%)
  - In practice, in complex nonlinear problems the parametric method yields better solutions due to more effective locating of global optimum
- Is the parameterization appropriate for all water uses and hydrosystems?
  - Yes, but different parameterizations may be needed for different components (e.g. aquifers)
  - Successful application to hydropower systems

{15,21}

Decision support tool integration
Data acquisition – Software systems – Management plans

Decision support tool structure



Measuring system

Central data base (archiving and processing of data)

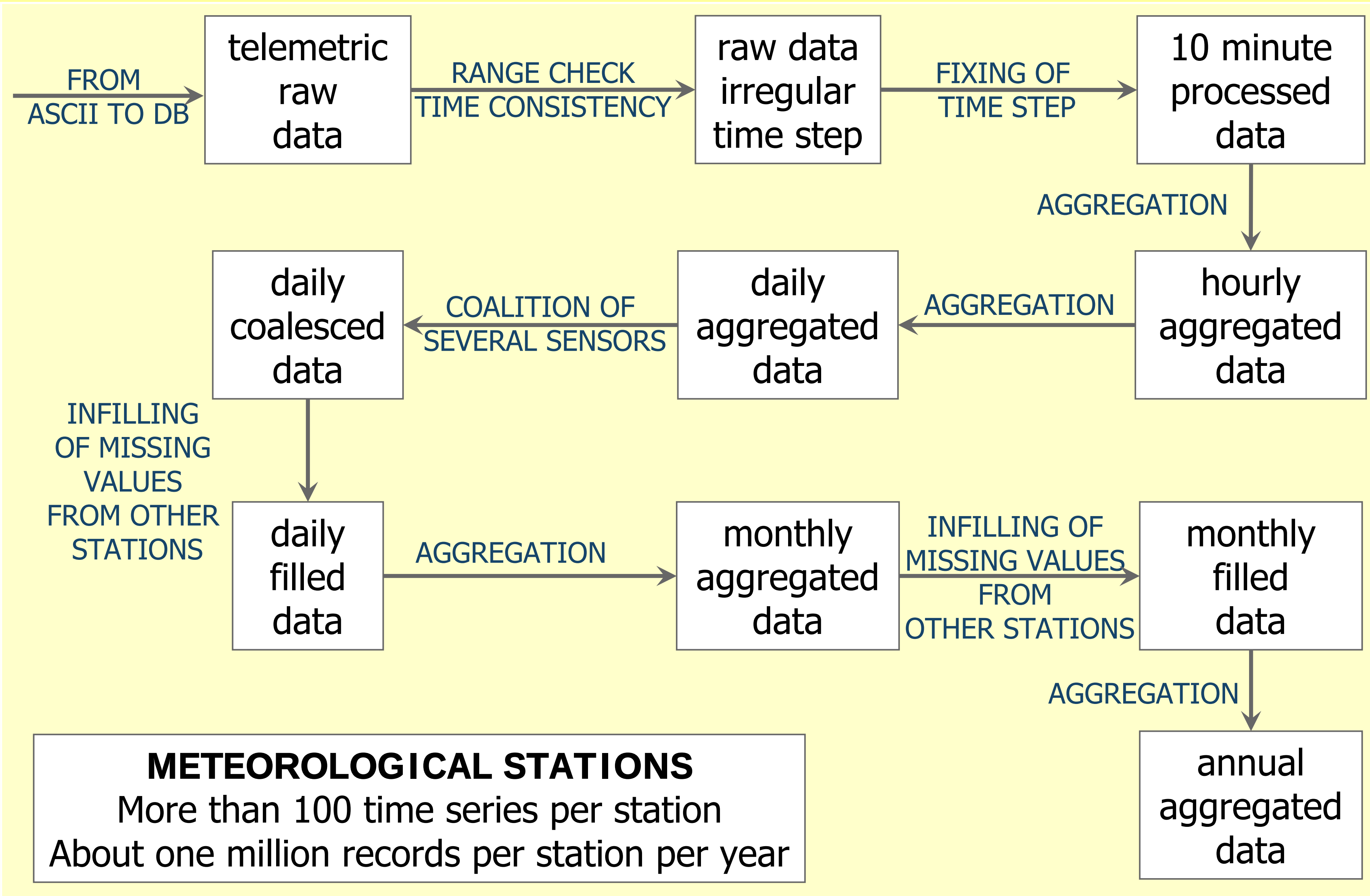
Central data collection unit (daily transmission by phone)

- Evinos Peripheral Data Center
- Mornos Peripheral Data Center
- Hylike Peripheral Data Center
- Marathon Peripheral Data Center

Evinos Peripheral Data Center:

- Meteorological station (10 min step)
  Rainfall, Temperature, Humidity, Wind, Radiation, Sunshine duration
- Reservoir elevation gage (1 h time step)
- River level gage (10 min time step)
- River flow measuring station (~once a month)

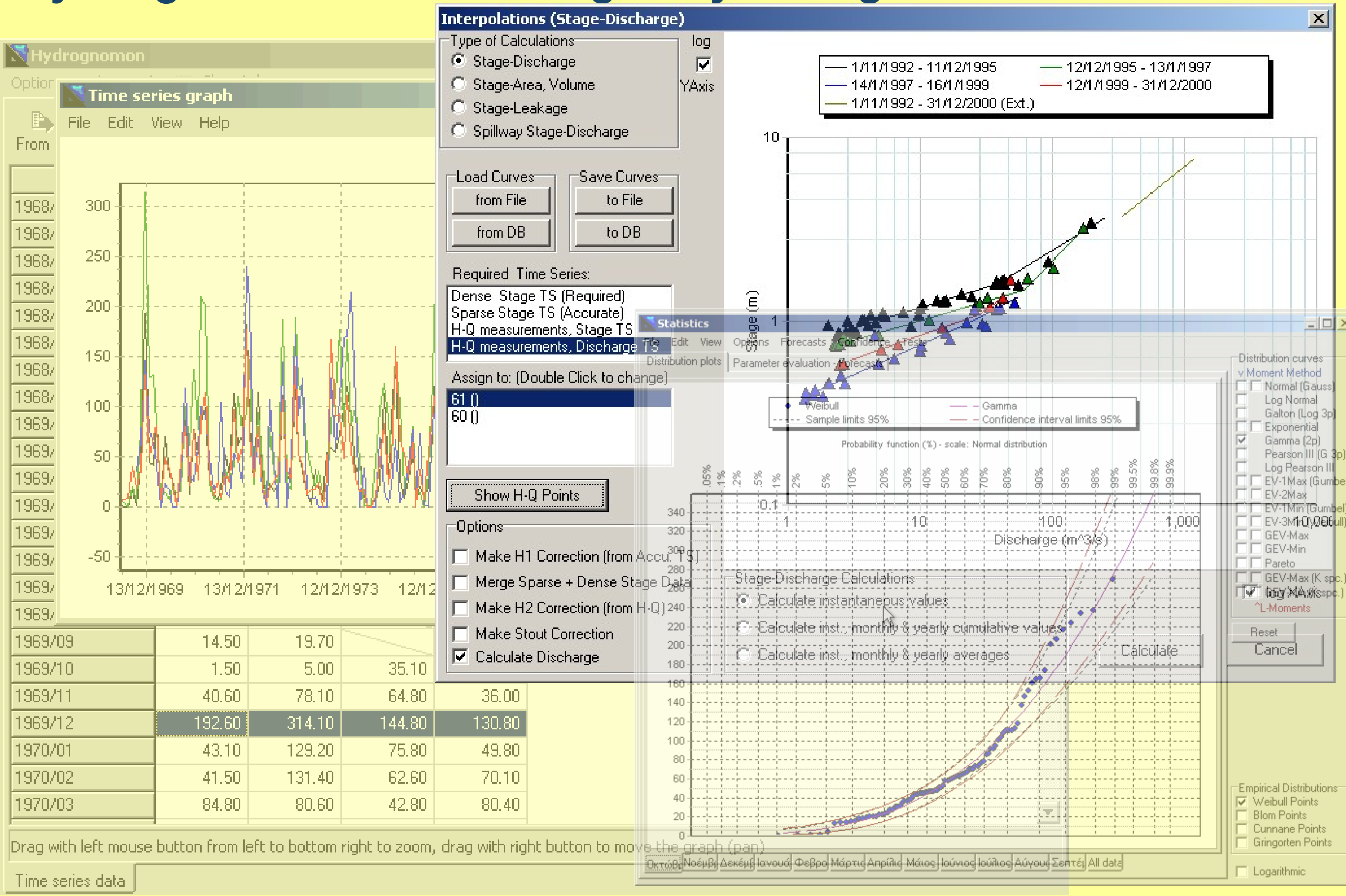
Data management and processing: Time series manipulation



