New insights on model evaluation inspired by the stochastic simulation paradigm (1)

EGU General Assembly 2011, Vienna, Austria, 4-8 April 2011

Session HS1.6: *Metrics and the Use of Data in Hydrology to Support Model Structure Improvement*

Andreas Efstratiadis

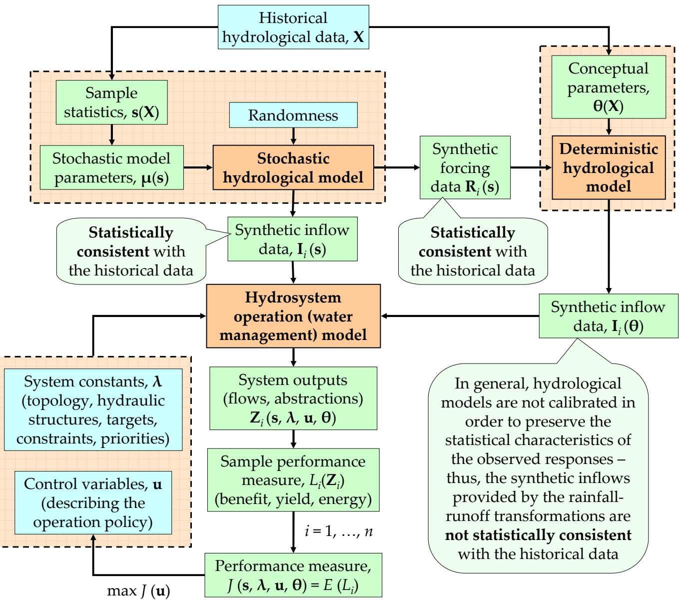
Department of Water Resources and Environmental Engineering, National Technical University of Athens, Greece (andreas@itia.ntua.gr)

1. Abstract

The working paradigm for evaluating the performance of practically any kind of mathematical model is based on metrics that assess an "average" departure between modelled outputs and observations (i.e. residuals). Yet, the outputs of hydrological, hydrogeological and climatic models are not deterministic responses against known or predictable inputs; they are stochastic variables, the interpretation of which should, consequently, be implemented in statistical terms. In addition, these processes exhibit multiple peculiarities (seasonality, long-term persistence, intermittency, skewness, spatial variability), which are rather impossible to be accounted for within a single measure (typically efficiency or other least square error expression). In this context, a comprehensive statistical framework is discussed for the evaluation of such models, seeking for the reproduction of a number of statistical characteristics of the observed data, instead of focusing to optimize an "overall" distance measure. This is inspired by the requirements of advanced stochastic simulation schemes, which are by definition built to preserve the essential statistics of the parent (i.e. historical) time series (marginal and joint statistics). This is a key concept, ensuring the generation of synthetic data that are statistically equivalent to the historical ones. The proposed framework emphasises the following issues: (a) the statistical comparison of computed and observed data at multiple time scales, to account for the variability of the modelled processes in both the short and the long term; (b) the preservation of the observed cross-correlations in multi-response calibration, to represent the interrelationship of the physical processes under study, and (c) the investigation of the model response under different stress conditions, preferably using synthetic data of appropriate length; this allows recognising structural deficiencies and irregular behaviours, which are hard to identify within the, typically short, period of observations. The above issues are analysed using examples from a number of modelling works, where initial calibration approaches, following typical hydrological practices, may result in misleading conclusions.

2. The stochastic simulation framework in water resources management and the key role of hydrological modelling

- In water resources management, the hydrosystem performance that corresponds to a specific operation policy (by means of economic benefit, safe yield, hydroelectric energy production, etc.), is by definition linked to a specific reliability or, equivalently, risk.
- The evaluation of risk requires simulation-based approaches to deal with hydrological uncertainty, thus handling all fluxes (flows, releases, abstractions, etc.) as random variables.
- In practice, stochastic simulation comprises a three-step procedure:
  - Generation of synthetic inflow time series, using a **stochastic hydrological model** that reproduces the statistical characteristics of the observed ones;
  - Running a **water management model** (typically implemented within a decision support framework) with synthetic forcing data, to simulate the hydrosystem operation;
  - Statistical analysis of model outputs to interpret the system responses in probabilistic terms.
- When either historical inflow data are not available or the natural flow regime is modified due to anthropogenic interventions, a **deterministic hydrological model** should run to provide the synthetic inflows, for given synthetic forcing data (e.g. rainfall, PET).



