European Geosciences Union General Assembly 2011
Vienna, Austria, 4-9 April 2011
Session HS7.5/NP6.7: Hydroclimatic stochastics

Multidimensional Hurst-Kolmogorov process for modelling temperature and rainfall fields

P. Dimitriadis[1], D. Koutsoyiannis[1], C. Onof[2] and K. Tzouka[1]
[1]Department of Water Resources and Environmental Engineering, National Technical University of Athens
[2]Department of Civil and Environmental Engineering, Imperial College of London
(www.itia.ntua.gr)

1. Abstract

A multidimensional (MD) stochastic simulation model is presented, which is a direct extension of the 1D simple scaling process, known as Hurst-Kolmogorov (HK) process following the analysis of the 2D extension of Koutsoyiannis et al. (2011). The MD HK (MHK) process can generate time-varying spatial geophysical fields (such as rainfall and temperature), consistent with the observed long-term spatiotemporal persistence (slowly decaying autocorrelation over spatial or temporal displacement). The MHK process is formulated assuming anisotropy, so as to take into account possibly different autocorrelation decay rates (Hurst coefficients) in each dimension of the field. The MHK process is also investigated through some applications based on observed temperature and rainfall fields.

2. Hurst phenomenon and the MHK process

"*High tendency of high/low values to occur in natural events*": Hurst (1951) ➔ Slowly decaying autocorrelation over scale➔ Power-law behaviour (Kolmogorov, 1940).

$$\left(\underline{Z}_v^{(k)} - \mu\right) =_d \left(\frac{k}{l}\right)^A \left(\underline{Z}_v^{(l)} - \mu\right), \text{ where } \mu = E[\underline{Z}_v], \ A = D(1-H), \ \underline{Z}_v^{(k)} = \frac{1}{k^D} \sum_{i_1=(v_1-1)k+1}^{v_1 k} \cdots \sum_{i_D=(v_D-1)k}^{v_D k} \underline{Z}_{i_1,\dots,i_D}$$

- $\underline{Z}$: random field of interest (assumed stationary and isotropic)
- $\underline{Z}_v$: mean aggregated field (at a spatio-temporal scale)
- v: vector index of random field indicating location in the field
- k,l: any aggregation scales of the process
- μ: mean of the process
- $=_d$: equal in distribution function
- A: power law exponent of autocorrelation over scale
- D: dimension of vector index space of random field (v)

3. Hurst coefficient (H) of the MHK process

• HK process depends on the characteristic parameter $0<H<1$. Here, the estimation of the H coefficient is done via the minimization of the square error (SE_H) of the empirical $(S^{(k)})^2$ and true ($\gamma^{(k)}$) variance over scale k of the process. A method of Tyralis and Koutsoyiannis (2010) for the estimation of H was extended to the MHK process (D dimensions).

$$SE_H = \sum_{k=1}^{k'} \left[2\,ln(\tilde{S}^{(k)}) - ln(\gamma^{(k)}) \right]^2 / k^p,\ p=2\ ,\ E\left[\left(\tilde{S}^{(k)} \right)^2 \right] = R\left(S^{(k)} \right)^2,\ R(k;H) = \frac{N/k^D - 1}{N/k^D - \left(N/k^D \right)^{2H-1}}$$

• The autocovariance γ (acvf) and autocorrelation function ρ (acrf) of the MHK are expressed as:

$$\gamma_{c(0)}^{(k)} = k^{-B}\gamma_{(0)}^{(1)} \text{ and } \gamma_{(r)}^{(k)} = L_D r^{-B}$$

$$\rho_{c(r)}^{(k)} = \gamma_{(r)}^{(k)} / \gamma_{(0)}^{(k)} \rightarrow \rho_{c(r)} = L_D r^{-B} / \gamma_{(0)}$$

where $B = 2D(1-H) = 2A$

c stands for continuous

L_D is a coefficient

r is the lag

- $0 < H < 0.5 \rightarrow$ Anticorrelated $(\rho < 0)$
- $H = 0.5 \rightarrow$ Independent $(\rho = 0)$
- $0.5 < H < 1 \rightarrow$ Correlated $(\rho > 0)$

Note: The continuous acvf and acrf become infinite for scale 0 and lag 0, respectively.

4. Field Normalization

MHK process generates random fields that follow the $N(0,1)$. Here the following transformation (Papalexiou et al., 2007) is used, where its coefficients p_i are estimated through the minimization of the square error of the transformed data and the $N(0,1)$ distribution function.

$$Z_N = \left(p_1 Z^{-p_5} + p_2 \right) \left(p_3 + \sqrt{(1 + 1/p_4)\ln\left[p_4 (Z - p_3)^2 + 1 \right]} \right)$$

5a. Discrete autocorrelation function of the MHK process

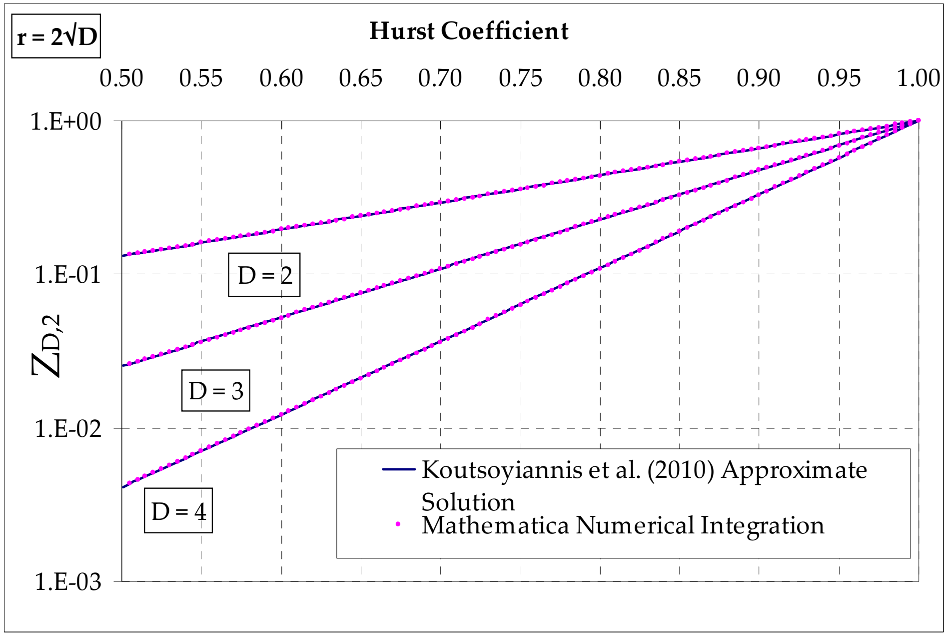
For the discrete acrf, one can adapt the Koutsoyiannis et al. (2010) approximate solution (KAS); this works well for $D \le 2$ and for $D \ge 3$ and lags greater than 1; for lags 0 and 1, a poly-line fit for lags can give better results.

