Simulation and optimization for the design and management of hydroelectric works

Andreas Efstratiadis, Nikos Mamassis & Demetris Koutsoyiannis
School of Civil Engineering
National Technical University of Athens

Presentation available online: http://www.itia.ntua.gr/1843/
Fundamental concepts

- **Storage**: Main function of reservoirs. Because of it—and unlike other works such as flood protection—reservoirs cannot be designed based on merely the marginal distribution of inflows. The time succession of inflows is of great importance and this requires a much more sophisticated probabilistic (or better stochastic) design methodology.

- **Firm yield**: Problematic (or even nonscientific) concept (because it implies elimination of risk), which, however, has constituted the design basis of most reservoirs worldwide.

- **Reliability**: The probability of achieving a target, which in the case of a reservoir is to satisfy the water demand. (Reliability = 1 – failure probability).

- **Reliable yield**: A constant withdrawal which can be satisfied for a specified reliability. It replaces the concept of firm yield.

- **Storage capacity-yield-reliability (SYR) relationship**: The relationship among these three concepts which constitutes the rational basis of reservoir design.

- **Monte Carlo or stochastic simulation**: Numerical mathematical method of solving complex problems, which was founded in Los Alamos (Metropolis and Ulam, 1949).

- **Optimization**: Mathematical methodology for locating the values of variables that maximize or minimize a function. In combination with simulation, it constitutes the rational basis for the design and management of reservoirs.

- **Hurst-Kolmogorov dynamics or long-term persistence**: Stochastic-dynamic behaviour that characterizes natural (as well as socio-economical and technological) processes. It is required to consider it in the design and management of reservoirs.

- **Generation of synthetic samples**: While stochastic simulation of a system is in principle possible if there exists a time series of observations with adequate length, in most problems observation periods are too short to support reliable results; therefore we resort to generating synthetic samples, which must have specified properties.
“Classical” methodology (Anglo-American School)

- **Ripple (1883)** Method of mass (cumulative) inflow-outflow curves: graphical method of reservoir design, based on the historical sample of inflows.

- **Hurst (1951)** Statistical study of the concept of *range* for reservoir design and its dependence on sample size. Important is the discovery of the eponymous behaviour.

- **Thomas and Burden (1963)** Sequent-peak method: tabulated version of Ripple’s method.

- **Schultz (1976)** (perhaps anticipated by others) A variant of Ripple’s method using synthetic (instead of observed) time series.

- The Anglo-American School’s methods, in spite of dominating in engineering education and handbooks for practitioners, do not have scientific consistency.

For more information on the chronicle of related research, see comprehensive review by Klemes (1987)
Systems-based methodology

- **Required**: Determination of the minimum net storage capacity $c$, so as to satisfy a constant demand $\delta$, given an inflow time series $x_t$ for a specific control horizon of length $n$, and an initial storage $s_0$.

- **Control variables**: Storage capacity $c$, (net) storage $s_t$ and losses due to spill $w_t$ for $n$ time steps ($2n + 1$ variables in total).

- **Mathematical formulation as a linear programming problem**:

  \[
  \begin{align*}
  \text{minimize} & \quad f = c \\
  \text{subject to} & \quad s_t = s_{t-1} + x_t - d - w_t \text{ for each } t = 1, \ldots, n \text{ (water balance)} \\
  & \quad s_t \leq c \text{ for each } t = 1, \ldots, n \\
  & \quad s_n = s_0 \text{ (steady state condition)} \\
  & \quad c, s_t, w_t \geq 0
  \end{align*}
  \]

- **Disadvantages**:
  - Very big number of control variables.
  - Inability to incorporate nonlinear relationships.
  - Fully deterministic formulation – reliability is not considered.
Stochastic methodology (Russian School)

- **Hazen (1914)** (American!) Introduction of the reliability concept and the SYR relationship.
- **Kritskiy & Menkel (1935, 1940) and Savarenskiy (1940)** Theoretical study and materialization of a practical methodology for reservoir design based on reliability and the SYR relationship.
- **Pleshkov (1939)** Construction of nomographs for facilitating practical application of the method.
- **Kolmogorov (1940)** Proposal of a mathematical model that represents the behaviour to be discovered 10 years after by Hurst. Kolmogorov was not involved in reservoir studies but in turbulence.
- **Moran (1954)** (Australian) Reinvention (perhaps independent) of the stochastic theory of reservoirs.

