

Uncertainty assessment of future hydroclimatic predictions: Methodological framework and a case study in Greece

Konstantine P. Georgakakos, Hydrologic Research Center, San Diego, CA., USA
 D. Koutsoyiannis and Andreas Efstratiadis, Department of Water Resources, National Technical University of Athens, Greece

1. Abstract

A stochastic framework for climatic variability and uncertainty is presented, based on the following premises: (1) a climatic variable is not a parameter constant in time but rather a variable representing the long-term (e.g. 30-year) time average of a certain natural process, defined on a fine scale; (2) the evolution of climate is represented as a stochastic process; (3) the distributional parameters of the process, marginal and dependence, are estimated from an available sample by statistical methods; (4) the climatic uncertainty is the result of at least two factors, the climatic variability and the uncertainty of parameter estimation; (5) a climatic process exhibits scaling behaviour, also known as long-range dependence or the Hurst phenomenon; (6) due to dependence, the uncertainty limits of the future are influenced by the available observations of the past. The last two premises differ from classical statistical considerations and produce uncertainty limits that are eventually much wider than those predicted by classical statistics. A combination of analytical and Monte Carlo methods is developed to determine uncertainty limits for the nontrivial scaling case. The framework developed is applied to a catchment in Greece with century-long temperature, rainfall and runoff data. The uncertainty limits are then compared to deterministic projections up to the year 2050, obtained for several scenarios from several climatic models combined with a hydrologic model.

5. Probabilistic quantification of hydroclimatic uncertainty

Parameter (or sampling) uncertainty: The estimate of a climatic parameter β (e.g. the mean annual rainfall at a certain location) has some uncertainty due to limited observation record $X_{0:n} = [X_1, \dots, X_n]$; this is defined assuming that $X_{0:n}$ is realization of a vector of identically distributed random variables $X_{0:n} = [X_1, \dots, X_n]$ and determined in terms of confidence limits for confidence coefficient α :

$$P(L \leq \beta \leq U) = \alpha$$

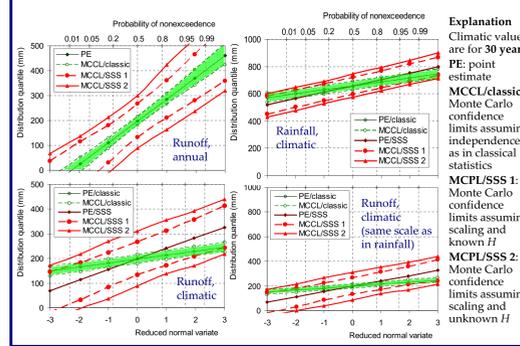
where L and U are estimators (functions of $X_{0:n}$) of the lower and upper limits with estimates l and u (functions of $x_{0:n}$)

Uncertainty due to variability: A climatic variable (e.g. the mean annual rainfall at a certain location for a 30-year period) in addition to parameter uncertainty, has also the uncertainty due to (natural) temporal variability; this is determined in terms of distribution quantiles for confidence coefficient α' :

$$P(y_L < X < y_U) = \alpha'$$

The quantiles y_L and y_U are parameters and entailed parameter uncertainty determined as above

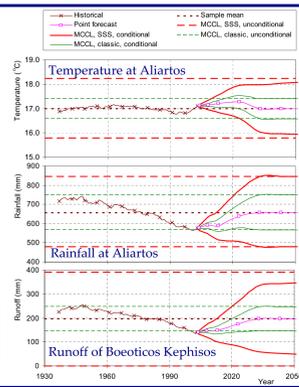
9. Demonstration of differences: classical vs. SSS case



13. Resulting SSS conditional confidence limits and comparison to classical statistical estimates

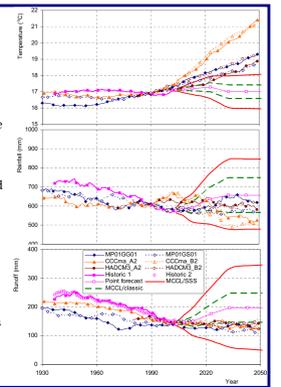
Historic climate and (conditional) point estimates and confidence limits of future climate (for $\alpha = \alpha' = 95\%$ and climatic time scale of 30 years)

Comparison shows that SSS confidence bands are 2.5-3 times wider than the classical statistical bands