Software system characteristics

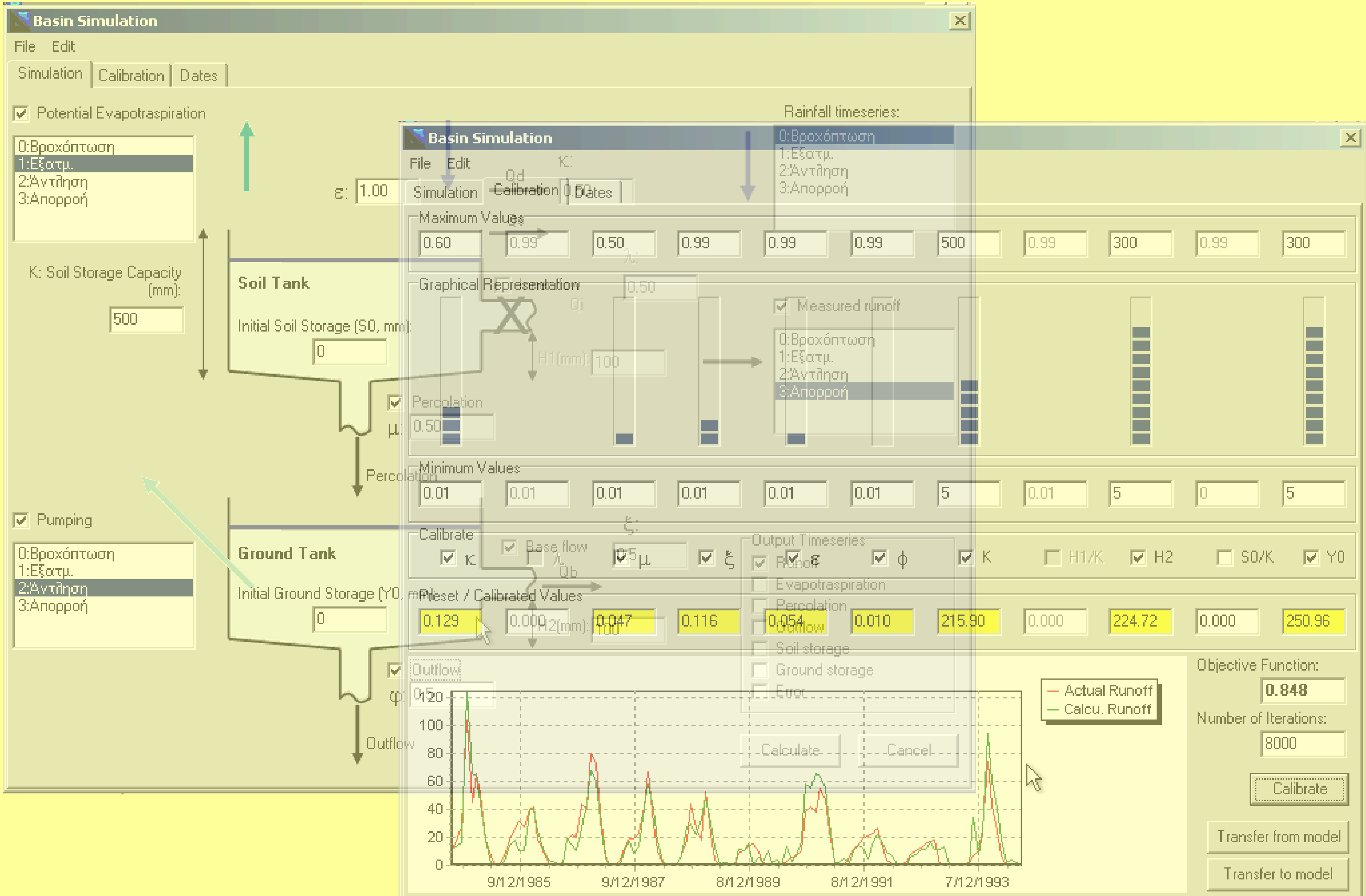
- All models written from scratch
- Basic development tool: **Delphi** (Object Pascal)
- Database: **Oracle** (more recently: **PostgreSQL**)
- Geographic system: **ArcView**
- Basic software units
  - **Hydrognomon**: Database management, processing of hydrologic data
  - **Castalia**: Stochastic hydrologic simulator
  - **Hydrogeios**: Simulation of surface and ground water processes
  - **Hydronomeas**: Hydrosystem control

Hydrognomon: Processing of hydrologic time series

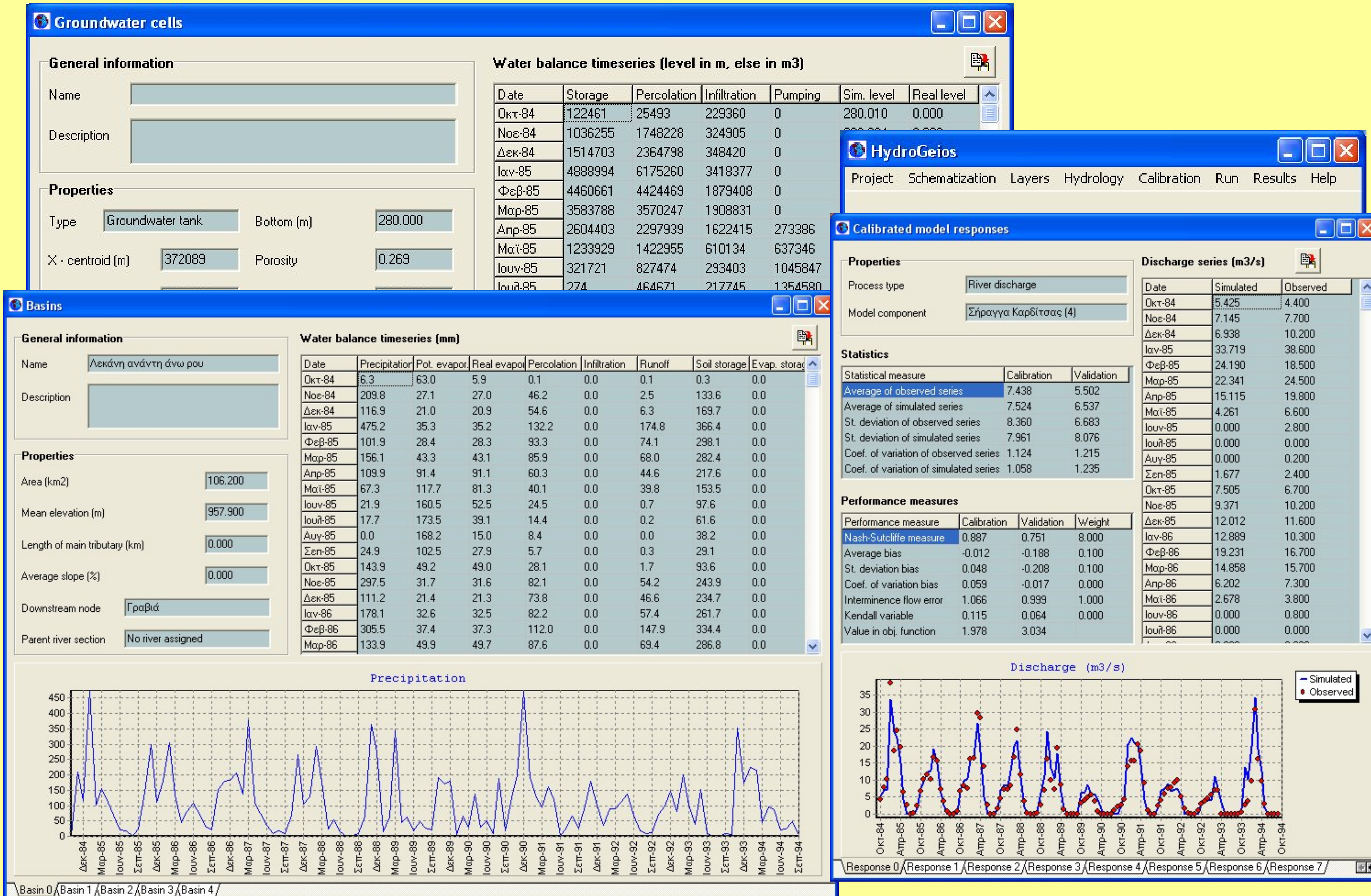


1969/09	14.50	19.70		
1969/10	1.50	5.00	35.10	
1969/11	40.60	78.10	64.80	36.00
1969/12	192.60	314.10	144.80	130.80
1970/01	43.10	129.20	75.80	49.80
1970/02	41.50	131.40	62.60	70.10
1970/03	84.80	80.60	42.80	80.40

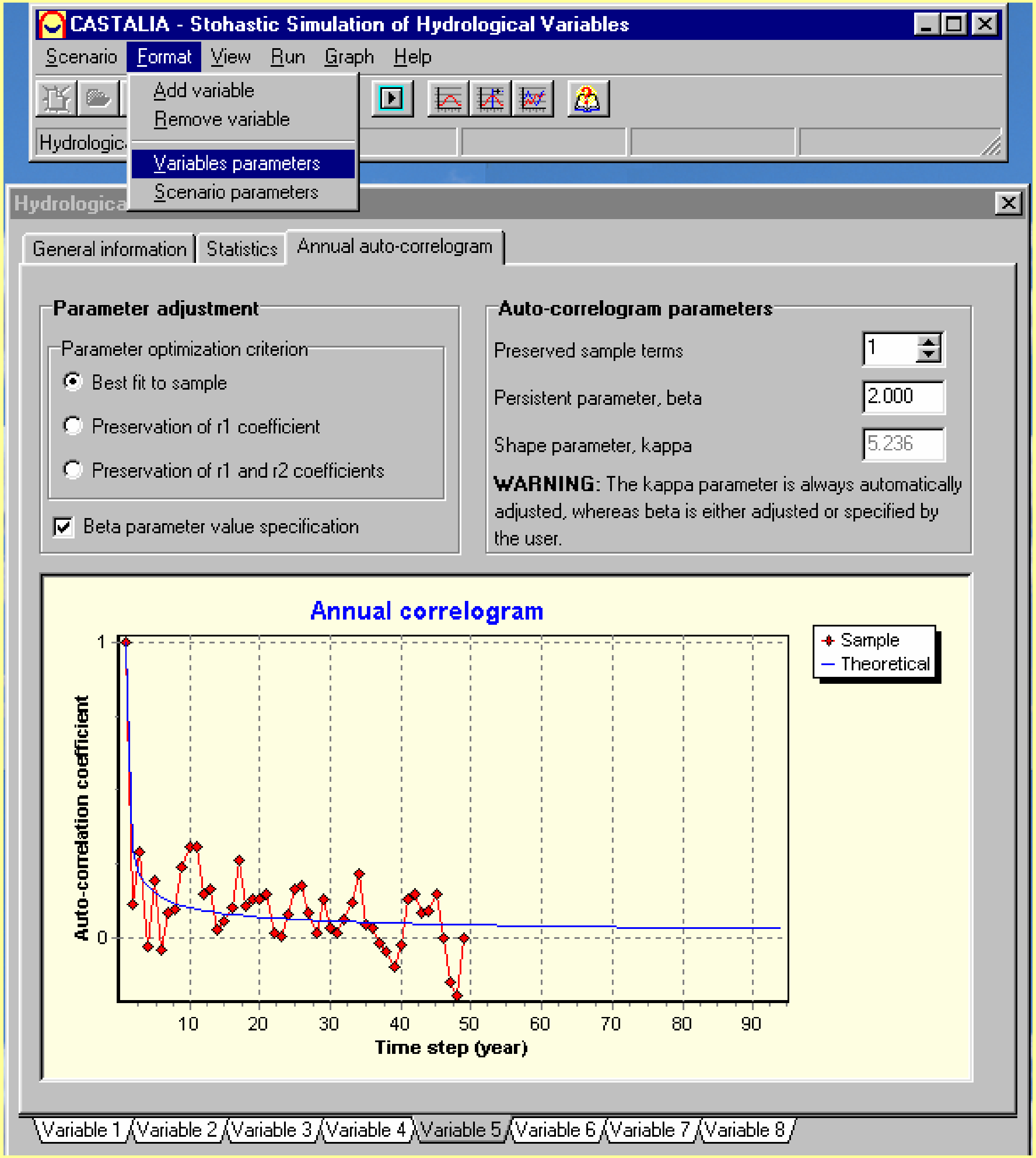
Hydrognomon: Automatic lumped hydrologic modeling



