Theoretical framework for the stochastic synthesis of the variability of global-scale key hydrological-cycle processes and estimation of their predictability limits under long range dependence.

S10: Forecasting and making decisions under uncertainty: ensemble approaches, evaluation methods and lessons learnt from post-event analyses

Location/Time of presentation: Room Barthez 1, 2 June 2022 (Thursday), 09:00

Convener: Shaun Harrigan | **Co-conveners**: James Bennett, Marie-Amélie Boucher, Céline Cattoën-Gilbert, Fernando Mainardi Fan, Ilias Pechlivanidis, Maria-Helena Ramos, Paolo Reggiani

Authors: Panayiotis Dimitriadis*, Theano Iliopoulou, 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: Uncertainty and change in geophysical processes can be robustly quantified by analyzing the observed variability. A challenging task in engineering studies is to introduce a framework that can simulate this observed variability while preserving only important stochastic attributes. An innovative methodology for genuine simulation of stochastic processes is presented based on the recent work by Koutsoyiannis and Dimitriadis (2021). The proposed algorithm includes the demanding task of simulating any second-order dependence structure of a process (with a focus on long-range dependence behaviour) and any marginal distribution (with focus on heavy tails) through the explicit preservation of its autocovariance function and its cumulants. The long-range dependence behaviour (i.e., power-law drop of variance vs. scale) and heavy-tails are known to be highly associated with the variability magnitude of a process, through which the range of its predictability-window can be also quantified. To estimate this range, an extensive global-scale network of stations of key hydrological-cycle processes (i.e., near-surface hourly temperature, dew point, relative humidity, sea level pressure, atmospheric wind speed, streamflow, and precipitation; for details see Dimitriadis et al., 2021) is analyzed using ensemble techniques and the proposed stochastic simulation algorithm. The limitations of existing methodologies for the stochastic simulation and estimation of the predictability-window, and how can they be tackled through the proposed approach, are discussed over applications in flood risk management.

IAHS2022-610, updated on 30 May 2022
IAHS-AISH Scientific Assembly 2022

1. Quantification of uncertainty through variability (I)

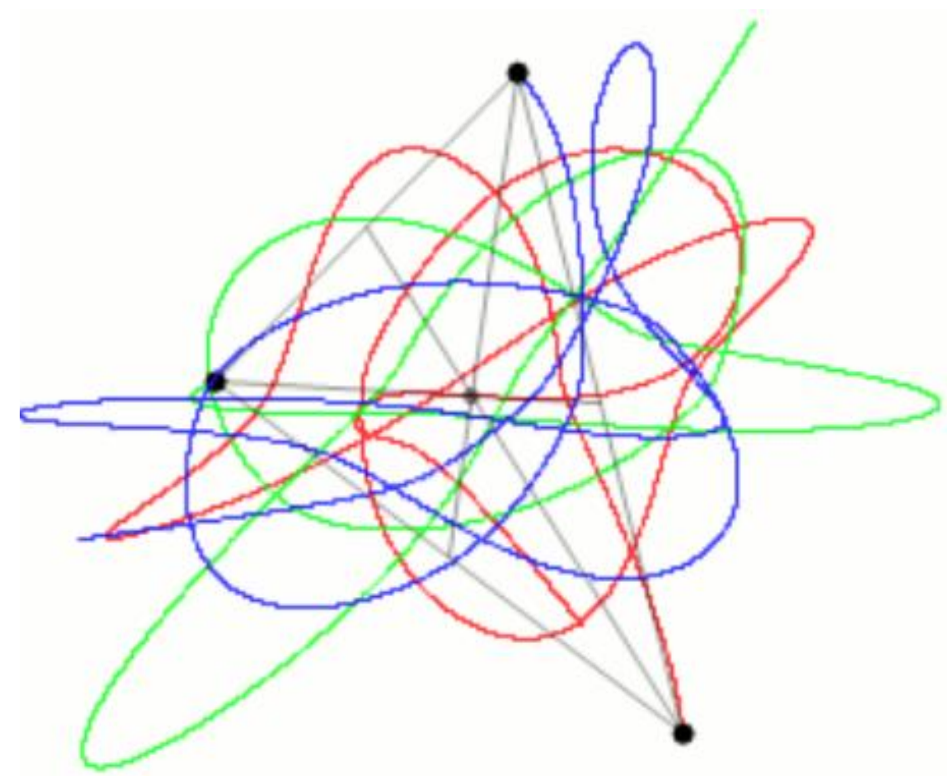
Complexity (non-linear interaction of numerous physical processes; e.g., consider the three-body problem by Poincare, 1980).

(intrinsic) ↑↓ (deterministic)

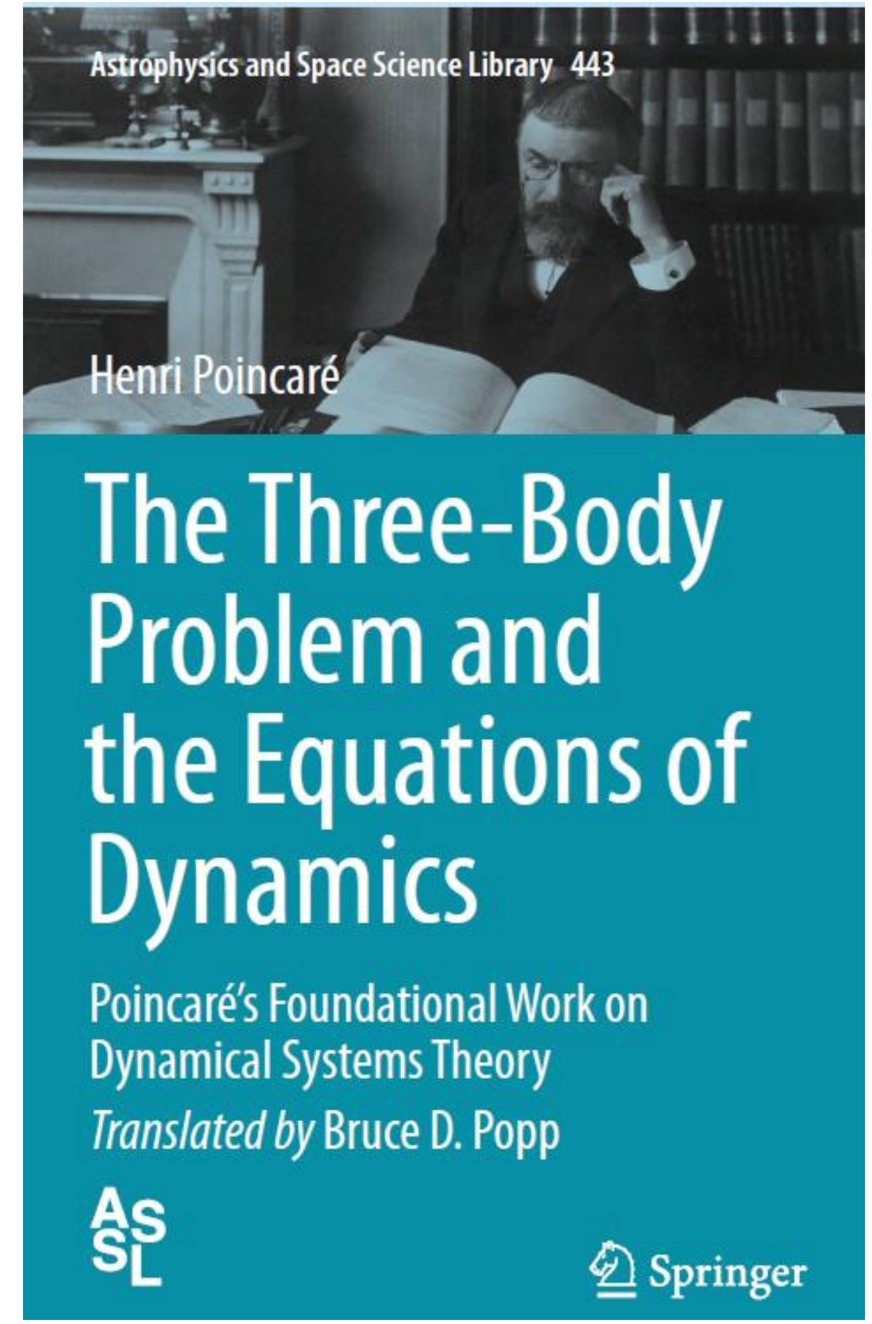
Uncertainty (properly defined through the Theory of Probability and Stochastics; Kolmogorov, 1933).