$$I_{D,r} = \frac{\rho_{discete}(j_1,\dots,j_D)}{L_D}\gamma_{(0)} = 2^D \int_{v_1=0}^{1} \dots \int_{v_D=0}^{1} \left[(j_1^{2} + \dots + j_D^{2}\right]^{-B/2} \left[\left(1-|j_1|\right)\dots\left(1-|j_D|\right)\right] dv_1 \dots dv_D,\ r = \sqrt{j_1^{2} + \dots + j_D^{2}}, j_i \in \mathbb{N}$$

$$\varrho_{\text{discete}}(r) = \rho_{D,r} \approx \min\left\{C_D\left(\rho_{1,r}/C_1\right)^D, \rho_{1,r}\right\},\ \rho_{1,r} = |r+1|^{2H}/2 + |r-1|^{2H}/2 - |r|^{2H}$$

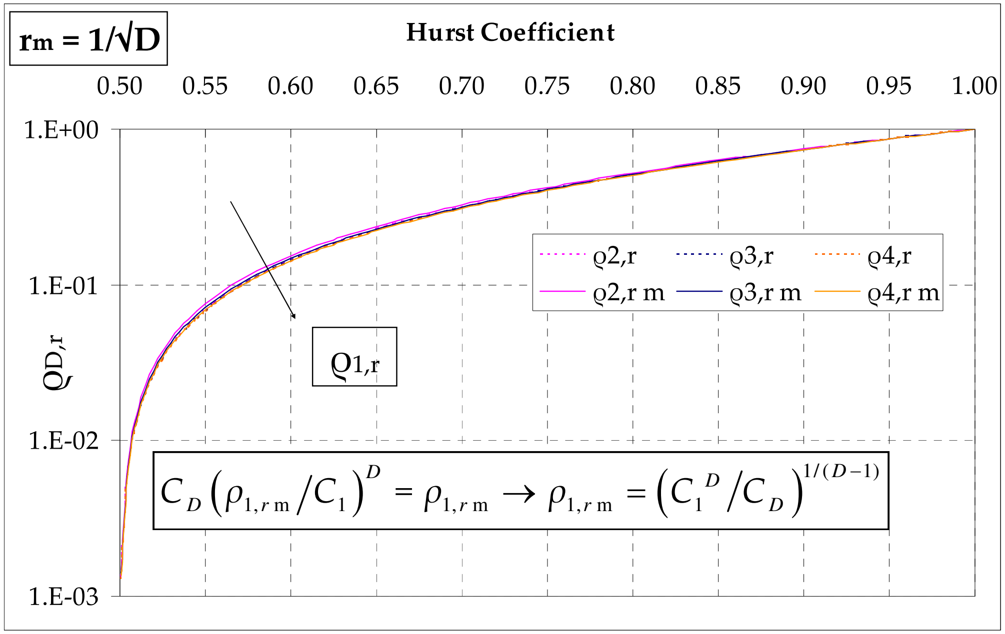
$$I_{D,0} = \frac{\rho_{D,0}}{L_D}\gamma_{(0)} \approx \frac{1}{C_D},\ C_D = \frac{(2H-1)[D(2H-1)+1]}{(D+1)}$$

Note that for great lags: $\rho_{D,r} \approx \dfrac{C_D}{r^B}$



$Z_{D,r} = C_D / \rho_{D,r} \approx 1 / I_{D,r}$

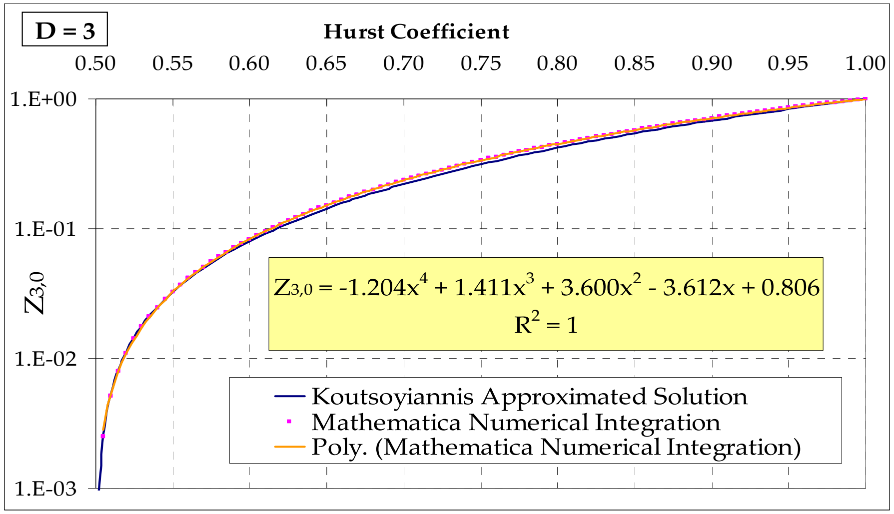
For lags r greater than 1 (thus r≥ 2), the KAS can be used.

