GEOPHYSICAL RESEARCH LETTERS, VOL. 37, L07403, doi:10.1029/2010GL042634, 2010

Scale-wise evolution of rainfall probability density functions fingerprints the rainfall generation mechanism

Annalisa Molini,[1,2] Gabriel G. Katul,[1,2] and Amilcare Porporato[1,2,3]

Received 28 January 2010; revised 8 March 2010; accepted 9 March 2010; published 7 April 2010.

[1] The cross-scale probabilistic structure of rainfall intensity records collected over time scales ranging from hours to decades at sites dominated by both convective and frontal systems is investigated. Across these sites, intermittency build-up from slow to fast time-scales is analyzed in terms of heavy tailed and asymmetric signatures in the scale-wise evolution of rainfall probability density functions (pdfs). The analysis demonstrates that rainfall records dominated by convective storms develop heavier-tailed power law pdfs toward finer scales when compared with their frontal systems counterpart. Also, a concomitant marked asymmetry build-up emerges at such finer time scales. A scale-dependent probabilistic description of such fat tails and asymmetry appearance is proposed based on a modified q-Gaussian model, able to describe the cross-scale rainfall pdfs in terms of the nonextensivity parameter q, a lacunarity (intermittency) correction and a tail asymmetry coefficient, linked to the rainfall generation mechanism. **Citation:** Molini, A., G. G. Katul, and A. Porporato (2010), Scale-wise evolution of rainfall probability density functions fingerprints the rainfall generation mechanism, *Geophys. Res. Lett.*, *37*, L07403, doi:10.1029/2010GL042634.

1. Introduction

[2] Numerous problems in contemporary hydrology routinely require explicit accounting for the scale-wise evolution of rainfall probability density functions (pdfs), as needed to link rainfall statistics computed from slowly evolving climatic fluctuations (predicted from climate models) to their highly intermittent counterpart at the storm event time scale. Understanding intermittency and skewness build-up across fine time scales and its connection with the rainfall generation mechanism is especially needed when constructing realistic disaggregation and cross-scale models of rainfall.

[3] Progressing from large to small time scales, rainfall intensity usually evolves from near-Gaussian to fat-tailed and sharp-peaked pdfs. This fact is mainly associated with strong clustering and large amplitude fluctuations that characterize such processes at fine scales, and is often interpreted as a signature of intermittency and anomalous scaling [e.g, see *Roux et al.*, 2009; *Lovejoy and Schertzer*, 1985, and references therein]. A similar scale-wise pdf phenomenology has been initially observed in many systems such as turbulent flows and stock market dynamics [*Frisch*, 1995; *Sornette*, 2003]. In turbulence, this scale-wise evolution of the pdfs was studied using the so-called propagator kernel, relating variations of velocity increments at small scales with those at larger scales [*Chillà et al.*, 1996]. The propagator kernel is not sensitive to the turbulence generation mechanism and may be quasi-universal [*Chevillard et al.*, 2006]. In addition, cross-scale pdfs are commonly assumed nearly symmetric - with positive and negative fluctuations (or gradients) possessing the same pdf tails, although a concurrent transition from symmetric to skewed scale-wise probability functions has been recently reported for turbulent velocity fluctuations [*Chevillard et al.*, 2006], active scalar turbulent fluctuations [*Zhang and Wu*, 2009] and atmospheric velocity fluctuations within forested canopies [*Bolzan et al.*, 2002].

[4] At event and sub-event time scales, the positive and negative fluctuations of rainfall are more likely to be asymmetric as may be evidenced by the shape of a hyetograph [*Huff*, 1967]. In addition, different rainfall generation mechanisms (e.g. convective versus frontal storms) may contribute to the intermittency and asymmetry build-up across scales in a distinct manner.

[5] In the following, we (1) explore how intermittency and asymmetry build-up can be detected in the probabilistic structure of rainfall across different time scales and whether discernible differences between convective versus frontal systems exist (surrogates for rainfall generation mechanisms), and (2) propose and test a family of probability distribution laws suitable for describing the scale-wise evolution of rainfall pdfs building on recent developments in turbulence research. Naturally, the joint exploration of these two points allows us to link the parameters describing the scale-wise probability distribution laws in (2) with the generation mechanism in (1).