(synthesis) ↑↓ (quantification)

Variability (expressed through second-order statistics).



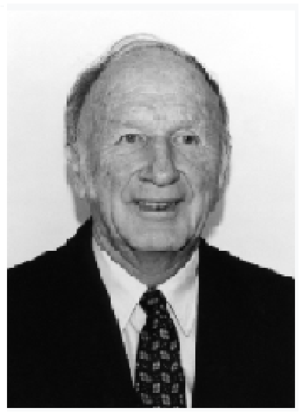
Source: https://en.wikipedia.org/wiki/Three – body_problem



I believed when I started this work that once the solution of the problem was found for the specific case that I dealt with it would be immediately generalizable without having to overcome any new difficulties outside of those which are due to the larger number of variables and the impossibility of a geometric representation. I was mistaken.

2. Quantification of uncertainty through variability (II)

JOURNAL OF THE ATMOSPHERIC SCIENCES



Deterministic Nonperiodic Flow[1]

EDWARD N. LORENZ

Massachusetts Institute of Technology

(Manuscript received 18 November 1962, in revised form 7 January 1963)

ABSTRACT

Finite systems of deterministic ordinary nonlinear differential equations may be designed to represent forced dissipative hydrodynamic flow. Solutions of these equations can be identified with trajectories in phase space. For those systems with bounded solutions, it is found that nonperiodic solutions are ordinarily unstable with respect to small modifications, so that slightly differing initial states can evolve into considerably different states. Systems with bounded solutions are shown to possess bounded numerical solutions.

A simple system representing cellular convection is solved numerically. All of the solutions are found to be unstable, and almost all of them are nonperiodic.

The feasibility of very-long-range weather prediction is examined in the light of these results.

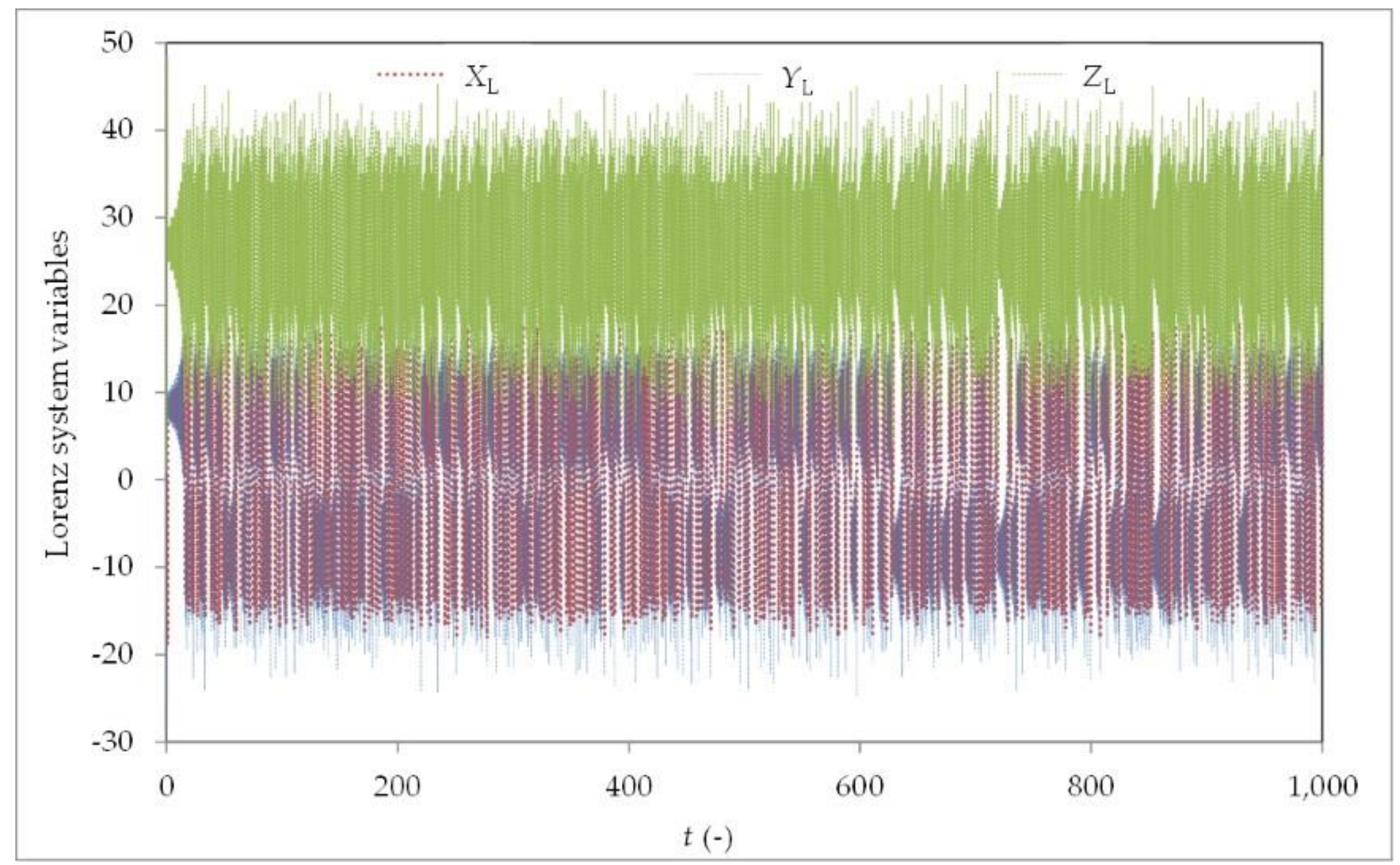
There remains the very important question as to how long is "very-long-range." Our results do not give the answer for the atmosphere; conceivably it could be a few days or a few centuries. In an idealized system,

Source: https://en. wikipedia. org/wiki/Edward_Norton_Lorenz

Lorenz-system (1963) dimensionless variables (denoted X_L, Y_L and Z_L), with randomly varying initial values between -1 and 1, a time step dt=Δ=0.01, time length 10^3, and σ_L=10, r_L=8 and b_L=8/3.

$$\left\{ \begin{aligned} \frac{dX_\mathrm{L}}{dt} &= \sigma_\mathrm{L}(Y_\mathrm{L} - X_\mathrm{L}) \\ \frac{dX_\mathrm{L}}{dt} &= r_\mathrm{L}X_\mathrm{L} - Y_\mathrm{L} - X_\mathrm{L}Z_\mathrm{L} \\ \frac{dX_\mathrm{L}}{dt} &= X_\mathrm{L}Y_\mathrm{L} - b_\mathrm{L}Z_\mathrm{L} \end{aligned} \right\}$$

Source: Dimitriadis et al., (2016).