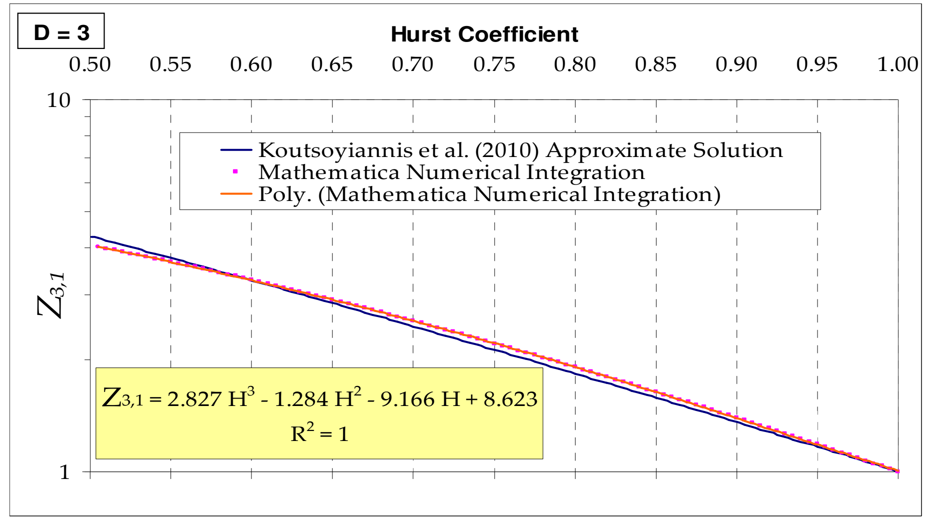


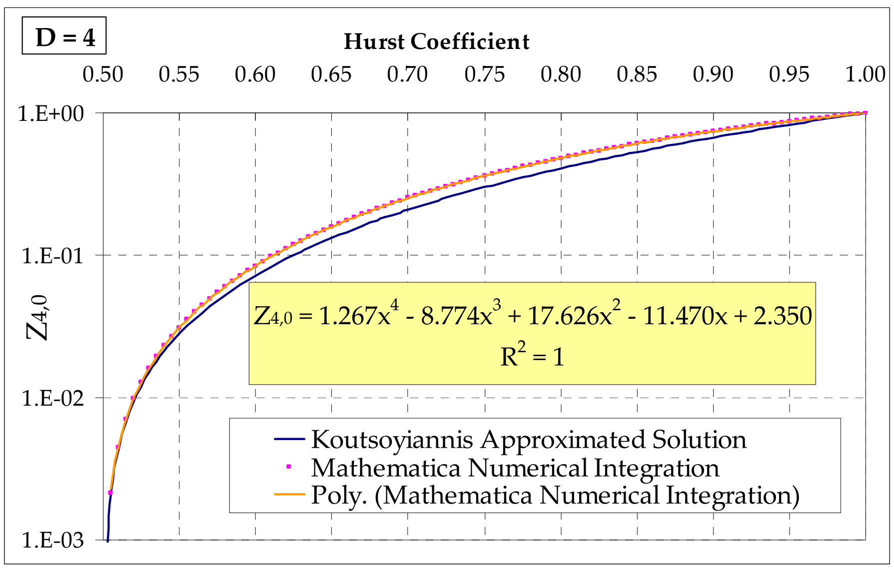
Comparison of equation $\rho_{D,r}$ with the KAS for lag 1/√D for a 2D, 3D and 4D field. It can be observed that the min function equals the first term when $r_m \ge 1/\sqrt{D}$.

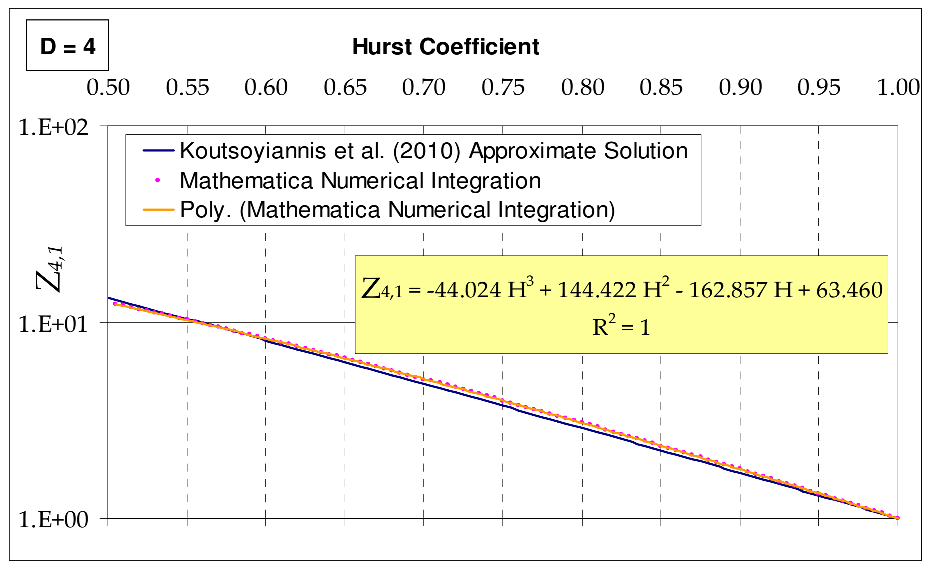
5b. Discrete autocorrelation function of the MHK process

Poly-line fits for 3D, 4D fields and lags 0 and 1, with the constrains $\rho_{D,r}$=1 for H=1 ($Z_{D,r} \to 1$) and $\rho_{D,0}$=0 for H=0.5 ($C_D \to 0$ more rapidly, so $Z_{D,r} \to 0$). Obviously, $\varrho_{D,r}$=0 for H=0.5 but it seems that this happens for H coefficients very close to 0.5 where the Mathematica software could not converge.