Most of these contributions, although theoretically consistent, often involve unrealistic assumptions, such as the independence of inflows over time, which make them unsatisfactory in practice.

For more information see Klemes (1987)
Which School is followed in Greece?

- Technical Universities mostly teach Anglo-American methods.
- However, consultants have been aware of the Russian School’s methods and have applied them in real-world studies.

Final design of the Iasmos dam (1971)
Differences in the behaviour of hydrological processes from that in simple random events

<table>
<thead>
<tr>
<th><strong>Roulette wheel</strong></th>
<th><strong>River discharge</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td>Discrete and finite set of possible values, {0, 1, ..., 36}</td>
<td>Infinite and continuous set of possible values, from 0 to (+\infty). The rate with which a value tends to (+\infty), for probability tending to 0, is not the minimum possible (Noah phenomenon)</td>
</tr>
<tr>
<td>Constant behaviour in time</td>
<td>Changing behaviour in time (regular seasonal changes–irregular changes in other time scales)</td>
</tr>
<tr>
<td><em>A priori</em> known probability of occurrence of each value (1/37)</td>
<td>Unknown probability distribution function which needs observations to infer</td>
</tr>
<tr>
<td>Each outcome does not depend of the previous ones</td>
<td>Each value depends on the previous values (persistence)</td>
</tr>
</tbody>
</table>
Change at different time scales in hydrological processes

Acheloos river basin upstream of the Kremasta dam

Mean daily discharge, hydrological year 1966-67 (m³/s)

Mean daily discharge, 1966-2008 (m³/s)

Mean monthly discharge, 1966-2008 (m³/s)

Mean annual discharge 1966-2008 (m³/s)
Difference in determination of probability of composite events

- Example for roulette wheel:
  What is the probability that in two consecutive throws the outcome be equal or smaller than 3?

- Analogous example for streamflow:
  If:
  
  (a) we characterize as dry any year in which the annual streamflow volume is less than or equal to 3 km$^3$, and
  
  (b) we know that the probability of a dry year is 1/10,

  what is the probability that two consecutive years are dry?

Reply: $(4/37)^2$

Reply: We need stochastic simulation to determine it
Scientific branches to enroll in order to reply the previous question

1. **Probability theory**: Foundation of calculations.
2. **Statistics**: Inference from data or induction (estimation of probability distribution function from the sample of observations).
4. **Simulation**: Numerical method that uses sampling to tackle difficult problems.

*All these are now known with the collective name* Stochastics
History of stochastic simulation (or Monte Carlo method)

- It is connected to the development of mathematics and physics in the mid-20th century as well as the development of computers.
- It was devised by the Polish mathematician Stanislaw Ulam (working in the Los Alamos team) in 1946 (Metropolis, 1989, Eckhardt, 1989).
- Immediately after, the method was used to solve neutron collision problems from the physicists and mathematicians in Los Alamos (John von Neumann, Nicholas Metropolis, Enrico Fermi) after being encoded in the first ENIAC computer.
- The “official” story of the method begins with a publication by Metropolis and Ulam (1949).
- Since the 1970s, simulation has been used in water resource problems (although the first steps were taken in the 1950s - Barnes, 1954).
- Research on stochastic methods in water resources continues and grows.
Perpetual change as seen in the Nilometer record - The Hurst-Kolmogorov behaviour

Each value is the minimum of $m=36$ roulette wheel outcomes. The value of $m$ was chosen so that the standard deviation be equal to the Nilometer series.

Nilometer data: Koutsoyiannis (2013a)
Hurst’s (1951) seminal paper

- The motivation of Hurst was the design of Nile River projects.
- However the paper was theoretical and explored numerous data sets from diverse fields.
- Hurst observed that: *Although in random events groups of high or low values do occur, their tendency to occur in natural events is greater*. This is the main difference between natural and random events.

Obstacles in the dissemination and adoption of Hurst’s finding:
- Its direct connection with reservoir storage.
- Its tight association with the Nile.
- The use of a complicated statistic (the rescaled range).
Kolmogorov (1940)

- Kolmogorov studied the stochastic process that describes the behaviour to be discovered a decade later in geophysics by Hurst.
- The proof of the existence of this process is important, because several researches, ignorant of Kolmogorov’s work, regarded Hurst’s finding as inconsistent with stochastics and as numerical artefact.

