Two-dimensional Hurst-Kolmogorov process and its application to rainfall fields

Demetris Koutsoyiannis[1], Athanasios Paschalis[1,2] and Nikos Theodoratos[1]

1. Department of Water Resources and Environmental Engineering, Faculty of Civil Engineering, National Technical University of Athens, Heroon Polytechneiou 5, GR-157 80 Zographou, Greece

2. Institute of Environmental Engineering, ETH - Swiss Federal Institute of Technology, ETH Hönggerberg, HIF CO 46.8, CH - 8093 Zürich, Switzerland

Corresponding author: D. Koutsoyiannis, dk@itia.ntua.gr

Abstract The Hurst-Kolmogorov (HK) dynamics has been well established in stochastic representations of the temporal evolution of natural processes, yet many regard it as a puzzle or a paradoxical behaviour. As our senses are more familiar with spatial objects rather than time series, understanding the HK behaviour becomes more direct and natural when the domain of our study is no longer the time but the two-dimensional space. Therefore, here we detect the presence of HK behaviour in spatial hydrological and generally geophysical fields including Earth topography, and precipitation and temperature fields. We extend the one-dimensional HK process into two dimensions and we provide exact relationships of its basic statistical properties and closed approximations thereof. We discuss the parameter estimation problem, with emphasis on the associated uncertainties and biases. Finally, we propose a two-dimensional stochastic generation scheme, which can reproduce the HK behaviour and we apply this scheme to generate rainfall fields, consistent with the observed ones.

Keywords Hurst-Kolmogorov dynamics; stochastic processes; stochastic simulation; random fields; rainfall fields; hydrometeorology

1. Introduction

Very few people know the beauty of the mountains (Osho)

In his seminal paper, H. E. Hurst (1951), after studying numerous geophysical time series, observed that: "Although in random events groups of high or low values do occur, their tendency to occur in natural events is greater. This is the main difference between natural and random events". This difference triggered coining the term "Hurst phenomenon", which even today many regard as an enigma or a puzzle, as also implied by the denotation "phenomenon" in this term. Koutsoyiannis and Cohn (2008) tried to show that the reluctance to recognize this behaviour and incorporate it into standard geophysical statistics may be related to the unfavourable terminology used, including "phenomenon" and the related term "long memory" introduced by the influential studies by Mandelbrot and Wallis (1968) and Mandelbrot and van Ness (1968). Specifically, "long memory" (a metaphoric term related to slowly decaying autocorrelation) gives a wrong message leading to incorrect understanding of the physical mechanism. "Long memory" is counterintuitive and not realistic. Klemes (1974) insisted that "the Hurst phenomenon is not necessarily an indicator of infinite memory of a process" and pointed out that it can be explained in terms of long-term change rather than long-term memory. He also compared the memory-based interpretation to the Ptolemaic planetary model, which worked well but hampered progress in astronomy for centuries. Koutsoyiannis (2002) demonstrated that changes occurring at multiple (e.g. three or more) time scales result in a process exhibiting Hurst behaviour and Koutsoyiannis (2005b) showed that the same process can be derived by applying the principle of maximum entropy (i.e., uncertainty) on multiple time scales. Following Koutsoyiannis and Cohn (2008) and to give proper credit to A. N. Kolmogorov (1940), who was the first to propose a mathematical model to describe this behaviour (for the study of turbulence), here we refer to this behaviour as the Hurst-Kolmogorov (HK) behaviour and to the stationary stochastic process that reproduces it as the HK process.

As our senses are more familiar with spatial objects rather than time series, the understanding of the HK behaviour becomes more direct and natural when the domain in

which we study a geophysical process is no longer the time but the two-dimensional (2D) space. We all have an experiential understanding of landscapes, like the one shown in Figure 1 (upper panel; part of the Alps). Considering a landscape as a realization of a stochastic field, with the altitude as a random variable and the 2D space as the index of the field, we can easily recognize visual similarities with a HK process. A cross section of the landscape will give us something visually similar to a time series. Such a cross section is shown in Figure 1 (middle panel) where, in addition to the graph based on each grid point of the digital elevation model, running averages with scale lengths 25 and 50 grid points are also shown. The fact that these averages exhibit large fluctuations, instead of tending to be constant as in purely random processes, is a prominent characteristics of the HK behaviour (Koutsoyiannis, 2010a). In his celebrated book "*The Fractal Geometry of Nature*", Mandelbrot (1977) emphasized the differences of natural objects with simple objects of Euclidean geometry, stating "*mountains are not cones, coastlines are not circle*s…". These differences refer to the shapes of individual natural elements. However, our emphasis in this study is on the overall macroscopic structure, rather than on individual elements. Inevitably the macroscopic view should involve stochastic descriptions. In this respect, we are more interested on the differences of natural with random processes or fields, originally pointed out by Hurst (the grouping or clustering of similar values), which is more prominent in natural than in random processes. In simple words, with reference to a landscape, this groupings are none other than the mountains and the valleys, which would not appear in a purely random landscape (see Figure 1, lower panel).

We can also observe the tendency of high and low values to cluster in space in hydrometeorological fields such as rainfall fields (Figure 2, Figure 3) and temperature fields (Figure 4). This clustering occurs also in time, but here we focus on the spatial occurrence. The stochastic descriptions of such fields is practically the only means to characterize the fields macroscopically and, simultaneously, to enable simulations of the fields, useful for such tasks as engineering design, hydrometeorological forecasting, and water resources management. The theory of stochastic fields is a direct extension of the theory of stochastic processes and can provide effective descriptions of the spatial and spatio-temporal structure of

hydrometeorological fields. In particular, for the rainfall fields, whose study is crucial for the rational design of flood protection works and the flow hydrograph prediction in real time, specialized models, e.g. based on the extension of point processes to spatial structures (e.g. Chandler *et al*., 2002) have been developed and effectively used. A more generalized framework for studying geophysical fields has been provided by the notion of multifractals (e.g. Schertzer and Lovejoy, 1987; Lovejoy and Schertzer, 1995; Pathirana and Herath, 2002; Gagnon *et al*., 2006; Veneziano and Langousis, 2010; Koutsoyiannis and Langousis, 2011), which has given impressive and operational results for the characterization and stochastic simulation of natural fields such as rainfall, cloud structure and Earth topography.

However, here we prefer to use the simpler notion of a 2D HK process (see e.g. Penttinen and Virtamo, 2003), which is a direct extension in a 2D space of the well-known 1D HK process (also known as fractional Gaussian noise—fGn, due to Mandlbrot and van Ness, 1968), based on the formalism that Kolmogorov and later (independently) Mandelbrot proposed. In fact the 2D HK process, which we describe next, is the simplest stochastic process that represents the above described natural behaviour. Its simplicity is a valuable characteristic when understanding is a principal target, and is also associated with parsimony of parameterization and mathematical ease: it uses a single parameter, the Hurst coefficient H, additional to those of classical statistics. Simplicity and parsimony make it realistic to derive all mathematical properties and also produce a full stochastic simulation model by easy analytical means.

This formalism has some additional advantages over some multifractal studies. Thus, it distinguishes the different types of scaling behaviours, i.e. the scaling in state and the scaling in time and/or space (Koutsoyiannis, 2005a), which some multifractal analyses confuse. Specifically, the scaling in state is a property of the marginal (order 1) distribution function (and here is handled using appropriate marginal distribution functions or appropriate normalizing transformation of the random variable of interest) whereas the scaling in space and time are properties characterizing the dependence structure of the field and are clearly associated with different scaling exponents, with different meanings. Furthermore the HK formalism allows understanding and characterizing the uncertainties and biases associated

with the estimation of the statistical properties and parameters of the process or field. As we will see next, such uncertainties and biases are prominent, substantially higher than in purely random processes, even in the estimation of the lowest moments, such as means and standard deviations. Some multifractal analyses miss this and do not properly account for the uncertainty and bias. However, the enhancement of uncertainty is perhaps the most important characteristic that a proper understanding of the scaling behaviour of nature should include. Neglecting uncertainties and biases (as often done in multifractal studies that, for example, rely on estimates of high order statistical moments treating them as if they were precisely known quantities), may have dramatic consequences at all modelling aspects including the identification of the appropriate model (Koutsoyiannis, 2010b; Papalexiou *et al.*, 2010).