17. Comparison of GCM projections with climatic confidence limits

- All GCM projections agree that sooner or later the temperature will depart from the natural uncertainty limits
- In contrast, GCM rainfall and the resulting runoff fall within the SSS uncertainty limits for the whole examined period up to 2049; this means that traces such as the ones projected by the GCM can be readily obtained by stochastic simulation assuming stationarity
- This more or less harmonizes with some earlier studies (e.g. a comprehensive climatic study for United States by Georgakakos and Smith, 2001)
- The proposed stochastic method suggests that the uncertainty of runoff, even assuming natural variability, is in fact much larger than projected by GCM



2. Rationale and main hypotheses

- A climatic characteristic at a certain place is not a parameter, constant in time, but rather a variable representing the long-term (e.g. 30-year) time average of a certain natural process, defined on a finer scale
- The evolution of climate is represented as a stochastic process; if X_t denotes an atmospheric or land surface variable at the annual scale, then climate is represented by the moving average process, with a typical time window $k = 30$: $X_t^{(k)} := (X_t + \dots + X_{t-k+1})/k$
- The distributional parameters of the process, marginal and dependence, are estimated from an available sample by statistical methods
- The climatic uncertainty is the result of at least two factors, the climatic variability and the uncertainty of parameter estimation (sampling uncertainty)
- A climatic process exhibits scaling behaviour, also known as long-range dependence, multi-scale fluctuation or the Hurst phenomenon (Koutsoyiannis, 2002, 2003, 2005); this has been verified both in long instrumental hydrometeorological (and other) series and proxy data
- Due to the long-range dependence, the uncertainty limits of the future are much wider than predicted by classical statistics and are also influenced by the available observations of the past
- Future climate trends suggested by general circulation models (GCM) should be viewed in the framework of uncertainty bands estimated by statistics accounting for scaling behaviour

6. The scaling property (the Hurst phenomenon)

In classical statistics we have the fundamental law $\sigma^{(k)} = \sigma/k^{1/2}$ where $\sigma^{(k)} := \text{Std}[X^{(k)}]$ is the standard deviation of the random variable $X^{(k)}$ at scale k and $\sigma = \sigma^{(1)}$ is the standard deviation of each X_t

As shown in the figure, the historical time series follow not this law, but the generalized law $\sigma^{(k)} = \sigma/k^{1-H}$ where the constant H is known as the Hurst coefficient

The latter law:

- defines the time scaling behaviour (or Hurst behaviour);
- defines a stochastic process known as simple scaling stochastic (SSS) process (or stationary interval of a self similar process)

10. Conditional uncertainty (for the observed past)

For observed past ($X_{0:n} = x_{0:n}$), the parameters whose conditional confidence limits are sought are the distribution quantiles $y_{i:n}^{(k)} = E[X_{i:n}^{(k)} | x_{0:n}] + \text{Std}[X_{i:n}^{(k)} | x_{0:n}]$ and $y_{i:n}^{(k)}$ that is given by a similar expression, where $X_{i:n}^{(k)} | x_{0:n} = \frac{i}{k} X_{i:n}^{(k)} | x_{0:n} + (1 - \frac{i}{k}) X_{i:n}^{(k)} | x_{0:n}$, $i \geq k$ and $X_{i:n}^{(k)} | x_{0:n} = \frac{i}{k} X_{i:n}^{(k)} | x_{0:n} + (1 - \frac{i}{k}) x_0^{(k)}$, $i < k$

The final equations for calculating the relevant quantities, derived in Koutsoyiannis et al. (2006) are

$$E[X_{i:n}^{(k)} | x_{0:n}] = \left[\frac{i}{k} \phi_{i:n}(H) + \left(1 - \frac{i}{k}\right) \phi_{i:n}(H) \right] \mu + \left[\frac{i}{k} \left[1 - \phi_{i:n}(H)\right] + \left(1 - \frac{i}{k}\right) \left[1 - \phi_{i:n}(H)\right] \right] x_0^{(k)}$$

$$E[X_{i:n}^{(k)} | x_{0:n}] = \frac{i}{k} \phi_{i:n}(H) \mu + \left(1 - \frac{i}{k}\right) x_0^{(k)}$$

$$\text{Std}[X_{i:n}^{(k)} | x_{0:n}] = k^{H-1} \sigma \sqrt{\psi_{i:n}(H)}, \quad i \geq k$$

$$\text{Std}[X_{i:n}^{(k)} | x_{0:n}] = \frac{i^H}{k} \sigma \sqrt{\psi_{i:n}(H)}, \quad i < k$$

where $\phi_{i:n}(H)$ is a function of the lag i , the sample size n and the Hurst coefficient H , whereas $\psi_{i:n}(H)$ is a function of the lag i and the Hurst coefficient H ; both are defined in Koutsoyiannis et al. (2006) and demonstrated in panel 11