2. Methods

[6] Five published historical rainfall time series records collected at various micro-climatic regimes are considered: the rainfall time series from Genova (north-western Italy), Chiavari (north-western Italy), Firenze (central Italy), Valentia (south-western coast of Ireland), and the Duke forest experimental site near Durham (North Carolina, USA). These records were collected using traditional tipping-bucket rain gauges and represent a variety of different micro-climatological conditions with convective events dominating time series recorded in Genova, Durham, Chiavari and Florence, and predominantly frontal events characterizing the Valentia's site. Further details about the data and the

[1]Nicholas School of the Environment, Duke University, Durham, North Carolina, USA.

[2]Department of Civil and Environmental Engineering, Pratt School of Engineering, Duke University, Durham, North Carolina, USA.

[3]Laboratory of Environmental Fluid Mechanics and Hydrology, École Polytechnique Fédérale de Lausanne, Lausanne, Switzerland.

0094-8276/10/2010GL042634

sites, together with a classical moment scaling analysis testing multi-fractality for the five sites can be found elsewhere [*Molini et al.*, 2009]. We also present a qualitative comparison with a sample turbulent air temperature time series (hereafter referred as *Te*), collected at the Duke forest site (see *Katul et al.* [1997] for description). The *Te* is representative of scalar turbulence in the atmospheric boundary layer and displays a quasi-symmetric evolution of cross-scale pdfs.

[7] Rainfall fluctuations $r(t)$ are treated as the realizations of a random process. Thus, the probabilistic structure of $r(t)$ across time scales T can be interpreted as a set of probability density functions $f_{\mathcal{R}_T}(r, T)$. We assume that $f_{\mathcal{R}_T}(r, T)$ is represented by a parametric family of distributions, $f_{\mathcal{R}}(r, \overline{\theta})$, where the dependence of the $f_{\mathcal{R}}$'s on the scale is 'absorbed' by a suitable ensemble of parameters $\overline{\theta} = \overline{\theta}(T)$. This is a crucial assumption since it allows us to describe the probability law representing rainfall variability across scales as invariant, while its scale-wise evolution can be captured in terms of scale-dependent parameters $\overline{\theta}$ [*Kumar and Foufoula-Georgiou*, 1993].

[8] A necessary first step in identifying $f_{\mathcal{R}_T}(r, T)$ is the scale-wise decomposition of the original rainfall series. Popular tools for time-frequency (or scale) decomposition are wavelet transforms, defined as the set of scalar products between the original rainfall intensity series $i(t)$ and a t-translated, T-dilated version of an analyzing mother wavelet $\psi(t)$, i.e $W_\psi\,[i](t, T) = \frac{1}{\sqrt{T}}\int_{-\infty}^{+\infty} i(t')\psi\left(\frac{t'-t}{T}\right)\,dt$, where $\frac{1}{\sqrt{T}}$ is a scale-dependent normalization factor and the W_ψ are the wavelet coefficients. The W_ψ are related (but not identical) to the fluctuations $r(t)$ (or differences) in the series [*Mallat*, 1989], so that the properties of $f_{\mathcal{R}_T}(r, T)$ (and then $\overline{\theta}$) can be inferred from those of $f_{\mathcal{W}}(W_\psi)$ at each scale T. We adopt an orthonormal wavelet transform (*OWT*) based on dilated and translated family of mother wavelets $\psi_{j,n}$ that are orthonormal, with discrete time and dyadic scale arrangements. The selected mother wavelet here is symmetric (e.g. Coiflet) so as to minimize the effects of the transform on the skewness of the scale-wise pdf's.