- Kolmogorov’s work did not become widely known.
- The process was named by Kolmogorov “Wiener’s Spiral” (Wienersche Spiralen) and later “Self-similar process”, or “fractional Brownian motion” (Mandelbrot and van Ness, 1968).
- Today it is called the Hurst-Kolmogorov (HK) process.
The Hurst-Kolmogorov (HK) process and its multi-scale stochastic properties

A natural process evolves in continuous time, \( t \):

\[ x(t) \]

We model it as a stochastic process in continuous time, \( t \):

\[ \tilde{x}(t) \]

... but we observe and study it by taking averages in discrete time \( i = 1, 2, \ldots \), for a convenient time scale \( k \):

\[ x_i^{(k)} := \int_{(i-1)k}^{ik} \tilde{x}(t) dt \]

<table>
<thead>
<tr>
<th>Properties of the HK process</th>
<th>At an arbitrary observation scale ( k = 1 ) (e.g. annual)</th>
<th>At any scale ( k )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Standard deviation</td>
<td>( \sigma \equiv \sigma^{(1)} )</td>
<td>( \sigma^{(k)} = \sigma / k^{1-H} ) (can serve as a definition of the HK process; ( H ) is the Hurst coefficient; ( 0.5 &lt; H &lt; 1 ))</td>
</tr>
<tr>
<td>Autocorrelation function (for lag ( j ))</td>
<td>( \rho_j \equiv \rho_j^{(1)} = \rho_j^{(k)} \approx H (2 H - 1)</td>
<td>j</td>
</tr>
<tr>
<td>Power spectrum (for frequency ( \omega ))</td>
<td>( s(\omega) \equiv s^{(1)}(\omega) \approx 4(1-H)\sigma^2 (2\omega)^{1-2H} )</td>
<td>( s^{(k)}(\omega) \approx 4(1-H)\sigma^2 k^{2H-2} (2\omega)^{1-2H} )</td>
</tr>
</tbody>
</table>

In classical statistics \( \sigma^{(k)} = \sigma/\sqrt{k} \)

All equations are power laws of scale \( k \), lag \( j \), frequency \( \omega \)

For detailed descriptions see Koutsoyiannis (2002, 2013)
Example 1: Clustering of floods

1845-90: Three floods greater than the 100-year flood in 45 years

1900-45: No flood greater than the 10-year flood in 40 years

Flood discharges of the Vltava river in Prague in the last 5 centuries (Brázdil et al., 2006)
Example 2: Annual minimum water levels of the Nile

- The longest time series of observations available (849 years).
- Hurst parameter $H = 0.87$.
- A similar value of $H$ is found from the simultaneous time series of maximum water levels and from a modern time series of annual discharge of Nile at Aswan (131 years).

For an HK process, the classical statistical estimator of the standard deviation entails bias, which has accounted for in the estimation of $H$. 

Simulation and optimization for the design and management of hydroelectric works
Example 3: The Moberg et al. proxy series of the Northern Hemisphere temperature

Suggests an HK behaviour with a very high Hurst coefficient: 

$H = 0.94$.

Estimation bias was determined by Monte Carlo simulation (200 simulations with length equal to the historical series).
Example 4: The Greenland temperature proxy during the Holocene

Reconstructed from the GISP2 Ice Core (Alley, 2000, 2004). Data from:
Example 4 (cont.): The Greenland temperature proxy on multi-millennial time scales

Example 4 (cont.): The Greenland temperature proxy on all scales

All three periods suggest an HK behaviour with a very high Hurst coefficient: $H \approx 0.94$. 

Estimation bias and 95% prediction limits were determined by Monte Carlo simulation (200 simulations with length equal to the historical series).
What we avoid in reservoir design

- Deterministic methods or pseudo-stochastic variants thereof (the Anglo-American methodology).
- Stochastic simulation methods incapable to reproduce Hurst-Kolmogorov dynamics.
- Software applications that are based on the above methods.
What we do for a preliminary reservoir design

1. We construct a SYR relationship from historical observations—if the record length is satisfactory.
   - Calculations are very simple and only require the water balance equation in the form:
     \[
     s_t = \max[0, \min(s_{t-1} + x_t - \delta_t, c)],
     \]
     where \( s_t \) is the storage at time \( t \), \( x_t \) the net inflow, \( \delta_t \) the water demand, and \( c \) the storage capacity. A failure is counted when \( s_t = 0 \).
   - The computational framework of a spreadsheet (OpenOffice, Excel) is enough.