6. Simulation scheme for generating MHK process

SMA stands for Symmetric Moving Average and it can be used to generate a stochastic process with any structure of autocorrelation or power spectrum (Koutsoyiannis, 2000). Here, the SMA scheme has been extended to D spatio-temporal dimensions (direct extension from 1D and 2D schemes).

$$Z_v = \sum_{y_D=-q}^{q} \dots \sum_{y_1=-q}^{q} \alpha_{\boldsymbol{y}} W_{\boldsymbol{v-y}}$$

- Z_v: generated normalized random field of interest
- W: discrete white noise (random field) with zero mean ($\mu_w = 0$) and unit standard deviation ($\sigma_w = 1$) (since $\underline{Z}$ has been normalized).
- q: finite limit for the range of coefficients α_y (for m, the desired number of autocorrelation coefficients that are to be preserved).
- α_y: field of coefficients that can be determined through the Fourier transform F_γ of the autocovariance field γ_Z (Koutsoyiannis, 2000, Koutsoyiannis et al. 2010).

7. Spectral density and α_y coefficients of SMA

The spectral density F_γ of the stochastic field can be determined via the Fourier transform of the discrete form of autocovariance $\gamma_{\text{discrete}}(r)$. It can be shown that the Fourier transform F_α, of the field α_y, is related to F_γ (for $q=\infty$), thus the α_y field can then be estimated.

$$F_\gamma(s) = \frac{(2\pi)^{D/2}}{s^{D/2-1}} \int_0^\infty |r|^{D/2} \gamma(r) J_{D/2-1}\{2\pi s r\} dr = \frac{(2\pi)^{D/2}}{s^{D/2-1}} L_D \int_0^\infty |r|^{D/2-B} J_{D/2-1}\{2\pi s r\} dr \rightarrow$$

$$\rightarrow F_\gamma(s) = L_D E |s|^{B-D}, \text{where } E = \pi^{-D/2+B} \frac{\Gamma\left[(D-B)/2\right]}{\Gamma\left[B/2\right]}, \text{ for } 1 < B < D \rightarrow \frac{1}{2} < H < 1 - \frac{1}{2D}, s \in \mathbb{R}$$

$$\text{Thus, it can be shown that, } F_{\gamma_{\text{discete}}} \overset{\text{great lags}}{\approx} C_D \gamma_0 E |s_{discrete}|^{B-D}, \text{ for } \frac{1}{2} < H < 1 - \frac{1}{2D}, s_{\text{discrete}} \in \mathbb{N}$$

$$\text{Also, it can be assumed that, } F_{\gamma_{\text{discrete}}} \approx K_D |s_{\text{discrete}}|^{B-D}, \text{for } 0<H<1 \text{ and } K_D \text{ a coefficient}$$

From the above equations and assumptions it can be derived that:

$$F_\alpha \underset{\text{of 1}D\text{ SMA}}{\overset{\text{extension}}{=}} \sqrt{F_{\gamma_{\text{discrete}}}} \rightarrow \alpha_y \approx \alpha_0 \rho(\|y\|; H'), \text{ where } B' - D = (B - D)/2 \rightarrow H' = (H + 0.5)/2$$

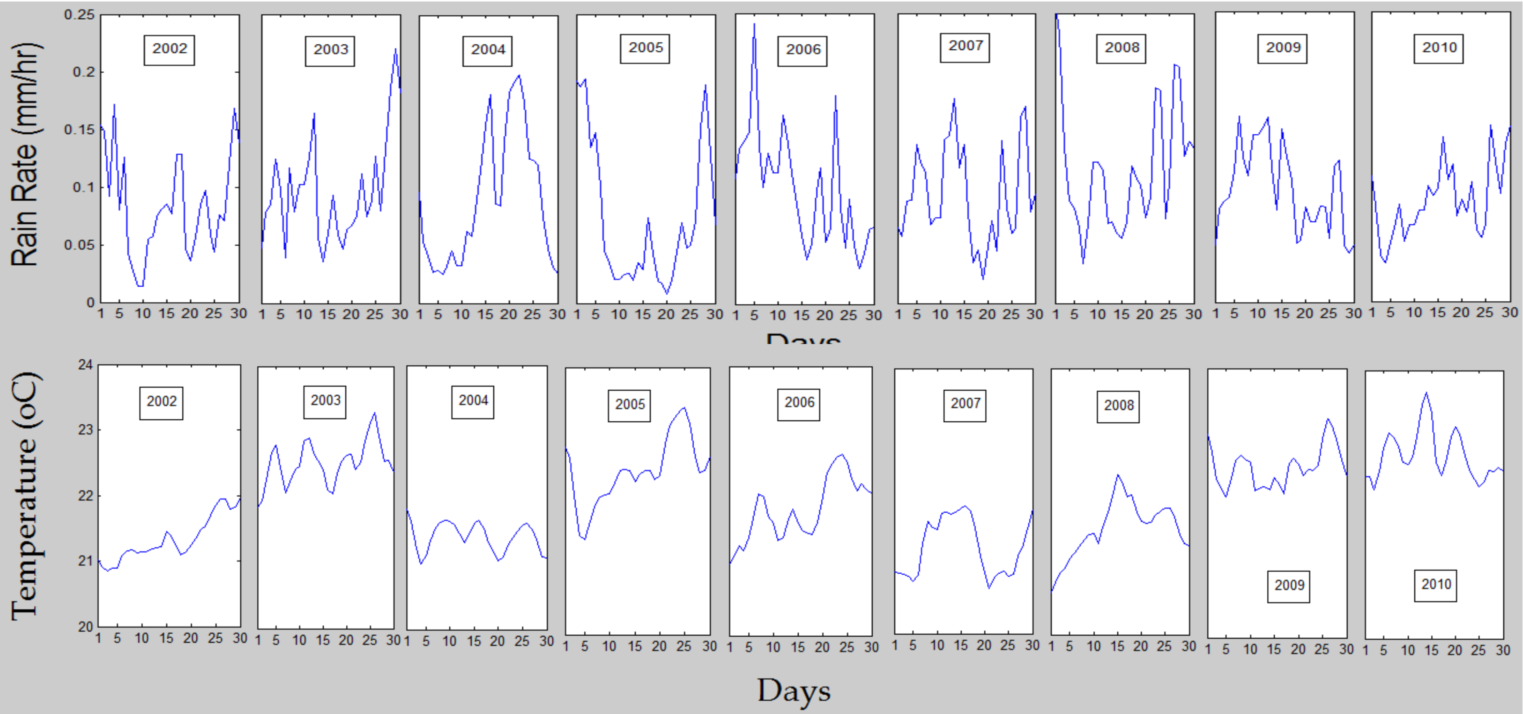
$$\gamma_0 = \sum_{y_D=-q}^{q} \cdots \sum_{y_1=-q}^{q} \alpha_{y_1,\ldots,y_D}^{\ 2} \rightarrow \alpha_0^{\ 2} = \gamma_0 \Big/ \sum_{\boldsymbol{y}=\underbrace{-q,-q,\ldots,-q}_{D}}^{q,q,\ldots,q} \rho^2(\|\boldsymbol{y}\|; H') \quad \text{and} \quad \alpha_{0,q=\infty} = \frac{\sqrt{\gamma_0 C_D(H)E(H)}}{C_D(H')E(H')}$$