3. Quantification of variability at the scale domain (I)

The observed (simulated) variability is suggested to be quantified (generated) at the scale domain through the climacogram (i.e., variance of the averaged process vs. scale; Koutsoyiannis, 2010), rather at the lag domain through the autocovariance function or the frequency domain through the power-spectrum (see limitations and discussion in Dimitriadis and Koutsoyiannis, 2015; and generalization in Stochastics at the scale domain in Koutsoyiannis, 2021).

The definition of the climacogram unbiased estimator is (underline quantities denote random variables and ∧ for estimation):

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

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.,

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

4. Quantification of variability at the scale domain (II)

For example, a model that fits adequately numerous geophysical processes (e.g., see review and applications in Dimitriadis et al., 2021) is (Koutsoyiannis, 2016):

$$\gamma(k) = \frac{\lambda}{(1 + (k/a)^{2M})^{(1-H)/M}}$$

where λ is the variance at scale 0, a is a scale parameter in units of the scale k, M is the dimensionless fractal parameter (Gneiting and Schlather, 2014) indicative of the roughness (M < 0.5) or smoothness (M > 0.5) of the fine scales (while the case M = 0.5 corresponds to the absence of fractal behaviour), and H is the Hurst parameter indicative of the strength of the long-range dependence (i.e.; for 0.5 < H < 1, while the case H = 0.5 corresponds to a white-noise behaviour, and 0 < H < 0.5 to an anti-persistence one). Note that other models may also capture the medium-scale drop of variance (Dimitriadis and Koutsoyiannis, 2015; 2018):

$$\gamma(k) = \frac{\lambda}{2(1 + (k/a)^{2M})^{\frac{1-H}{M}}} + \frac{\lambda\left(k/a + \mathrm{e}^{-k/a} - 1\right)}{(k/a)^2}$$

while for additional attributes, more generalized models are introduced (Koutsoyiannis, 2021).

5. Quantification of variability at the scale domain (III)

For the quantification of the joint-effect between the marginal $F(\underline{x})$ and the second-order dependence structure, the Knowable (K-) moments are proposed (and here, particularly, the hyper-central ones), which are shown to have additional merits as compared to the classical, L-moments, etc., and thus, enabling more reliable estimations from data (Koutsoyiannis, 2021). For example, the hyper-central K-kurtosis and K-skewness can be expressed as:

$$S = \frac{K_{32}^{+}}{K_{22}^{+}} = 2\frac{K_{32}}{K_{22}} - 2 \quad \text{and} \quad K = \frac{K_{42}^{+}}{K_{22}^{+}} = 4\frac{K_{42}}{K_{22}} - 6\frac{K_{32}}{K_{22}} + 3$$

where for $p \geq q$

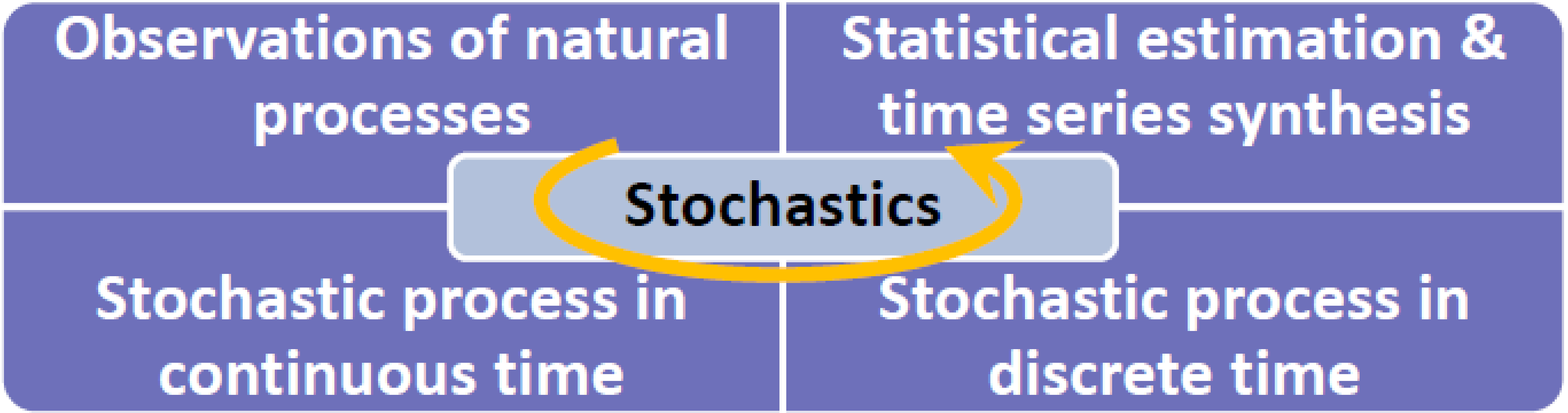
$$K_{pq}^{+} \coloneqq (p-q+1)\mathrm{E}\left[\left(2F\left(\underline{x}\right)-1\right)^{p-q}\left(\underline{x}-\mu\right)^{q}\right]$$

$$K_{pq} \coloneqq (p-q+1)\mathrm{E}\left[\left(F\left(\underline{x}\right)\right)^{p-q}\left(\underline{x}-\mu\right)^{q}\right]$$

$$\underline{\hat{K}}_{p2} = \frac{n}{n-1}\sum_{i=p}^{n}\frac{p-1}{n}\frac{\Gamma(n-p+2)}{\Gamma(n)}\frac{\Gamma(i)}{\Gamma(i-p+2)}\left(\underline{x}_{(i:n)}-\underline{\hat{\mu}}\right)^{2}$$

and n is the length of the sample, $\underline{\hat{\mu}} = \sum_{i=1}^{n}\underline{x}_i/n$ is the estimator of the mean, and $\underline{x}_{(i:n)}$ is the observed sample rearranged in ascending order.

6. Stochastic synthesis preserving several aspects of variability (I)



The steps for a stochastic analysis (after having removed before the analysis and added back after the analysis any known deterministic behaviour). Source: Koutsoyiannis and Dimitriadis (2016).

A challenging task in engineering studies is to introduce a framework that can simulate the observed variability while preserving only important stochastic attributes, such as:

1. Any second-order dependence structure of a process with focus on long-range dependence (Hurst, 1951; Kolmogorov, 1940) and small-scale fractal behaviour at the scale domain.
2. Any marginal distribution with focus on heavy-tails (e.g., Pareto-Burr-Feller distributions).
3. Explicit preservation of the autocovariance function and cumulants (including intermittency, Koutsoyiannis 2016; joint-moments up to 4 in Dimitriadis and for any number of moments in Koutsoyiannis and Dimitriadis, 2021; and time-irreversibility; Koutsoyiannis, 2019; expansion up to the 2nd scale in Vavoulogiannis et al., 2021).

