Maximum Entropy Distributions of Scale-Invariant Processes

Veronica Nieves,^{*} Jingfeng Wang,[†] and Rafael L. Bras[‡]

Department of Civil Environmental Engineering, University of California, Irvine, Irvine, California, USA

Elizabeth Wood[§]

Department of Civil Environmental Engineering, Massachusetts Institute of Technology, Cambridge, Massachusetts, USA.

(Received 15 July 2010; published 9 September 2010)

Organizations of many variables in nature such as soil moisture and topography exhibit patterns with no dominant scales. The maximum entropy (ME) principle is proposed to show how these variables can be statistically described using their scale-invariant properties and geometric mean. The ME principle predicts with great simplicity the probability distribution of a scale-invariant process in terms of macroscopic observables. The ME principle offers a universal and unified framework for characterizing such multiscaling processes.

DOI: 10.1103/PhysRevLett.105.118701

PACS numbers: 89.75.Da, 05.65.+b, 89.70.Cf, 92.40.Zg

Self-similarity is one of the common statistical properties of many natural phenomena encompassing a wide range of space or time scales. Scale invariance is a form of self-similarity, where these properties are scale-independent. The concept was first introduced in the context of turbulence by Kolmogorov [1], who used probabilistic tools to describe the variability of the moments of velocity increments across a wide range of spatial scales. Success of this type of analysis was experimentally confirmed years later [2–4]. Yet the scaling properties were not analyzed locally until recent years by detecting spatial transitions in some satellite-derived scalar variables over extensive domains [5–9]. One of the outstanding issues is the theoretical prediction of the statistical behavior based on scale-invariant properties. The question is: would it be possible to use the known multiscaling moments to determine probability distributions? This study considers a possible approach, i.e., the principle of maximum entropy (ME), to answer that question.

First developed for statistical mechanics by Jaynes [10], the ME theory has been successfully applied to many scientific and engineering problems [11–13]. Its application in hydrology started in the early 1960s [14] and has attracted growing attention [15–20]. In this study we attempt to use the ME theory to characterize some important variables in hydrometeorology: the drainage area of a river network, soil moisture, and topography.

As an inference algorithm based on Bayesian probability theory, the ME theory selects the “most probable” microscopic configuration of a system among all possibilities consistent with macroscopic observables. According to the ME theory, the probability representing the current state of knowledge is the one associated to the maximum information entropy. The information entropy is a measure of uncertainty or lack of information about a physical system [21]. The generality and power of ME is rooted in the fact that the Bayesian interpretation of probability allows

probability theory to be applicable to any situation where statistical inference based on incomplete information is sought. Although it is not a physical law *per se*, the ME theory has physical significance when applied to natural processes.

The ME formalism may be summarized as follows. The Shannon information entropy S_I [22] of a probability density function $p(x)$ of a continuous variable x is defined as

$$S_I = - \int p(x) \log \left[\frac{p(x)}{m(x)} \right] dx. \quad (1)$$

Consider that some macroscopic properties of a system are expressed in terms of constraints in the form,

$$\int f_k(x) p(x) dx = F_k, \quad k = 0, \dots, M, \quad (2)$$

where f_k is an arbitrary function of x , F_k the given parameters, and $m(x)$ a given prior distribution.

Maximizing S_I in Eq. (1) subject to the constraints of Eq. (2) leads to a (normalized) probability density function,

$$p(x) = \frac{m(x)}{Z} \exp \left\{ - \sum_{k=0}^M \mu_k f_k(x) \right\} \quad (3)$$

where μ_k is the Lagrangian multiplier associated with the constraint F_k through,

$$F_k = \frac{\partial \log Z}{\partial \mu_k}, \quad (4)$$

and Z is the well-known partition function,

$$Z(\mu_0, \dots, \mu_M) = \int m(x) \exp \left\{ - \sum_{k=0}^M \mu_k f_k(x) \right\} dx. \quad (5)$$

$p(x)$ in Eq. (3) is referred to as the “ME distribution” herein.

Next, we demonstrate that the distributions of some scale-invariant variables can be derived using the ME

formalism. Here we present two cases: (1) drainage area of river network (DARN), and (2) surface soil moisture field (SSM) and topography as captured in a digital elevation map (DEM).