8. Case study on observed rainfall fields

The application presented is based on an observed rainfall and temperature field South of the Indian Ocean (coordinates: 0N-30S, 55E-85E). The data were acquired from NASA satellite system (available on-line):

http://disc2.nascom.nasa.gov/Giovanni/tovas/TRMM_V6.3B42.shtml (rainfall)
http://gdata1.sci.gsfc.nasa.gov/daac-bin/G3/gui.cgi?instance_id=neespi_daily (temperature)

The sample consists of a spatial grid 31 x 31 points (of a 1° x 1° spatial resolution, approx. 110 km x 110 km) and a temporal 270-days grid for September 2002 to 2010 (October was also needed for the calculation of the acrf).

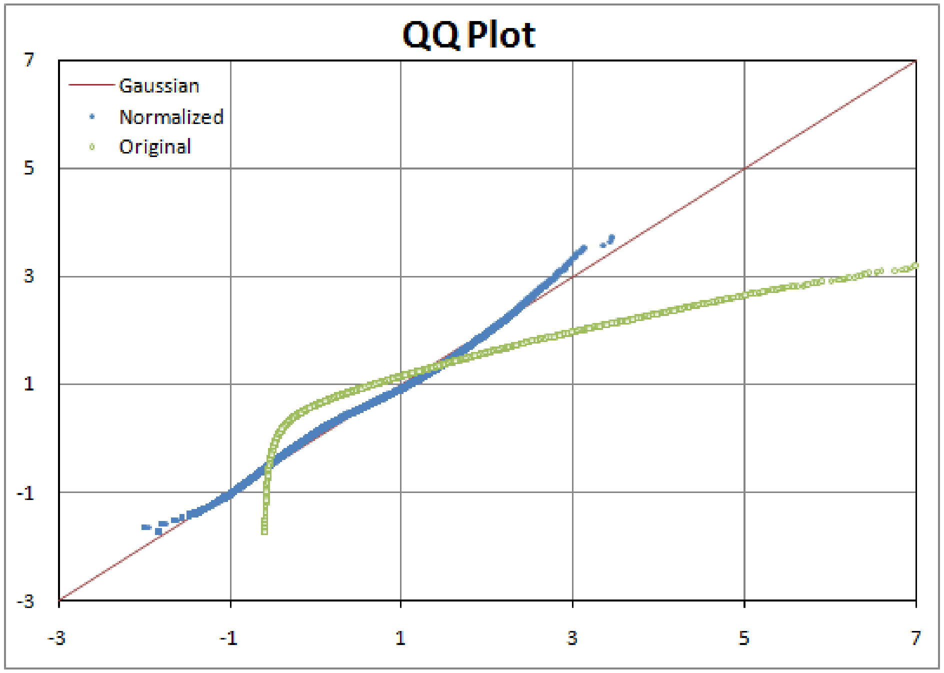


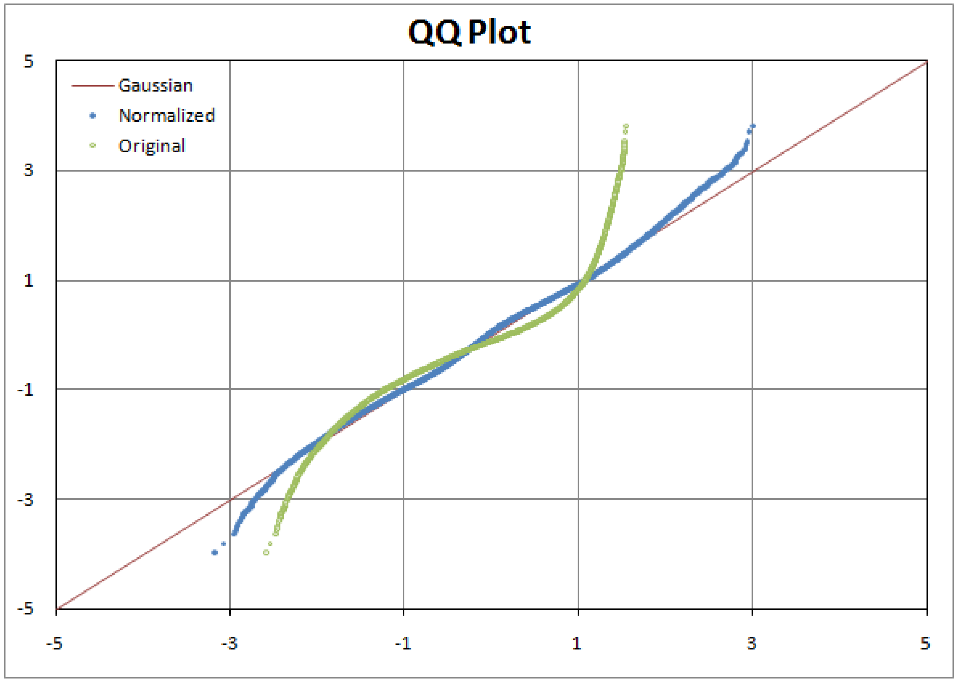
Spatial averaged rainfall and temperature fields. The rainfall field contained no missing values. The temperature field contained a lot of missing values which were supplemented assuming linear regression.

