How to adjust the fGn stochastic model for statistical bias when handling a single time series; application to annual flood inundation

Panayiotis Dimitriadis[1*], Naya Gournary[1], Amalia Petsiou[1], and Demetris Koutsoyiannis[1**]

[1] National Technical University of Athens; Heroon Polytechneiou 5-9, Athens 15773, Greece
[*]corresponding author
[**]presenting author
pandim@itia.ntua.gr

Abstract

A simple scaling model known as the fractional Gaussian noise is often chosen for the description of several annual (and of larger) scale hydroclimatic processes exhibiting the Hurst phenomenon. An important characteristic of such model is the induced large statistical bias, i.e. the deviation of a statistical characteristic (e.g. variance) from its theoretical discretized value. Most studies in literature perform stochastic modelling by equating the sampling second order dependence structure with the expected value of the estimator of a stochastic model. However, this is justified only when many realizations (i.e. many time series) of a single process are available. In case where we have a single realization we should model the mode estimator of the dependence structure of the desired stochastic model instead, otherwise we may overestimate the extremeness of a realization, e.g. flood event. In this study, we show an innovative way of handling the statistical bias for an fGn process when analyzing one time series. Particularly, we conduct a thorough Monte-Carlo analysis based on the climacogram (i.e., marginal distribution of a scaled process, with focus on the second central moment of variance that is shown to be the least uncertain from the rest central moments) of an fGn process and we propose to equate the 25% quartile (and not the expected) value of the modeled climacogram with the sampling one to correctly adjust the model for bias.

Keywords: flood inundation; fractional-Gaussian-noise; Hurst phenomenon; mode climacogram; statistical bias; time series analysis

1 Introduction

Many hydrological and geophysical phenomena cannot be adequately simulated using solely deterministic processes. The debate of long term dependence structure in geophysical processes rather than short term has raised the scientific interest in the last decades [1]. Also, normally distributed processes that are of high significance in Hydrology and beyond, since the annual (or larger) scale of hydroclimatic processes (whose distribution is very close to the Gaussian one based on the Central Limit Theorem) is often used for climatic analyses and water resources management. A simple stochastic model that incorporates both the power-law drop of correlation vs. scale (else known as long-term persistence if viewed through lags [2] or Hurst-Kolmogorov behaviour -HKb- if viewed through scales [3]) and Gaussian distribution in state is known as the fractional Gaussian noise [4].

The detection of long term persistence is sometimes difficult since it requires many data in order to be adequately examined. In order to estimate the HKb a stochastic metric is employed known as the climacogram, i.e. variance of the time-averaged process over averaging time scale [5], which is preferred over the autocorrelation and power spectrum due to its smaller statistical uncertainty [6]. The analysis presented here can be easily applied to the autocorrelation, power spectrum or any other stochastic metric of the dependence structure. An important note is that the concept of climacogram can be applied to other type of moments such as the K-moments introduced and applied in [7] (or other such as raw, L-moments etc., as mentioned in [1]), but also to other order of moments (such as the mean, skewness, kurtosis etc.). Nevertheless, it has been shown that for HK processes the 2nd-order moment exhibits the lowest statistical uncertainty as compared to the 1st moment (mean) and higher-order ones, as illustrated in [8] through a cautious sensitivity analysis of an fGn process for the marginal (1st) scale, as investigated in [6,9] through the sample distribution and confidence intervals of the climacogram vs. scale (metrics that obviously are related to high-order climacograms), and as employed in a grid-turbulence application in [8].

An important characteristic of the fGn model is the induced large statistical bias, i.e. the deviation of a statistical characteristic (e.g. variance) from its theoretical discretized value [1]. Most studies in literature perform stochastic modelling by equating the sampling second-order dependence structure with the expected value of a stochastic model. However, this is justified only when many realizations (i.e. many time series) of a single process are available and thus, the mean of realizations can be validly estimated for each scale. In case where a single realization (i.e. a single time series of a physical process) is available the mode dependence structure of the desired stochastic model should be modeled instead of its expectation. In this study, we show an innovative way (based on the Monte-Carlo technique) of handling the statistical bias for an fGn process when solely one time series is known. This method can be applied to other distributions and dependence structures (see in [8] for such an example in daily precipitation).