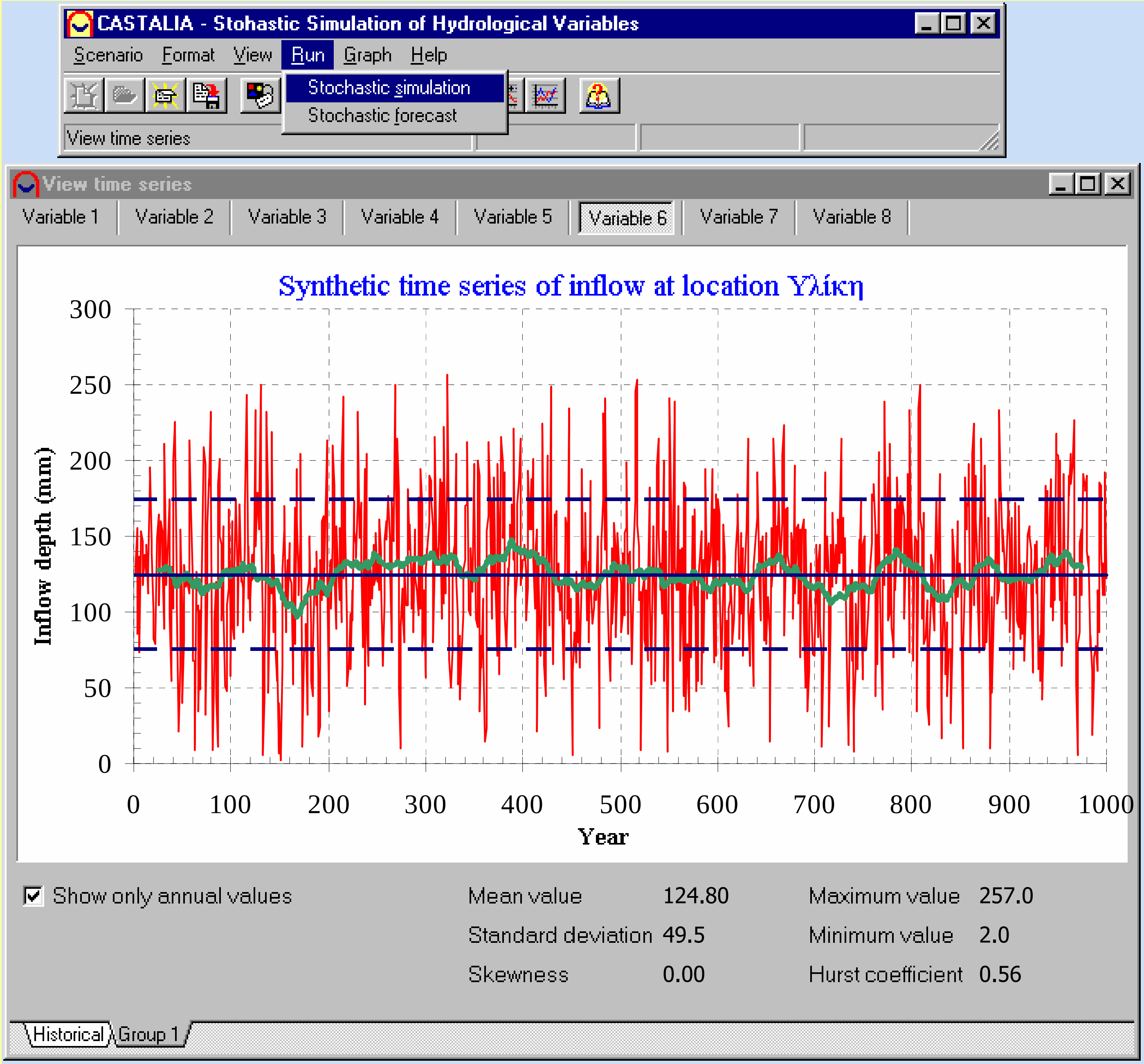
Hydrogeios: Detailed geo-hydrologic modeling



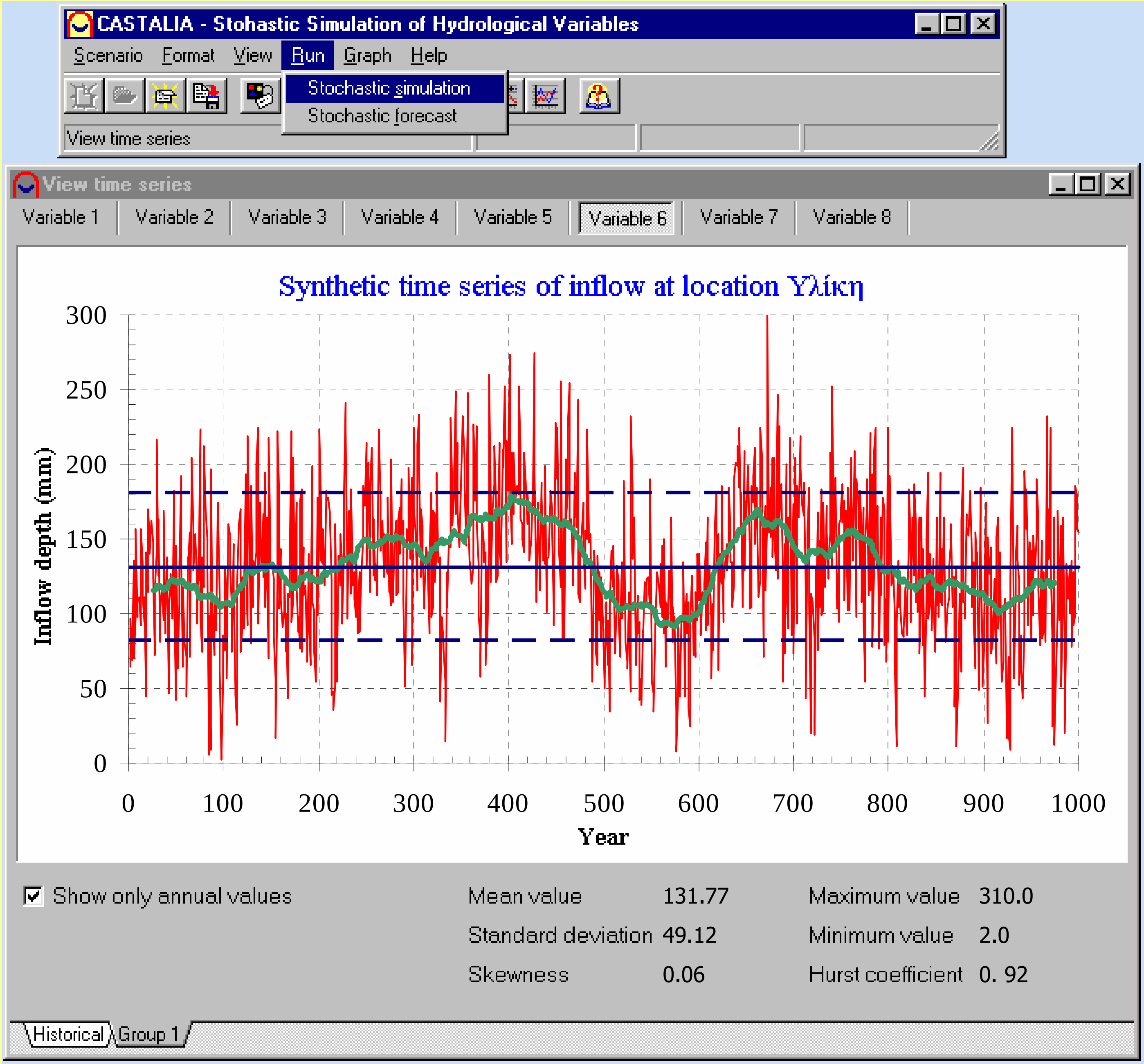
Castalia:
Parameter estimation-Parameters of autocorrelation and persistence