3. Generating statistically consistent synthetic inflows via conjunctive use of stochastic and deterministic hydrological models

- The well-known peculiarities of all hydrological processes (seasonality, long-term persistence, intermittency, skewness, spatial variability) gave rise to substantial research that resulted in numerous stochastic tools appropriate for applications in hydrosystems.
- Advanced multivariate stochastic models are designed to represent all the essential statistical characteristics of the observed data, at multiple time scales (monthly, annual, over-annual), i.e.:
  - the marginal statistics up to third order (mean, variance, skewness);
  - the joint second order statistics (auto- and cross-correlations);
  - the long-term persistence, also known as Hurst-Kolmogorov dynamics.
- The specific mathematical structure of the stochastic models and the procedures for estimating their parameters (analytical or numerical) ensure an explicit preservation of the above statistics.
- On the other hand, the typical calibration approaches for deterministic hydrological models pay few or even no attention to the overall statistical consistency of the simulated responses; thus, when models are fed with synthetic forcing data to run in stochastic simulation mode, fail to generate synthetic inflows that reproduce, in statistical terms, the hydrological regime of the historical data with satisfactory accuracy.

Conclusion: Hydrological models should preserve the observed statistics, to provide statistically equivalent synthetic inflows to decision support tools that run in stochastic simulation setting.

4. Interpretation of flow characteristics in statistical terms

Statistical metric	**Link with flow characteristics**
Mean	Seasonality of flows (monthly scale), allocation of water balance components (annual scale)
Standard deviation	Variability of basin responses under different forcing conditions
Skewness	Statistical representation of high and low events
Auto-correlation	Time-lagged components of runoff (storage, recession, baseflow, etc.)
Cross-correlation	Interrelationships between cause-effect mechanisms (e.g. rainfall-runoff) and interacted hydrological processes
Hurst coefficient	Over-year scaling behaviour of annual flow, e.g. long-term fluctuations, non-systematic trends, persistent wet and drought periods

Conclusion: The statistical characteristics of the observed flows encompass all the fundamental macroscopic information about the hydrological regime of the basin and its process interactions.

5. Accounting for the observed statistics as soft data within hydrological calibration: a means to reduce parameter uncertainty?

- In hydrological modelling, the principle of consistency (i.e. building models that represent as close as possible the behaviour of the physical system) may be in contrast with the principle of optimality (i.e. ensuring the best fitting of the simulated responses to observations).
- The efficiency index, which is typically used as the overall evaluation criterion in calibrations, is rather insufficient to capture the multiple aspects, at multiple time scales, of the flow regime that are systematically represented in the statistical characteristics of the historical data.
- A consistent (and at the same time robust) calibration framework should ask for reproducing, as close as possible, not the observations themselves but their statistical properties, which is in fact the principal requirement of all stochastic hydrological schemes.
- This approach significantly increases the information contained in calibration, since instead of optimizing against a single criterion (e.g. efficiency), multiple objectives are now accounted for, by means of statistical metrics; for instance, assuming a lumped model that is calibrated against monthly runoff, the number of the related monthly statistical metrics (mean, standard deviation, skewness, first order autocorrelation and cross-correlation with rainfall) are 12 × 5 = 60.
- This artificial increase of information is rather related to the "soft" data concept, given that the amount of hard data, based on observations, remains constant.

Conclusion: The use of multiple statistical metrics as soft data within calibration, which represent different aspects of the basin responses, helps reducing parameter uncertainty (which is straightforwardly linked to equifinality), thus providing schemes of improved predictive capacity.

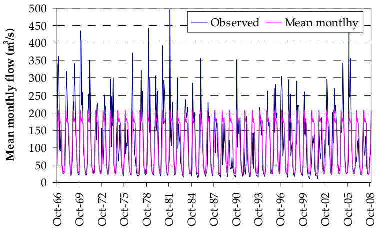


Fig 1: An elementary prediction using mean monthly values, which ensures 54% efficiency (Acheloos basin at Kremasta, West Greece)

New insights on model evaluation inspired by the stochastic simulation paradigm (2)

EGU General Assembly 2011, Vienna, Austria, 4-8 April 2011

Session HS1.6: *Metrics and the Use of Data in Hydrology to Support Model Structure Improvement*

Andreas Efstratiadis

Department of Water Resources and Environmental Engineering, National Technical University of Athens, Greece (andreas@itia.ntua.gr)