2 Methodology

The fGn process is defined as follows:

$$\left(\underline{x}_i^{(k)} - \mu\right) =_d k^{2(1-H)}\left(\underline{x}_j - \mu\right) \tag{1}$$

where x is the process (underlined quantities denote random variables), k is the scale, $\underline{x}^{(k)} = \int_0^k \underline{x}(t)dt$ is the time-averaged process at scale k, H is the Hurst parameter, μ is the mean value of the process and subscript d denotes equality in distribution, and i and j are subscripts.

The (second-order) climacogram is defined as follows:

$$\gamma(k) = \mathrm{Var}\left[\underline{x}^{(k)}\right] / k^2 \qquad (2)$$

It can be proven that in the case of a Gaussian white noise process (i.e. $H = 0.5$) the climacogram follows the chi-square distribution, whereas the Hurst parameter increases so does the skewness of the distribution and the climacogram is close to the two-parameter gamma distribution (which is a generalization of the chi-square distribution).

In the next section, we perform thorough Monte-Carlo experiments by generating fGn time series and we estimate the climacogram for a wide range of Hurst parameters, sample lengths and scales. For the generating algorithm we use the symmetric moving average (SMA) scheme [10] that is defined as follows:

$$\underline{x}_i = \sum_{i=-l}^{i=l} a_{|j|} V_{i+j} \qquad (3)$$

where l is the number of correlations we wish to preserve (usually equals the process length), V is a N(0,1) white noise process and the coefficients a_j can be calculated analytically for an fGn process by [10]:

$$a_j = \frac{\Gamma(2H+1)\sin(\pi H)\gamma(1)}{\Gamma^2(2H+1)(1+\sin(\pi H))}\left(|j+1|^{H+0.5} + |j-1|^{H+0.5} - 2|j|^{H+0.5}\right) \qquad (4)$$

where $\gamma(1)$ is the process variance and $\Gamma(x)$ the gamma function.

3 Monte-Carlo analysis

Here, we present the results from the Monte-Carlo experiments over the fGn model for a wide range of Hurst parameters. In particular, we produce through the SMA model an adequate number (N) of synthetic timeseries required for the sample mean value to reach the expected one at scale k = 10%n [6], with a fixed standardized error of 1‰. In this way, the mode is also preserved with an even smaller error since the mode value is more probable to occur than the mean one (raising a dispute on using the expression "expected" for the mean rather than the most expected value which is the mode by mathematical definition). From preliminary analysis we find that N is around 107/n2-H. The expectation of the climacogram estimator for the 2nd central moment with an unknown mean is given by [3]:

$$E[\underline{\hat{\gamma}}(k)] = (\gamma(k) - \gamma(n))/(1 - k/n) \qquad (5)$$