2. Stochastic model

We denote by $\underline{z}$ a random variable representing the modelled quantity of interest (e.g. topographic elevation, rainfall, temperature, or nonlinear transformations thereof) and by z (without underscore) any realization (numerical value) of the variable. We assume that $\underline{z}$ is defined on a 2D space denoted by the continuous (real) variables (x, y) or the discrete (integer) variables (i, j) which define the index space of the stochastic field $\underline{z}(x, y)$ or $\underline{z}_{i,j}$; the two are related as

$$\underline{z}_{i,j} := \frac{1}{\Delta^2} \int_{(i-1)\Delta}^{i\Delta} \int_{(j-1)\Delta}^{j\Delta} \underline{z}(x, y)\, \mathrm{d}x\, \mathrm{d}y \tag{1}$$

where Δ is a fixed interval representing a "unit" scale for conversion of continuous to discrete spatial representation. The discrete space representation has several advantages, both practical (simulations are necessarily done in gridded space) and theoretical (as we will see, the continuous space representation involves some infinities). Therefore our analysis will be based on discretized space, while we will use the continuous-space representation in an auxiliary role. For simplicity and in accord to the explanatory character of this study, we assume that the field $\underline{z}$ is stationary and isotropic and we denote its mean as $\mu := E[\underline{z}_{i,j}]$, its autocovariance as $\gamma_{k,l} := \mathrm{Cov}[\underline{z}_{i,j}, \underline{z}_{i+k,j+l}]$ and its autocorrelation as $\rho_{k,l} := \mathrm{Corr}[\underline{z}_{i,j}, \underline{z}_{i+k,j+l}] = \gamma_{k,l}/\gamma_0$, where γ_0 is the variance, i.e. $\gamma_0 := \mathrm{Var}[\underline{z}_{i,j}] = \gamma_{0,0} \equiv \sigma^2$ (σ denotes standard deviation).

Furthermore we define the average process at a spatial scale which is an integer (k = 1, 2, …) multiple of Δ, i.e.,

$$\underline{z}_{i,j}^{(k)} := \frac{1}{k^2} \sum_{m=(i-1)k+1}^{ik} \sum_{n=(j-1)k+1}^{jk} \underline{z}_{m,n} \tag{2}$$

and denote its autocovariance as $\gamma_{l,m}^{(k)}$, its variance as $\gamma_0^{(k)}$ and its autocorrelation as $\rho_{l,m}^{(k)}$. The notation implies that a superscript equal to 1 can be omitted (i.e., $\underline{z}_{i,j}^{(1)} \equiv \underline{z}_{i,j}$, $\gamma_{l,m}^{(1)} \equiv \gamma_{l,m}$, etc.).

A 2D HK process can be defined as a stochastic process, which for any indices i, j, m, n, and any scales k and l, has the property

$$(\underline{z}_{i,j}^{(k)} - \mu) =_d \left(\frac{k}{l}\right)^{2H-2} (\underline{z}_{m,n}^{(l)} - \mu) \tag{3}$$

where $=_d$ means that the two random variables have the same finite-order joint distribution functions. The constant H in the exponent is the so called Hurst coefficient and takes on values in the interval (0, 1). Setting l = 1 in (3) and taking variances, we easily obtain that the variance at scale k is

$$\gamma_0^{(k)} = k^{4H-4} \gamma_0 \tag{4}$$

For comparison, we recall that in the standard, 1D, HK process, defined by

$$(\underline{z}_i^{(k)} - \mu) =_d \left(\frac{k}{l}\right)^{H-1} (\underline{z}_m^{(l)} - \mu) \tag{5}$$

the variance at scale k is

$$\gamma_0^{(k)} = k^{2H-2} \gamma_0 \tag{6}$$

In the 1D case, the autocorrelation function can be easily derived from (5) and is

$$\rho_j = g_1(j; H) := |j + 1|^{2H} / 2 + |j - 1|^{2H} / 2 - |j|^{2H} \approx H(2H - 1) |j|^{2H-2} \tag{7}$$

where j is the lag. This however, is more difficult to determine in the 2D process. Initially, similar to the 1D process, we can observe that (3) defines three very different types of processes. Specifically, for H = 0.5, it can be easily verified that (4) characterizes a purely random process with zero autocorrelation. For $0 < H < 0.5$, it can be demonstrated from (4) (e.g., for scales 1 and 2) that the autocorrelation is negative even for the smallest

displacements or lags (e.g. a displacement equal to one). While this is mathematically possible and describes a special type of an antipersistent process, it is not physically realistic: Consistency with physical reality demands that at neighbouring times or locations the process should be positively correlated (Koutsoyiannis, 2010a). Finally, the case $0.5 < H < 1$ defines a physically realistic process which has the clustering or grouping behaviour discussed in the Introduction. For the latter case, for the continuous space representation of a stationary and isotropic process we can write the autocovariance function as

$$\gamma(l_x,l_y) := \mathrm{Cov}[\underline{z}(x, y), \underline{z}(x + l_x, y + l_y)] = f(r) = f(\sqrt{l_x^2 + l_y^2}) \tag{8}$$

where l_x and l_y denote spatial displacements in the directions x and y, respectively, f is a function to be specified and $r := \sqrt{l_x^2 + l_y^2}$ is the radial displacement. As shown in Appendix A, a specific form of the function $f(r)$ consistent with (4) is

$$f(r) = a\, r^{4H-4} \tag{9}$$

where a is any constant. Based on this, in Appendix A we derive the exact variance as a function of H and the autocorrelation function for the discrete space representation, which is most useful in applications. These functions do not have a closed form. Nonetheless, simple closed approximations are possible and are derived in Appendix A from numerical solutions, i.e.,

$$\gamma_0^{(k)} \approx \frac{3\, a}{(2H-1)\,(4H-1)}\, k^{4H-4}\, \Delta^{4H-4} \tag{10}$$

$$\rho_{l,m}^{(k)} \approx \rho_d^{(k)} = g_2(d; H) := \min\left\{\frac{(4H-1)[g_1(d; H)]^2}{3H^2\,(2H-1)}, g_1(d; H)\right\} \tag{11}$$

where $d = \sqrt{l^2 + m^2}$ is the radial displacement in the discrete space representation corresponding to spatial displacements in the directions x and y equal to l and m, respectively (notice that d is not necessarily an integer, even if l and m are, but (11) can be applied even for non-integer d, l and m). It is seen in Figure 5 that the approximations of the variance $\gamma_0^{(k)}$ by (10) (upper panel) and the autocorrelation $\rho_{l,m}^{(k)}$ by (11) are very good for all values of H and all displacements. A further simplification of (11) can be obtained by observing that the two terms in the curly brackets in (11) become almost equal for $d = 1$ and thus

$$\rho_d^{(k)} \approx \min\left\{\frac{[g_1(d;\, H)]^2}{g_1(1;\, H)},\, g_1(d;\, H)\right\} = g_1(d;\, H)\, \min\left\{\frac{g_1(d;\, H)}{g_1(1;\, H)},\, 1\right\} \tag{12}$$

Comparing the continuous-space with the discrete-space representation, we can observe that in the former the variance (obtained from $f(r)$ in (9) for $r = 0$) is infinite, whereas in the latter it is always finite (tending to infinity as the scale $k \to 0$). In addition, in the continuous-space representation the autocorrelation function cannot be defined (because the variance is infinite), whereas in the discrete-space representation, not only can be it defined, but it is invariant for any scale and thus is the most characteristic function of the stochastic field.

The spectral density $s_\gamma^c(u_x, u_y)$ of the stochastic field (where u_x and u_y are frequencies at the two directions) for the continuous space representation is presented in Appendix A and can also provide an approximation of the spectral density for the discrete space representation, which is a power function of the radial frequency $p := \sqrt{u_x^2 + u_y^2}$, i.e.,

$$s_\gamma(u_x, u_y) = s_\gamma(p) = c\, p^{2-4H} \tag{13}$$

where c depends on the variance γ_0 and the Hurst parameter H.

