Spatial and temporal long-range dependence in the scale domain

HS3.3: Spatio-temporal and/or (geo)statistical analysis of hydrological events, floods, extremes, and related hazards
Convener: Yunqing Xuan | **Co-conveners**: Emmanouil Varouchakis, Gerald A Corzo P, Vitali DiazECS, Francisco Munoz-Arriola, Adrian Almoradie

Chairpersons: Yunqing Xuan, Emmanouil Varouchakis, Vitali Diaz
Time of presentation: 13:20–13:30 (ECS; Solicited; Virtual)

Authors: Panayiotis Dimitriadis*, Theano Iliopoulou, G.-Foivos Sargentis, and Demetris Koutsoyiannis
* presenting author (email: pandim@itia,ntua.gr)

Affiliation: Department of Water Resources and Environmental Engineering
School of Civil Engineering
National Technical University of Athens

Abstract: Long-range dependence (LRD) estimators are traditionally applied in the lag domain (e.g., through the autocovariance or variogram) or in the frequency domain (e.g., through the power-spectrum), but not as often in the scale domain (e.g., through variance vs. scale). It has been contended that the latter case introduces large estimation bias and thus, corresponds to "bad estimators" of the LRD. However, this reflects a misrepresentation or misuse of the concept of variance vs. scale. Specifically, it is shown that if the LRD estimator of variance vs. scale is properly defined and assessed (see literature studies for the so-called climacogram estimator), then the stochastic analysis of variance in the scale domain can be proven to be a robust means to identify and model any LRD process ranging from small scales (fractal behavior) to large scales (LRD, else known as the Hurst-Kolmogorov dynamics) for any marginal distribution. Here, we show how the above definitions can be applied both in spatial and temporal scales, with various applications in geophysical processes, key hydrological-cycle processes, and related natural hazards.

EGU22-13051, updated on 23 May 2022
https://doi.org/10.5194/egusphere-egu22-13051
EGU General Assembly 2022

1. Definitions of Long-Range Dependence (LRD)

The LRD process is often defined through the discrete autocovariance function c (vs. lag τ) or power-spectrum s (vs. frequency w) of a stationary (spatial or temporal; continuous or discrete) process satisfying (i.e., Kolmogorov, 1940; Mandelbrot and Wallis, 1968; Beran, 1992):

$$\lim_{\tau \to \infty} c(\tau) \sim \tau^{2H-2} \text{ and } \lim_{w \to 0} s(w) \sim w^{1-2H}$$

where H is the so-called Hurst parameter (Hurst, 1951).

Note that:

1) in the above expressions the LRD process is defined through the lag or frequency domain,

2) the LRD process is often assumed to be Gaussian in literature (e.g., consider the fractional-Gaussian-noise, fGn; arising from the increment process of the fractional Brownian motion; Mandelbrot and van Ness, 1968),

3) if the LRD process is applied for the whole range of lags or frequencies (e.g., for a continuous fGn, we get $c(h) = \sigma^2 H(2H-1)|h/\alpha|^{2H-2}$, where h is the continuous-time lag, a a time-scale parameter, and σ^2 the process variance at $h = \alpha$), then at zero lag the variance tends to infinity (and thus, resulting to a physically inconsistent process; Koutsoyiannis, 2021).

2. 'Bad Estimator' of Long-Range Dependence

Several estimators have been tested based on the power-spectrum (e.g., through the periodogram) and maximum-likelihood (see discussion and comparisons in Beran, 1992), R/S (Hurst, 1951; Mandelbrot and Taqqu, 1979), autocovariance (e.g., Mandelbrot and Wallis, 1968) or even preliminary attempts at the scale domain, stating that (Beran, 1988):

> In contrast to $\bar{X}_n$ the classical scale estimator $s^2 = (n-1)^{-1} \sum_{i=1}^{n} (X_i - \bar{X}_n)^2$ is very bad. It has a large bias and loses much efficiency. For $H \geq 3/4$ the efficiency is even equal to zero.