14. Scenario-based analysis of uncertainty: IPCC scenarios, models and data sets

- IPCC scenarios for future climatic projections by GCM
 - SRES A2: high population growth (15.1 billion in 2100), high energy and carbon intensity, and correspondingly high CO₂ emissions (concentration 834 cm³/m³ in 2100)
 - SRES B2: lower population (10.4 billion in 2100), energy system predominantly hydrocarbon-based but with reduction in carbon intensity (CO₂ concentration 601 cm³/m³ in 2100)
 - IS92a (older): in between the above two (population 11.3 and CO₂ concentration 708 cm³/m³ in 2100)
- GCM coupled atmosphere-ocean global models
 - ECHAM4/OPYC3: developed in co-operation between the Max-Planck-Institute for Meteorology and Deutsches Klimarechenzentrum in Hamburg, Germany; mean resolution 2.81° both in latitude and longitude (a total of 64 latitudes × 128 longitudes)
 - CGCM2: developed at the Canadian Centre for Climate Modeling and Analysis; resolution 3.75° both in latitude and longitude (a total of 48 latitudes × 96 longitudes)
 - HADCM3: developed at the Hadley Centre for Climate Prediction and Research; resolution 2.5° in latitude and 3.75° in longitude (a total of 73 latitudes × 96 longitudes)
- GCM outputs (from http://ipcc-ddc.cru.uea.ac.uk/ddc_gmdata.html)
 - MPIRG01: output of ECHAM4/OPYC3 with historical inputs for 1860-1989 and inputs from IS92a beyond 1990
 - MPIRG01B2: same as in 1 but also considering the sulphate concentration
 - CCGMA_A2: output of CGCM2 with historical inputs for 1900-1989 and inputs from scenario A2 beyond 1990
 - CCGMA_B2: same as in 3 but for scenario B2 beyond 1990
 - HADCM3_A2: output of CGCM2 with historical inputs for 1950-1989 and inputs from scenario A2 beyond 1990
 - HADCM3_B2: same as in 5 but for scenario B2 beyond 1990

18. Further validation

- Further validation of the proposed method is done against the potential arguments that:
 - the 20th century data used in this study are already affected by anthropogenic influences,
 - the natural variability would be less than observed in the 20th century, and
 - as a consequence, the uncertainty limits estimated by the proposed method are artificial (not representing the natural variability) and too wide
 - To put light to these arguments, a longer data set, not related to the case study, was used: the mean annual temperature record of Berlin/Tempelhof, one of the longest instrumental meteorological records going back to 1701 (with missing data before 1756)
 - The data of the period 1908-2003 (as in the Boeotikos Kephisos case) were used to estimate the parameters of the scaling model ($H = 0.78$) and then climatic hindcasts were calculated in terms of conditional point estimates and confidence bands (for $\alpha = \alpha' = 95\%$ and climatic time scale of 30 years)
 - This was done for both the scaling and the classic statistical model; the non used part of the series was compared with the confidence limits
-

3. Test case and data sets

- Boeotikos Kephisos River basin: a closed basin (i.e. without outlet to the sea), with an area of 1955.6 km², mostly formed over a karstic subsurface; an important basin in Greece, part of the water supply system of Athens whose history, as regards hydraulic infrastructure and management, extends backward to at least 3500 years
- Data availability extends for about 100 years (the longest data set in Greece) and modelling attempts with good performance have already been done on the hydrosystem (Rozos et al., 2004).
- The relatively long records made possible the identification of the scaling behaviour of rainfall and runoff (Koutsoyiannis, 2003)

Sample statistic	Temperature (C)	Rainfall (mm)	Runoff (mm)
Size, n	96	96	96
Mean, $x_0^{(k)}$	17.0	658.4	197.6
Standard deviation, s	0.72	158.9	87.6
Variation, $C_v = s/m$	0.04	0.24	0.44
Skewness, C_s	0.34	0.44	0.36
Lag-1 autocorrelation, r_1	0.31	0.10	0.34
Hurst coefficient, H	0.72	0.64	0.79