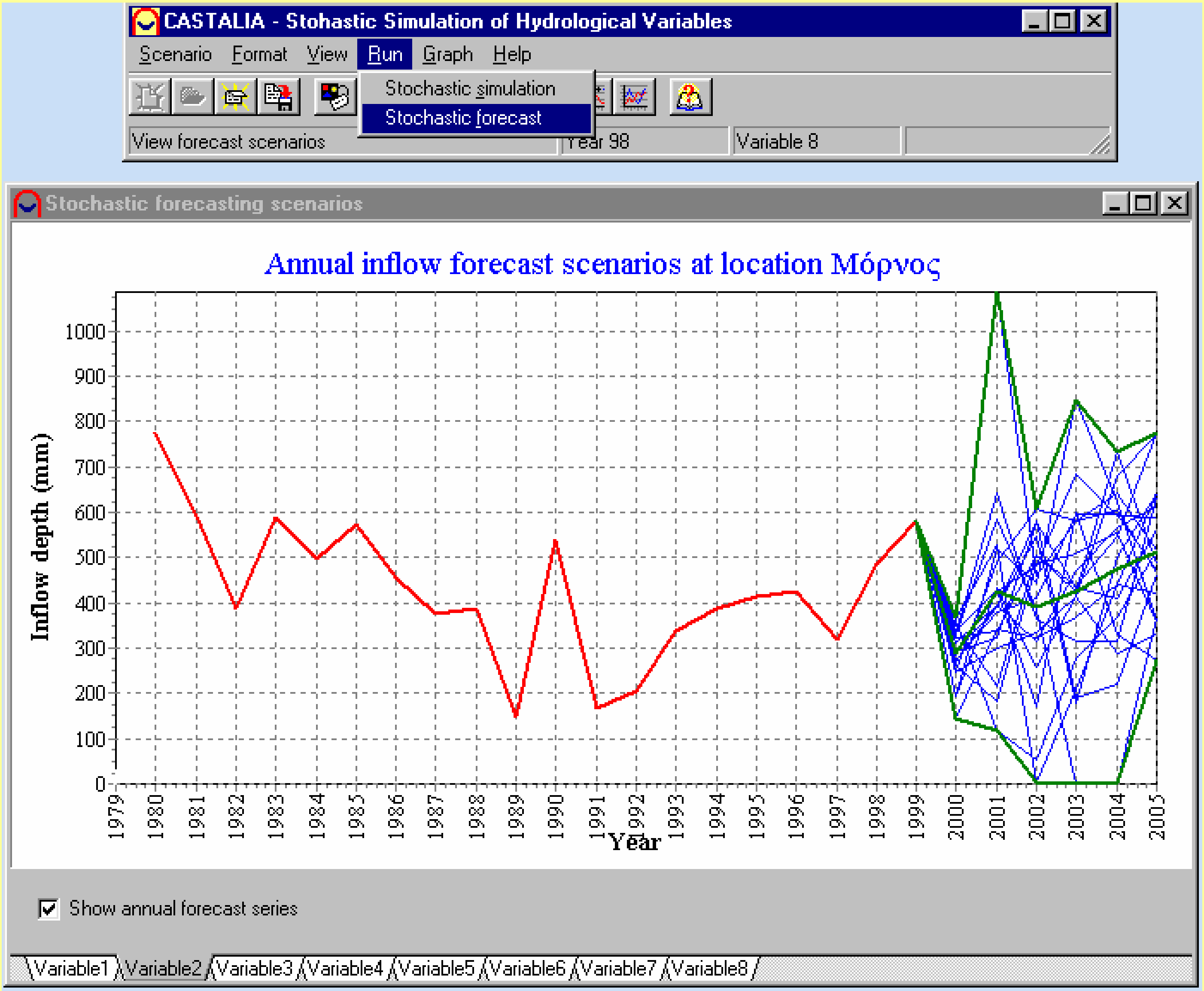
Castalia: Stochastic simulation without long term persistence



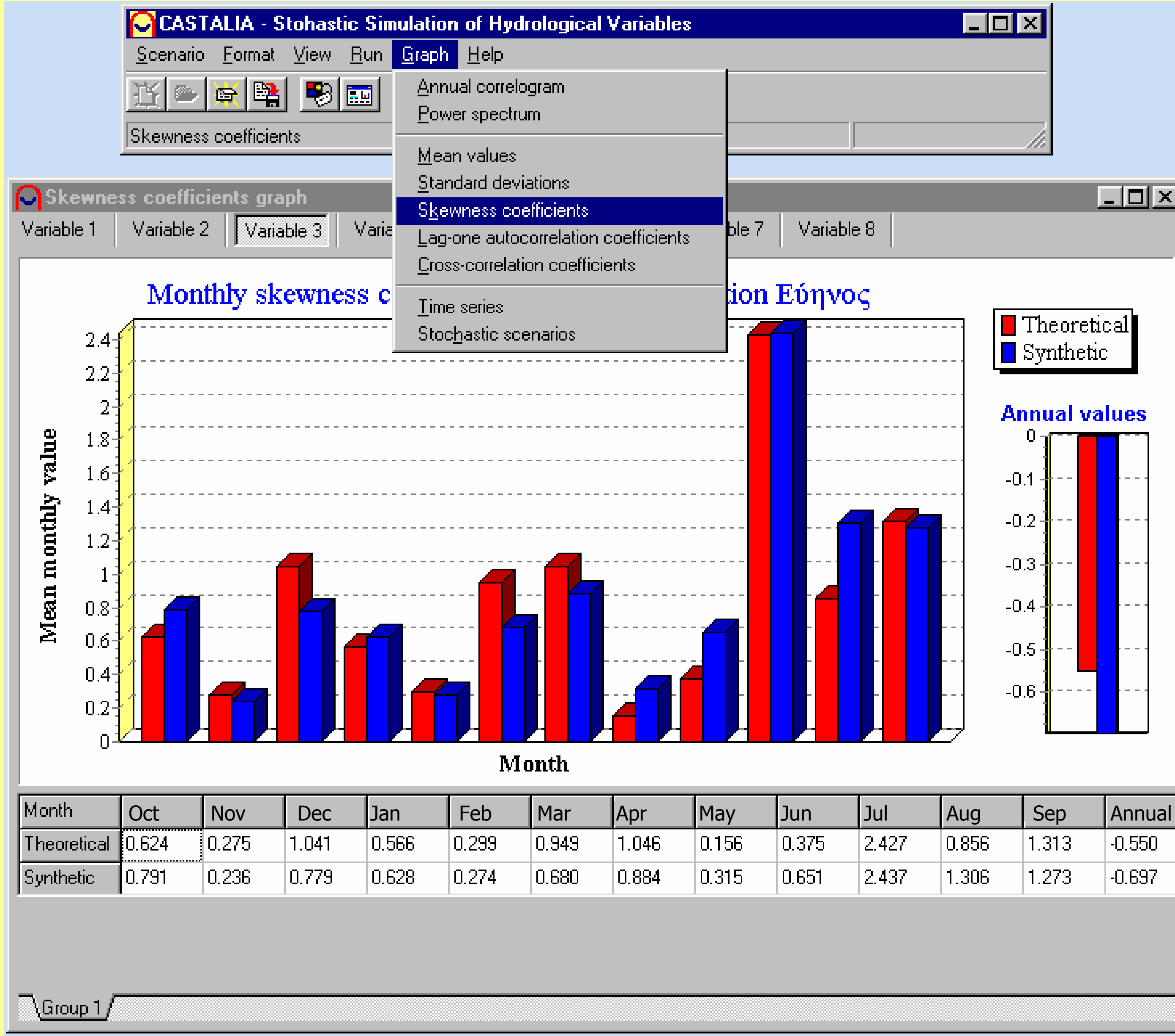
Castalia: Stochastic simulation with long term persistence