2. We construct a «lower envelope» SYR relationship based on standardized relationships, expressed in terms of nomographs or equations and based on stochastic simulation.
   - The results give the storage capacity required for long-term (over-annual) regulation. An additional storage of about 50%-80% of annual demand must be added for sub-annual regulation (the higher percentage corresponds to irrigation reservoirs).

3. We estimate the design storage capacity by optimization considering technical, economical, and environmental data.
Typical results of the consistent method (storage-yield-reliability relationship)

Characteristic quantities
- $\mu$: mean inflow
- $\sigma$: standard deviation of inflow
- $\alpha$: reliability
- $T := 1 / (1 - \alpha)$: return period of reservoir emptying
- $\delta$: demand
- $c$: reservoir storage capacity
- $\kappa := c / \sigma$ : standardized reservoir storage capacity
- $\epsilon := (\mu - \delta) / \sigma$ : standardized mean loss

Assumptions
- Annual time scale (seasonal variation neglected) with constant withdrawal rate.
- Inflows independent identically distributed with normal distribution.

Approximate mathematical expressions (for $T > 2$ or $\alpha > 0.5$)
\[
\ln(T - 1) = 2 (\epsilon + 0.25) (\kappa + 0.5)^{0.8} \quad \text{or} \quad \ln(T - 1) = -\ln(1/\alpha - 1) = (2/\sigma^{1.8}) (\mu + 0.25\sigma - \delta) (\lambda + 0.5\sigma)^{0.8}
\]

For details see Koutsoyiannis (2005)
Extensions of results for more complex stochastic structure of inflows

Effect of skewness (Results for independent gamma distributed inflows)

- While the case presented is simple, the method is fully generic and can perform with any type of system dynamics and stochastic structure of inflows.
- While there exist in the literature different approaches (e.g. the formulation by Moran, 1954, based on Markov chains, as well as recent attempts) these involve radical simplifications (e.g. discretization of the reservoir space) and their usefulness is questionable.
- For details see Koutsoyiannis (2005).
What we do for a final reservoir design

1. We construct a SYR curve as in step 1 of the preliminary design but now using a synthetic time series (with length of thousands of years) in monthly time scale.
   - The synthetic time series should be generated with a method that reproduces HK dynamics.
   - The simplest methods reproducing HK dynamics are those by Koutsoyiannis (2003) and Langousis & Koutsoyiannis (2006); these can easily be materialized in spreadsheets (OpenOffice, Excel).
   - More sophisticated methods require appropriate software applications (e.g. Castalia).

2. We estimate the design storage capacity by optimization considering technical, economical, and environmental data.
Algorithmic application of simulation: Introduction to random numbers

- A *sequence of random numbers* is a sequence of numbers $x_i$ whose every one statistical property is consistent with realizations from a sequence of independent identically distributed random variables $x_i$ (adapted from Papoulis, 1990).

- A *random number generator* is a device (typically computer algorithm) which generates a sequence of random numbers $x_i$ with given distribution $F(x)$.

- Random number generation is also known as Monte Carlo sampling.

- Most algorithms are purely deterministic, and generate the same sequence of numbers if we start from the same initial condition, often referred to as *seed*. If we change the seed we get another sequence (more precisely another part of a periodic sequence with very large period). Yet the numbers are random because if we do not know the algorithm and the initial condition ($q_0$ or $q_{i-1}$) we cannot predict these numbers.
Generation of independent random numbers with specified distribution function