7. Global-scale analysis (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 see in Dimitriadis et al., 2021).

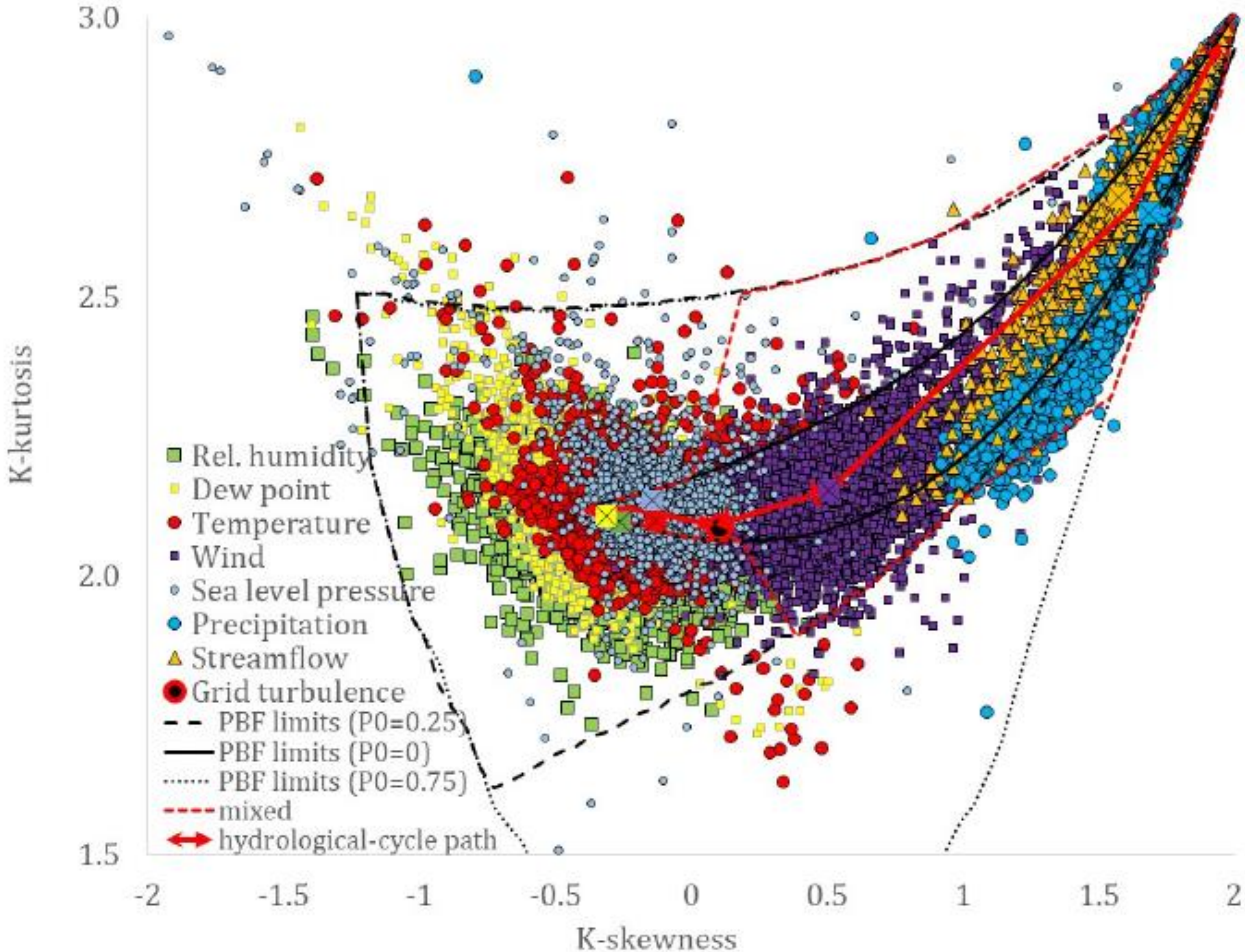
	Near-Surface Temperature	Dew Point	Humidity	Sea Level Pressure	Wind Speed	Precipitation	Streamflow
Temporal resolution	Hourly	hourly	hourly	hourly	hourly	hourly/daily	hourly/daily
Total number of stations/time series	6613	5978	4025	4245	6503	93,904	1815
Total number of data values ($\times 10^6$)	907.1	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.

8. Global-scale analysis (II)

The K-skewness vs. K-kurtosis, for the key hydrological-cycle and grid-turbulence processes, and the empirically calculated limits of the mixed Pareto-Burr-Feller distribution for probabilities of zero values at 25% and 75%:

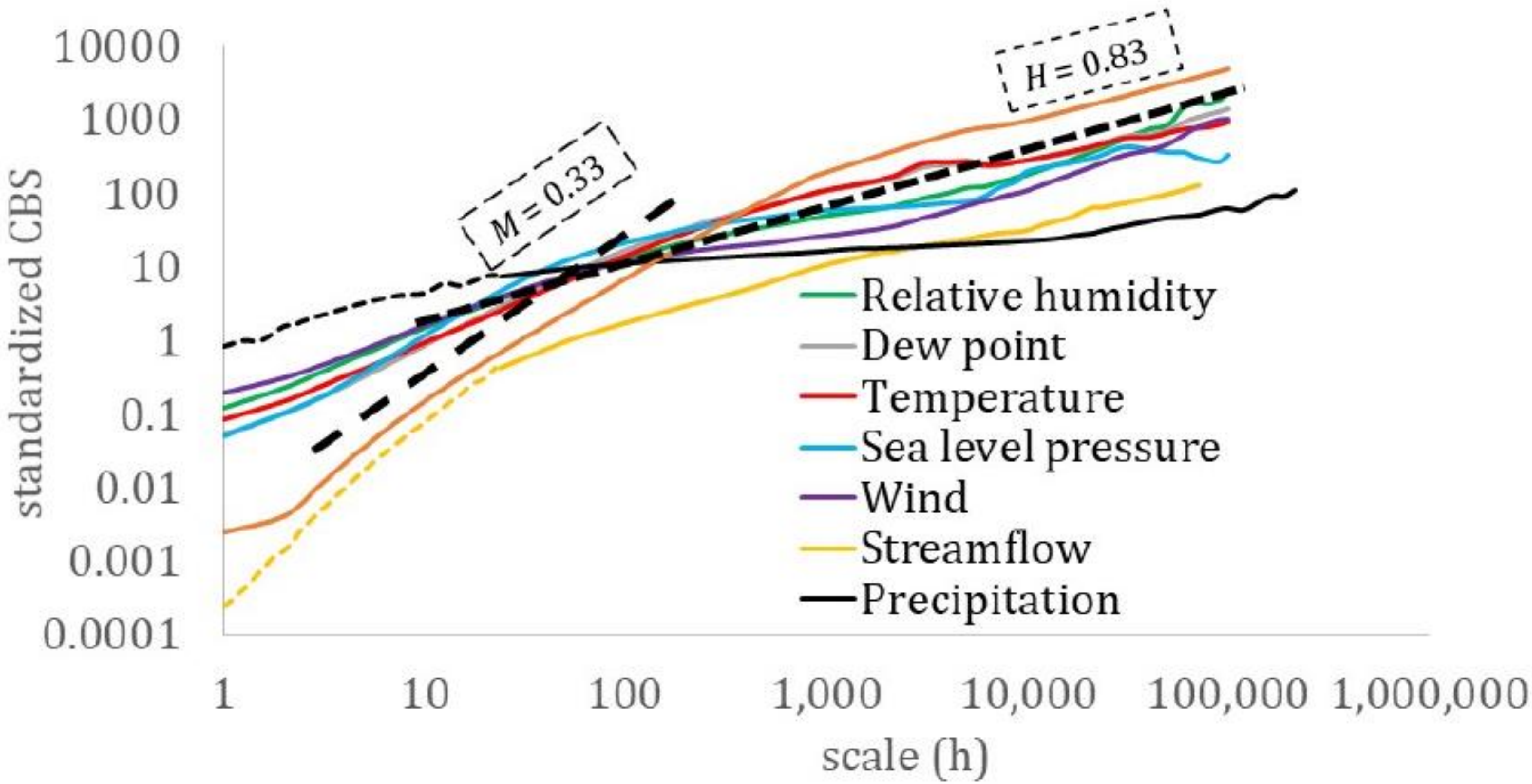
$$F(x) = 1 - \left(1 + \xi\zeta\left(\frac{x}{\lambda}\right)^{\zeta}\right)^{-\frac{1}{\xi\zeta}}$$