[9] To select a probabilistic model describing the rainfall pdf's across scales, we first analyze the W_ψ's using statistics sensitive to the tail, peak and asymmetry. Classical measures of peakness, tail heaviness and asymmetry are based on statistical moments such skewness $\gamma_1(T)$ and flatness factor $\gamma_2(T)$ computed from W_ψ's at each scale T. One must exercise caution when adopting such statistics to rainfall data due to three inter-related factors: (1) for highly intermittent processes higher-order moments (including γ_1 and γ_2) may not be theoretically defined (or finite); (2) γ_2 cannot separately assess contributions from fat tails and peakedness; (3) both γ_1 and γ_2 exhibit sensitivity to outliers due to their "zero breakdown" value and the unbounded influence function [*Hogg*, 1972]. Thus, quantile-based statistical measures known for being less sensitive to outliers and theoretical existence of higher-order moments are employed here [*David*, 1981; *Schmid and Trede*, 2003]. We re-define such quantile-based statistics in the wavelet domain so that measures such as Hogg's tail index Θ (assessing the tail-fatness of a pdf), the peakedness index Π and the asymmetry index (or left-right tail ratio) A can be respectively expressed as $\Theta(T) = \frac{w_{97.5}(T)-w_{2.5}(T)}{w_{87.5}(T)-w_{1.25}(T)}$, $\Pi(T) = \frac{w_{87.5}(T)-w_{1.25}(T)}{w_{75}(T)-w_{2.5}(T)}$ and $A(T) = \frac{w_{50}(T)-w_{10}(T)}{w_{90}(T)-w_{50}(T)}$ (where the w subscript represents the percentile). The A can be computed for different distributions in closed form or numerically, and for the standardized Gaussian distribution, it is equal to 1 [*Schmid and Trede*, 2003]. If a phenomenon displays increasing (or decreasing) values of A as a function of scale T, then intermittent patterns of positive and negative oscillations propagate differently across scales and a skewed $f_{\mathcal{R}_T}(r, T)$ must be adopted.

3. Probability Models

[10] A wide spectrum of probability models - ranging from α-stable to q-Tsallis distributions - have been used (especially in image analysis and in the turbulence literature) to mimic the $f_{\mathcal{R}_T}(r, T)$ across scales and are here tested on the rainfall time series described in Section 2. For example, *Mallat* [1989] proposed to analyze the probability features of W_ψ extracted from images by a scale-dependent family of stretched exponential distributions $\mathcal{SE}(b, m)$ with pdf $f_{\mathcal{SE}}(r) = K_{\mathcal{SE}} e^{-(|r|/b)^m}$ where b and m are scale and shape parameters and $K_{\mathcal{SE}}$ ($= m/[2b\Gamma(1/m)]$) the normalization constant. The $\mathcal{SE}(b, m)$-model is commonly adopted when a phenomenon displays both a characteristic scale and a fat tail behavior. $\mathcal{SE}$ random variables can be seen as a result of multiplicative processes like classical random multiplicative cascades $\mathcal{RMC}$ [*Frisch and Sornette*, 1997], a down-scaling model widely adopted in the hydro-meteorological literature [*Waymire*, 2006].

[11] Four parameters $S_\alpha(\sigma, \beta, \mu)$ (α-stable) variates with the stability index $0 < \alpha \leq 2$ constitute an alternative model to $\mathcal{SE}$ with faster growing fat tails and peakedness [*Kumar and Foufoula-Georgiou*, 1993, 1996]. Despite their ability to represent both symmetric and skewed pdfs, α-stable distributions are prone to a number of different drawbacks among which: (1) they require the estimation of an elevated number of parameters, and (2) they posses finite variance only in the limit case $\alpha = 2$ (Gaussian distribution). Assuming the precipitation scale-wise pdf structure is the result of a superposition of different intermittent dynamics, possible models can be derived from mixtures of distributions - also known as super-statistics. A well known mixture - largely adopted in the field of complex systems science due to its connections to nonextensive thermodynamics - is the q-Gaussian (or Tsallis) model [*Tsallis et al.*, 1995]. This probability model results from a scale mixture of $\mathcal{SE}(b, m)$ with the scale parameter b distributed as an inverse Γ [*Shi et al.*, 2005]. Also, the q-Gaussian (formally a Student-t distribution) resembles the Gaussian, Cauchy and Levý distributions for particular values of the nonextensivity parameter q (i.e. α-stable distributions are a sub-parametric space of the q-Gaussians and the stability index α is related to q by $q = \frac{3+\alpha}{1+\alpha}$). Since the q-Gaussian is symmetric, several q-modified models have been proposed to account for asymmetries in the tails. One such model - here referred to as the Modified-Tsallis ($\mathcal{MT}$) model - was proposed by *Beck et al.* [2001] to reproduce fat tails and their asymmetry in a high Reynolds number Taylor-Couette flow (a flow consisting of a viscous fluid confined in a gap between two rotating cylinders). The model is given by:

$$f_{\mathcal{MT}}(x) = \frac{1}{Z_q}\left[1 + \beta(q-1)\varepsilon(x)\right]^{1/(1-q)} \tag{1}$$