- The basis of practically all random generators is the uniform distribution in [0,1]. A typical procedure is the following:
  - We generate a sequence of integers \( q_i \) from the recursive algorithm
    \[ q_i = (kq_{i-1} + c) \mod m \]
    where \( k, c \) and \( m \) are appropriate integers (e.g. \( k = 69069, c = 1, m = 2^{32} = 4294967296 \) or \( k = 7^5 = 16807, c = 0, m = 2^{31} - 1 = 2147483647 \); Ripley, 1987, p. 39).
  - We calculate the sequence of random numbers \( u_i \) with uniform distribution in [0,1] by \( u_i = q_i / m \).
- For any probability distribution \( F(x) \) the following procedure works always (but sometimes is time demanding):
  - If \( F^{-1}(u) \) is the inverse function of \( F(x) \) and \( u_i \) are random numbers with uniform distribution in [0,1], then the required random numbers are given by
    \[ w_i = F^{-1}(u_i) \]

In spreadsheets, the function `rand()` generates random numbers with uniform distribution in [0,1] and the function `normsinv(rand())` generates random numbers with normal distribution \( \mathcal{N}(0, 1) \).
Generation of random numbers from the HK process; the Symmetric Moving Average Method

The symmetric moving average (SMA) scheme, introduced by Koutsoyiannis (2000), transforms a sequence of independent random numbers (white noise) $v_i$ to a sequence of dependent ones $x_i$ using the equation

$$x_i = \sum_{j=-q}^{q} a_{|j|} x_{i+j} = a_q v_{i-q} + \cdots + a_1 v_{i-1} + a_0 v_i + a_1 v_{i+1} + \cdots + a_q v_{i+q}$$

where $a_j$ are weights whose number $q$ is theoretically infinite but in practice is chosen finite with a large value.

In the case of the HK process (else known as fractional Gaussian noise—FGN) it is shown (Koutsoyiannis, 2002) that the weights are:

$$a_j \approx \sqrt{\frac{(2 - 2H)}{3 - 2H}} \gamma_0 \left[ |j + 1|^{H+0.5} + |j - 1|^{H+0.5} - 2 \ |j|^{H+0.5} \right]$$
Generation of synthetic samples using the Castalia software

A synthetic series with length of 1000 years for inflows at lake Hylike
The need for redesign and adaptation of management

- In a first stage, several reservoirs are designed as individual hydraulic works of a single purpose.

- In the course of their operation, increased needs require that they be complemented by new projects.
  - Characteristic example: Evinos projects to boost the water supply of Athens from Mornos river.
  - New projects were studied from the outset as components of a system rather than as individual projects (system redesign).

- In other cases, changes in social and economic priorities make it necessary to adapt their management to new (multiple) purposes.
  - The new management policy recognizes the need for a minimum ecological limit on the level of the reservoir without neglecting the importance of water supply and the economic and social benefit of irrigation and energy.

See details in Koutsoyiannis et al. (2003), Christofides et al. (2005) and Koutsoyiannis (2011).
Example: The Acheloos hydropower and irrigation hydrosystem

- 5 reservoirs in the Acheloos river system
- 2 additional reservoirs at the Thessaly area
- 1 more reservoir (Plastiras) out of the system
- 8 hydropower plants
- Conveyance network
- Main water use: Energy production
- Secondary water uses: Irrigation, Water supply
- Environmental constraints
Example: The Acheloos hydrosystem structure

- Irrigation requirements
  - Main irrigation nodes at Stratos and Mavromati (450 and 600 hm³ per year, respectively)
  - Local demand 4 hm³ per year at Pyli

- Environmental constraints
  - Minimum environmental preservation discharge at Acheloos river 1.5 m³/s downstream of Mesohora, 5 m³/s downstream of Sykia and 21 m³/s at the estuary
  - Minimum discharge downstream of Pyli and Mouzaki 0.15 m³/s
  - Additional 0.35 m³/s downstream of Pyli for the aquifer recharge

Method of choice: Parameterization-Simulation-Optimization
For more information see Nalbantis & Koutsoyiannis (1997), Koutsoyiannis & Economou (2002) for the methodology and Κουτσογιάννης (1996) for the application.
References

References (2)

- Κουτσογιάννης, Δ., Μελέτη λειτουργίας ταμιευτήρων, Γενική διάταξη έργων εκτροπής Αχελώου προς Θεσσαλία, Ανάδοχος: Ειδική Υπηρεσία Δημοσίων Έργων Αχελώου - Γενική Γραμματεία Δημοσίων Έργων - Υπουργείο Περιβάλλοντος, Υπουργείο Περιβάλλοντος, Καλαούζης, Καλαούζης, Καλαούζης, 420 σελίδες, 1996.