(1) *DARN*.—We denote the drainage area [16] by a with $A_1 < a < A_2$ with A_1 and A_2 as the upper and lower limit of a , respectively. It is straightforward to show that the ME distribution of a , according to Eq. (3), subject to the constraint of a given geometric mean, A_g ,

$$\int_{A_1}^{A_2} p(a) \log a da = \log A_g, \quad (6)$$

assuming $m(a) = 1$, is the power-law distribution [23],

$$p(a) = \left\{ \frac{1 - \mu}{A_2^{1-\mu} - A_1^{1-\mu}} \right\} a^{-\mu}, \quad (7)$$

where the Lagrangian multiplier μ is related to the constraint A_g through,

$$\frac{1}{1 - \mu} + \log A_g = \frac{A_2^{1-\mu} \log A_2 - A_1^{1-\mu} \log A_1}{A_2^{1-\mu} - A_1^{1-\mu}}. \quad (8)$$

For the case of $A_2 \gg A_1$, Eq. (8) leads to an explicit function of μ in terms of A_g ,

$$\mu = 1 + \left(\log \frac{A_g}{A_1} \right)^{-1} \quad \text{or} \quad \mu = 1 - \left(\log \frac{A_2}{A_g} \right)^{-1}, \quad (9)$$

depending on whether $\mu > 1$ or $\mu < 1$.

The ME predicted power-law exponent μ has been validated using observations of drainage areas derived from a DEM map of the Grand Canyon from the U.S. Geological Survey (USGS) data archive. The ME predicted $\mu = 1.32$, according to Eq. (9), is in close agreement with $\mu = 1.30$ obtained by direct curve fitting of the empirical probability distribution of a according to Eq. (7). Equation (9) has also been tested for the river networks of Puerto Rico using the USGS DEM data with similar results (not shown). The close agreement between predicted and observed power-law exponents reveals the link between statistics and the macroscopic properties of a physical system.

(2) *SSM and DEM*.—We denote a SSM field or a DEM by $z(\vec{x})$ where $\vec{x}$ is a two-dimensional location vector, and $p(z_1, z_2)$ the joint distribution of $z_1 \equiv z(\vec{x}_1)$ and $z_2 \equiv z(\vec{x}_2)$. Consider the following constraints: (i) Given multiscaling moments of the incremental process $|z_1 - z_2|$,

$$\iint p(z_1, z_2) |z_1 - z_2|^q dz_1 dz_2 = a_q r^{\tau(q)}, \quad 1 \leq q \leq M, \quad (10)$$

where $r \equiv |\vec{x}_1 - \vec{x}_2|$ is the separation distance, $\tau(q)$ refers to the multiscaling exponent independent of r , and a_q is a constant parameter. (ii) Given the geometric mean of $|z_1 - z_2|$,

$$\iint p(z_1, z_2) \log |z_1 - z_2| dz_1 dz_2 = \frac{1}{n} \sum_{i,j=1}^n \log |z_j - z_i| \quad (11)$$

where z_i is observed value of z at location $\vec{x}_i$ and n is the number of sampling locations.

These constraints are based on some previous studies showing that variables such as drainage area, discharge and energy release, soil moisture, increments of topography, etc., are self-similar characterized by the multiscaling moments or power-law distributions [16,23–27].

The ME distribution according to Eq. (3), subject to constraints Eqs. (10) and (11) assuming $m(z_1, z_2) = 1$, is

$$p(z_1, z_2) = \frac{1}{Z} |z_1 - z_2|^{-\mu_0} \exp \left\{ - \sum_{q=1}^M \mu_q |z_1 - z_2|^q \right\} \quad (12)$$

with

$$Z(\mu_0, \dots, \mu_M) = \iint |z_1 - z_2|^{-\mu_0} \times \exp \left\{ - \sum_{q=1}^M \mu_q |z_1 - z_2|^q \right\} dz_1 dz_2, \quad (13)$$

where μ_q is the Lagrangian multiplier associated with the q th moment constraint. μ_q is related to $\tau(q)$ according to Eq. (4),

$$a_q r^{\tau(q)} = \frac{\partial \log Z}{\partial \mu_q}, \quad 1 \leq q \leq M. \quad (14)$$

For the case of $M = 1$, Eq. (12) reduces to a Gamma distribution with Z expressed as an explicit function of μ_1 and μ_0 :

$$Z = \frac{\Gamma(1 - \mu_0)}{\mu_1^{1-\mu_0}}. \quad (15)$$

Equations (14) and (15) lead to

$$a_q r^{\tau(q)} = \frac{1 - \mu_0}{\mu_1}. \quad (16)$$

The right-hand side of Eq. (16) is simply the expected value of a Gamma distribution.