Castalia:
Stochastic forecasting with long term persistence

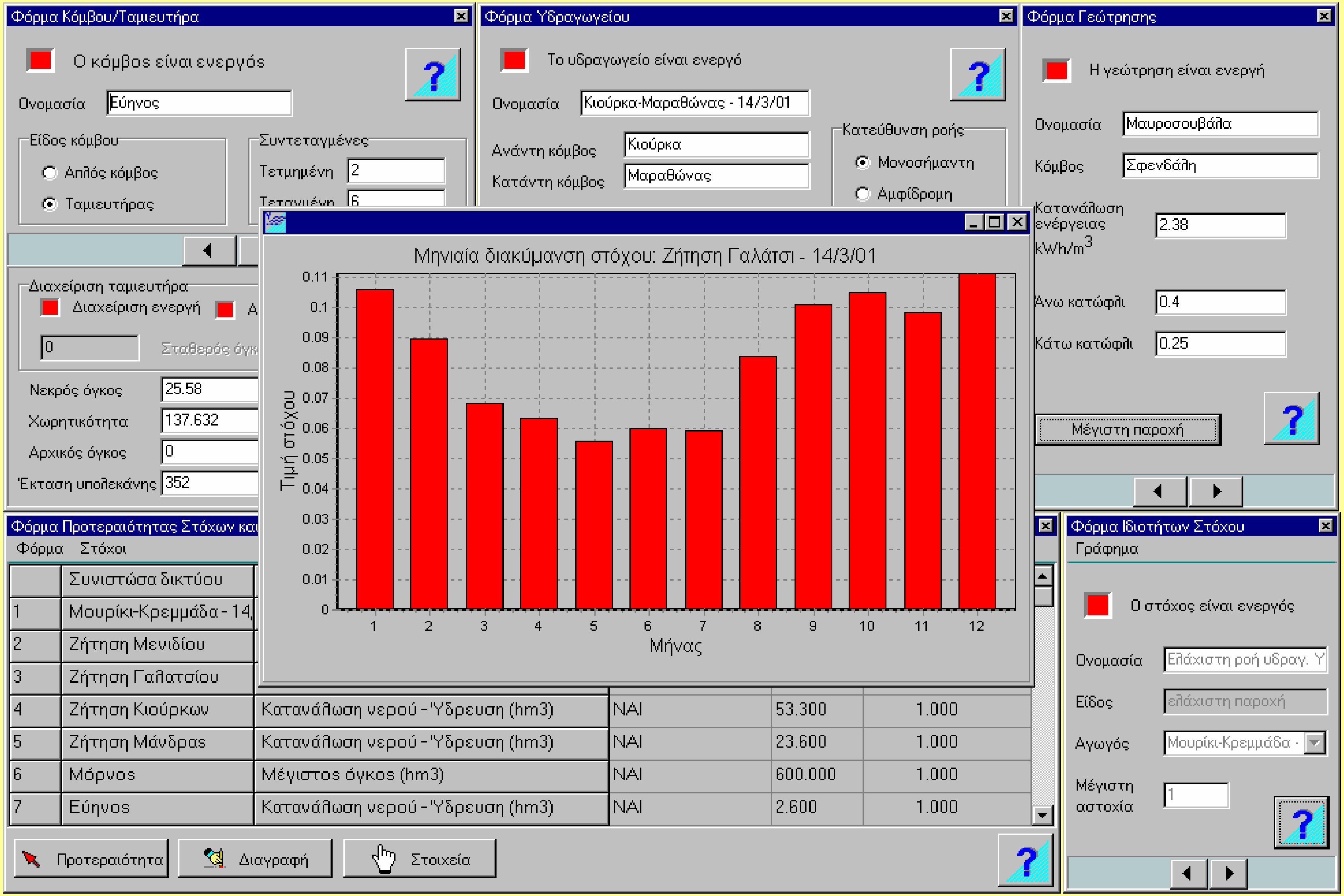


Castalia: Preservation of marginal statistics – Skewness

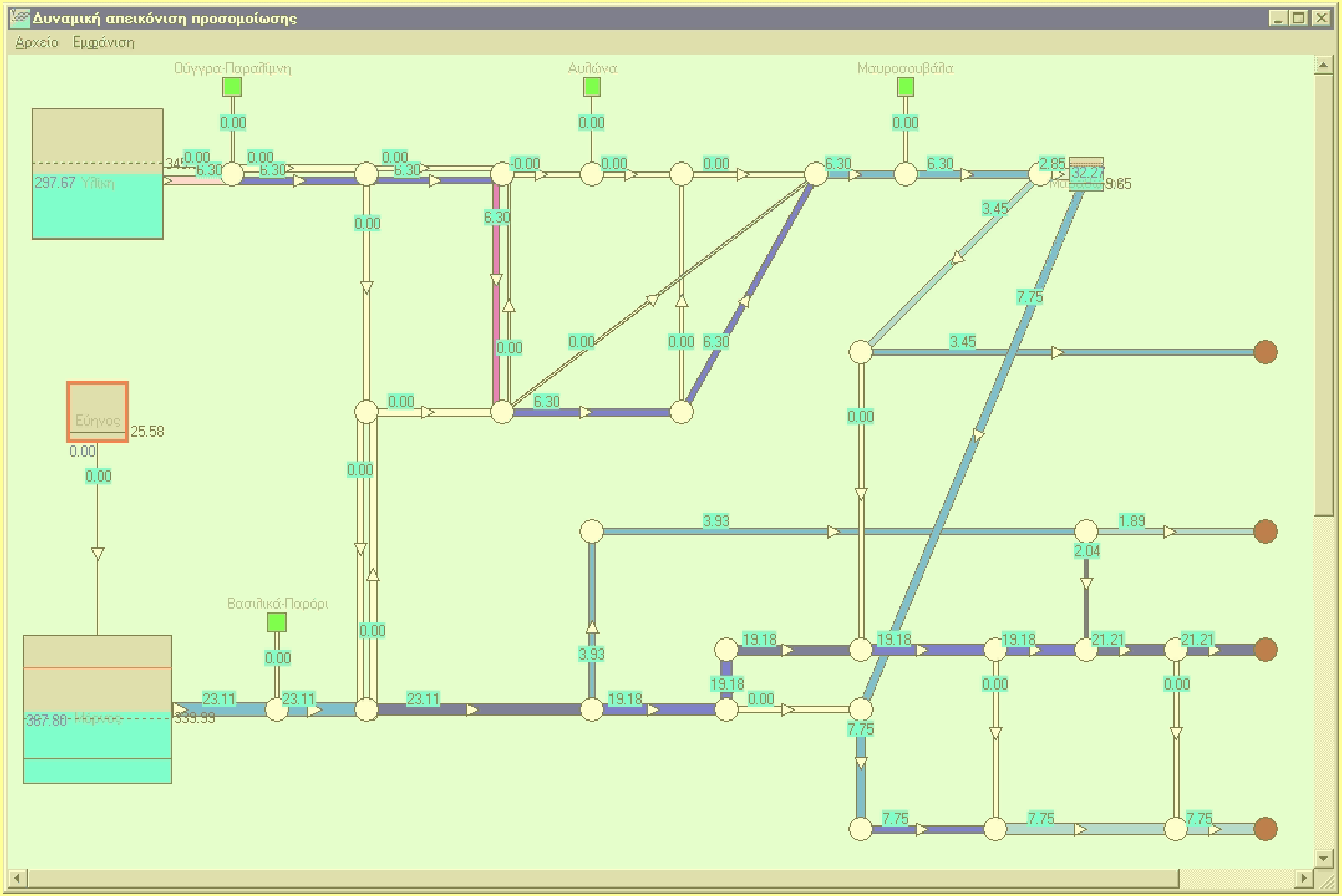


Month	Oct	Nov	Dec	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Annual
Theoretical	0.624	0.275	1.041	0.566	0.299	0.949	1.046	0.156	0.375	2.427	0.856	1.313	-0.550
Synthetic	0.791	0.236	0.779	0.628	0.274	0.680	0.884	0.315	0.651	2.437	1.306	1.273	-0.697