6. Case study: Lumped simulation of Acheloos flows at Kremasta dam

- Basin characteristics: wet climate, extended areas of low permeability (domination of flysch)
- Historical data: monthly precipitation, potential evapotranspiration and runoff at the basin outlet (upstream of Kremasta dam) for 22 years (Oct. 1970 to Sep. 1992)
- Modelling approach: lumped, two interconnected tanks (representing the unsaturated and saturate zones), three runoff components (direct, overland, baseflow), four parameters (direct runoff percentage, soil moisture capacity, recession rate for percolation and baseflow).
- Calibration approaches:
  1. Maximization of efficiency (observed vs. simulated runoff, single criterion);
  2. Minimization of average departure from all essential statistics (60 values, standardized).
- The first approach, which is typical in hydrological calibration, although ensures an efficiency of 86%, it fails to reproduce significant properties of the basin mechanisms during summer, which are imprinted in statistics such as skewness and cross-correlations with rainfall.
- The second approach required two more parameters to be added, to better reproduce the exceptionally high nonlinear behavior of the summer runoff, while the overall efficiency was slightly only affected (from 86% to 83%).
- Paying attention to the statistical characteristics of flow was a guide for an improved model structure and a physically-consistent calibration.
- The increase of the number of parameters did not contrast the principle of parsimony, due to the augmented information offered to calibration.

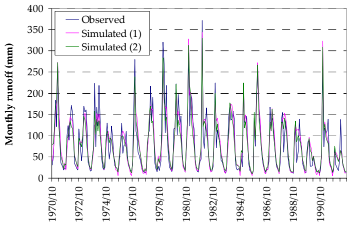
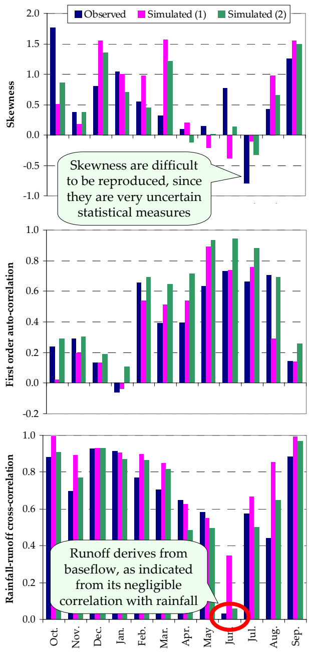


Fig 2: Comparison of observed and simulated runoff data with calibration approaches 1 and 2.



Skewness are difficult to be reproduced, since they are very uncertain statistical measures

Runoff derives from baseflow, as indicated from its negligible correlation with rainfall

Fig 3: Comparison of skewness coefficients, first order autocorrelations and cross-correlations with rainfall, between observed and simulated runoff.

7. What about the preservation of statistics in long-term?

- A satisfactory efficiency achieved at the monthly time scale does not necessary ensure a reliable reproduction of the observed flows at the annual and over-annual time scales, which are very important, especially when dealing with hydrosystems of overyear regulation, where decisions have long-term impacts to the water resources management and the system reliability.
- **Case study**: monthly simulation of Boeoticos Kephisos basin, employing various lumped approaches with different modeling structures, up to 9 parameters (Zygos model)
- **Basin characteristics**: semi-arid climate, extended areas of high permeability (domination of karst), modified status (due to both surface and groundwater abstractions), extended yet unknown groundwater losses to the sea
- **Historical data**: monthly precipitation, potential evapotranspiration, runoff at the basin outlet, water demand for irrigation and water supply for 96 years (Oct. 1907 to Sep. 2003)
- Two approaches are detected with the following efficiency values at three time scales:

	Monthly	**Annual**	**20-y moving average**
Model 1	0.765	0.646	0.542
Model 2	0.676	0.682	0.833

Reproduction of over-year scaling behaviour

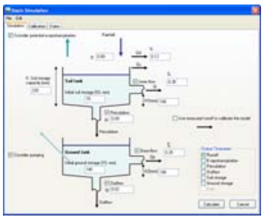


Fig 4: Main form of the Zygos software, supporting multiple modelling configurations.

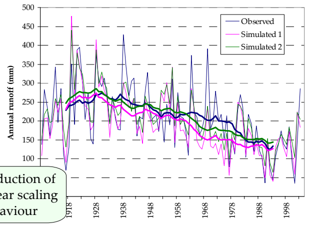


Fig 5: Comparison of observed and simulated runoff at the annual and 20-year time scales.