Note that the sample mode cannot be estimated algebraically as in the case of other metrics such as the mean, variance etc. The sample mode can be easily estimated from data if a fixed number of accuracy (in terms of decimal places) is set. Under this concept, we here choose three digits of accuracy and we estimate the most probable value from the timeseries after rounded up each value of the timeseries to their third decimal place. The sample mode value (M) for the first scale as well as its difference to the expected one can be seen in Fig. 1 and 2, respectively.
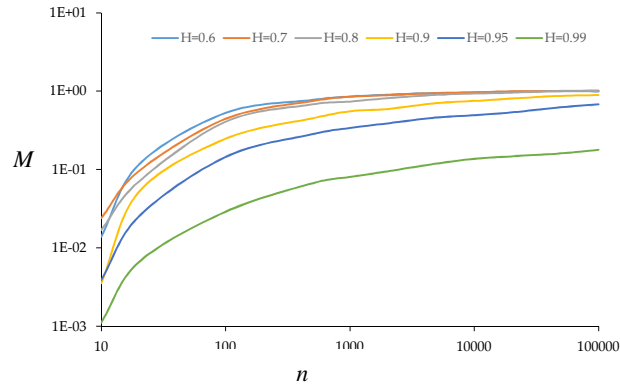
Figure 1: Sample mode value (*M*) from the Monte-Carlo experiments vs. the length of an fGn process for a range of Hurst parameters.
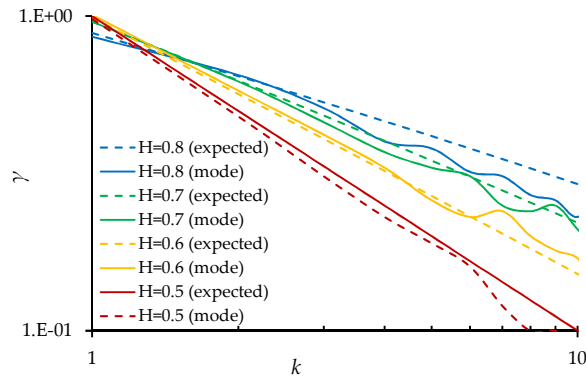

Figure 2: Sample mode values and expected ones as estimated from the Monte-Carlo experiments for a range of Hurst parameters.

A general conclusion from the Monte-Carlo analysis is that the gamma distribution adequately fits the climacogram distribution for all the examined scenarios but with different distribution parameters, and that the mode climacogram is very close to the sample Q25 value for all cases (for such examples see in Fig. 3 and 4, respectively).
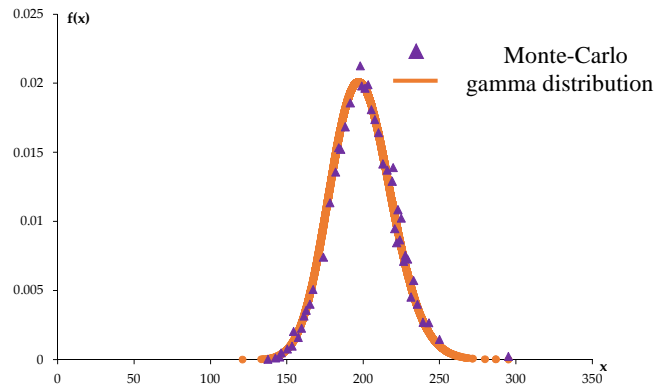
Figure 3: Theoretical gamma probability density distribution function and modelled (from the Monte-Carlo analysis) for the case of $H = 0.6$ and $n = 200$ (scale $k = 1$).
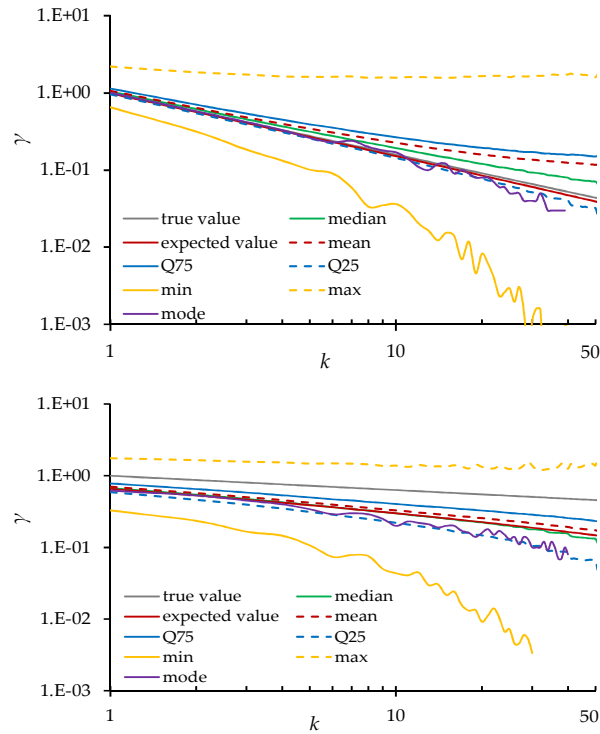

Figure 4: Theoretical and modelled (from the Monte-Carlo analysis) climacograms (mean, median, mode, max, min, Q25 and Q75) for $H = 0.6$ [up] and $H = 0.9$ [down], both for $n = 500$.