For the case of $M = 2$, no simple analytical solution of Z , as a function of μ_0 , μ_1 , and μ_2 , like that in Eq. (15) for the case of $M = 1$, is available.

Two data products are used to validate the ME distributions Eqs. (12) and (16): the AMSR-E soil moisture L2B product and the USGS National Elevation Dataset (NED). The AMSR-E L2B soil moisture product provides estimates of soil moisture of the top 1 cm derived from brightness temperature at 10.7 GHz with resolution of 38 km, regridded into a global cylindrical 25 km Equal-Area Scalable Earth Grid. The NED raster elevation data with

a resolution of one arc second (30 m) over the continental U.S.A. are derived from multiple satellite sensors. NAD83 and NAV88 are consistently used as horizontal and vertical datums, respectively, and all the data are regridded using a geographic projection. Soil moisture accuracy (root-mean-square difference) is about 60 kg m^{-3} , which almost holds the typical incremental values of 100 kg m^{-3} . The vertical accuracy (absolute error) of the NED product is on the order of 7–15 m. Equations (12) and (16) have been tested over 15 regions. Results for four of them are shown in this paper and referred to as R1SSM and R2SSM in Fig. 1, and R1DEM and R2DEM in Fig. 2. The first two are selected from the SSM maps and the other ones from the DEMs. The top left and bottom right corners of the four regions are: $[64N, -104E]$ $[54N, -92.5E]$; $[70N, 15E]$ $[65.5N, 32E]$; $[36.28N, -120.79E]$ $[36.24N, -120.75E]$; $[35.3, -83.1E]$ $[35.25N, -83.04E]$, respectively.

FIG. 1. From top to bottom. Left panels: L2B AMSR-E soil moisture map for October 18, 2009 and region R1SSM, associated empirical (P_e) and the ME distributions for $M = 1, 2$ ($P_{t,M=1}$ and $P_{t,M=2}$) according to Eq. (12). Right panels: same for region R2SSM. Maps are represented in longitude λ and latitude θ degrees. All probabilities are plotted versus the absolute value of the increments $\Delta z = |z(\vec{x}_1) - z(\vec{x}_2)|$ for different separation distances $r = |\vec{x}_1 - \vec{x}_2|$.

Figures 1 and 2 compare the ME distributions of SSM and DEM increments, according to Eq. (12), with the empirical ones for different separation distances r . It is evident that the Gamma distribution for $M = 1$ agrees closely with the data (log-log regression coefficients of ~ 0.97 and 0.99 for SSM and DEM, respectively). However, the predicted $\tau(q)$ according to Eq. (16) does not agree with the observed values when the ME distribution is given in Eq. (12) for most of the regions under study. Relative errors are from 0.9% to 32% for the case of SSM, and up to 94% for the case of DEM. This finding suggests that the first moment constraint alone is not sufficient to characterize the multiscaling processes such as SSM and DEM. It is the ME distribution for $M = 2$ that is almost identical to the empirical distributions (compare top and bottom panels). Note that the value of Z in Eq. (12) has been also estimated by fitting the empirical values with log-log regression coefficients of around 0.99, as no analytical solution of Z is available. It is also worthwhile to mention that the data [shown in Figs. 1 and 2] are positively skewed (i.e., right skewed), implying that the logarithmic transformation of the processes would be appropriate in data analysis.

FIG. 2. Same as in Fig. 1 for regions R1DEM and R2DEM, which corresponds to the 1 arc second NED data.

We also compared the ME distribution subject to constraint (i) only, Eq. (10), with the empirical distribution (figure not shown),

$$p(z_1, z_2) = \frac{1}{Z} \exp \left\{ - \sum_{q=1}^M \mu_q |z_1 - z_2|^q \right\}. \quad (17)$$

It is found that the ME distribution, Eq. (17), for $M = 1$ only fits the tails of the empirical distributions. Equation (17) for $M = 2$, a Gaussian distribution, does not capture all features of the empirical ones either. It turns out that the combination of the constraints (i) and (ii) is crucial to yield the best agreement between the ME and empirical distributions for SSM and DEM fields. We conclude that the geometric mean contains crucial information about the multiscaling processes of SSM and DEM. As a result, the geometrical mean may offer important clues regarding the physical mechanisms underlying the observed statistical properties.

In summary, we have demonstrated that the ME theory allows the statistical behavior of some scale-invariant processes such as DARN, SSM, and DEM to be characterized by a small number of macroscopically observable quantities. Such scale-invariant behavior resulting from self-organization emerges as the most probable and macroscopically reproducible state. It turns out that the geometric mean provides essential information for shaping river networks. The geometric mean is also identified as an important parameter, in addition to the moments, in characterizing multiscaling incremental processes of soil moisture and topography. More importantly, this analysis supports the assertion that the ME theory is a universal and unified framework to characterize those processes governed by scale-invariant laws.