3. Parameter estimation

We assume that the observed sample $z_{i,j}$ extends over a rectangular area with edge size k_m, and we regard $z_{i,j}$ as a realization of the HK field $\underline{z}_{i,j}$. The sample size is $n = k_\mathrm{m}^2$. Using the notation of section 2, the sample average is, clearly, $\overline{\underline{z}} \equiv \underline{z}_{1,1}^{(k_\mathrm{m})}$. It can be readily shown that $\overline{\underline{z}}$ is an unbiased estimator of the mean μ in HK statistics (HKS) as in classical statistics (CS) in which the sample is random, i.e. all its elements are independent and identically distributed (iid). By virtue of (4), the variance of the estimator of $\overline{\underline{z}}$ is

$$\mathrm{Var}[\overline{\underline{z}}] = \frac{\sigma^2}{n^{2-2H}} \tag{14}$$

which is identical to that in the 1D HK process but, for high H, departs notoriously from the corresponding law in classical statistics, i.e.,

$$\mathrm{Var}[\overline{\underline{z}}] = \frac{\sigma^2}{n} \tag{15}$$

The latter equation can be written as $n = \sigma^2/\mathrm{Var}[\overline{\underline{z}}]$ and, by analogy, this allows us to define an "equivalent" or "effective" sample size n' of an HKS sample of size n, such that a sample size

n' in CS results in the same uncertainty of the mean as a sample with size n in HKS. Specifically, we define

$$n' := \sigma^2/\mathrm{Var}[\underline{\bar{z}}] \tag{16}$$

so that for an HK process

$$n' = n^{2-2H} \tag{17}$$

The estimator of the variance in CS is

$$\underline{s}^2 = \frac{1}{n-1}\sum_{i=1}^{k}\sum_{j=1}^{k}(\underline{z}_{i,j} - \underline{\bar{z}})^2 \tag{18}$$

(with s standing for standard deviation) and is unbiased for independent $\underline{z}_{i,j}$. For dependent $\underline{z}_{i,j}$, it can be easily inferred that the expected value of $\underline{s}^2$ is

$$E[\underline{s}^2] = \frac{n}{n-1}\{E[\underline{z}_{i,j}^2] - E[\underline{\bar{z}}^2]\} = \frac{n}{n-1}\{\mathrm{Var}[\underline{z}_{i,j}] - \mathrm{Var}[\underline{\bar{z}}]\} \tag{19}$$

or

$$E[\underline{s}^2] = \frac{1-1/n'}{1-1/n}\sigma^2 \tag{20}$$

This indicates that the classical estimator is biased. The bias may be very high for high H: for example, for $H = 0.99$ (and, as we will see below, this value is not unrealistically high) and for $k = 100$, so that $n = k^2 = 10\,000$, $n' = 1.20$ (only!), so that $E[\underline{s}^2] = 0.17\,\sigma^2$ (a 83% bias!). Notably, equations (17) and (20) are precisely the same as in the 1D HK process (Koutsoyiannis, 2003; Koutsoyiannis and Montanari, 2007).

The parameters σ and H are dependent to each other, whereas μ is independent to both (Tyralis and Koutsoyiannis, 2010). Thus, μ can be easily estimated by $\bar{z}$. However, the estimation of σ and H should be done simultaneously and should take into account the bias in the estimation of standard deviation. Here we follow an algorithm similar to that proposed by Koutsoyiannis (2003) for the 1D HK process, which is based on the CS estimates $s^{(k)}$ at several scales k from $k = 1$ to a maximum scale k' chosen such that the sample size $n_{k'}$ at that scale is at least 10. In the 2D case, the edge length at scale k is $[k_\mathrm{m}/k]$ (where the square brackets denote the integer part of a real number), so that $n_k = [k_\mathrm{m}/k]^2$.

From (20) we obtain

$$E[(\underline{s}^{(k)})^2] = \frac{1-1/n'_k}{1-1/n_k}(\sigma^{(k)})^2 \tag{21}$$

or, combining (4) and (17),

$$E[(\underline{s}^{(k)})^2] = c_k(H)\, k^{4H-4}\, \sigma^2 \tag{22}$$

where $c_k(H) := (1 - 1/n'_k) / (1 - 1/n_k) = (1 - [k/k_m]^{4-4H}) / (1 - [k/k_m]^2)$ is a bias correction factor. Therefore, we try to minimize the fitting error of the estimated variance $(s^{(k)})^2$ to the right-hand side of (22), as detailed in Appendix B.

Equation (22), which is the basis of this method of parameter estimation, has an easy graphical depiction, which is very useful to assess whether the HK model is appropriate or not. This is made by a logarithmic plot of the estimated variance $(s^{(k)})^2$ (or the estimated standard deviation $s^{(k)}$) versus scale k, also known as climacogram (from the Greek climax, meaning scale; Koutsoyiannis, 2010a). In a purely random (iid) process, where $H = 0.5$, the double logarithmic plot of $(s^{(k)})^2$ *vs.* k will form a straight line with slope –2. In a positively autocorrelated HK process, if we plot the true (population) variance $\gamma_0^{(k)} \equiv (\sigma^{(k)})^2$, as given in (4), not considering the bias correction factor $c_k(H)$, we will have a straight line with a milder slope, $4H - 4$. However, because for large H the factor $c_k(H)$ is significantly lower than 1 and cannot be neglected, the plot of $E[(\underline{s}^{(k)})^2]$ given in (22) *vs.* k would be a concave curve, with slope becoming steeper as k increases. The plot of the sample estimate of variance $(s^{(k)})^2$ should comply with that of $E[(\underline{s}^{(k)})^2]$ rather than with that of $\gamma_0^{(k)}$.

Figure 6 depicts such climacograms for the example fields discussed above and given in Figures 1-4. The estimated parameters for all fields are shown in Table 1 and it can be seen that in all cases the Hurst coefficient is exceptionally high ($0.99 < H < 1$). It is important that $H < 1$, because in the case $H = 1$ the theoretical variance becomes infinite (even in the discrete space representation), whereas the case $H > 1$ is mathematically meaningless. Figure 6 shows that the HK framework provides a better alternative in modelling all examined examples, in comparison to the classical, iid model, which is fully inappropriate.

4. Simulation scheme

A simple scheme for generation of realizations of the 2D HK process can be derived as an extension of the 1D symmetric moving average (SMA) scheme, introduced by Koutsoyiannis (2000), which is a stochastic model that can generate time series with any autocorrelation

structure. Extending the 1D scheme, we introduce the 2D SMA scheme as

$$\underline{z}_{i,j} = \sum_{l=-\infty}^{\infty} \sum_{m=-\infty}^{\infty} \alpha_{l,m}\, \underline{v}_{i-l,j-m} \tag{23}$$

where $\alpha_{l,m}$ is a field of coefficients to be determined and $\underline{v}_{i,j}$ is a discrete white noise random field with zero mean and unit variance (assuming also, without loss of generality, that $\mu = E[\underline{z}_{i,j}] = 0$). Numerically, this can be implemented using a finite number (q) of coefficients, i.e.,

$$\underline{z}_{i,j} = \sum_{l=-q}^{q} \sum_{m=-q}^{q} \alpha_{l,m}\, \underline{v}_{i-l,j-m} \tag{24}$$

By extending the 1D analysis in Koutsoyiannis (2000) it can be shown without difficulties that the Fourier transform of the field α, $s_\alpha(u_x, u_y) = s_\alpha(p)$, is related to that of the autocovariance $s_\gamma(u_x, u_y)$ by

$$s_\alpha(u_x, u_y) = \sqrt{s_\gamma(u_x, u_y)} \tag{25}$$

By virtue of (13),

$$s_\alpha(u_x, u_y) = \sqrt{c}\, p^{1-2H} \tag{26}$$

Comparing (13) and (26), we observe that the latter is equivalent to the Fourier transform of the autocovariance of a HK process with Hurst parameter H', so that $2 - 4H' = 1 - 2H$ or $H' = (0.5 + H)/2$. By inverting the Fourier transform, we conclude that the coefficients $\alpha_{l,m}$ should be proportional to the autocorrelations $\rho_{l,m}$ of an HK process with Hurst parameter H', i.e.,

$$\alpha_{l,m} = c'\, \rho_{l,m} = c'\, g_2(\sqrt{l^2 + m^2};\, 1/4 + H/2) \tag{27}$$

where the function $g_2(\,)$ is defined in (11) and c' is an unknown constant. The latter could be determined by observing that (24) results in

$$\gamma_0 = \sum_{l=-q}^{q} \sum_{m=-q}^{q} \alpha_{l,m}^2 \tag{28}$$

or

$$\gamma_0 = c'^2 \sum_{l=-q}^{q} \sum_{m=-q}^{q} [g_2(\sqrt{l^2 + m^2};\, 1/4 + H/2)]^2 \tag{29}$$

which can be readily solved for c'.