9. Normalization of fields

The zero values of the natural field are replaced with the small value of 1e-5.

The simulated field should be converted to natural units by solving arithmetically the inverted transformation.



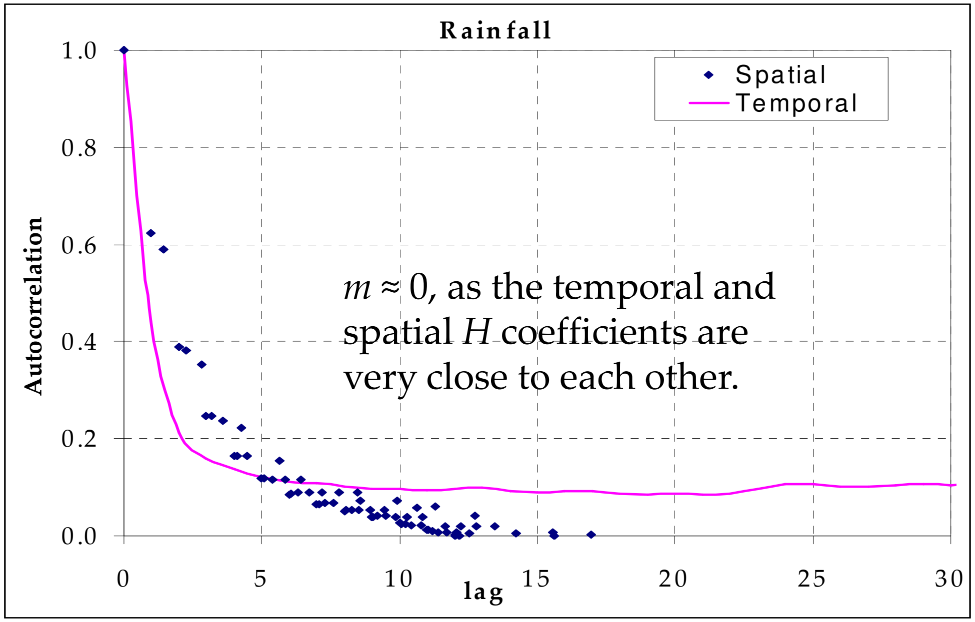


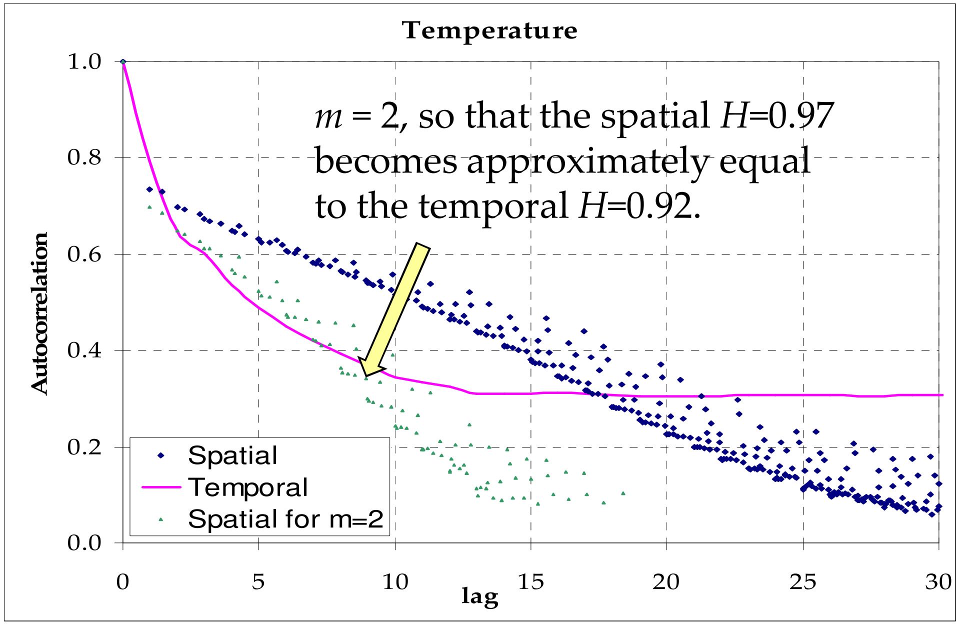
Rainfall:
$p_1 = 77.0$, $p_2 = 6,.0$ $p_3 = -1.0$, $p_4 = -2.2E\text{-}5$, $p_5 = 0.073$, $SE = 54.0$

Temperature:
$p_1 = 8.0E\text{-}7$, $p_2 = 1E4$, $p_3 = 7.4$, $p_4 = -35.2$, $p_5 = 0.00165$, $SE = 97.0$

10a. Dealing with anisotropy

A separate Hurst coefficient should be assigned to the quantities that are non isotropic to each other. It is still not applicable to create a multi-dimensional model that can synthesize time-series assuming anisotropy (thus assuming a different autocorrelation behavior in each dimension) and that is why hydrologists tend to use multi-variate models. A proposed solution is to omit intermediate data of the field grids, so as the multi-acrfs decay at the same rate (at least for the first lags). So, omitting factor (omfc) m means that the $(m*c)_{th}$ cell is omitted in the model, where c is 0,1,…,maximum number of cells in each direction of the sub-field and in the diagonally ones.