The mean values of the K-skewness and K-kurtosis for each process are depicted by the square markers with the x-symbol inside. (Source: Dimitriadis et al., 2021).

9. Global-scale analysis (III)

Both fractal and long-range dependence behaviour are traced in all key hydrological-cycle processes through the mean standardized climacospectrum, i.e. $\zeta(k)/\zeta(1)$. Dashed and continuous lines in streamflow and precipitation correspond to hourly and daily stations (Source: Dimitriadis et al., 2021).



10. Global-scale analysis (IV)

Summary statistics of the scale, fractal and Hurst parameters of the second-order dependence structure adjusted for bias based on the climacogram and CBS estimation, with the 5% and 95% quantiles in parentheses, and for each key hydrological-cycle process of hourly resolution (Source: Dimitriadis et al., 2021).

	a (h)	Fractal Parameter (M)	LRD Parameter (H)
Near-surface temperature	135.1 (9.2–323.1)	0.16 (0.01–0.22)	0.81 (0.61–0.82)
Relative humidity	17.4 (5.6–57.3)	0.23 (0.2–0.27)	0.83 (0.62–0.85)
Dew point	120.3 (16.4–213.2)	0.23 (0.15–0.46)	0.77 (0.58–0.79)
Sea level pressure	36.5 (10.0–67.2)	0.36 (0.25–0.55)	0.7 (0.53–0.77)
Wind speed	9.1 (0.1–25.9)	0.15 (0.07–0.3)	0.85 (0.69–0.86)
Streamflow	96.5 (16.8–533.1)	0.43 (0.2–0.46)	0.78 (0.67–0.86)
Precipitation	2.1 (0.1–3.0)	0.25 (0.18–0.67)	0.61 (0.52–0.69)

11. Predictability window (I)

For the estimation of the local future mean at period length κ (conditional on the present and past values of the discrete process $\underline{x}_i$), i.e.,

$$\underline{\mu}_\kappa := \mathrm{E}\left[\frac{1}{\kappa}\left(\underline{x}_1 + \cdots + \underline{x}_\kappa\right) | \underline{x}_0, \underline{x}_{-1}, \ldots\right]$$

we follow Koutsoyiannis (2021) approach, by selecting only the past $0 \leq \nu \leq n$ values, i.e.,

$$\underline{\hat{\mu}}_\nu := \frac{1}{\nu}\left(\underline{x}_0 + \underline{x}_{-1} + \cdots + \underline{x}_{-\nu+1}\right)$$

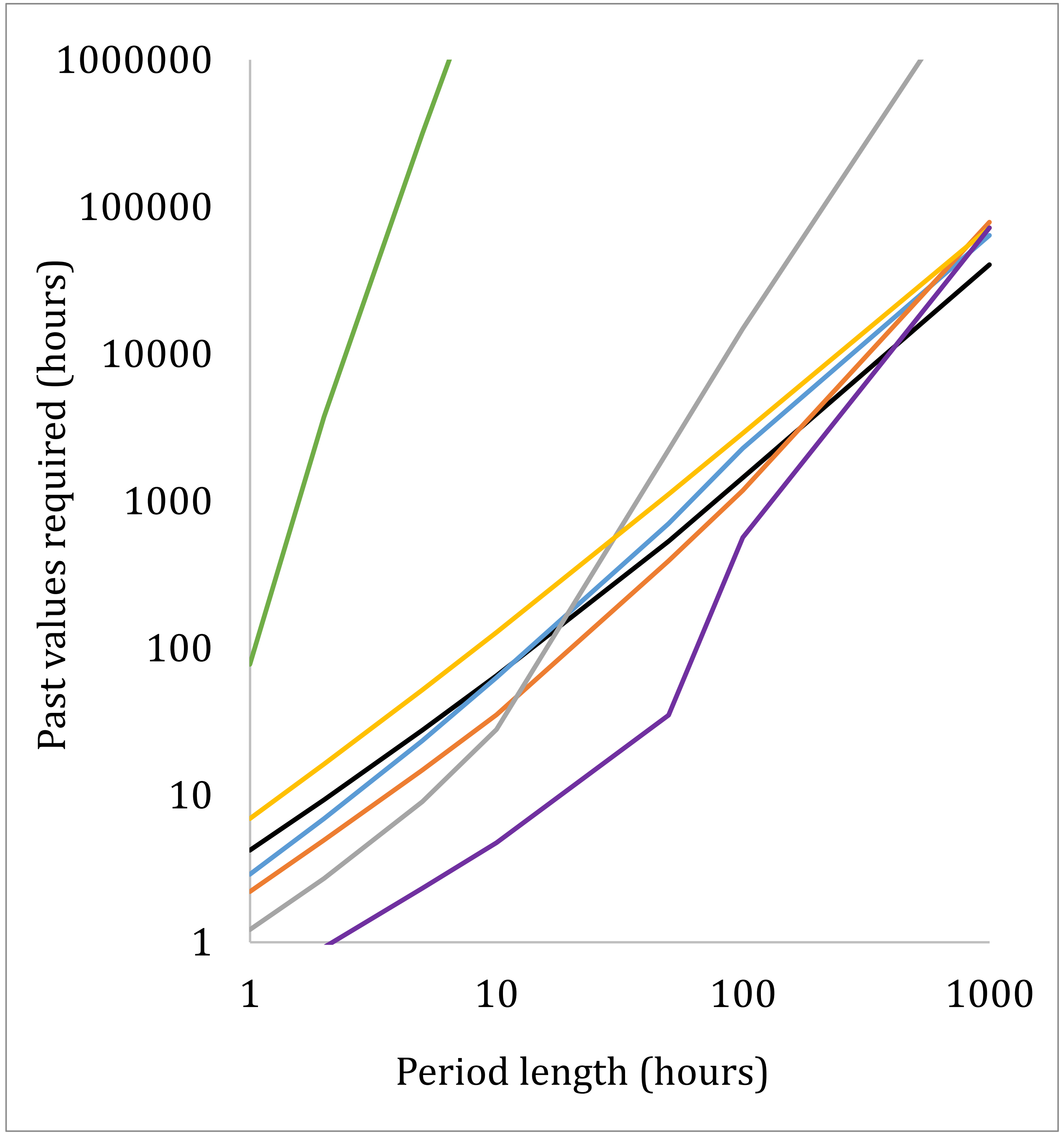
that minimize the square error between these estimations, i.e., $A(\kappa, \nu) := \mathrm{E}\left[\left(\underline{\hat{\mu}}_\nu - \underline{\hat{\mu}}_\kappa\right)^2\right]$.