An example demonstration of the simulation scheme has been done for the rainfall field of

Figure 3. The model described above can easily generate Gaussian random fields. However, the examined rainfall field (as well as several geophysical processes and fields) is not Gaussian and, thus, we need to normalize the original field before applying the procedure. Several families of transformations can be found in the bibliography, e.g., the Box-Cox family of transformations (Box and Cox, 1964), which in many cases give satisfactory results. However, this normalizing transformation is not appropriate for rainfall, because it produces exponential distribution tails, while the rainfall distribution tails are more likely of power type (Koutsoyiannis, 2004, 2005a), thus reflecting an asymptotic scaling behaviour in state. Therefore, here we used the transformation

$$z = g(x) = \lambda \left[1 + \left(\frac{x}{v} \right)^{-\theta} \right] \sqrt{\left(1 + \frac{1}{\kappa} \right) \ln \left[1 + \kappa \left(\frac{x}{\lambda} \right)^2 \right]} \tag{30}$$

which has a power-type tail. In (30), θ and κ are dimensionless parameters, and λ and v are parameters with dimensions identical to those of x, so that z has also dimensions identical to those of x. This transformation is derived from a similar one introduced by Koutsoyiannis *et al.* (2008), which aimed to produce a power-type right tail in the distribution of x. The quantity $[1 + (x/v)^{-\theta}]$ was proposed by Papalexiou *et al.* (2007) because it provides a better fit on the left tail of the distribution of x. The parameters of the transformation are $\theta = 1.0715$, $\kappa = 3.6016$, $\lambda = 3.0429$ mm, $v = 1.7039$ mm, and the normalized field is shown in Figure 3 (right).

The Hurst coefficient of the normalized field, is estimated, using the algorithm of Appendix B, to $H = 0.997$. A normalized stochastic field generated using the above method is shown in Figure 7 with statistical estimates given in Table 2, along with the de-normalized (natural) synthetic rainfall field, obtained by inverting the transformation (30) (this can be done only numerically). The comparison of the statistical characteristics of the synthetic field with the original one of Figure 3 is given in Table 2 and Figure 7 and indicates a good performance of the simulation algorithm.

5. Summary, conclusions and discussion

The so-called Hurst phenomenon, detected in many geophysical processes, has been regarded by many as a puzzle. The "infinite memory", often associated with it, has been regarded as a counterintuitive and paradoxical property. However, it may be easier to perceive the Hurst-Kolmogorov behaviour if one detaches the "memory" interpretation and associates it with the rich patterns apparent in real world phenomena, which are absent in purely random processes. Furthermore, as our senses are more familiar with spatial objects rather than time series, understanding the Hurst-Kolmogorov behaviour becomes more direct and natural when the domain, in which we study a geophysical process, is no longer the time but the 2D space. In this respect, this study offers an extension of the one-dimensional HK process into two dimensions. We provide exact relationships of its basic statistical properties and closed approximations thereof. We discuss the parameter estimation problem, with emphasis on the increased uncertainties and biases of the classical statistical estimators, when applied to a Hurst-Kolmogorov process. These are very similar, as in the one-dimensional HK process. Finally, we study a stochastic generation scheme, which can reproduce the HK behaviour. The scheme is an extension of the symmetric moving average algorithm introduced by Koutsoyiannis (2000), which can perform with any arbitrary autocorrelation function.

This study has an exploratory and explanatory character and, therefore, the focus is on the simplest and most parsimonious, yet sufficiently realistic, representation of real world fields, such as terrain, rainfall and temperature. It does not aim to provide detailed modelling tools for such fields. Therefore, the stochastic model relies on a single parameter, the Hurst coefficient, to describe the spatial dependence of the field. The assumption of homogeneous and isotropic fields is, therefore, appropriate for the scope of this study, although in a more detailed representation of natural fields it may be relaxed (e.g. to take account of the dependence of precipitation intensity on altitude and geographical characteristics). Evidently, further research is needed for the case where this assumption is relaxed.

Acknowledgment We thank Tim Cohn and another reviewer for their kind and encouraging comments and their useful suggestions.

Appendix A: Exact statistical properties of the 2D HK process and derivation of their approximations

Combining (1) and (8) we obtain that the discrete-space autocovariance is

$$\gamma_{l,m} = \mathrm{Cov}[\underline{z}_{1,1}, \underline{z}_{1+l,1+m}] = \frac{1}{\Delta^4} \int_{x=0}^{\Delta} \int_{y=0}^{\Delta} \int_{x'=l\Delta}^{(l+1)\Delta} \int_{y'=m\Delta}^{(m+1)\Delta} f(\sqrt{(x-x')^2 + (y-y')^2})\, \mathrm{d}x\mathrm{d}y\mathrm{d}x'\mathrm{d}y' \quad \text{(A1)}$$

After algebraic manipulations which are typical for stochastic processes, the above integral can be reduced to a double integral, i.e.,

$$\gamma_{l,m} = \int_{-1}^{1} \int_{-1}^{1} f(\Delta \sqrt{(l-\xi)^2 + (m-\psi)^2})\, (1-|\xi|)\, (1-|\psi|)\mathrm{d}\xi\mathrm{d}\psi \quad \text{(A2)}$$

For $l = m = 0$, we obtain the variance of the process, i.e.

$$\gamma_0 = \int_{-1}^{1} \int_{-1}^{1} f(\Delta \sqrt{\xi^2 + \psi^2})\, (1-|\xi|)\, (1-|\psi|)\mathrm{d}\xi\mathrm{d}\psi \quad \text{(A3)}$$

which obviously simplifies to

$$\gamma_0 = 4 \int_{0}^{1} \int_{0}^{1} f(\Delta \sqrt{\xi^2 + \psi^2})\, (1-|\xi|)\, (1-|\psi|)\mathrm{d}\xi\mathrm{d}\psi \quad \text{(A4)}$$

Likewise, the variance at scale k is

$$\gamma_0^{(k)} = 4 \int_{0}^{1} \int_{0}^{1} f(k\Delta \sqrt{\xi^2 + \psi^2})\, (1-|\xi|)\, (1-|\psi|)\mathrm{d}\xi\mathrm{d}\psi \quad \text{(A5)}$$

It can be easily verified that a sufficient condition to make (A5) consistent with (4) is the one given in (9). By virtue of (9), the discrete-space variance from (A4) for any spatial scale $k\Delta$ becomes

$$\gamma_0^{(k)} = a\, k^{4H-4}\, \Delta^{4H-4}\, I_0(H) \quad \text{(A6)}$$

where

$$I_0(H) := 4 \int_{0}^{1} \int_{0}^{1} (\xi^2 + \psi^2)^{2H-2}\, (1-|\xi|)\, (1-|\psi|)\mathrm{d}\xi\mathrm{d}\psi \quad \text{(A7)}$$

Likewise,

$$\gamma_{l,m}^{(k)} = a\, k^{4H-4}\, \Delta^{4H-4}\, I_{l,m}(H) \quad \text{(A8)}$$

where

$$I_{l,m}(H) := \int_{-1}^{1}\int_{-1}^{1} [(l-\xi)^2 + (m-\psi)^2]^{2H-2} (1-|\xi|)(1-|\psi|)\mathrm{d}\xi\mathrm{d}\psi \tag{A9}$$

We observe that the quantity $I_0(H)$ depends only on the Hurst coefficient H, whilst $I_{l,m}(H)$ depends also on the displacements l and m. The autocorrelation function is

$$\rho_{l,m}^{(k)} = I_{l,m}(H) / I_0(H) \tag{A10}$$

and is independent of the scale k and the constant a.

The integral $I_0(H)$ has a rather complicated and inconvenient analytical expression whereas $I_{l,m}(H)$ is difficult to express in a closed form. However, numerical integration is very easy and, from its results, simple analytical approximations can be derived. Specifically, $I_0(H)$ can be calculated as

$$I_0(H) \approx \frac{3}{(2H-1)(4H-1)} \tag{A11}$$

which allows the approximation of $\gamma_0^{(k)}$ given in (10).

For $I_{l,m}(H)$ we first observe that for l and m much larger than 1, the term in square brackets in (A9) can be approximated as $(l^2 + m^2)^{2H-2}$, or d^{4H-4}. Thus, it is easy to see that

$$I_{l,m}(H) = I_d(H) \approx d^{4H-4}, \quad \text{for } d >> 1 \tag{A12}$$

so that

$$\rho_{l,m}^{(k)} \approx \rho_d^{(k)} \approx (1/3)(2H-1)(4H-1)\, d^{4H-4}, \quad \text{for } d >> 1 \tag{A13}$$

For smaller values of d down to zero, we can make use of the function $g_1(d; H)$ that defines the autocorrelation of the 1D process, to find an approximation valid for all displacements. From (A13) and (7) we observe that, for high d, the autocorrelation ρ_d in the 2D case is proportional to the square of ρ_d in the 1D case with proportionality coefficient equal to $[(1/3)(2H-1)(4H-1)] / [H(2H-1)]^2$. We extend this proportionality to smaller d, also disallowing the autocorrelation of the 2D case to exceed that of the 1D case. This results in (11).