Hydronomeas: Hydrosystem data management

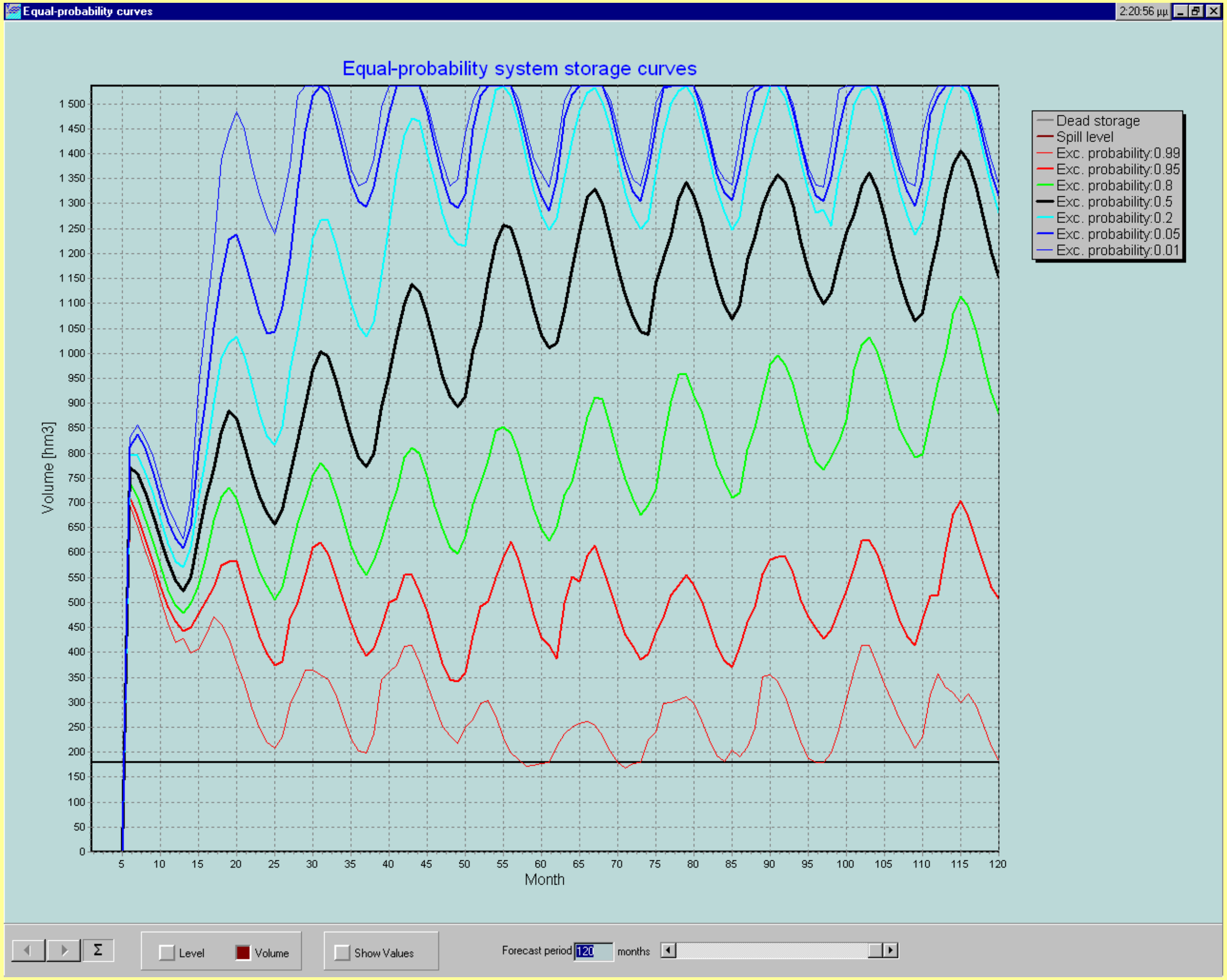


Hydronomeas: Visualization of hydrosystem simulation



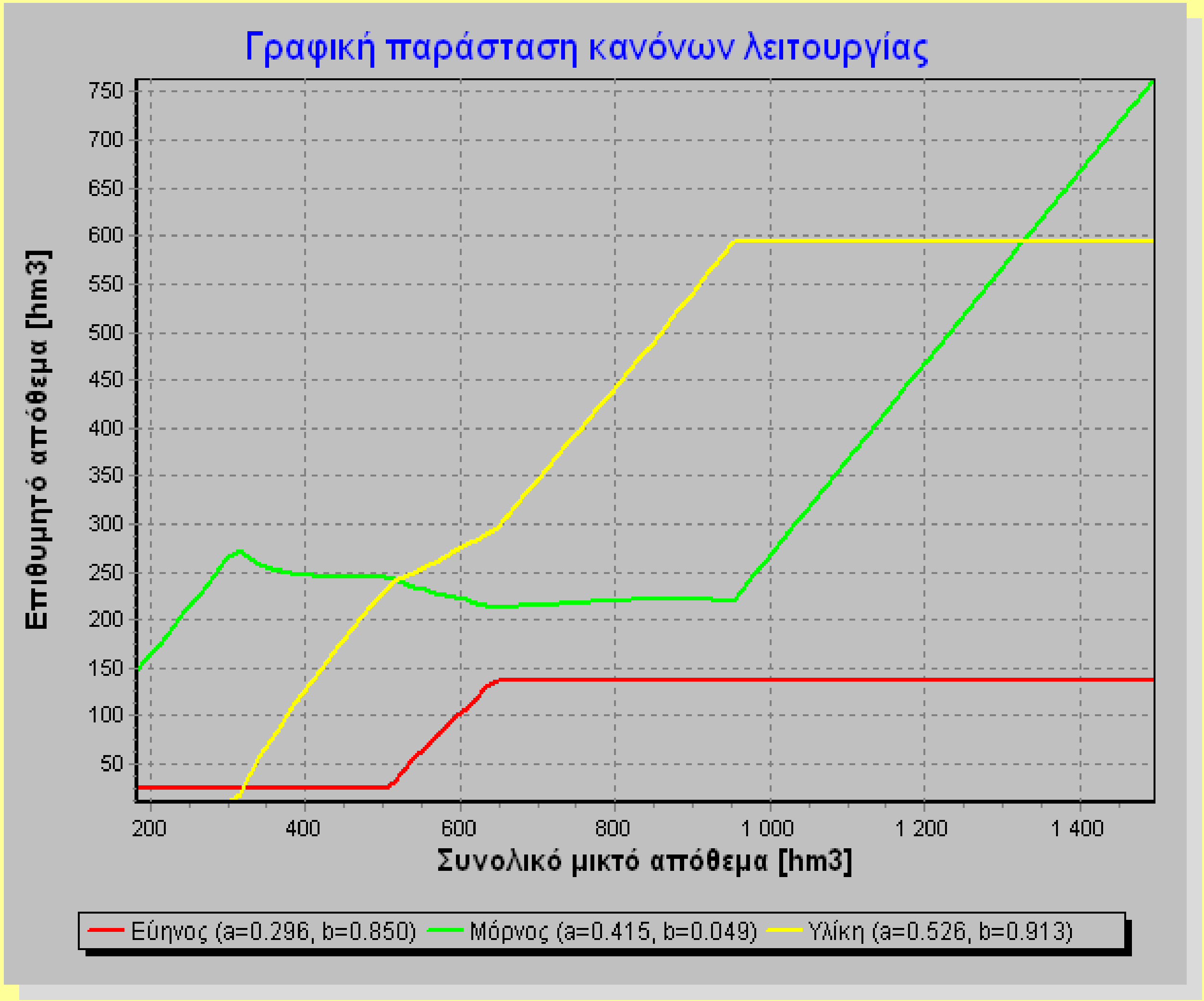
Hydronomeas: Stochastic forecast of hydrosystem storage

Evolution of quantiles of system storage (for several levels of probability of exceedance) for the next 10 years as a result of 200 terminating simulations with long-term persistence