Each method presented several limitations, and since, no clear superiority of any method was achieve, it was concluded that (Beran, 1992):

> More research is needed, however, both to deal with more complex situations and to refine the methods in use. Central limit theorems for processes with long-range dependence are rather different from the classical type of theorems, so that many standard results in statistics do not hold. The classical methods should be investigated under this aspect, and methods that also perform well under long-range dependence should be developed. Also, in many applications spatial long-range dependence as well as multivariate time series (or multivariate spatial data) with long-range dependence occur. The development of statistical methods for such data will certainly be a rewarding task for future research.

3. Resurrection of the ‘Bad Estimator’

“If we confess our sins, He is faithful and just to forgive us our sins and to cleanse us from all unrighteousness.” (John 1:9)

After identifying the limitations of the previous methods (e.g., see discussion and comparisons in Dimitriadis and Koutsoyiannis, 2015), a proper definition of the “variance of the averaged process vs. scale” estimator can be introduced, and also, baptized (at the scale domain) to the so-called *climacogram* (Koutsoyiannis, 2010), by linking it to the concept of scale (i.e., climax in Greek), so as not to be confused with the already established term in literature of scale(o)gram. The definition of the climacogram unbiased estimator is (underline quantities denote random variables and ∧ for estimation):

$$\underline{\widehat{\gamma}}\,(\kappa\Delta) = \frac{1}{[n/\kappa]}\sum_{i=1}^{[n/\kappa]}\left(\underline{x}_i^{(\kappa)} - \underline{\widehat{\mu}}\right)^2 + \gamma\left([n/\kappa]\kappa\Delta\right)$$

where $\kappa = k/\Delta$ is the dimensionless scale, k the continuous-scale, Δ the time-space resolution of the continuous-process $\underline{x}$, $[n/\kappa]$ the integer part of n/κ, n the length of the discrete-process $\underline{x}_i$ with mean $\underline{\mu}$, and $\underline{x}_i^{(\kappa)}$ is the i-th element of the averaged sample of the process at scale κ, i.e.,

$$\underline{x}_i^{(\kappa)} = \frac{1}{\kappa}\sum_{j=(i-1)\kappa+1}^{i\kappa}\underline{x}_j \quad \text{with} \quad \gamma\,(k) := \mathrm{Var}\left[\int_0^k \underline{x}\,(y)\,\mathrm{d}y\right]/k^2 \quad .$$

4. Definition of LRD with the new estimator

The definition of the LRD (through the Hurst parameter) can be now defined as:

$$H \coloneqq 1 + \frac{1}{2} \lim_{k \to \infty} \gamma^{\#}(k)$$

where for any function $f(x)$ we set $f^{\#}(x) := \frac{\mathrm{d}\ln(f(x))}{\mathrm{d}\ln x} = \frac{x}{f(x)} \frac{\mathrm{d}f(x)}{\mathrm{d}x}$.

Note that based on the above climacogram estimator, other estimators can be also defined as for example the climacogram-based variogram $v(k) := \gamma(0) - \gamma(k)$, and the climacogram-based spectrum (Koutsoyiannis, 2021),

$$\zeta(k) := \frac{k\left(\gamma(k) - \gamma(2k)\right)}{\ln 2}$$

with similar properties to the classical variogram and power-spectrum, respectively, but with more robust corresponding estimators.