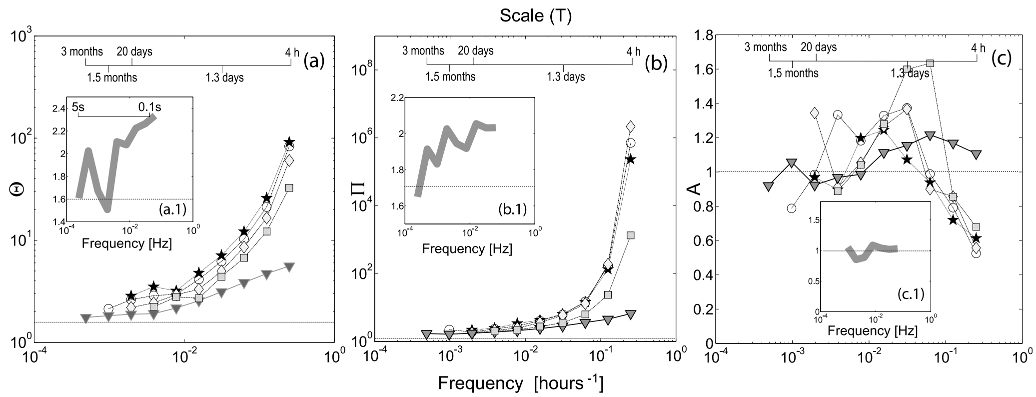

Figure 1. (a) Θ, (b) Π and (c) A as a function of frequency and time scale T (top of the plots) for all five sites: Genova (stars), Chiavari (circles), Florence (diamonds), Valentia (triangles) and the Durham (squares). Insets in Figure 1a to Figure 1c show the same statistics for the *Te* time series. Dotted lines represent respectively the values of Θ and Π for the Gaussian distribution in Figures 1a (and inset) and 1b (and inset), and A = 1 (symmetric case) in Figure 1c (and inset).

and

$$\varepsilon(x) = \frac{1}{2}|x|^{2\gamma} - c\,\mathrm{sgn}(x)\left(|x|^{\gamma} - \frac{1}{3}|x|^{3\gamma}\right) \qquad (2)$$

where $\varepsilon(x)$ represents the effective energy associated with velocity differences x in the turbulent flow, q depends on the distance separation (or scale) in an a priori unknown manner and the empirical parameter $c = 0.124(q - 1)$ is proportional to the relaxation time of the stationary state and the typical time scale of the forcing. The $\gamma = 2 - q$ (also empirically obtained for turbulence) is a lacunarity (intermittency) exponent conceptually similar to the intermittency correction in the classical mono-fractal models of turbulence. This model can be applied to rainfall time series by assuming that r follows a more complex relation than the classical q-Gaussian squared-law. Moreover, the amplification of rainfall intensities along scales due to clustering effects can be parameterized in terms of γ and the asymmetry build-up in terms of c. What is particularly appealing in this model is that both γ and c can be expressed as a function of q thereby reducing the number of parameters.

4. Results and Discussion

[12] Using the *OWT* coefficients from the measured rainfall time series, we computed Θ, Π and A as a function of frequency (1/*T*). For reference, we repeated the same calculations for *Te*.

[13] In Figures 1a and 1b, the tail index and peakedness components are shown. Comparing Figures 1a and 1b, it is evident that the extreme tails grow slower than the peak in the quantile-based estimation. Both Θ and Π increase with increasing frequency (or decreasing time scale) for all five rainfall time series. But, sites characterized by convective rainfall events (e.g. Chiavari and Genova) are more intermittent (higher Θ and Π) when compared to a site dominated by frontal systems (Valentia). Finally, when comparing rainfall to scalar turbulence (insets in Figures 1a and 1b), Θ and Π for the *Te* series present a slow increase with frequency when compared to rainfall.