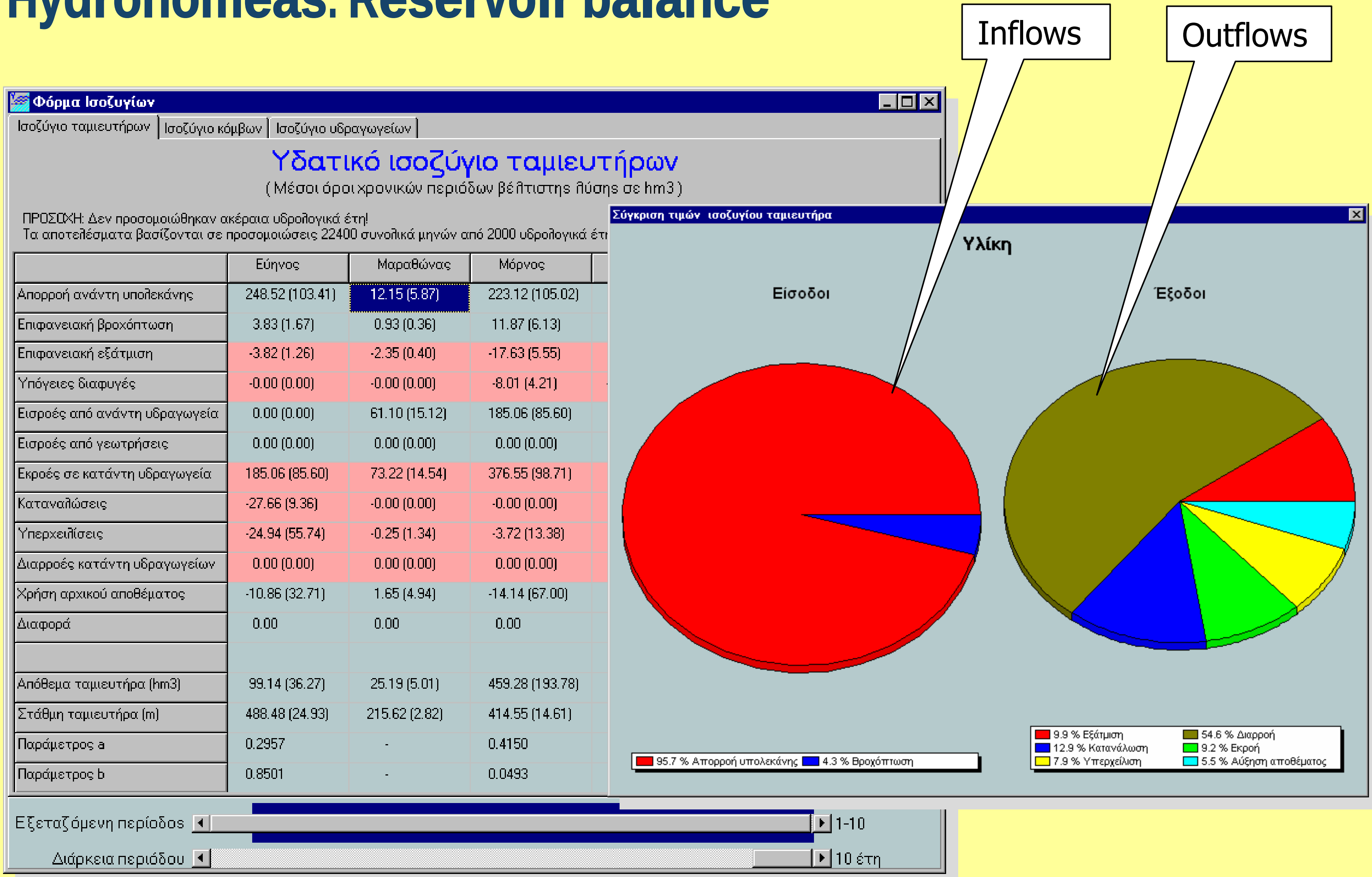


Hydronomeas: Optimal hydrosystem control rules

Target allocation of total reservoir storage per each reservoir



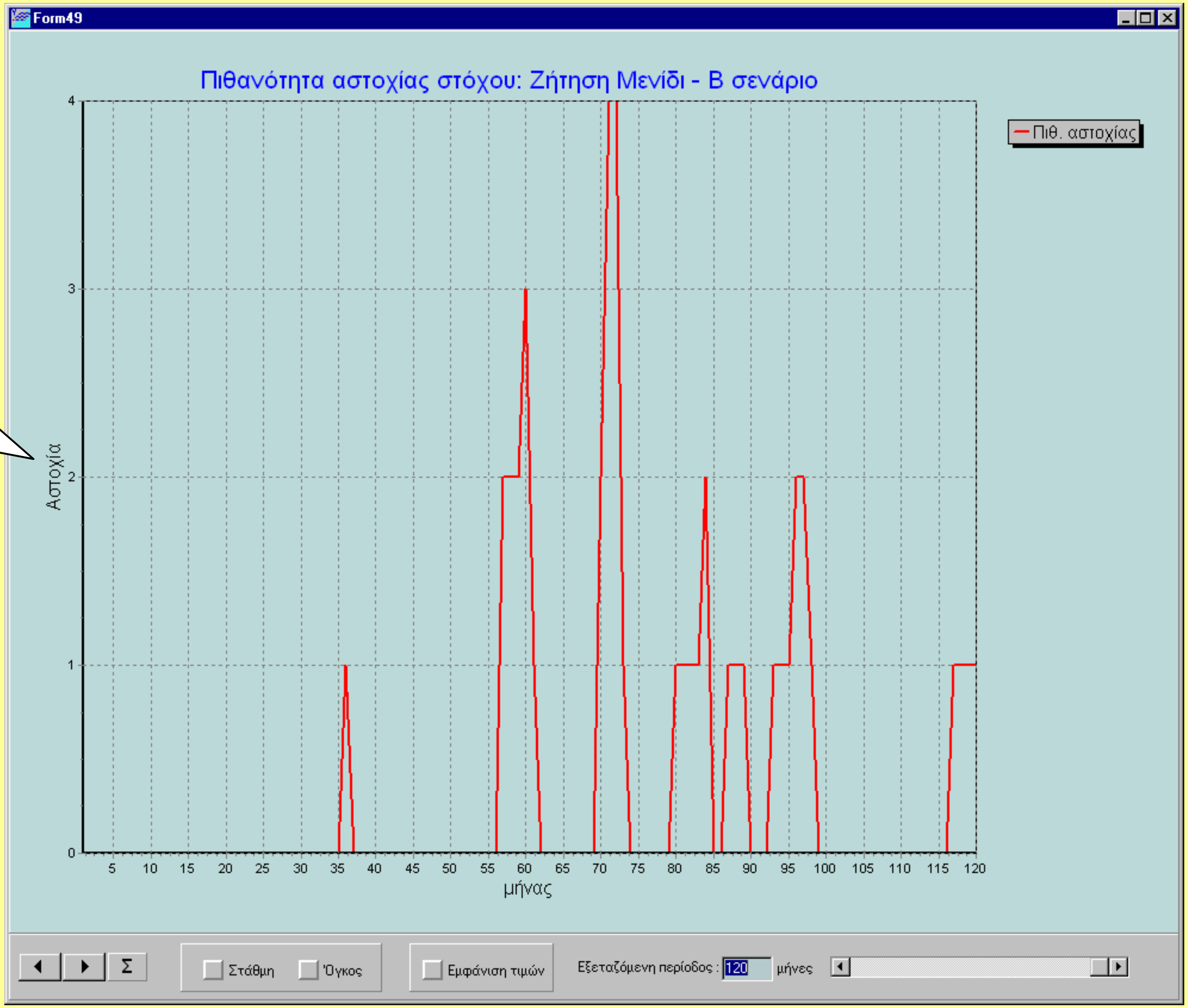
Hydronomeas: Reservoir balance



	Εύηνος	Μαραθώνας	Μόρνος
Απορροή ανάντη υπολεκάνης	248.52 (103.41)	12.15 (5.87)	223.12 (105.02)
Επιφανειακή βροχόπτωση	3.83 (1.67)	0.93 (0.36)	11.87 (6.13)
Επιφανειακή εξάτμιση	-3.82 (1.26)	-2.35 (0.40)	-17.63 (5.55)
Υπόγειες διαφυγές	-0.00 (0.00)	-0.00 (0.00)	-8.01 (4.21)
Εισροές από ανάντη υδραγωγεία	0.00 (0.00)	61.10 (15.12)	185.06 (85.60)
Εισροές από γεωτρήσεις	0.00 (0.00)	0.00 (0.00)	0.00 (0.00)
Εκροές σε κατάντη υδραγωγεία	185.06 (85.60)	73.22 (14.54)	376.55 (98.71)
Καταναλώσεις	-27.66 (9.36)	-0.00 (0.00)	-0.00 (0.00)
Υπερχειλίσεις	-24.94 (55.74)	-0.25 (1.34)	-3.72 (13.38)
Διαρροές κατάντη υδραγωγείων	0.00 (0.00)	0.00 (0.00)	0.00 (0.00)
Χρήση αρχικού αποθέματος	-10.86 (32.71)	1.65 (4.94)	-14.14 (67.00)
Διαφορά	0.00	0.00	0.00
Απόθεμα ταμιευτήρα (hm3)	99.14 (36.27)	25.19 (5.01)	459.28 (193.78)
Στάθμη ταμιευτήρα (m)	488.48 (24.93)	215.62 (2.82)	414.55 (14.61)
Παράμετρος a	0.2957	-	0.4150
Παράμετρος b	0.8501	-	0.0493

Hydronomeas: Time profile of failure probabilities

Number of failures in a total of 200 stochastic scenarios



Hydronomeas: Reporting

Print Preview

Close

Υδρονομέας 2.1 – Εθνικό Μετσόβιο Πολυτεχνείο

ΑΣΤΟΧΙΑ ΕΝΕΡΓΩΝ ΣΤΟΧΩΝ

(με σειρά προτεραιότητας)

Ονομασία στόχου	Περιγραφή	Αστοχία υδρολογικών περιόδων	Ανώτατη αποδεκτή αστοχία	Αστοχία μηνών	Αστοχία όγκου
Ελάχιστη ροή υδραγ. Υλίκης	Ελάχιστη ροή υδραγωγείου Υλίκης τους πρώτους μήνες προσομοίωσης για υλοποίηση υποχρεωτικής απόληψης	0/ 2000	1.00	0/ 24000	0.00 / 0.00
Ζήτηση Μενίδι - 14/3/01	Σταθερός σε ετήσια βάση και εποχιακά κυμαινόμενος στόχος ζήτησης νερού για ύδρευση (Σύνολο Αθήνας 440 hm3)	17 / 2000	1.00	43 / 24000	0.17 / 224.27
Ζήτηση Γαλάτσι - 14/3/01	Σταθερός σε ετήσια βάση και εποχιακά κυμαινόμενος στόχος ζήτησης νερού για ύδρευση (Σύνολο Αθήνας 440 hm3)	23 / 2000	1.00	50 / 24000	0.23 / 124.36
Ζήτηση Κιούρκα - 14/3/01	Σταθερός σε ετήσια βάση και εποχιακά κυμαινόμενος στόχος ζήτησης νερού για ύδρευση (Σύνολο Αθήνας 440 hm3)	23 / 2000	1.00	56 / 24000	0.13 / 51.61
Ζήτηση Μάνδρα - 14/3/01	Σταθερός σε ετήσια βάση και εποχιακά κυμαινόμενος στόχος ζήτησης νερού για ύδρευση (Σύνολο Αθήνας 440 hm3)	25 / 2000	1.00	63 / 24000	0.07 / 22.77
Μεγιστος Όγκος Μόρνου	Σταθερός μέγιστος όγκος για αποφυγή υπερχειλίσεων	864 / 2000	1.00	7304/ 24000	0.00 / 0.00
Παροχή Ευήνου	Σταθερή περιβαλλοντική παροχή 2,6 hm3, εισάγεται ώς κατανάλωση	959 / 2000	1.00	2918/ 24000	2.77 / 29.90
Μέγιστος όγκος Μαραθώνα	Στόχος μέγιστου όγκου για αποφυγή υπερχειλίσεων	862 / 2000	1.00	4553/ 24000	0.00 / 0.00
Ελαχιστος όγκος Μαραθώνα	Εποχικά κυμαινόμενος στόχος ελάχιστου όγκου για τη διατήρηση αποθέματος ασφαλείας	1415/ 2000	1.00	2820/ 24000	0.11 / 17.07
Άρδευση Κωπαΐδα	Ζήτηση νερού για άρδευση της Κωπαΐδας. Παρουσιάζει έντονη εποχιακή διακύμανση.	16 / 2000	1.00	46 / 24000	0.15 / 35.00

Μόνο οι ενεργοί στόχοι χρησιμοποιούνται κατά την προσομοίωση/βελτιστοποίηση
Η πιθανότητα αστοχίας υδρολογικής περιόδου (έτος) ορίζεται από το λόγο των υδρολογικών περιόδων με αστοχία προς το σύνολο των προσομοιωμένων υδρολογικών περιόδων.
Αστοχία του συστήματος ως προς τη υδρολογική περίοδο διαπιστώνεται κατά την αστοχία ενός τουλάχιστον χρονικού βήματος (μήνας) της περιόδου.
Η προτεινόμενη λύση τηρεί τον περιορισμό της ανώτατης αποδεκτής πιθανότητας αστοχίας για κάθε στόχο.

16/3/2001 5:31:54 μμ

33

Page 33 of 33

Management plans and every day operation of the hydrosystem

- Every five years a master plan of the water supply of Athens is elaborated (the first was issued in 2000)
- Every year the master plan is revised based on current data and model runs
- Every three months the annual plan is reassessed and, if necessary, updated by new model runs
- Meanwhile, the every day management is based on optimal parametric operation rules
- Models are run for a 10-year lead time to account for long-term effects of today’s decisions
- The general management targets are:
  - Adequacy of water resources
  - Adequacy of conveyance system
  - Cost effectiveness
- All management is based on a probabilistic approach of forecasts/risk/reliability assuming:
  - Acceptable reliability 99% on an annual basis
  - Potential for further increase of reliability taking into account elasticity of demand and emergency measures in case of impending failure
- So far, the decision support tool and its modules (thoroughly tested for the Olympics 2004) exhibited good performance