The spectral density $s_\gamma^c(u_x, u_y)$ of the stochastic field for its continuous-space representation

can be calculated by taking the 2D Fourier transform $s^{c}_{\gamma}(u_x, u_y)$ of the autocovariance $\gamma(l_x,l_y) = a\,(l_x^2 + l_y^2)^{2H-2}$. Due to circular symmetry ($\gamma(l_x,l_y) = \gamma(r) = a\,r^{4H-4}$ for $r := \sqrt{l_x^2 + l_y^2}$), the 2D Fourier transform equals the Henkel transform for the radial frequency $p := \sqrt{u_x^2 + u_y^2}$ (Bracewell, 2000, p. 336), i.e.,

$$s^{c}_{\gamma}(u_x, u_y) = s^{c}_{\gamma}(p) = 2\pi \int_0^{\infty} r\,\gamma(r)\,J_0(2\pi p r)\,\mathrm{d}r \tag{A14}$$

where $J_0(\,)$ is the Bessel function of the first kind. This results in

$$s^{c}_{\gamma}(u_x, u_y) = s^{c}_{\gamma}(p) = a\,\pi^{3-4H}\frac{\Gamma(2H-1)}{\Gamma(2-2H)}\,p^{2-4H} \tag{A15}$$

where $\Gamma(\,)$ is the gamma function. We can then conclude that an approximation of the spectral density for the discrete space representation will also be a power function of the frequency p, as given by (13).

Appendix B: Fitting error and its minimization

As stated, the basis of the fitting method is the minimization of the fitting error of the estimated variance $(s^{(k)})^2$ to the right-hand side of (22). In logarithmic terms and using weights equal to $1/k^2$ to the partial error of each scale k, the fitting error is

$$e^2(\sigma, H) := \sum_{k=1}^{k'} \frac{\{\ln[(s^{(k)})^2] - \ln[c_k(H)\,k^{4H-4}\,\sigma^2]\}^2}{k^2} \tag{B1}$$

which, after algebraic manipulations, becomes

$$e^2(\sigma, H)/4 = \sum_{k=1}^{k'} \frac{[\ln(k^2\,s^{(k)}) - \ln c_k(H)/2 - H\ln k^2 - \ln\sigma]^2}{k^2} \tag{B2}$$

The minimization of (B2) can be done only numerically and results in the optimal estimates of σ and H simultaneously (see the algorithmic details in Koutsoyiannis, 2003, and Tyralis and Koutsoyiannis, 2010).

References

Box, G. E. P. and D. R. Cox (1964) An analysis of transformations, *Journal of the Royal Statistical Society B*, 26 (2), 211-252.

Bracewell, R. N. (2000) *The Fourier Transform and Its Applications*, 3rd Edition, McGraw-Hill, Boston, Mass., USA.

Chandler, R. E., H. S. Wheater, V. S. Isham, C. Onof, S. M. Bate, P. J. Northrop, D. R. Cox, and D. Koutsoyiannis (2002) Generation of spatially consistent rainfall data, *Continuous river flow simulation: methods, applications and uncertainties, BHS Occasional Paper No. 13*, 59–65, British Hydrological Society, London.

Gagnon, J.-S., S. Lovejoy, and D. Schertzer (2006) Multifractal earth topography, *Nonlin. Processes Geophys.* 13, 541-570.

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

Klemes, V. (1974) The Hurst phenomenon: A puzzle?, *Water Resour. Res.* 10 (4) 675-688.

Kolmogorov, A. N. (1940) Wienersche Spiralen und einige andere interessante Kurven in Hilbertschen Raum, *Dokl. Akad. Nauk URSS* 26, 115–118.

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

Koutsoyiannis, D. (2002) The Hurst phenomenon and fractional Gaussian noise made easy, *Hydrological Sciences Journal* 47 (4), 573–595.

Koutsoyiannis, D. (2003), Climate change, the Hurst phenomenon, and hydrological statistics, *Hydrological Sciences Journal*, 48(1), 3-24.

Koutsoyiannis, D. (2004) Statistics of extremes and estimation of extreme rainfall, 2, Empirical investigation of long rainfall records, *Hydrological Sciences Journal* 49 (4), 591–610.

Koutsoyiannis, D. (2005a) Uncertainty, entropy, scaling and hydrological stochastics, 1, Marginal distributional properties of hydrological processes and state scaling, *Hydrological Sciences Journal* 50 (3), 381–404.

Koutsoyiannis, D. (2005b) Uncertainty, entropy, scaling and hydrological stochastics, 2, Time dependence of hydrological processes and time scaling, *Hydrological Sciences Journal* 50 (3), 405–426.

Koutsoyiannis, D. (2010a) A random walk on water, *Hydrology and Earth System Sciences* 14, 585–601.

Koutsoyiannis, D. (2010b) Some problems in inference from time series of geophysical processes (solicited), *European Geosciences Union General Assembly 2010, Geophysical Research Abstracts, Vol. 12*, Vienna, EGU2010-14229, European Geosciences Union (http://www.itia.ntua.gr/en/docinfo/973/).

Koutsoyiannis, D., and T. A. Cohn (2008) The Hurst phenomenon and climate (solicited), *European Geosciences Union General Assembly 2008, Geophysical Research Abstracts, Vol. 10*, Vienna, 11804, European Geosciences Union (http://www.itia.ntua.gr/en/docinfo/849/).

Koutsoyiannis, D., and A. Langousis (2011) Precipitation, *Treatise on Water Science*, edited by S. Uhlenbrook, Elsevier (in press).

Koutsoyiannis, D., and A. Montanari (2007) Statistical analysis of hydroclimatic time series: Uncertainty and insights, *Water Resources Research* 43 (5), W05429, doi:10.1029/2006WR005592.

Koutsoyiannis, D., H. Yao, and A. Georgakakos (2008) Medium-range flow prediction for the Nile: a comparison of stochastic and deterministic methods, *Hydrological Sciences Journal*, 53 (1), 142–164.

Lovejoy, S., and D. Schertzer (1995), Multifractals and rain, in *Uncertainty Concepts in Hydrology and Hydrological Modelling*, edited by A. W. Kundzewicz, pp. 62–103, Cambridge Univ. Press, New York.

Mandelbrot, B. B. (1977), *The Fractal Geometry of Nature*, Freeman, New York.

Mandelbrot, B. B., and J. W. van Ness (1968) Fractional Brownian motions, fractional noises and applications, *SIAM Review* 10 (4) 422-437.

Mandelbrot, B. B., and J. R. Wallis (1968) Noah, Joseph, and operational hydrology, *Water Resour. Res.* 4 (5), 909-918.

Papalexiou, S.M., D. Koutsoyiannis, and A. Montanari, Mind the bias! (2010) *STAHY Official Workshop: Advances in statistical hydrology*, Taormina, Italy, International Association of Hydrological Sciences (http://www.itia.ntua.gr/en/docinfo/985/).

Papalexiou, S.M., A. Montanari, and D. Koutsoyiannis (2007) 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 (http://www.itia.ntua.gr/en/docinfo/751/).

Pathirana A., Herath S. (2002) Multifractal modelling and simulation of rain fields exhibiting spatial heterogeneity, *Hydrology and Earth System Sciences* 6(4), 695-708.

Penttinen, A., and J. Virtamo (2003) Simulation of Two-Dimessional Fractional Gaussian Noise, *Methodology and Computing in Applied Probability* 6, 99-107.

Schertzer, D., and S. Lovejoy (1987), Physical modeling and analysis of rain and clouds by anisotropic scaling of multiplicative processes, *J. Geophys. Res.* 92, 9693–9714.

Tyralis, H., and D. Koutsoyiannis (2010) Simultaneous estimation of the parameters of the Hurst-Kolmogorov stochastic process, Stochastic Environmental Research & Risk Assessment, DOI: 10.1007/s00477-010-0408-x.

Veneziano, D., and A. Langousis (2010) Scaling and fractals in hydrology, In: *Advances in Data-based Approaches for Hydrologic Modeling and Forecasting*, Ed. Sivakumar B, World Scientific, Chapter 4, 145 pp.