[14] Focusing on the asymmetry index A (Figures 1c), results suggest a peak of positive asymmetry for $T \sim 1$ day and a slightly negative asymmetry at fine scales (see also Figure 2). This peak of asymmetry is connected with the well-known asymmetric shape of hyetographs at these daily scales [*Huff*, 1967]. On the other hand, scalar turbulence remains quasi-symmetric across scales (inset in Figure 1c). Despite the strong intermittency of rainfall, similar results were obtained via the classical moment-based statistics γ_1 and γ_2 (Figure not shown).

[15] As evident in Figure 2, to adopt the generalized ($\mathcal{MT}$) distribution in modeling rainfall provides satisfactory results when compared with the $\mathcal{SE}$ model (representing the best fit for the scalar turbulence series). Also, we fitted an α-stable distribution and the generalized ($\mathcal{MT}$) model offers superior fitting due to its ability in recovering both the finite ($q < 5/3$) and infinite ($5/3 \leq q \leq 3$) variance sub-cases. For the Taylor-Couette flow experiments, *Beck et al.* [2001] found that q not only depends on scale but also on the Reynolds number. These experiments also reveal consistent empirical expressions for c and γ, namely $c = c_1(q - 1)$ with c_1 = 0.124 and $\gamma = 2 - q$ for all Reynolds numbers. A similar - but less physically based - parametrization of c and γ can be employed for the rainfall series. For example, whether c and γ differ between sites dominated by convective storms versus sites dominated by frontal rainfall events, and whether they appear coherent across sites dominated by similar convective storms can be readily tested with the data here. The results obtained from our analysis are encouraging. At all the sites with the exception of Valentia (dominated by frontal systems), we found a γ between 2.4-q and 2.5-q across all the considered scales. For Valentia, we found a $\gamma = 2.2 - q$. Also, c_1 assumed values close to −0.1 for all the rainfall series here, revealing the tendency of positive rainfall fluctuations to decay slower with respect to the negative ones. For *Te* series collected in the atmospheric surface layer, c was negligible (i.e. the pdfs across scales are near symmetric). The introduction of a lacunarity index $\gamma = 2.25 - q$ allows for a better description of the tails at fine scales when compared with the $\mathcal{SE}$ model for turbulence (see Figure 2f). The 2.25 value obtained for *Te* exceeds the one derived by *Beck et al.* [2001] suggesting