Sample statistics of the three long time series of the case study on an annual basis

7. The importance of the scaling behaviour in typical hydrological tasks

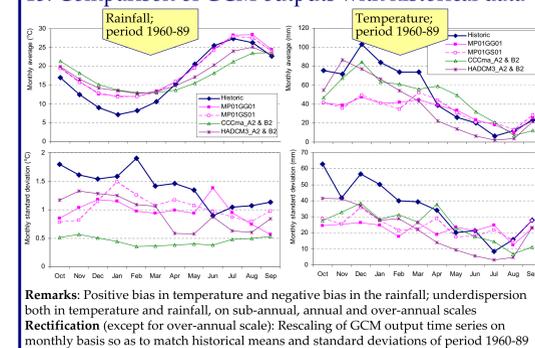
- In the observed 96-year flow record of Boeotikos Kephisos, there are multi-year periods of high flows and low flows (persistent droughts, such as a recent drought that lasted 7 years); this is observed in flow records of other rivers as well
- If such behaviour is modelled with classical statistics, return periods of 10³-10⁶ years are obtained, whereas for the given record length we would expect return periods of the order of 10² years
- In contrast, SSS statistics (Koutsoyiannis, 2003) give reasonable results (close to expectation)

Return periods of the minimum and maximum of the average, over scale $k = 1$ to 10 years, runoff of Boeotikos Kephisos from the 96-year runoff record; the return periods were calculated assuming that the distribution is normal for all scales and that the standard deviation over scale k is given by the classical and SSS laws (panel 6), respectively

11. Demonstration of quantities involved in the estimation of conditional climatic confidence limits

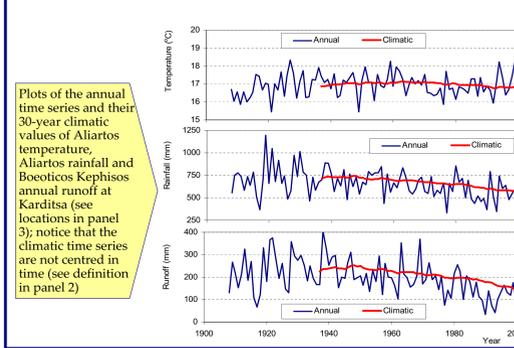
- The function $\phi_{i:n}(H)$ was evaluated for sample size $n = 100$ (rounding off 96, which is the historical sample size)
- Both $\phi_{i:n}(H)$ and $\psi_{i:n}(H)$ were evaluated numerically for different values of i and H and then approximate analytical expressions were established, which are $\phi_{i:n}(H) = 1 - (2H - 1)^2 [1 - c_2(1 - H)]$, $c_1 = 0.75 + 0.1 \ln i$, $c_2 = 2 - 3.3 \exp[-(0.18 \ln i)^2]$ and $\psi_{i:n}(H) = 1 - (2H - 1)^2 \ln^2 [1 - (2 - 1.28 / i^{0.25})(1 - H)]$

15. Comparison of GCM outputs with historical data



Remarks: Positive bias in temperature and negative bias in the rainfall; underdispersion both in temperature and rainfall, on sub-annual, annual and over-annual scales
Rectification (except for over-annual scale): Rescaling of GCM output time series on monthly basis so as to match historical means and standard deviations of period 1960-89

4. Observed behaviour in the test basin



8. Unconditional uncertainty for the SSS case

- For confidence coefficient α' , the distribution quantiles of $X_{0:n}^{(k)}$ defining the uncertainty due to variability are $y_{0:n}^{(k)}$ and $y_{0:n}^{(k)}$ where $b = (1 - \alpha')/2$, $a = (1 + \alpha')/2$; for the SSS case assuming normal distribution, these are given as $y_{0:n}^{(k)} = \mu + \zeta_b \sigma / k^{1-H}$ where ζ_b is the b quantile of the standard normal distribution; a similar expression is obtained for $y_{0:n}^{(k)}$
- The parameter uncertainty is due to uncertainty in the estimation of μ , σ and H
- Assuming that the true values of σ and H are known without uncertainty (the former being equal to the sample estimate s of standard deviation), the following semi-analytical expressions have been derived for the upper and lower α -confidence limits $u(y_{0:n}^{(k)})$ and $l(y_{0:n}^{(k)})$ for the SSS case (Koutsoyiannis, 2003): $u(y_{0:n}^{(k)}) = x_0^{(k)} + \zeta_a s / k^{1-H} \pm \zeta_{1+\alpha/2} s$ and $l(y_{0:n}^{(k)}) = x_0^{(k)} - \zeta_b s / k^{1-H} \pm \zeta_{1-\alpha/2} s$ where $\zeta_b = \frac{1}{n} \sqrt{\frac{\phi(n, H)}{2n^{H-1} (1 - \frac{\zeta_b^2}{n})}}$, $\phi(n, H) = (0.1 + 0.8) \frac{0.088(4H^2 - 1)^2}{n}$
- It can be verified that if $H = 0.5$ (independence case), these switch to known expressions in classical statistics; as H grows away from 0.5 the differences from the classical statistics increase drastically
- In the more realistic case that all μ , σ and H are unknown and estimated from the sample, the confidence limits can be estimated numerically by Monte Carlo simulation