Figures

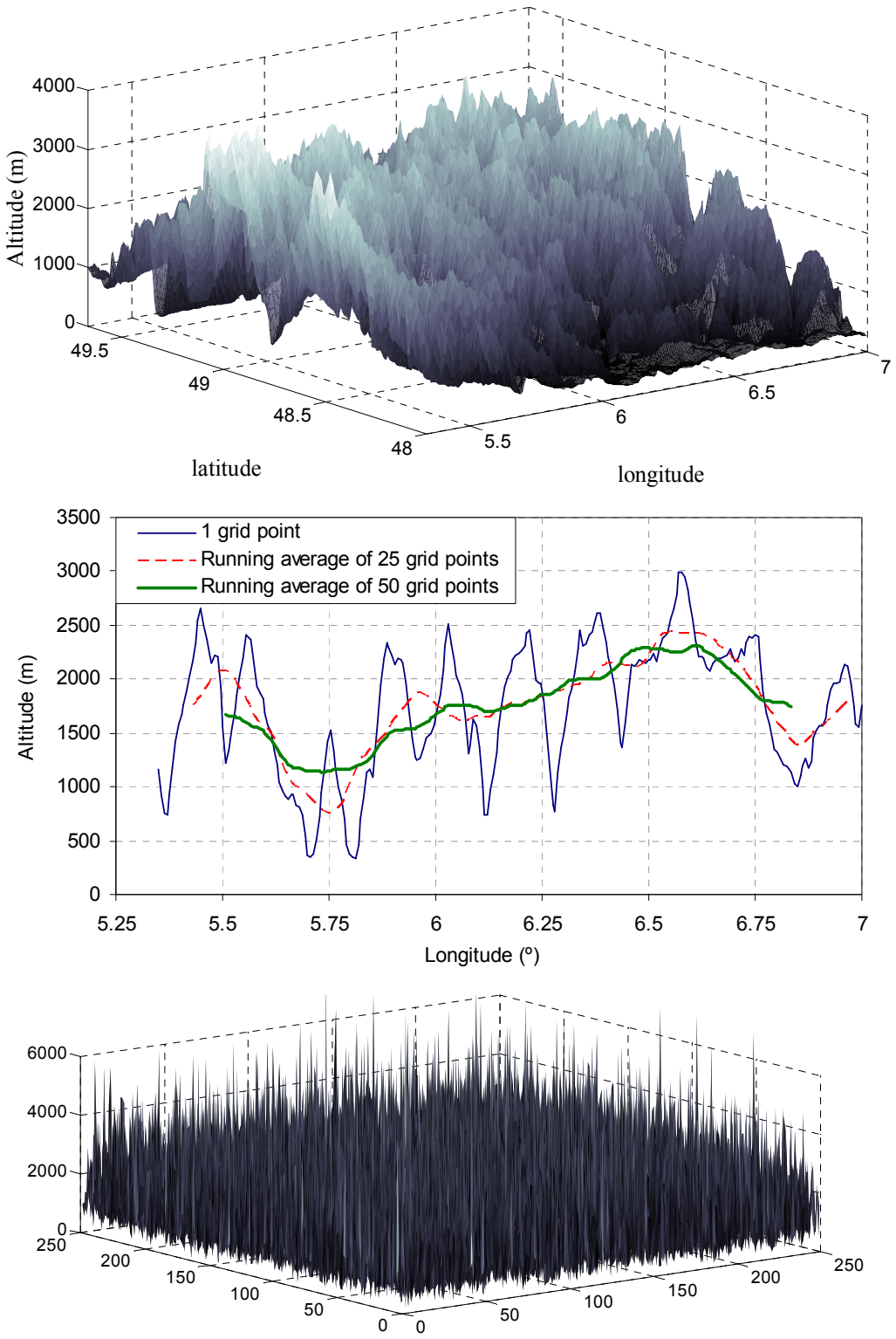

Figure 1 (upper) Elevation data from NASA's shuttle radar topography mission (SRTM) in an area in the Alps (coordinates 48ºN-49.67ºN, 5.35ºE-7ºE; http://srtm.csi.cgiar.org/SELECTION/listImages.asp); (**middle**) cross-section at 48.835ºN; (**lower**) a purely random landscape generated from the Gamma distribution with number of grid points, mean and variance equal to those of the real world landscape of the upper panel.

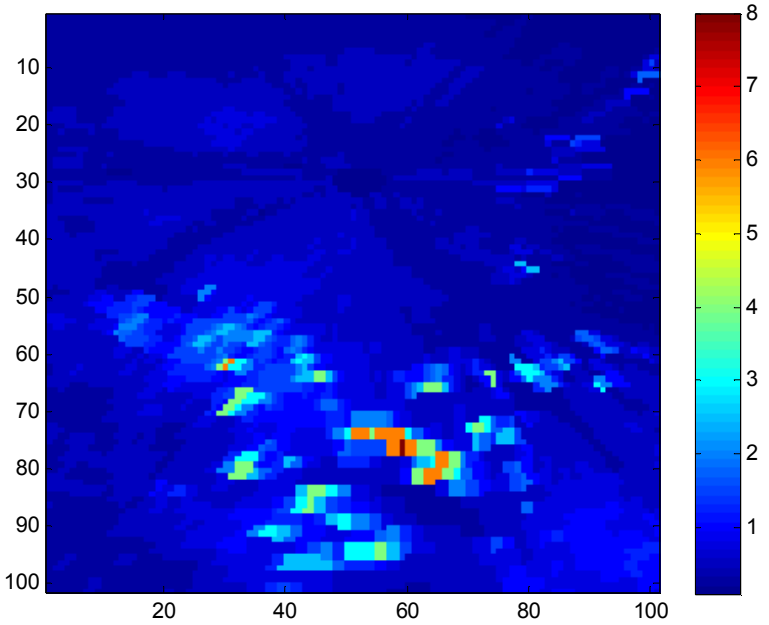

Figure 2 NEXRAD (radar-Doppler) measurements of rainfall depth (mm) in the state of Alaska, USA (PACG coordinates 56.51ºN-135.31ºW) at 2005-11-18, 21:00-23:00; data from the National Oceanic and Atmospheric Administration; http://www.ncdc.noaa.gov/nexradinv/; grid size 0.0066º.

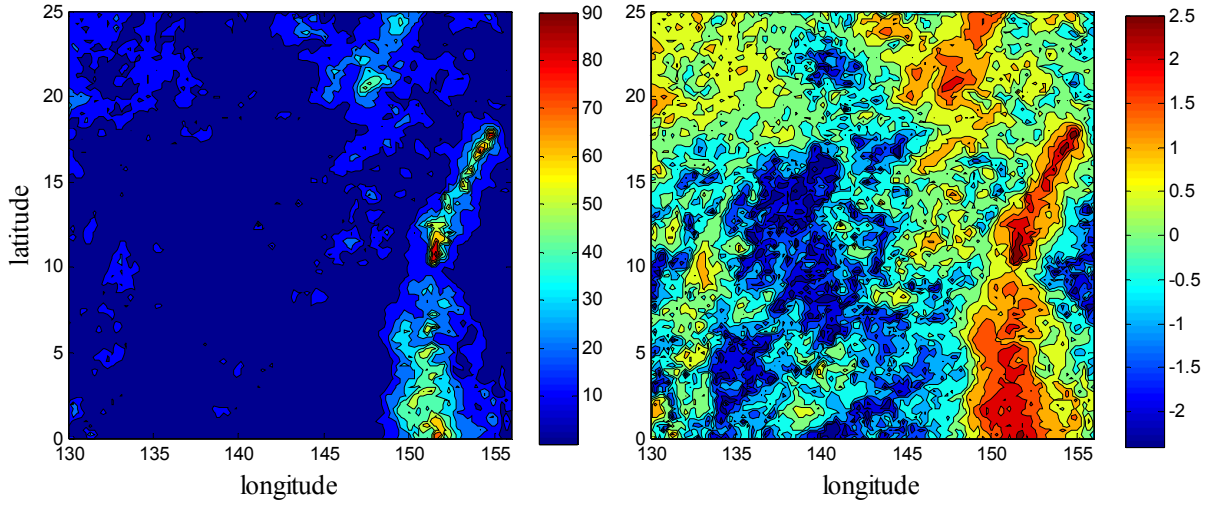

Figure 3 Satellite data for rainfall depth field in the Pacific ocean, north of New Zealand (coordinates 0ºN-25ºN, 130ºE-155ºE) for the time period 13 to 16 Jul 2005; data from NASA's TRMM programme (http://disc2.nascom.nasa.gov/Giovanni/tovas/TRMM_V6.3B42.shtml); the sample consists of 101 × 101 grid points, at resolution 0.25º × 0.25º; (**left**) natural field (mm); (**right**) normalized field.