References

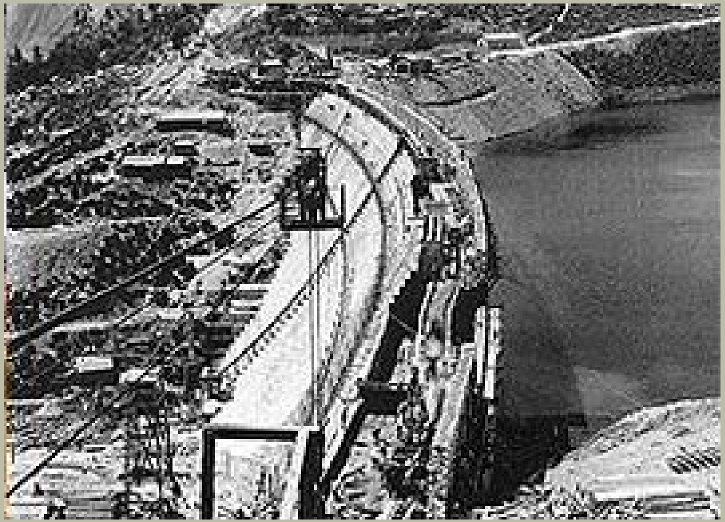
1. Koutsoyiannis, D. A. Efstratiadis and K. P. Georgakakos, Uncertainty assessment of future hydroclimatic predictions: A comparison of probabilistic and scenario-based approaches, 2005 (prepared).
2. Koutsoyiannis, D., An entropic-stochastic representation of rainfall intermittency: The origin of clustering and persistence, *Water Resources Research*, 2005 (in press).
3. Koutsoyiannis, D., Nonstationarity versus scaling in hydrology, *Journal of Hydrology*, 2005 (in press).
4. Rozos, E., and D. Koutsoyiannis, A multicell karstic aquifer model with alternative flow equations, *Journal of Hydrology*, 2005 (in press).
5. Koutsoyiannis, D., A toy model of climatic variability with scaling behaviour, *Journal of Hydrology*, 2005 (in press).
6. Langousis, A., and D. Koutsoyiannis, A stochastic methodology for generation of seasonal time series reproducing overyear scaling, *Journal of Hydrology*, 2005 (in press).
7. Koutsoyiannis, D., Uncertainty, entropy, scaling and hydrological stochastics, 1, Marginal distributional properties of hydrological processes and state scaling, *Hydrological Sciences Journal*, 50(3), 381-404, 2005.
8. Koutsoyiannis, D., Uncertainty, entropy, scaling and hydrological stochastics, 2, Time dependence of hydrological processes and time scaling, *Hydrological Sciences Journal*, 50(3), 405-426, 2005.
9. Koutsoyiannis, D., Statistics of extremes and estimation of extreme rainfall, 1, Theoretical investigation, *Hydrological Sciences Journal*, 49(4), 575-590, 2004.
10. Koutsoyiannis, D., Statistics of extremes and estimation of extreme rainfall, 2, Empirical investigation of long rainfall records, *Hydrological Sciences Journal*, 49(4), 591-610, 2004.
11. Rozos, E., A. Efstratiadis, I. Nalbantis, and D. Koutsoyiannis, Calibration of a semi-distributed model for conjunctive simulation of surface and groundwater flows, *Hydrological Sciences Journal*, 49(5), 819-842, 2004.
12. Efstratiadis, A., D. Koutsoyiannis, and D. Xenos, Minimising water cost in the water resource management of Athens, *Urban Water Journal*, 1(1), 3-15, 2004.
13. Koutsoyiannis, D., G. Karavokiros, A. Efstratiadis, N. Mamassis, A. Koukouvinos, and A. Christofides, A decision support system for the management of the water resource system of Athens, *Physics and Chemistry of the Earth*, 28(14-15), 599-609, 2003.
14. Koutsoyiannis, D., Climate change, the Hurst phenomenon, and hydrological statistics, *Hydrological Sciences Journal*, 48(1), 3-24, 2003.
15. Koutsoyiannis, D., and A. Economou, Evaluation of the parameterization-simulation-optimization approach for the control of reservoir systems, *Water Resources Research*, 39(6), 1170, 1-17, 2003.
16. Koutsoyiannis, D., The Hurst phenomenon and fractional Gaussian noise made easy, *Hydrological Sciences Journal*, 47(4), 573-595, 2002.
17. Koutsoyiannis, D., A. Efstratiadis, and G. Karavokiros, A decision support tool for the management of multi-reservoir systems, *Journal of the American Water Resources Association*, 38(4), 945-958, 2002.
18. Koutsoyiannis, D., Coupling stochastic models of different time scales, *Water Resources Research*, 37(2), 379-392, 2001.
19. Koutsoyiannis, D., A generalized mathematical framework for stochastic simulation and forecast of hydrologic time series, *Water Resources Research*, 36(6), 1519-1533, 2000.
20. Koutsoyiannis, D., Optimal decomposition of covariance matrices for multivariate stochastic models in hydrology, *Water Resources Research*, 35(4), 1219-1229, 1999.
21. Nalbantis, I., and D. Koutsoyiannis, A parametric rule for planning and management of multiple reservoir systems, *Water Resources Research*, 33(9), 2165-2177, 1997.
22. Koutsoyiannis, D., and A. Manetas, Simple disaggregation by accurate adjusting procedures, *Water Resources Research*, 32(7), 2105-2117, 1996.

Early stage

The Hadrianean aqueduct

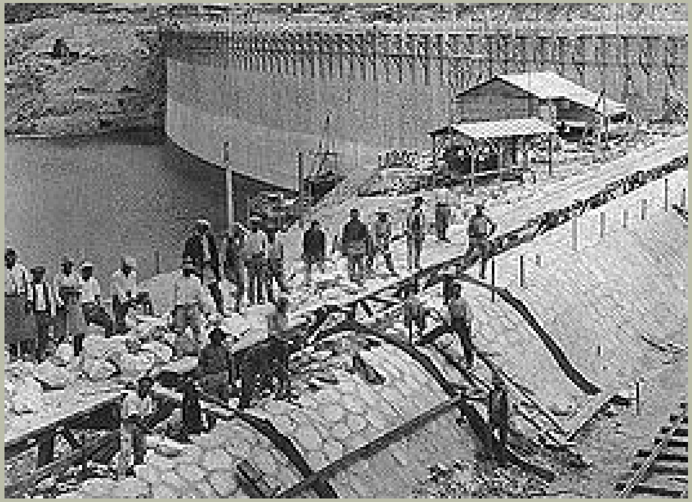
Supplementary water collection and distribution in Athens (early 20th century until 1930s)

Marathon dam



Construction of dam, 1928

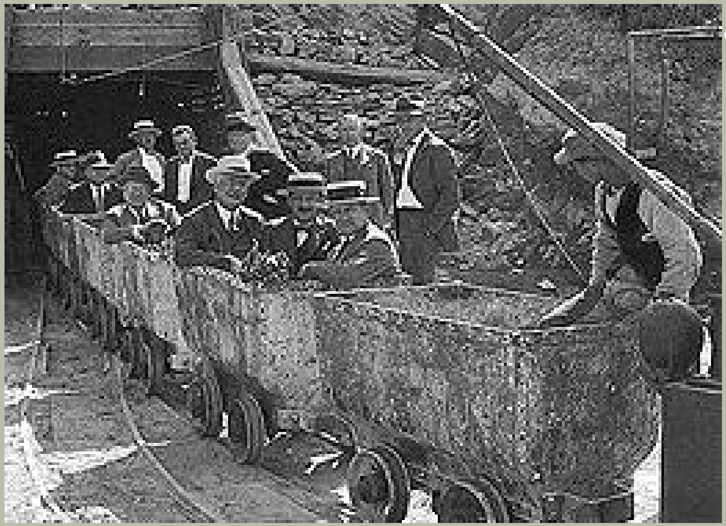
Today



Construction of spillway, 1928

Marathon dam (2)

Devastating flood, 1926



Inauguration of Boyati tunnel, 1928

Marathon spillway in action, 1941

Hylike lake and pumping stations

Hylike lake

Hylike, floating pumping stations

Hylike, main pumping station

Kiourka pumping station

Mornos reservoir and aqueduct

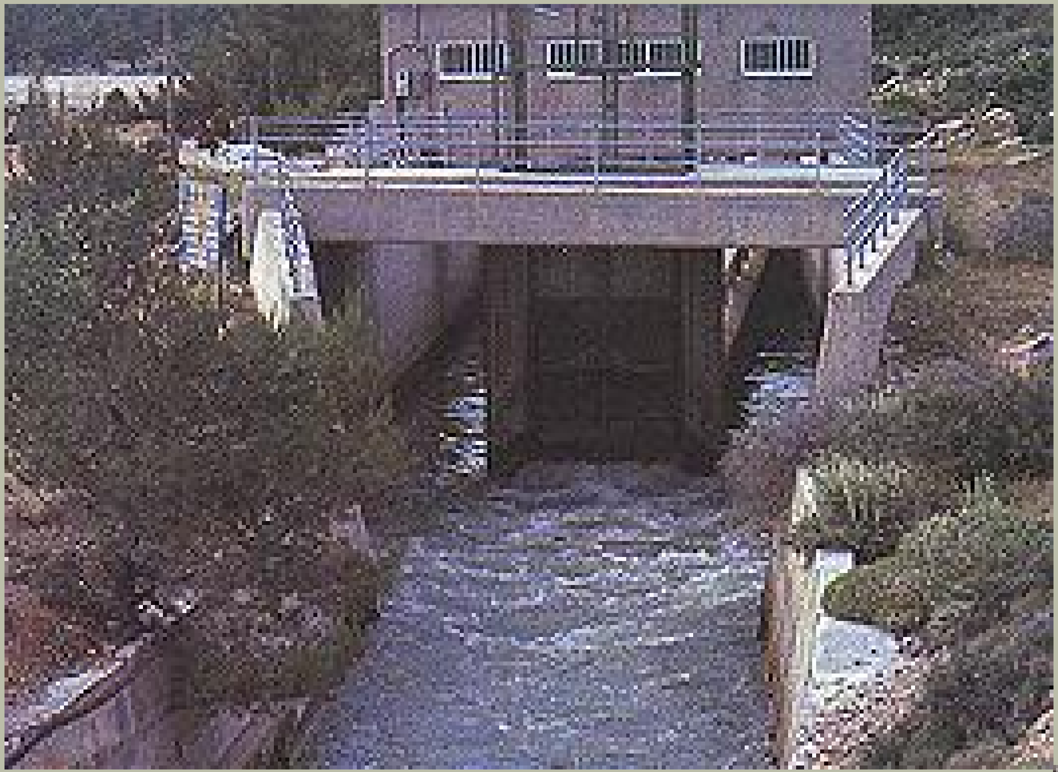
Mornos canal at Delphi

Mornos reservoir

Mornos canal at Thebes plain

Siphon at Distomo

Control of Mornos aqueduct



Canal flow control construction

Aqueduct supervizing & control centre

Evinos dam and tunnel

Evinos dam during construction

Construction of the Evinos-Mornos connection tunnel

Treatment plants

Perissos water treatment plant

Aspropyrgos water treatment plant