12. Variation of conditional statistics with lead time

- In the calculation of the conditional mean, three terms involved (panel 10)
 - Term 1 - coefficient of the true mean μ ; it is an increasing function of the lead time i
 - Term 2 - coefficient of the sample mean $x_0^{(k)}$; it is an increasing function of the lead time i up to $i = k = 30$ and then becomes a decreasing function
 - Term 3 - the last (constant) term in the second equation in panel 10: it is a decreasing function of i and vanishes off at $i = k = 30$
- Even for lead time as high as 100, the influence of the known sample average is significant and the influence of the unknown true mean is smaller than 60%; this has an attenuating effect to the width of the confidence band
- The conditional standard deviation increases rapidly with lead time, reaching 85% at $i = k = 30$; then it increases gradually and becomes 93% at a lead time 100

Ratios of conditional mean (each of its three additive components) and standard deviation to the unconditional quantities; the true parameters were assumed equal to their sample estimates

16. Generation of runoff from GCM outputs

- The hydrologic response that is used in this climatic study is runoff but, in order to estimate it, the entire water cycle had to be simulated
- Given the great extend of karst, both surface and ground water processes had to be modelled simultaneously
- Given that the basin is not in natural condition, the model had to take into account both natural processes and anthropogenic influences on the catchment
- All these requirements were handled with an advanced hydrologic modelling scheme, called Hydrogeos (Efstratiadis et al., 2005)
- As shown in the diagram, the model performance in comparison to historical data is excellent, even though its calibration was done on a very small (6-year) period with detailed data (Efstratiadis, 2006)

20. References

- Efstratiadis, A. (2006). Strategies and algorithms for multicriteria calibration of complex hydrological models, *Presentation of research activities of the Department of Water Resources, Department of Water Resources, Hydraulic and Maritime Engineering - National Technical University of Athens, Athens*
- Efstratiadis, A., E. Rozos, A. Koukouvios, I. Nalbantis, G. Karavokiros, and D. Koutsoyiannis (2005). An integrated model for conjunctive simulation of hydrological processes and water resources management in river basins, *2nd General Assembly of the European Geosciences Union, Geophysical Research Abstracts, Vol. 7, Vienna, (3360), European Geosciences Union*
- Georgakakos, K.P. and D.E. Smith (2001). Soil moisture tendencies into the next century for the conterminous United States, *Journal of Geophysical Research - Atmosphere*, 106(D21), 27367-27382
- Koutsoyiannis, D. (2002). The Hurst phenomenon and fractional Gaussian noise made easy, *Hydrologic Sciences Journal*, 47(4), 573-595
- Koutsoyiannis, D. (2003). Climate change, the Hurst phenomenon, and hydrologic statistics, *Hydrologic Sciences Journal*, 48(1), 3-24
- Koutsoyiannis, D. (2005). Uncertainty, entropy, scaling and hydrologic stochasticity, 2. Time dependence of hydrologic processes and time scaling, *Hydrologic Sciences Journal*, 50(3), 405-426
- Koutsoyiannis, D., A. Efstratiadis and K. P. Georgakakos (2006). Uncertainty assessment of future hydroclimatic predictions: A comparison of probabilistic and scenario-based approaches (under review)
- Rozos, E., A. Efstratiadis, I. Nalbantis, and D. Koutsoyiannis (2004). Calibration of a semi-distributed model for conjunctive simulation of surface and groundwater flows, *Hydrologic Sciences Journal*, 49(5), 819-842