4 Application

In this section, we apply a stochastic analysis following the methodology presented in [11] at a fraction of the Penios river located at Thessaly in Greece and extending from the Ali Efenti (upstream) to Amygdalia (downstream) gauging station [12]. Particularly, we conduct a Monte-Carlo analysis of 1000 synthetic annual-discharge timeseries of length $n = 200$, following an fGn process ($H = 0.8$, $\mu = 750$ cms, $\sigma = 100$ cms), and for the flood inundation of the generated discharges we use the quasi-2d LISFLOOD-FP model [13] in an unsteady-flow regime so as to take into account the effect of the past floods inundation. From the 1000 sample we estimate the mode climacogram and we fit a theoretical fGn process assumed that it coincides to the expected climacogram of the new model (H is estimated approximately 0.7), we then generate 1000 samples for the new model, and we estimate both flood inundation maps for the original and the new models (Fig. 6). We find that the flood inundation corresponding to the new model underestimates the uncertainty of the discharge process and, as a consequence, that of the flood inundation, a fact that could have a non-negligible effect to the effective damage cost caused by floods. Also, if the uncertainty is underestimated a regular cluster of annual floods could be erroneously regarded as a more extreme cluster.
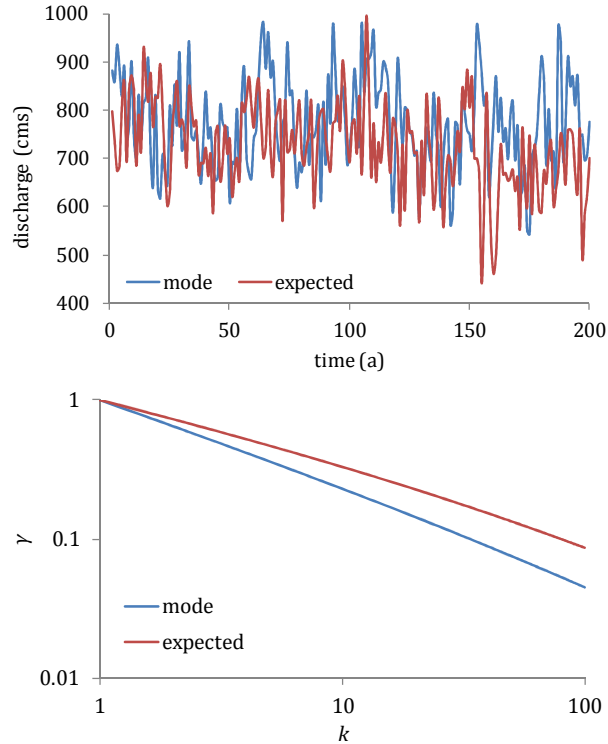

Figure 5: [up] Sample timeseries from the mode and expected models; [down] climacograms (adjusted for bias) for the mode (assumed as expected value for the new model) and expected (original) model.
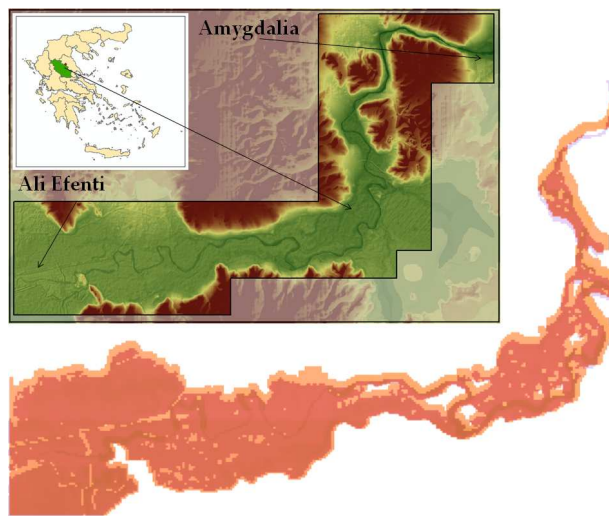
Figure 6: Mean flood inundation map as derived from the mode (dark) and expected (light) Monte-Carlo analyses.