5. Global-scale application (I)

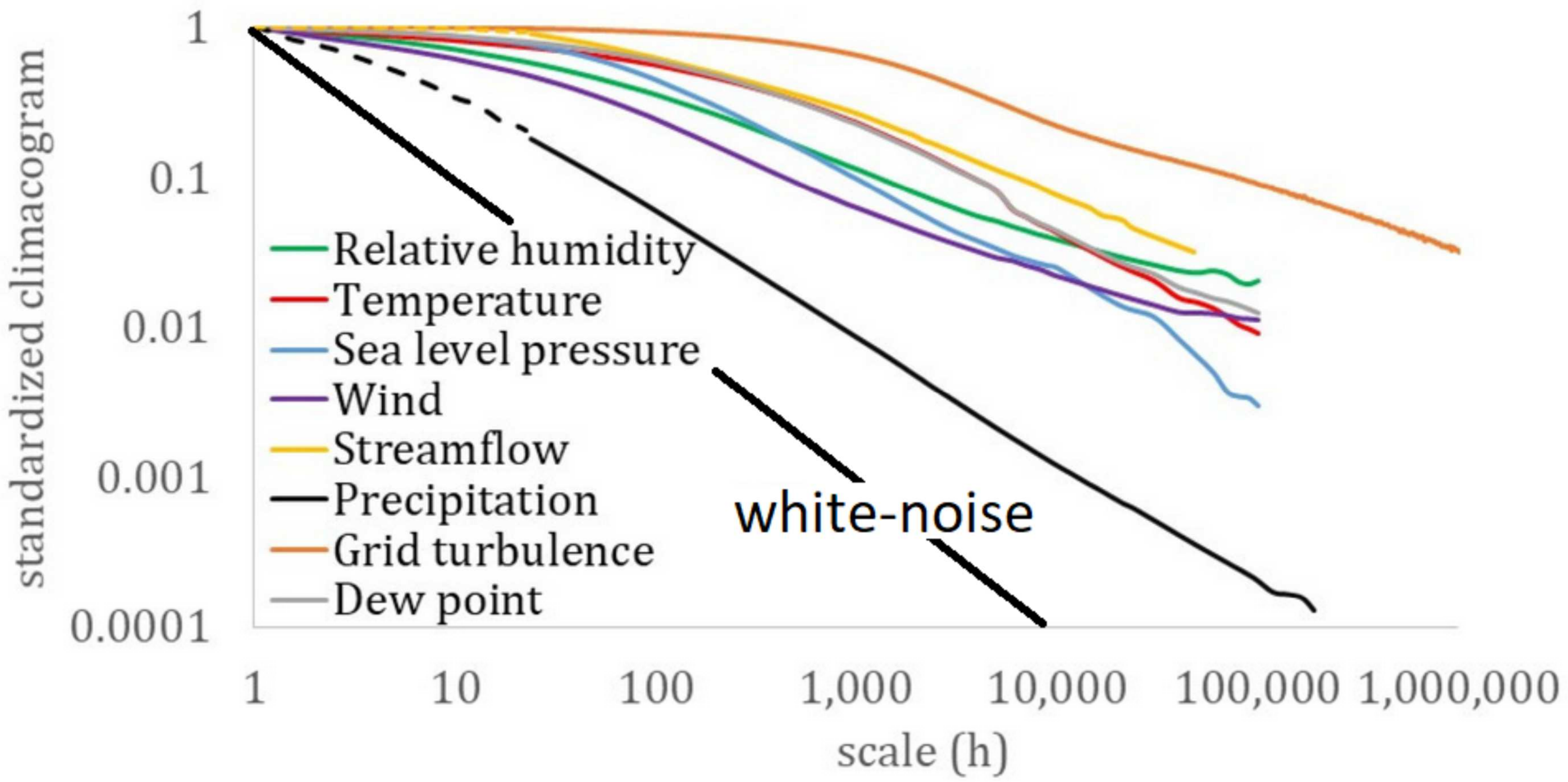
Application of the above estimators to an hourly and daily resolution massive database of global-scale ground stations of key hydrological-cycle natural processes (i.e., near-surface temperature, dew-point, relative humidity, sea level pressure, wind-speed, precipitation and streamflow; more details and sources in Dimitriadis et al., 2021).

	Near-Surface Temperature	Dew Point	Humidity	Sea Level Pressure	Wind Speed	Precipitation	Streamflow
Temporal resolution	Hourly	hourly	hourly	hourly	hourly	hourly/daily	hourly/daily
Total number of stations/time series	6613	5978	4025	4245	6503	93,904	1815
Total number of data values ($\times 10^6$)	907.1	730.0	540.2	364.9	781.7	938.7	13.5
Time period	1938–today	1938–today	1940–today	1939–today	1939–today	1778–today	1900–today

Note that, in total, approximately 50×10^{10} data values are extracted and handled from over 2×10^5 stations.