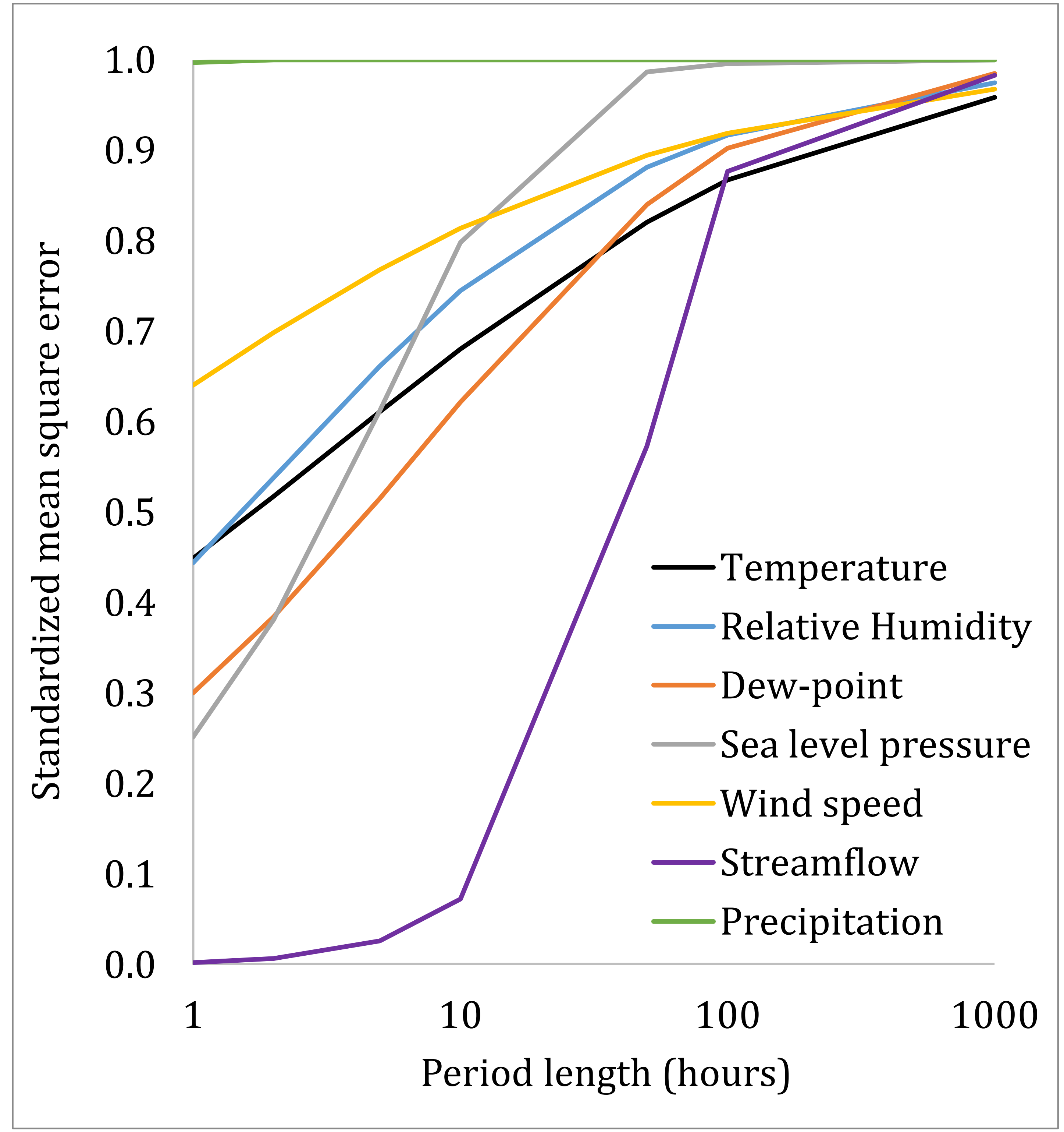
It can be shown that the standardized mean square error is (Koutsoyiannis, 2021):

$$\frac{A(\kappa, \nu)}{\lambda} = \left(\frac{1}{\kappa} + \frac{1}{\nu}\right)\left(\kappa + \nu\frac{\gamma(\nu)}{\lambda} - (\nu + \kappa)\frac{\gamma(\nu + \kappa)}{\lambda}\right)$$

12. Predictability window (II)

For each key hydrological-cycle process and for a range of period lengths, we estimate the predictability limit up to the variance of the process.