8. Stochastic simulation as a crash-test for model evaluation

- The typical "split-sample" procedure for evaluating hydrological models (i.e., calibration and validation based on historical data) may hide severe structural deficiencies, given that this approach is too much depended on the historical data and the related uncertainties (e.g. data errors, unknown initial and boundary conditions, shortcomings of optimization, etc.).
- Stochastic simulation allows for examining the model responses under realistic hypothetical conditions, using synthetic forcing data that are statistically consistent with the historical ones.
- This allows checking for inconsistencies of the modelling structure or misuse calibration practices, by means of unreasonable trends, abnormal patterns and other "hidden" artefacts.
- Given that it is usually impossible to recognize such issues in the short period encountered in calibration and validation, this kind of empirical evaluation strategy constitutes an essential supplementary verification of the model credibility.
- Obviously, in a stochastic simulation setting, it is impossible to use quantitative criteria to assess the model performance, as in calibration; thus, the evaluation is rather based on the grounds of common sense, taking advantage of the hydrological experience.

9. Case study: Testing the Hydrogeios model against synthetic inputs for the stochastic simulation of Boeoticos Kephissos hydrosystem

- **Model components**:
  - *Surface system*: 15 sub-basins, 15 river segments, 6 hydrological response units (6 × 6 = 36 parameters);
  - *Groundwater system*: 38 cells, 4 outflows, 6 springs, 53 boreholes (16 parameters);
  - *Water management network*: 7 irrigated areas (represented as nodal demands), 7 borehole groups, 16 aqueducts, multiple water uses fulfilled through both surface water abstractions and pumping.
- **Hybrid calibration components**:
  - Efficiency and bias of monthly hydrographs at the basin outlet and downstream of six karst springs;
  - Penalties for not reproducing flow intermittency;
  - Penalties for the generation of unrealistic trends regarding the groundwater levels.
- **Evaluation test through stochastic simulation**:
  - Generation of synthetic point rainfall data of 1000 years length through a multivariate stochastic model (Castalia), preserving the essential statistics and the Hurst phenomenon;
  - Spatial aggregation of point rainfalls;
  - Formulation of two simulation states:
    - Steady state simulation of 1000-year length;
    - Terminating simulation, 100 stress scenarios of 10-year length, same initial conditions;
  - Formulation of three management scenarios:
    - *Zero abstractions*: represents a hypothetical "unmodified" state of the system;
    - *Normal abstractions*: represents the actual demand policy for irrigation;
    - *Intensive abstractions*: pumping of 50 hm^3/y for water supply from the Vassilika-Parori boreholes (typically used for emergency);

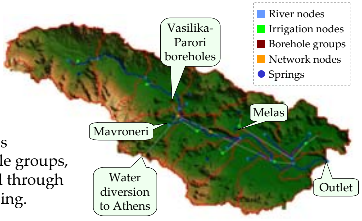
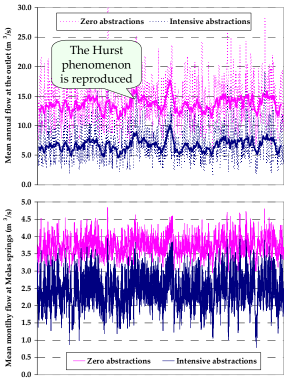


Fig 6: The Boeoticos Kephisos hydrosystem.



The Hurst phenomenon is reproduced

Fig 7: Simulated flows at the outlet (annual scale, up) and at Melas springs (monthly scale, down).

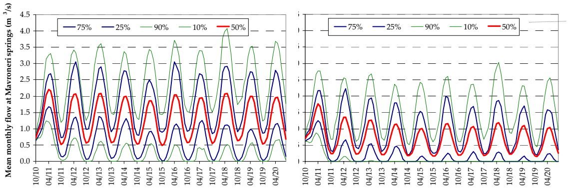


Fig 8: Simulated monthly flows at Mavroneri springs for the 10-year horizon (mean, 50% and 80% prediction limits) under the normal (left) and the intensive (right) abstraction policy.

Under intensive pumping for the water supply of Athens, the risk of flow interruption is notably increased, while under the actual abstraction policy the karst system generates, as expected, steady-state responses (i.e. spring runoff).

Further information

- Efstratiadis, A., and D. Koutsoyiannis, One decade of multiobjective calibration approaches in hydrological modelling: a review, *Hydrological Sciences Journal*, 55(1), 58-78, 2010.
- Nalbantis, I., A. Efstratiadis, E. Rozos, M. Kopsiafti, and D. Koutsoyiannis, Holistic versus monomeric strategies for hydrological modelling of human-modified hydrosystems, *Hydrology and Earth System Sciences*, 15, 743-758, 2011.

The poster is available online at http://www.itia.ntua.gr/en/docinfo/1116/