6. Global-scale application (II)

The LRD behaviour is traced in all processes and compared to the white-noise climacogram as well as to a laboratory-scale experiment of grid-turbulence for illustration.



7. Global-scale application (III)

The Hurst Parameters through the climacogram estimator and the 5% and 95% quantiles (in parentheses) are shown below.

	Parameter (*H*)
Near-surface temperature	0.81 (0.61–0.82)
Relative humidity	0.83 (0.62–0.85)
Dew point	0.77 (0.58–0.79)
Sea level pressure	0.7 (0.53–0.77)
Wind speed	0.85 (0.69–0.86)
Streamflow	0.78 (0.67–0.86)
Precipitation	0.61 (0.52–0.69)

8. Concluding Remarks

1) The estimator for the LRD based on the variance at the scale domain is not a 'bad estimator' and should be further investigated.

2) As compared to the classical lag and frequency domains, the second-order temporal and spatial statistics is suggested (see definitions, discussion and applications in Koutsoyiannis, 2021, and references therein) to be analyzed and studied in the scale domain (through the climacogram metric), which shows great potential (as compared to the lag and frequency domains, through the autocovariance and power-spectrum, respectively).

3) The LRD behaviour is traced (after estimated through the more robust estimator of climacogram) in key hydrological-cycle processes (this is also verified from other studies in literature; see review and references in Dimitriadis et al., 2021).

4) Interesting stochastic similarities of the second-order dependence structure ranging from hourly to climatic scales are revealed when using the climacogram metric.

Thank you!

For questions please also consider sending an email to initiate a fruitful discussion.

References

Beran, J., *Estimation, Testing and Prediction for Self-Similar and Related Processes*, Ph.D. Thesis, ETH, Zurich, Switzerland, 1986.

Beran, J., *Statistical Aspects of Stationary Processes with Long-Range Dependence*, Mimeo Series 1743, Department Statistics University: Chapel Hill, NC, USA, 1988.

Beran, J., Statistical Methods for Data with Long-Range Dependence, *Stat. Sci.*, 7, 404–416, 1992.

Dimitriadis, P., and D. Koutsoyiannis, Climacogram versus autocovariance and power spectrum in stochastic modelling for Markovian and Hurst–Kolmogorov processes, *Stoch. Environ. Res. Risk Assess.*, 29, 1649–1669, 2015.

Dimitriadis, P., D. Koutsoyiannis, T. Iliopoulou, and P. Papanicolaou, A global-scale investigation of stochastic similarities in marginal distribution and dependence structure of key hydrological-cycle processes, *Hydrology*, 8 (2), 59, doi:10.3390/hydrology8020059, 2021.

Hurst, H.E., Long term storage capacities of reservoirs, *Transactions of the American Society of Civil Engineers*, 116, 776–808, 1951.

Mandelbrot, B.B., and J.R. Wallis, Noah, Joseph, and operational hydrology, *Water Resources Research*, 4 (5), 909–918. doi:10.1029/WR004i005p00909, 1968.

Mandelbrot, B.B., and J.W. Van Ness, Fractional Brownian motions, fractional noises and applications, *J. Soc. Ind. Appl. Math.*, 10, 422–437, 1968.

Mandelbrot, B.B., and M.S. Taqqu, *Robust R/S analysis of long run serial correlation*, Proceedings 42$^{\text{nd}}$ Session of the ISI, Manila Book, 2, 69-100, 1979.

Kolmogorov, A.N., Wiener Spirals and Some Other Interesting Curves in a Hilbert Space, 1940, *Selected Works of Mathematics and Mechanics,* Kolmogorov, A.N., Tikhomirov, V.M., Eds.; Kluwer: Dordrecht, The Netherlands, pp. 303–307, 1991.

Koutsoyiannis, D., HESS opinions, A random walk on water, *Hydrol. Earth Syst. Sci.*, 14, 585–601, 2010.

Koutsoyiannis, D., *Stochastics of Hydroclimatic Extremes—A Cool Look at Risk,* Edition 0, National Technical University of Athens: Athens, Greece, 330p., 2021.