5 Conclusions and discussion

In this study, we present an innovative recipe for handling the statistical bias in case where a single time series of an fGn process is available (which is the usual case). We conduct exhaustive Monte-Carlo experiments and we estimate several statistical characteristics (such as the mode and mean) of the dependence structure through the (second-order) climacogram. We observe that the two-parameter gamma distribution can adequately approximate the distribution of the climacogram, and that the mode climacogram is close to the Q25 value. Therefore, in case where we have a single realization (i.e. a single time series) we should model the mode dependence structure of the desired stochastic model rather than its expectation. Finally, we apply our analysis for the construction of a probabilistic flood inundation map in a fraction of the Peinios river in Greece, and we observe that if bias is not treated properly the uncertainty of the discharge process and the flood inundation are underestimated.

References

[1] Dimitriadis, P., *Hurst-Kolmogorov dynamics in hydrometeorological processes and in the microscale of*
turbulence, PhD thesis, 167 pages, National Technical University of Athens, Athens, 2017.
[2] Beran, J., Statistics for Long-Memory Processes, 315 pp., CRC Press, Boca Raton, Fla., 1994.
[3] Koutsoyiannis, D., The Hurst phenomenon and fractional Gaussian noise made easy, *Hydrological Sciences Journal*, 47 (4), 573–595, 2002.
[4] Mandelbrot, B.B., Wallis, J.R., Noah, Joseph, and operational hydrology, *Water Resour. Res.*, 909–918, 1968.
[5] Koutsoyiannis, D., A random walk on water, *Hydrology and Earth System Sciences*, 14, 585–601, 2010.

[6] Dimitriadis, P., and D. Koutsoyiannis, Climacogram versus autocovariance and power spectrum in stochastic modelling for Markovian and Hurst–Kolmogorov processes, *Stochastic Environmental Research & Risk Assessment*, 29 (6), 1649–1669, 2015.

[7] Koutsoyiannis, D., Knowable moments for high-order characterization and modelling of hydrological processes. 10.13140/RG.2.2.14876.59529, 2018 (submitted).

[8] Dimitriadis, P., and D. Koutsoyiannis, Stochastic synthesis approximating any process dependence and distribution, *Stochastic Environmental Research & Risk Assessment*, 2018.

[9] Dimitriadis, P., D. Koutsoyiannis, and P. Papanicolaou, Stochastic similarities between the microscale of turbulence and hydrometeorological processes, *Hydrological Sciences Journal*, 61 (9), 1623–1640, 2016.

[10] Koutsoyiannis, D., A generalized mathematical framework for stochastic simulation and forecast of hydrologic time series, *Water Resources Research*, 36 (6), 1519–1533, 2000.

[11] Dimitriadis, P., A. Tegos, A. Oikonomou, V. Pagana, A. Koukouvinos, N. Mamassis, D. Koutsoyiannis, and A. Efstratiadis, Comparative evaluation of 1D and quasi-2D hydraulic models based on benchmark and real-world applications for uncertainty assessment in flood mapping, *Journal of Hydrology*, 534, 478–492, 2016.

[12] Dimitriadis, P., A. Tegos, A. Petsiou, V. Pagana, I. Apostolopoulos, E. Vassilopoulos, M. Gini, A.D. Koussis, N. Mamassis, D. Koutsoyiannis, and P. Papanicolaou, Flood Directive implementation in Greece: Experiences and future improvements, European Water, 57(35-41), 2017.

[13] Bates, P.D., and A.P.J. De Roo, A simple raster-based model for flood inundation simulation, *Journal of Hydrology*, 236(1), 54–77, 2000.