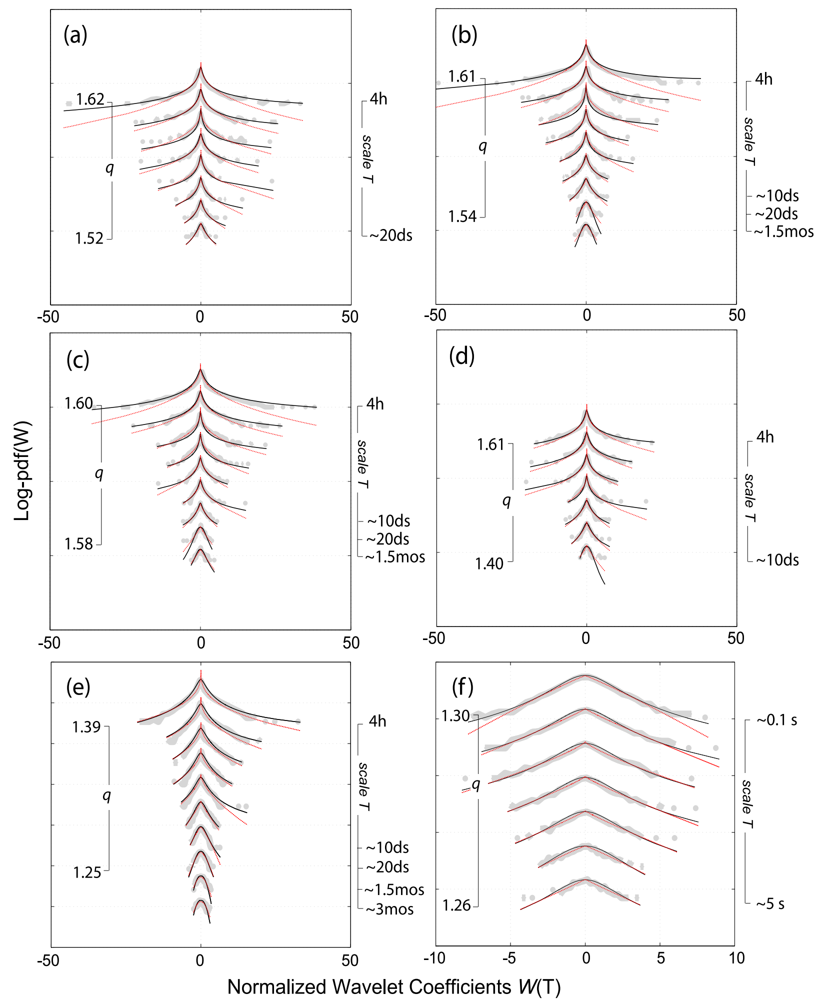

Figure 2. Comparison between the measured (grey dots), the $\mathcal{MT}$ (solid line) and the $\mathcal{SE}$ $f_{\mathcal{W}}(T)$ for T ranging between 4 hours and 10 days to ~3 months (depending on the length of the considered series): (a) Genova, (b) Chiavari, (c) Florence, (d) Durham and (e) Valentia. A similar comparison is reported for (f) *Te* at scales within the inertial subrange (0.1 to 5 seconds). Time-scales T are reported on the right side of each graph, while the range of fitted $\mathcal{MT}$ model q is shown on the left side. For clarity, the pdf's are shifted upwards with increasing resolution. Parameter fitting is obtained via quantile based techniques.

some sensitivity to the turbulence generation mechanism (i.e. boundary layer vs. Taylor-Couette flow).

5. Conclusions

[16] A crucial issue in rainfall modeling research is understanding how fat tails and intermittency are encoded in the probability structure of the process across different time-scales, and whether the rainfall generation mechanism can further informs us about their scale-wise evolution. Multiplicative cascade models commonly adopted in rainfall downscaling assume a symmetric probabilistic structure for the positive and negative fluctuations, although the validity of this symmetry assumption has been scarcely investigated in the context of rainfall. We approached this problem by decomposing an ensemble of historic rainfall time series in their time - scale domain through *OWT* filtering. The evolution of the pdfs of wavelet coefficients across scales - mimicking the dynamics of rainfall scale-wise fluctuations - was then analyzed.

[17] Rainfall pdfs exhibit faster developing heavy-tails and peakedness as well as an asymmetry build-up when progressing from slow to fine temporal scales. For scalar fluctuations in atmospheric turbulent flows, the development of asymmetry across scales has been noted at the finest scales and heuristically linked to possible interactions between these and larger scales that are impacted by the production mechanism of turbulence (and hence inherently skewed [*Katul et al.*, 2006]). A similar short-circuit of the cascade may offer an explanation for skewness build-up across rainfall time scales, considering that the maximum observed asymmetry is reached at daily scales, where the "resonance" between climatological and micro-meteorological scales may be maximum. When explicitly including a lacunarity (or intermittency) correction γ and an asymmetry parameter c to the traditional q-Tsallis model, the main features of the rainfall pdfs across scales were well reproduced. We also showed that γ and c can be expressed as a function of the nonextensivity parameter q. Using this revised model, sites dominated by convective systems assume higher values of q when compared with sites dominated by frontal systems for

a given time scale T. These higher q values reveal stronger intermittency signature in the former when compared to the latter. On the other hand, asymmetric patterns (through c) appear to be rather consistent across sites, pointing out a tendency of positive oscillations to display a slower decay than negative ones.

[18] **Acknowledgments.** This study was supported by the National Science Foundation (NSF-EAR 0628342, NSF-EAR 0635787 and NSF-ATM-0724088).

References

Beck, C., G. S. Lewis, and H. L. Swinney (2001), Measuring nonextensitivity parameters in a turbulent Couette-Taylor flow, *Phys. Rev. E*, *63*(3), 035303, doi:10.1103/PhysRevE.63.035303.

Bolzan, M. J. A., F. M. Ramos, L. D. A. Sá, C. Rodrigues Neto, and R. R. Rosa (2002), Analysis of fine-scale canopy turbulence within and above an Amazon forest using Tsallis' generalized thermostatistics, *J. Geophys. Res.*, *107*(D20), 8063, doi:10.1029/2001JD000378.