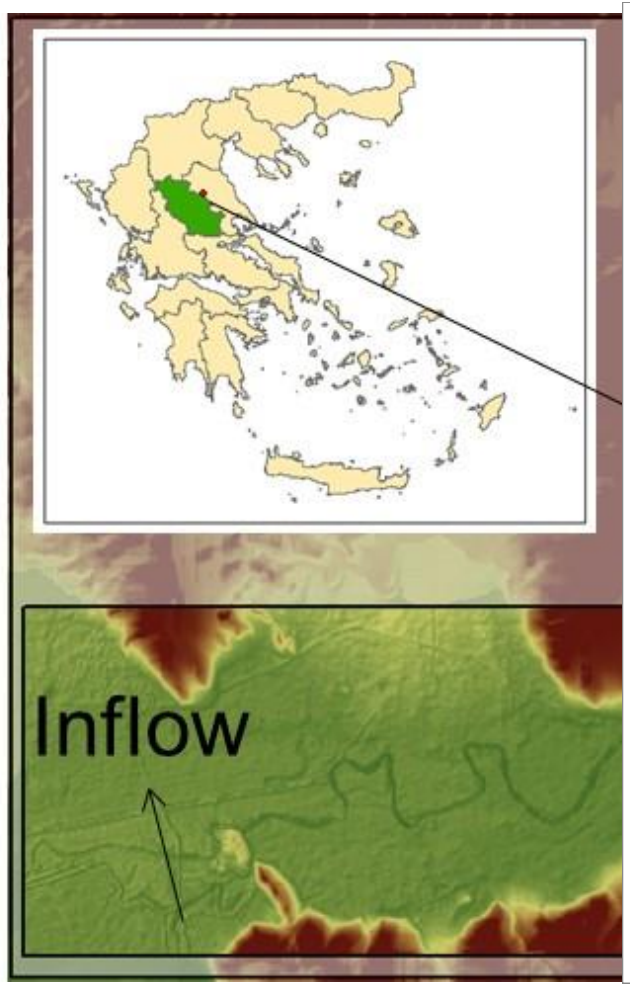


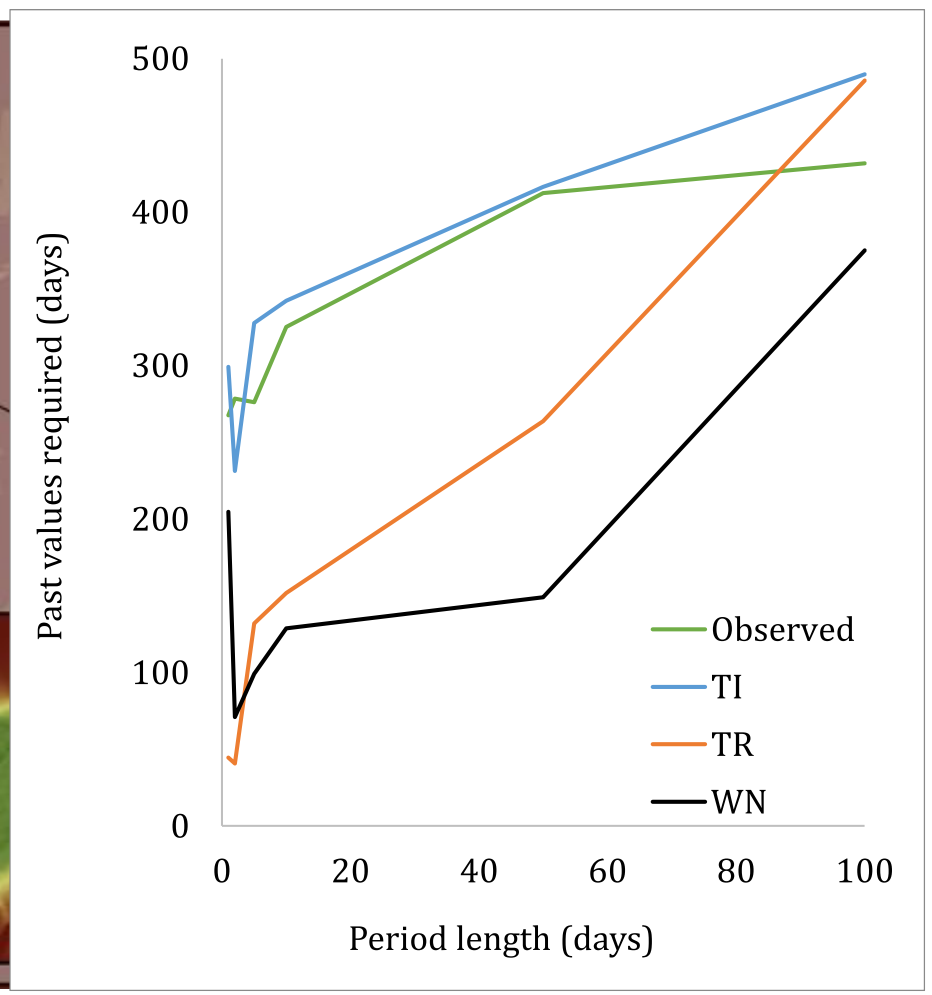


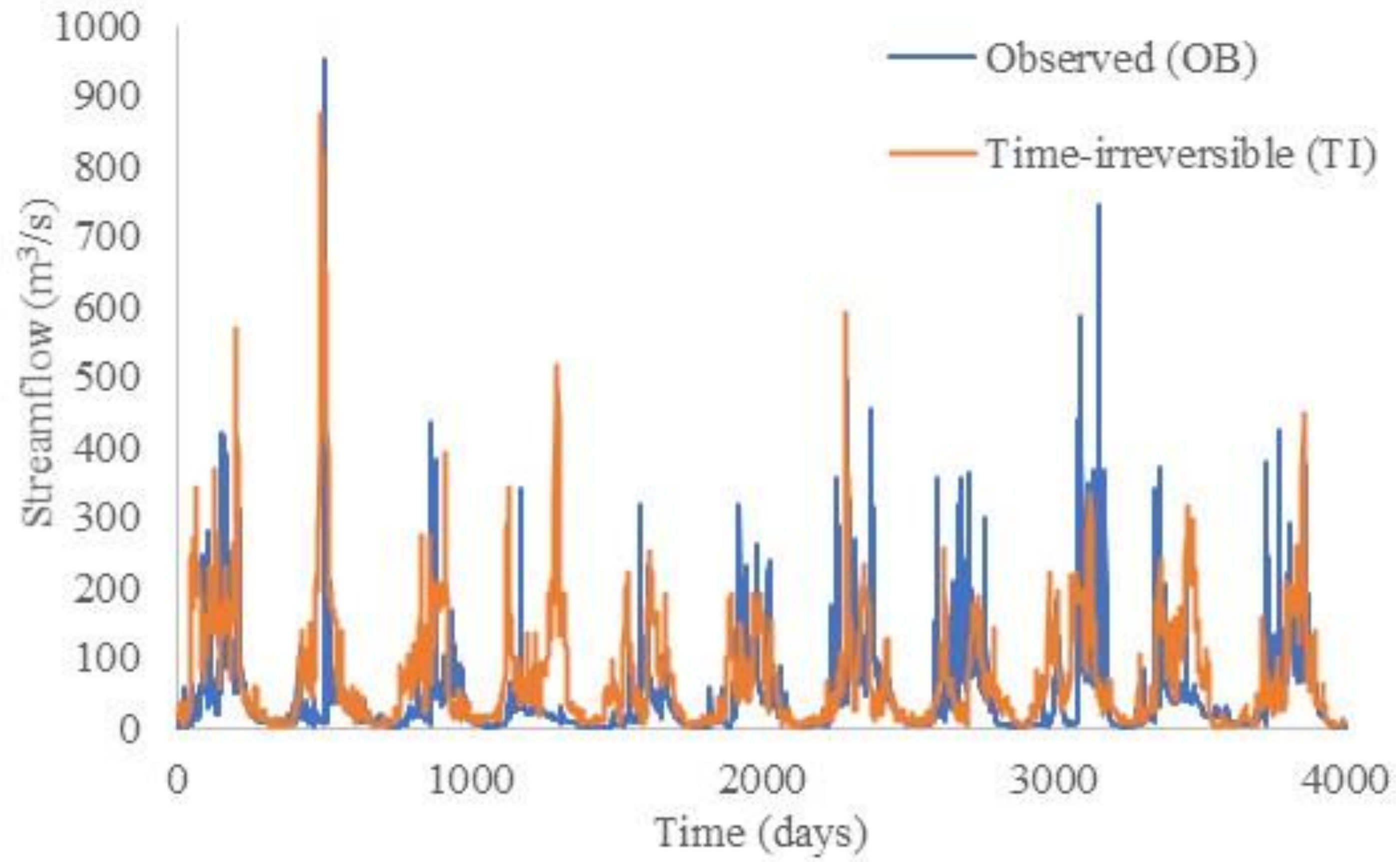
13. Application

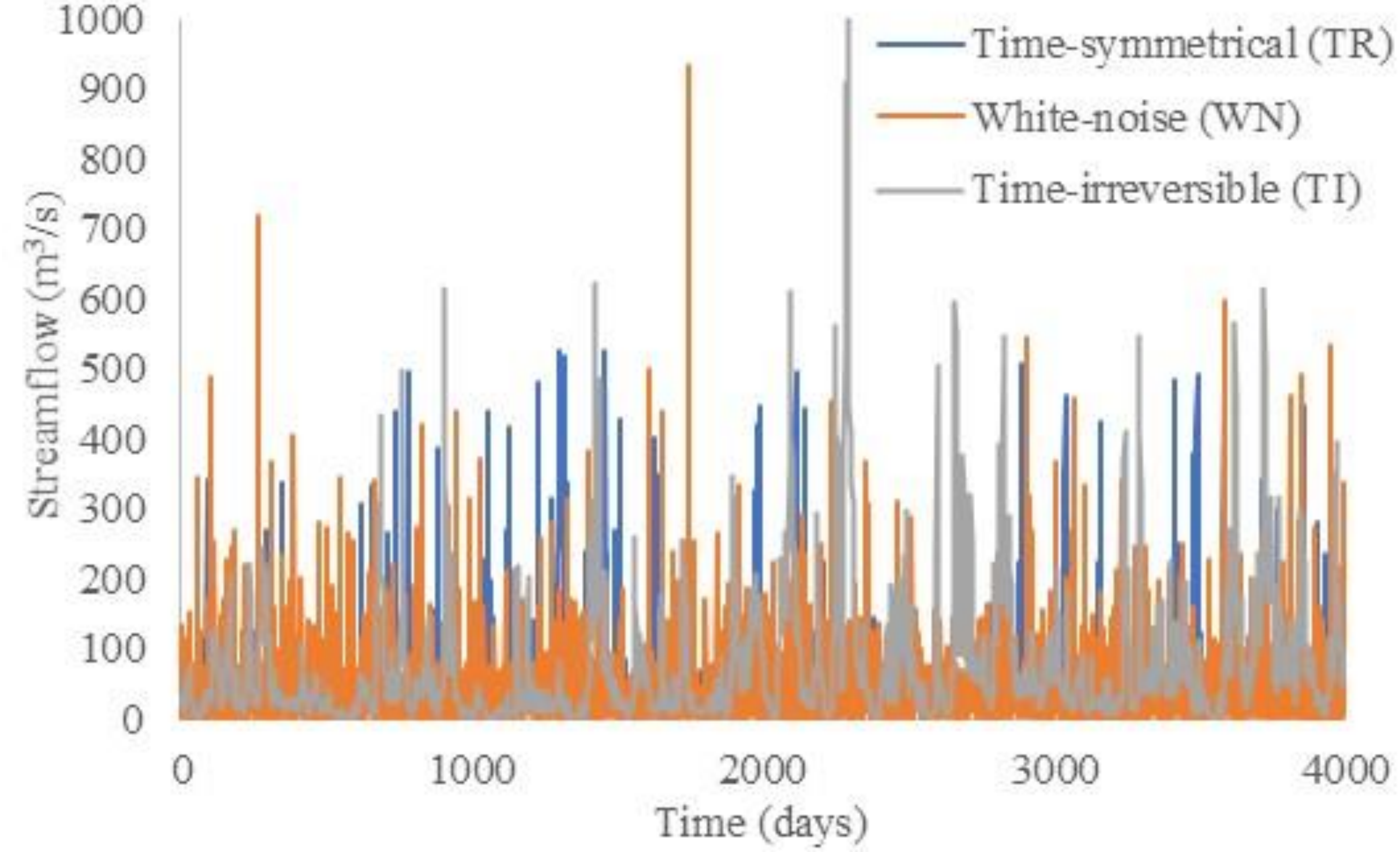
To compare the predictability time windows, we consider a recorded outflow of 4000 days of streamflow at Ali Efenti in Thessaly.

To compare the observed (OB) timeseries with the synthetic one adjusted for marginal function, long-range dependence and time-irreversibility (TI), as well as without time-irreversibility (TR) and also with a white-noise (WN) behaviour.









14. Concluding Remarks for Discussion

1) For estimating the predictability limits through the stochastic approach, both the marginal function and the second-order dependence structure are required to be analyzed.

2) The marginal function is quantified through the K-moments and modelled with the Pareto-Burr-Feller distribution, while the second-order dependence structure is quantified through the climacogram at the scale domain and modelled with an LRD-type model (both identified in numerous key hydrological-cycle processes via ensemble techniques).

3) For the stochastic synthesis of the recorded variability, an explicit stochastic scheme is suggested, since any transformation from/to Gaussian processes may underestimate certain observed aspects such as time-irreversibility, intermittency, joint-effects, etc.

4) The required past values and the mean square error (standardized with the variance) both increase with future period length.

5) Precipitation is considered by far the most difficult to accurately predict.

Thank you!

For questions please also consider sending an email (pandim@itia.ntua.gr) to initiate a fruitful discussion.

References

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, and K. Tzouka, Predictability in dice motion: how does it differ from hydrometeorological processes?, Hydrological Sciences Journal, 61 (9), 1611–1622, doi:10.1080/02626667.2015.1034128, 2016.

Dimitriadis, P., and D. Koutsoyiannis, Stochastic synthesis approximating any process dependence and distribution, Stochastic Environmental Research & Risk Assessment, 32 (6), 1493–1515, doi:10.1007/s00477-018-1540-2, 2018.

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.

Gneiting, T., and M. Schlather, Stochastic models that separate fractal dimension and the Hurst effect, SIAM Rev. 46, 269–282, 2004.

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.