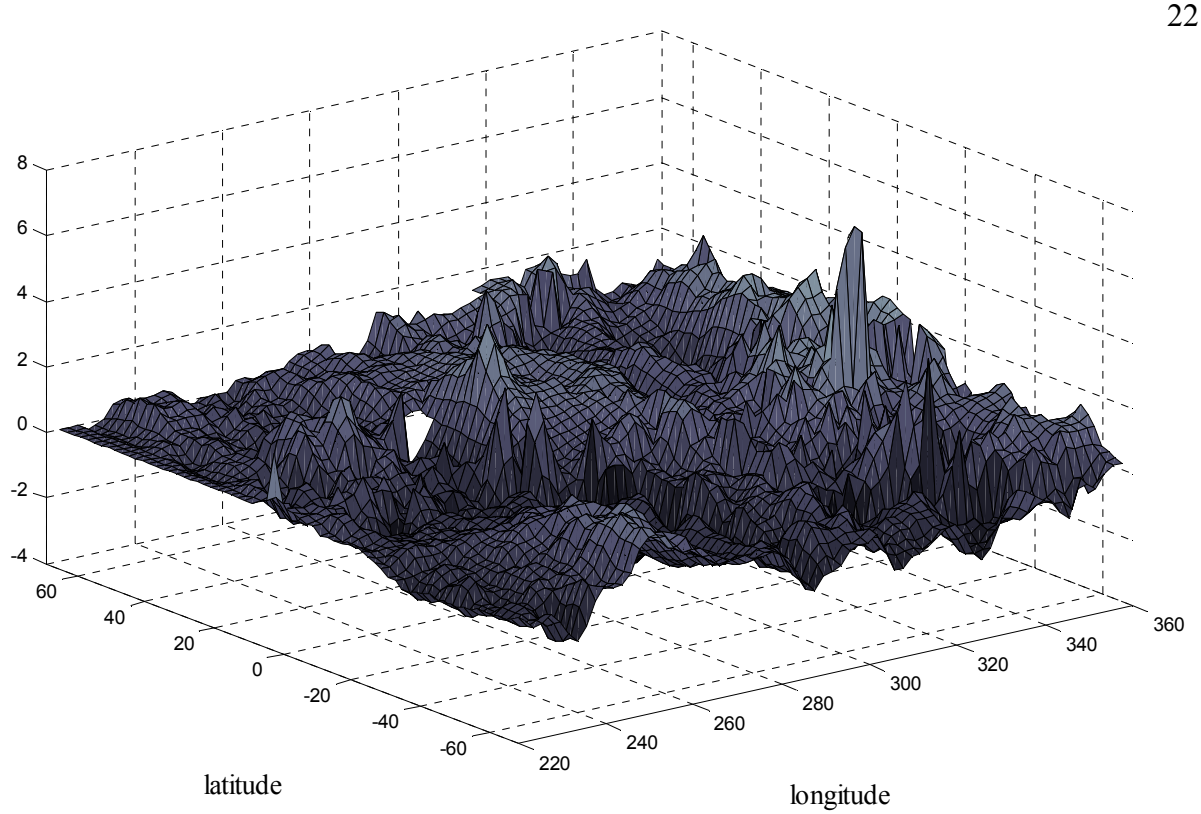

Figure 4 Temperature departure from historic mean in ºC on the part of the globe between 60°S-60°N and 220°E-360°E at 1 July 2010; data from http://iridl.ldeo.columbia.edu.

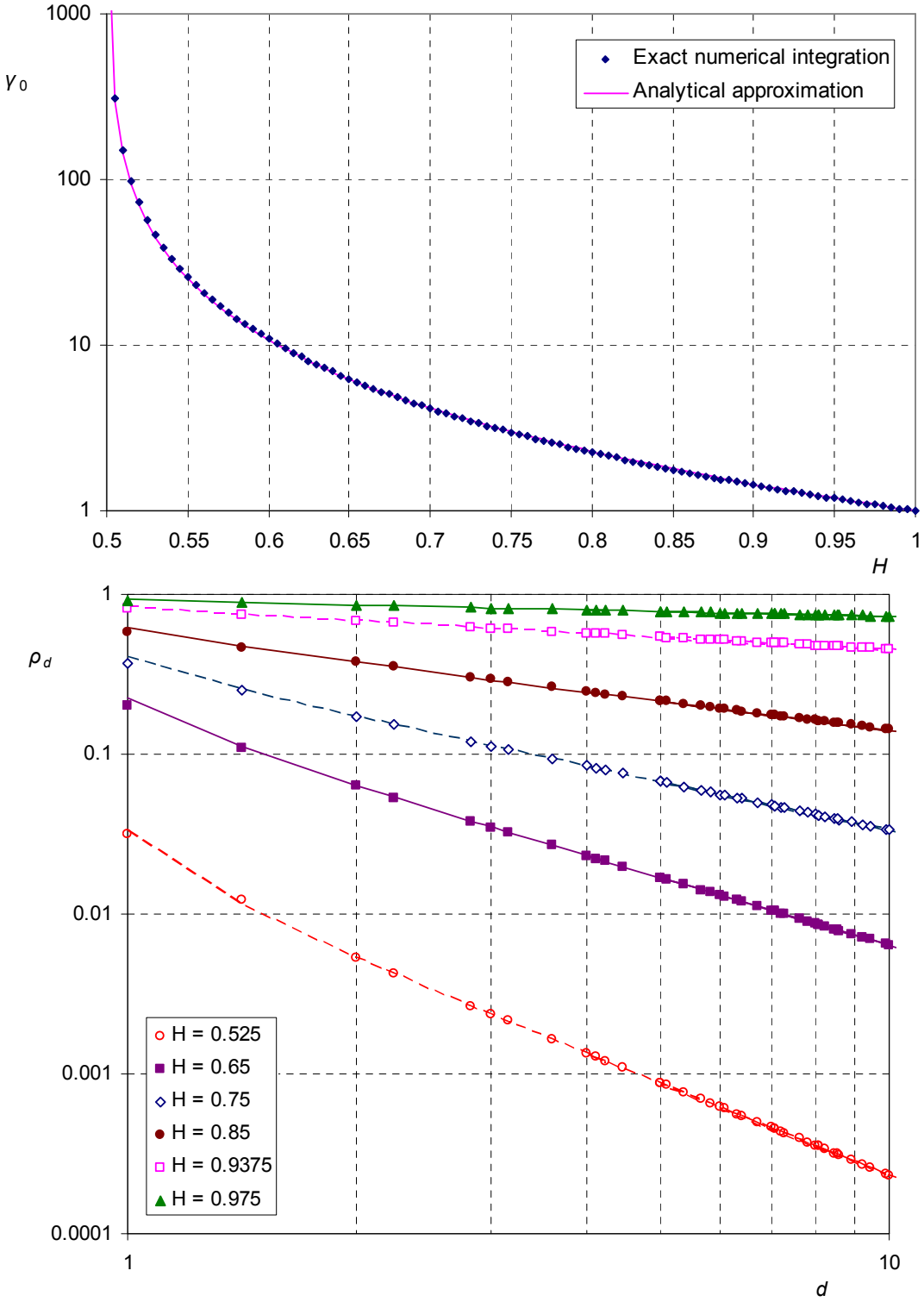

Figure 5 Comparison of exact results for the 2D HK process with approximations given by closed relationships of this study: (**upper**) variance $\gamma_0^{(k)}$, with exact relationship and approximation given by (A5) and (10), respectively, for $a = \Delta = k = 1$; (**lower**) autocorrelation $\rho_{l,m}^{(k)} \approx \rho_d^{(k)}$ with exact relationship and approximation given in (A10) and (11), respectively.