This work was supported by the U.S. Army RDECOM ARL Army Research Office under Grant No. W911NF-07-1-0126.

*vnieves@uci.edu

†jingfenw@uci.edu

‡rlbras@uci.edu

§ewood@mit.edu

- [1] A.N. Kolmogorov, Dokl. Akad. Nauk SSSR **30**, 301 (1941).
- [2] G. Parisi and U. Frisch, in *Turbulence and Predictability in Geophysical Fluid Dynamics. Proc. Intl. School of*

Physics E. Fermi, edited by M. Ghil, R. Benzi, and G. Parisi (North Holland, Amsterdam, 1985), pp. 84–87.

- [3] R.H. Kraichnan, *Phys. Rev. Lett.* **72**, 1016 (1994).
- [4] Z.S. She and E. Leveque, *Phys. Rev. Lett.* **72**, 336 (1994).
- [5] A. Turiel, J. Isern-Fontanet, E. García-Ladona, and J. Font, *Phys. Rev. Lett.* **95**, 104502 (2005).
- [6] V. Nieves, C. Llebot, A. Turiel, J. Solé, E. García-Ladona, M. Estrada, and D. Blasco, *Geophys. Res. Lett.* **34**, L23602 (2007).
- [7] A. Turiel, H. Yahia, and C. Pérez-Vicente, *J. Phys. A* **41**, 015501 (2008).
- [8] V. Nieves and A. Turiel, *J. Marine Syst.* **77**, 482 (2009).
- [9] A. Turiel, J. Solé, V. Nieves, J. Ballabrera-Poy, and E. García-Ladona, *Remote Sensing of Environment* **112**, 2246 (2008).
- [10] E.T. Jaynes, *Phys. Rev.* **106**, 620 (1957).
- [11] G.W. Paltridge, *Q. J. R. Meteorol. Soc.* **101**, 475 (1975).
- [12] N. Wu, *Maximum Entropy Method* (Springer, New York, 1997).
- [13] E.T. Jaynes, *Probability Theory—the Logic Of Science* (Cambridge University Press, Cambridge, England, 2003).
- [14] J.H. Lienhard, *J. Geophys. Res.* **69**, 5231 (1964).
- [15] M. Fiorentino, P. Claps, and V.P. Singh, *Water Resour. Res.* **29**, 1215 (1993).
- [16] I. Rodríguez-Iturbe and A. Rinaldo, *Fractal River Basin—Chance and Self-Organization* (Cambridge University Press, Cambridge, England, 1997).
- [17] J. Wang, G.D. Salvucci, and R.L. Bras, *Water Resour. Res.* **40**, W09303 (2004).
- [18] A. Kleidon and K. Fraedrich, *Biotic Entropy Production And Global Atmosphere-Biosphere Interaction* (Springer-Verlag, Berlin, 2004).
- [19] R.C. Dewar, *J. Phys. A* **38**, L371 (2005).
- [20] J. Wang, R.L. Bras, M. Lerda, and G.D. Salvucci, *J. Geophys. Res.* **112**, G03010 (2007).
- [21] C.E. Shannon, *Bell Syst. Tech. J.* **27**, 379 (1948).
- [22] P.C. Gregory, *Bayesian Logical Data Analysis For The Physical Sciences* (Cambridge University Press, Cambridge, England, 2005).
- [23] D.G. Tarboton, R.L. Bras, and I. Rodríguez-Iturbe, *Water Resour. Res.* **25**, 2037 (1989).
- [24] I. Rodríguez-Iturbe, E. J. Ijjasz-Vasquez, R.L. Bras, and D.G. Tarboton, *Water Resour. Res.* **28**, 1089 (1992).
- [25] I. Rodríguez-Iturbe, G.K. Vogel, R. Rigon, D. Entekhabi, F. Castelli, and A. Rinaldo, *Geophys. Res. Lett.* **22**, 2757 (1995).
- [26] D. Veneziano and J.D. Niemann, *Water Resour. Res.* **36**, 1937 (2000).
- [27] C.D. Peters-Lidard, F. Pan, A. Y. Hsu, and P.E. O'Neill in *Proceedings of the Geoscience and Remote Sensing Symposium* (IEEE, Sydney, Australia, 2001) p. 1297.