Chevillard, L., B. Castaing, E. Leveque, and A. Arneodo (2006), Unified multifractal description of velocity increments statistics in turbulence, *Physica D*, *218*(1), 77–82.

Chillà, F., J. Peinke, and B. Castaing (1996), Multiplicative process in turbulent velocity statistics: A simplified analysis, *J. Phys. II*, *6*(4), 455–460.

David, H. A. (1981), *Order Statistics*, 2nd ed., Wiley, New York.

Frisch, U. (1995), *Turbulence: The Legacy of A.N. Kolmogorov*, Cambridge Univ. Press, New York.

Frisch, U., and D. Sornette (1997), Extreme deviations and applications, *J. Phys. I*, *7*, 1155–1171.

Hogg, R. V. (1972), More light on the kurtosis and related statistics, *J. Am. Stat. Assoc.*, *67*(33), 422–424.

Huff, F. A. (1967), Time distribution of rainfall in heavy storms, *Water Resour. Res.*, *3*, 1007–1019, doi:10.1029/WR003i004p01007.

Katul, G. G., C. Hsieh, and J. Sigmon (1997), Energy-inertial scale interactions for velocity and temperature in the unstable atmospheric surface layer, *Boundary Layer Meteorol.*, *82*(1), 49–80.

Katul, G. G., A. Porporato, D. Cava, and Mario Siqueira (2006), An analysis of intermittency, scaling, and surface renewal in atmospheric surface layer turbulence, *Physica D*, *215*(2), 117–126.

Kumar, P., and E. Foufoula-Georgiou (1993), A multicomponent decomposition of spatial rainfall fields 1. Segregation of large- and small-scale features using wavelet tranforms, *Water Resour. Res.*, *29*, 2515–2532, doi:10.1029/93WR00548.

Kumar, P., and E. Foufoula-Georgiou (1996), A multicomponent self-similar characterization of rainfall fluctuations, *Inst. Math. Appl.*, *79*, 239.

Lovejoy, S., and D. Schertzer (1985), Generalized scale invariance in the atmosphere and fractal models of rain, *Water Resour. Res.*, *21*(8), 1233–1250.

Mallat, S. (1989), A theory for multiresolution signal decomposition: The wavelet representation, *IEEE Trans. Pattern Anal.*, *11*(7), 674–693.

Molini, A., G. G. Katul, and A. Porporato (2009), Revisiting rainfall clustering and intermittency across different climatic regimes, *Water Resour. Res.*, *45*, W11403, doi:10.1029/2008WR007352.

Roux, S. G., V. Venugopal, K. Fienberg, A. Arneodo, and E. Foufoula-Georgiou (2009), Evidence for inherent nonlinearity in temporal rainfall, *Adv. Water Resour.*, *32*(1), 41–48.

Schmid, F., and M. Trede (2003), Simple tests for peakedness, fat tails and leptokurtosis based on quantiles, *Comput. Stat. Data Anal.*, *43*(1), 1–12.

Shi, B., B. Vidakovic, G. G. Katul, and J. D. Albertson (2005), Assessing the effects of atmospheric stability on the fine structure of surface layer turbulence using local and global multiscale approaches, *Phys. Fluids*, *17*(5), 055104, doi:10.1063/1.1897008.

Sornette, D. (2003), *Critical Phenomena in Natural Sciences Chaos, Fractals, Selforganization and Disorder: Concepts and Tools*, 2nd ed., Springer, New York.

Tsallis, C., S. Levy, A. Souza, and R. Maynard (1995), Statistical-mechanical foundation of the ubiquity of Levy distributions in nature, *Phys. Rev. Lett.*, *75*(20), 3589–3593.

Waymire, E. C. (2006), Two highly singular intermittent structures: Rain and turbulence, *Water Resour. Res.*, *42*, W06D08, doi:10.1029/2005WR004492.

Zhang, J., and X. L. Wu (2009), Persistent skewness of a strongly active scalar, *Phys. Rev. E*, *79*(4), doi:10.1103/PhysRevE.79.045301.

G. G. Katul, A. Molini, and A. Porporato, Nicholas School of the Environment, Duke University, Box 90328, Durham, NC 27708-0328, USA. (annalisa.molini@duke.edu)