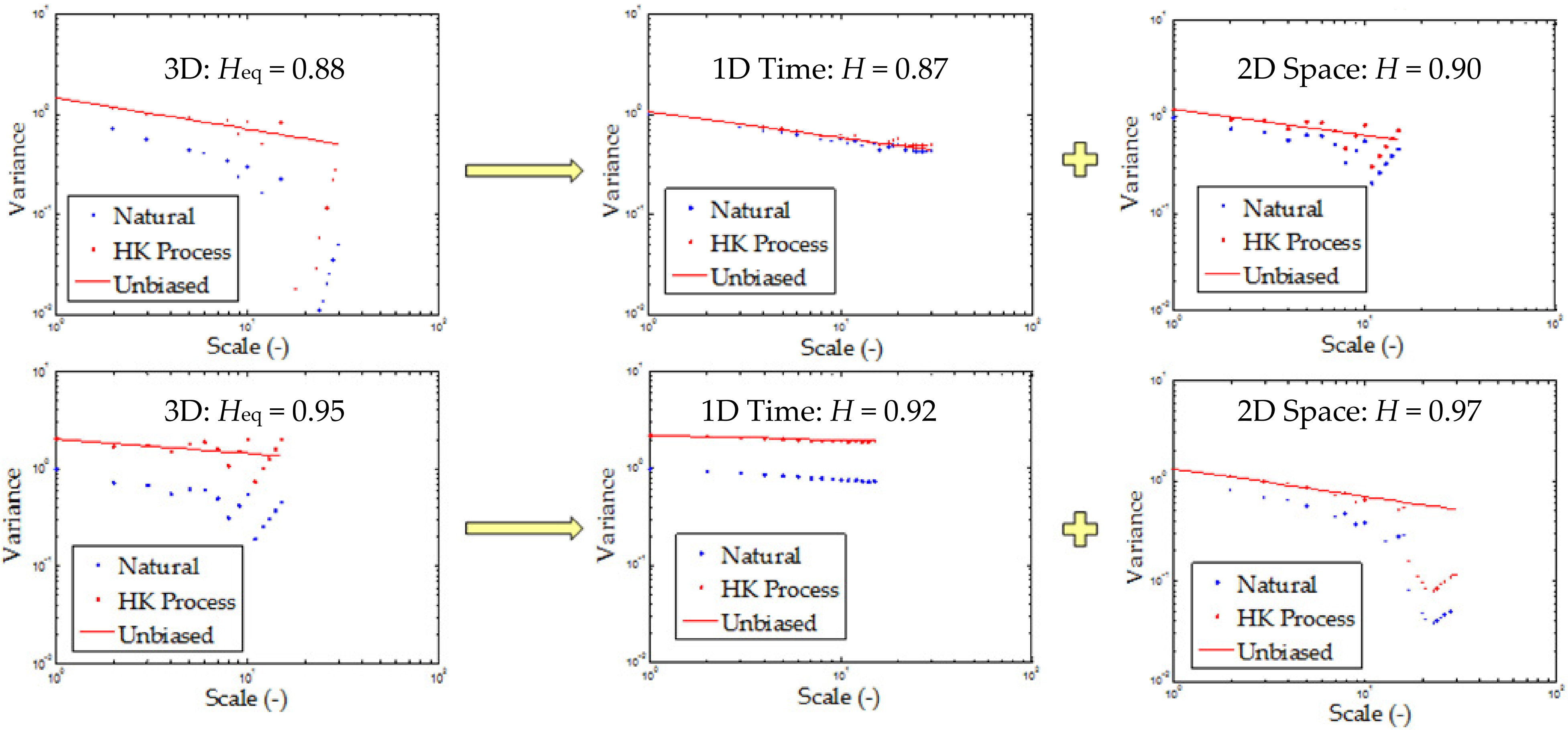




Since, the autocorrelagram is more sensible than the climacogram (as the first is the second derivative of the second, Koutsoyiannis, 2010), it is rather more appropriate to work with the second. So, one should change the m until the minimum Hurst coefficient is reached.

10b. Dealing with anisotropy

For this application, m is found 0 for the rainfall field and 2 for the temperature one, so that the Hurst coefficients of 0.87 and 0.92 are approximately reached for all the sub-fields, respectively.

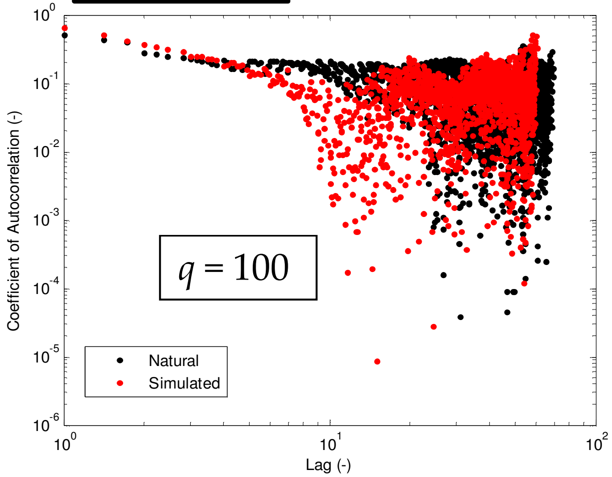


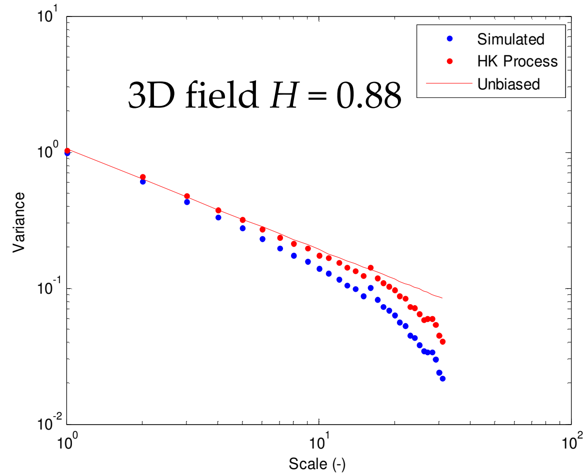
The equivalent Hurst coefficient can be determined by the equation: $H_{eq} = 1 - A / D$