Kolmogorov, A.N., 1933, Grundbegriffe der Wahrscheinlichkeitsrechnung, Ergebnisse der Math. (2), Berlin (2nd English Edition: Foundations of the Theory of Probability, 84 pp. Chelsea Publishing Company, New York, 1956).

Lorenz, E.N., Deterministic nonperiodic flow, J. Atmospheric Sci., 20, 130–141, 1963.

Poincare, H., The Three-body Problem and the Equations of Dynamics (Astrophysics and Space Science Library, 443), Springer, Cham, 2017. Poincare's foundational work on dynamical systems theory, Translated from the 1890 French original and with a preface by Bruce D. Popp.

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

Koutsoyiannis, D., Revisiting the global hydrological cycle: is it intensifying?, *Hydrology and Earth System Sciences*, 24, 3899–3932, doi:10.5194/hess-24-3899-2020, 2020.

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

Koutsoyiannis, D., and Dimitriadis, P., From time series to stochastics: A theoretical framework with applications on time scales spanning from microseconds to megayears, in: Orlob Second International Symposium on Theoretical Hydrology, University of California Davis, 2016.

Koutsoyiannis, D., Generic and parsimonious stochastic modelling for hydrology and beyond, Hydrological Sciences Journal, 61 (2), 225–244, doi:10.1080/02626667.2015.1016950, 2016.

Koutsoyiannis, D., Time's arrow in stochastic characterization and simulation of atmospheric and hydrological processes, Hydrol. Sci. J., 64, 1013–1037, 2019.

Vavoulogiannis, S., T. Iliopoulou, P. Dimitriadis, and D. Koutsoyiannis, Multiscale temporal irreversibility of streamflow and its stochastic modelling, Hydrology, 8, 63, 2021.

Funding information: This research is carried out / funded in the context of the project "Development of Stochastic Methods for Extremes (ASMA): identification and simulation of dependence structures of extreme hydrological events" (MIS 5049175) under the call for proposals "Researchers' support with an emphasis on young researchers—2nd Cycle". The project is co-financed by Greece and the European Union (European Social Fund; ESF) by the Operational Programme Human Resources Development, Education, and Lifelong Learning 2014–2020.

European Union
European Social Fund

Operational Programme
Human Resources Development, Education and Lifelong Learning
Co-financed by Greece and the European Union

ΕΣΠΑ
2014-2020
ανάπτυξη - εργασία - αλληλεγγύη