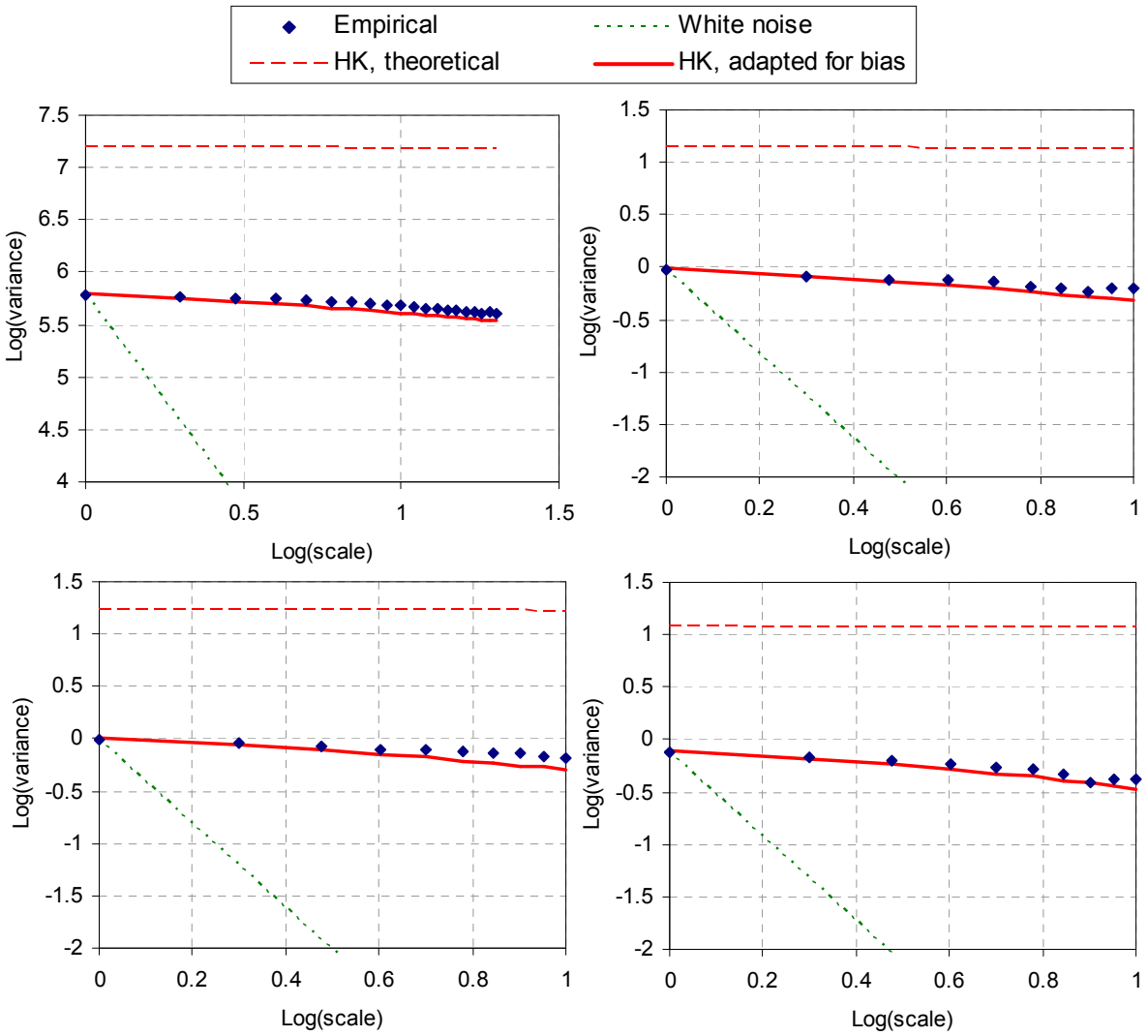

Figure 6 Climacograms of the example fields shown in Figures 1-4; scale denotes the number of grid points and variances are, for each of the panels: (**upper left**, corresponding to Figure 1) in m^2; (**upper right**, corresponding to Figure 2, after normalization) in mm^2; (**lower left**, corresponding to Figure 3, right) in mm^2; (**lower right**, corresponding to Figure 4) in $ºC^2$.

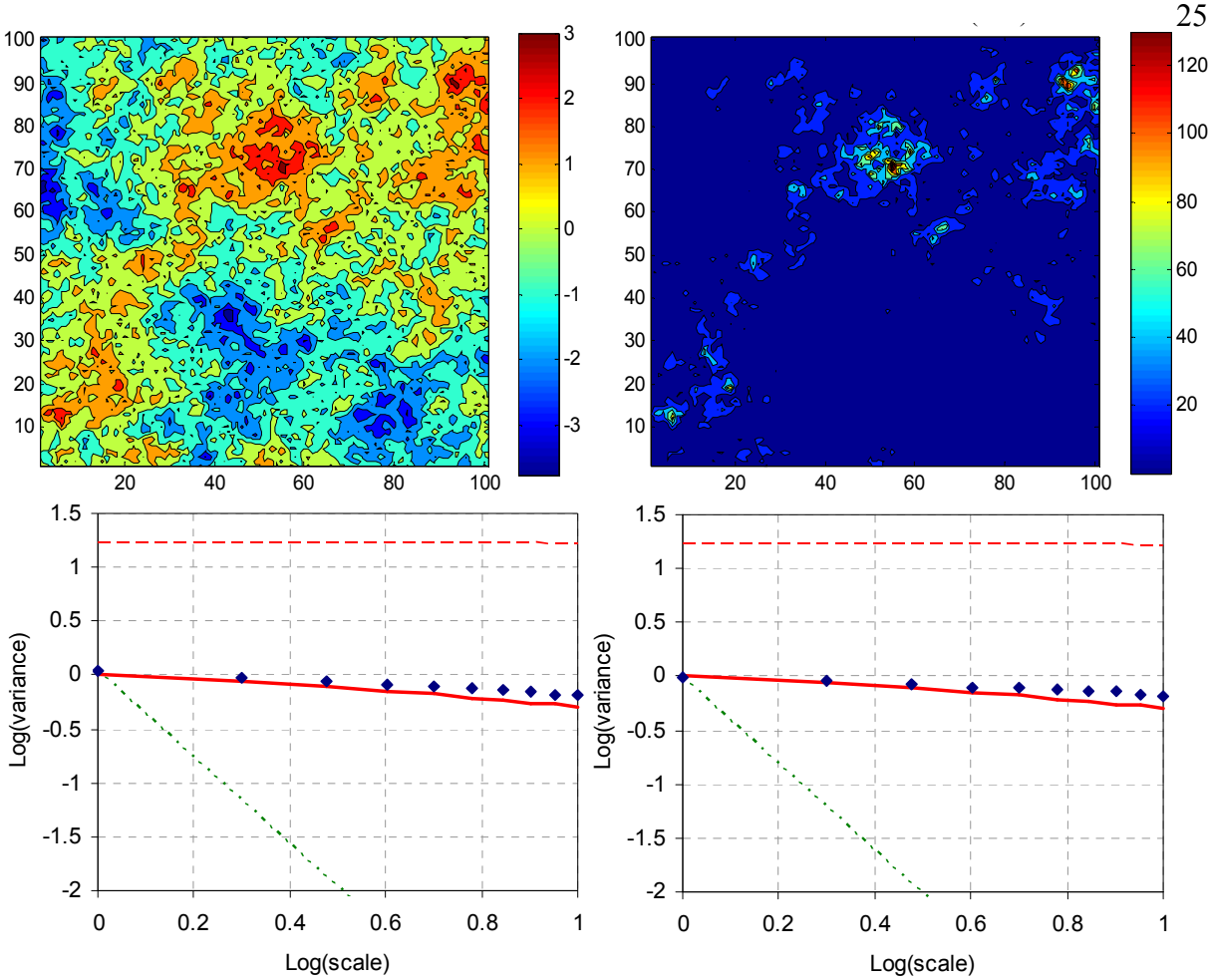

Figure 7 Synthetic rainfall field, (**upper left**) normalized, and (**upper right**) natural (mm), along with (**lower left**) the climacogram of the normalized synthetic field, which is virtually indistinguishable from that of the normalized actual field (**lower right**, copied from Figure 6, lower left, for comparison).

Tables

Table 1 Statistics of the example fields.

Field	Terrain (Fig. 1)	Rainfall from radar (Fig. 2)	Rainfall from satellite (Fig. 3, left)	Normalized rainfall from satellite data (Fig. 3, right)	Tempera-ture departures (Fig. 4)
Sample size, n	562 500	10 000	10 201	10 201	4761
Average, $\overline{z}$	1490	0.505	8.980	-0.005	0.258
CS estimate of variance	600 062	0.464	101.4	0.982	0.757
CS estimate of skewness	0.108	1.995	2.773	-0.140	0.394
Hurst coefficient, H	0.998	0.993	0.994	0.997	0.996
HK estimate of variance	15 557 780	3.944	966.53	17.055	11.940
Equivalent sample size n'	1.042	1.135	1.117	1.064	1.071

Table 2 Comparison of the statistics of the observed (Figure 3) and the simulated rainfall field (Figure **7**).

Field	Observed, natural	Observed, normalized	Synthetic, natural	Synthetic, normalized
Sample size, n	10 201	10 201	10 201	10 201
Average, $\overline{z}$	8.98	–0.005	10.10	0.065
CS estimate of variance, s^2	101.4	0.982	143.77	1.074
CS estimate of coefficient of skewness	2.773	–0.140	3.40	–0.120
Hurst coefficient, H	0.994	0.997	0.994	0.997
HK estimate of variance	966.53	17.055	1395.49	17.055
Equivalent sample size, n'	1.117	1.064	1.117	1.064