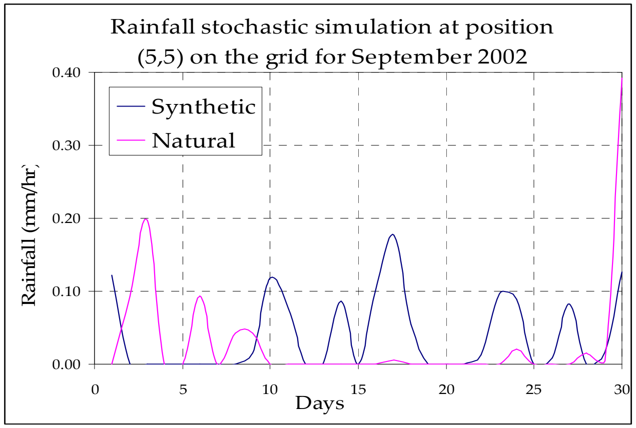
$A = D - \sum_{i=1}^{n} N_i H_i$, $N_i \leq D$, $\sum_{i=1}^{n} N_i = D$, N_i the number of dimensions of the sub-field i

11. Stochastic simulation model

Rainfall

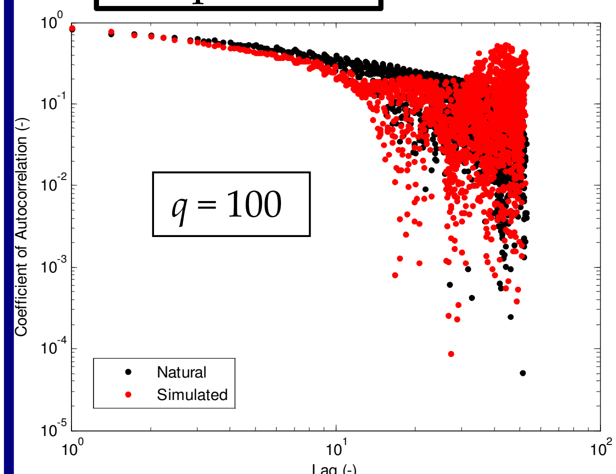


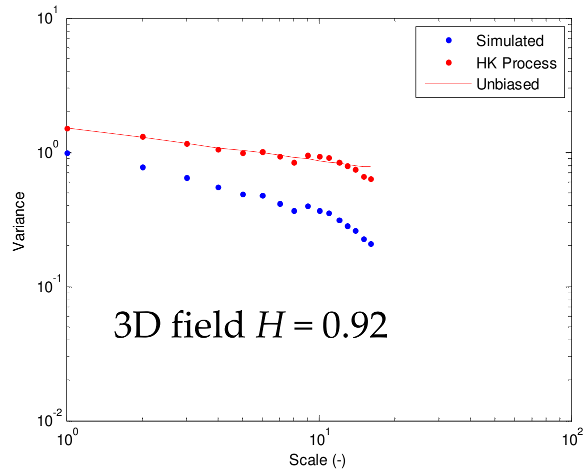


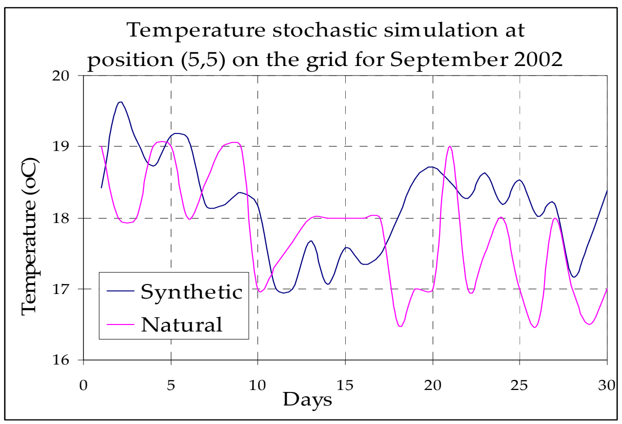


Note: All the negative values of the synthetic rainfall field are set equal to zero.

Temperature







Note: The simulated autocorrelations seem to be smaller than the natural ones. This is due to the small q parameter that is chosen, as larger values would enormous increase the numerical simulation time.

12. Conclusions

- A multi-dimensional (MD) stochastic simulation model is proposed, which is a direct extension of the 1D simple scaling process (HK or FGN).
- The HK process and the SMA generation algorithm are extended for any dimension *D* of the field (the autocorrelation function is extended for D≤4 and a methodology is proposed for greater dimensions).
- The MHK process is formulated assuming anisotropy through a methodology of changing the *m* omitting factor of the grids until the minimum Hurst coefficient of the sub-fields is reached.
- A 3D spatio-temporal model is applied based on an observed rainfall and temperature field.

References

Hurst, H.E. (1951) Long term storage capacities of reservoirs. Trans. ASCE 116, 776-808.

Kolmogorov, A.N., WienerscheSpiralenund einigeandereinteressanteKurvenin HilbertschenRaum, Dokl. Akad. NaukURSS, 26, 115–118, 1940.

Koutsoyiannis, D. (2000) A generalized mathematical framework for stochastic simulation and forecast of hydrologic time series. Wat. Resour. Re. 36(6), 1519-1534.

Koutsoyiannis, D. (2010), A random walk on water, Hydrol. Earth Syst. Sci., 14, 585–601, 2010.

Koutsoyiannis et al., (2011), Two-dimensional Hurst-Kolmogorov dynamics and its application to the study of rainfall fields (in preparation).

Papalexiou, S.M., A. Montanari, and D. Koutsoyiannis, Scaling properties of fine resolution point rainfall and inferences for its stochastic modelling, European Geosciences Union General Assembly 2007, Geophysical Research Abstracts, Vol. 9, Vienna, 11253, European Geosciences Union, 2007.