SINGLE AND MULTISCALE REMOTE SENSING TECHNIQUES,
MULTIFRACTALS AND MODIS DERIVED VEGETATION AND SOIL
MOISTURE

S. Lovejoy(1), A. M. Tarquis(2), H. Gaonac’h(3), D. Schertzer (4)

(1) Physics, McGill, 3600 University St., Montreal, Que. H3A 2T8. (2) Dept. of Applied
Mathematics to Agriculture Engineering, E.T.S. de Ingenieros Agrónomos - U.P.M. Ciudad
Universitaria s.n. 28040 Madrid –SPAIN. (3) GEOTOP, UQAM. (4) CEREVE, Ecole Nationale
des Ponts et Chaussées, 6-8, avenue Blaise Pascal, Cité Descartes, 77455 MARNE-LA-
VALLEE Cedex

Shaun Lovejoy  lovejoy@physics.mcgill.ca  voice + 15143986537
Ana M. Tarquis anamaria.tarquis@upm.es  voice +34913365826
Helene Gaonac’h gaonach.helene@uqam.ca  voice + 15149873000
Daniel Schertzer Daniel.Schertzer@enpc.fr  voice +33164153633
Abstract

Geophysical and remotely sensed fields are typically variable over huge ranges of scale. In the last 25 years much progress has been made in understanding scaling processes which generically generate such variability. These processes are increasingly understood to be the result of a nonlinear dynamical mechanism repeating scale after scale from large to small scales leading to non-classical resolution dependencies. This means that the statistical properties systematically vary in strong, power law ways with the resolution. This implies that when (classical) single scale remote sensing algorithms are used to determine surrogates of various (in situ) geophysical fields, that they can at most be correct at the calibration resolution. This is unsatisfactory if only because this unique resolution is usually the finest resolution available; a rather subjective choice.

In this paper we systematically apply scaling analysis and modeling techniques to MODIS TERRA channels 1-7 and the standard derived vegetation and soil moisture indices, this includes (conventional) spectra, trace moments, structure functions and probability distribution multiple scaling methods. We show how to characterize the wide range scaling using various multifractal data analysis techniques. At first sight, the scaling exponents we find are not so large, however they are exponents so that when they act over large scale ranges (here, a factor of 512, in the earth’s surface, it could be many orders of magnitude larger), and they imply large effects. We explicitly show that if we use the index algorithms on lower and lower resolution satellite data, we obtain indices with significantly different statistical properties than if the same algorithm is used at the finest resolution and then degraded to an intermediate value. This shows that the algorithms can at best be accurate at the unique calibration scale and points to the need to develop resolution independent algorithms based on the scaling exponents. Finally, we show how to make multifractal simulations which reproduce the observed scaling behaviors.
Introduction:

Soil moisture is an important hydrologic state variable critical to successful hydro-climatic and environmental predictions. Soil moisture varies both in space and time because of spatio-temporal variations in precipitation, soil properties, topographic features, and vegetation characteristics (Corwin et al., 2006). Monitoring all these processes across multiple scales is a fundamental component of many environmental and natural resource issues and remote sensing has long been identified as a technology capable of supporting such development (Viña et al., 2004; Vereeckena et al., 2007).

The geophysical and biological processes which determine vegetation, soil moisture and other surface characteristics are highly nonlinear; they involve interacting structures from planetary down to millimeter scales. A consequence is that the surface radiance fields are statistically inter-related over comparable ranges. Such statistical inter-relations (e.g. correlations) are the basis of most remote sensing algorithms. A typical approach identifies wavebands particularly sensitive to the surface features of interest - such as the vegetation or soil moisture - then semi-empirical algorithms are developed to relate them to the radiances. Typically, the algorithms assume that satellite data are homogenous or roughly homogeneous at scales below the resolution.

The problem with this approach is that it ignores systematic scale/resolution dependencies due to strong sub pixel variability and singles out a single scale for the calibration. This scale is essentially subjective; it is primarily determined by the available technology. A symptom of this scale problem is that when new, higher resolution satellite data is used in algorithms calibrated at earlier, lower resolutions, it is found that the algorithm must be re-calibrated.

In the last twenty-five years an understanding of these strong resolution dependencies has started to emerge, and systematic techniques are now available for analyzing and modeling.
such singular behavior. Motivated by the fractal geometry of sets (Mandelbrot, 1983), and by
the development of (multifractal) cascade models in turbulence, the origin of this behavior has
been traced to nonlinear dynamical mechanisms which repeat scale after scale from large to
small scales. Paralleling this revolution in our understanding of the consequences of wide range
(space – time) scaling, is the generalization of the notion of scale itself to encompass systems
which are scaling but highly anisotropic (“generalized scale invariance”, GSI, Schertzer and
Lovejoy, 1985b). In the following paper, we illustrate some of these ideas using MODIS data
and vegetation and soil moisture surrogates derived from them. Although we briefly review the
basic multifractal notions, our aim is to perform systematic scale by scale analyses on the
MODIS data and surrogates. This allows us to explicitly demonstrate and quantify the strong
scale dependencies in the current algorithms.

2. Spectral analyses:

The data we used are calibrated MODIS TERRA radiances (product level B1), with
wavelengths indicated in table 1, some of the images are shown in fig. 1. The vegetation and soil
surface moisture indices are standard products described in Rouse et al. (1973) and Lampkin and
Yool (2004); the formulae given in eq. 1:

\[
\sigma_{vi} = \frac{I_2 - I_1}{I_2 + I_1}; \quad \sigma_{sm} = \frac{I_6 - I_7}{I_6 + I_7}
\]

(1)

where the \(\sigma\) are the indices \((vi, sm\) for vegetation and soil surface moisture indices respectively)
and the \(I\)’s the radiances (the band number is indicated in the subscript); relation 1 is usually
stated without reference to any particular scale. We may immediately note that although the
MODIS TERRA data has a resolution of 500 m (channels 1, 2 were degraded from 250m), and
the variability of the radiances and surface features continues to much smaller scales, that the
surrogates are defined at a single (subjective) resolution equal to that of the sensor. One of the
applications of our analyses will be to investigate how the relations between the surrogates and
the bands used to define them change with scale: we expect that since the scaling properties are
different, making the surrogates with data at different resolutions will produce fields with
different properties.

Before describing more sophisticated (multifractal) scale by scale analysis techniques of,
we first use one which is fairly familiar to geoscientists: spectral analysis. In addition to its
familiarity, spectral analysis has the advantage of being very sensitive to scale breaks; it is also
useful for studying anisotropy. In fig. 2, we show the spectral density $P(k)$ for the single scale
surrogates (sss) i.e. the vegetation index and the soil moisture index estimated from the
algorithm at the highest resolution:

$$P(k) = \langle |\tilde{I}(k)|^2 \rangle; \quad \tilde{I}(k) = \int e^{ik \cdot r} I(r) dr$$

where $r$ is a position vector, $k$ is a wavevector. Since $P$ is quadratic in $I$, it is a second order
statistic. In the definition, we have included the theoretically motivated ensemble average
(denoted “$\langle .\rangle$”) although in fact, below we estimate $P$ from a single realization using a fast
Fourier algorithm and a standard Hanning window. To make the contours clearer, we smoothed
log$P$ (with a four point Gaussian smoother). In the figure, we see that contours are fairly
roundish indicating that to some approximation, the second order statistics are isotropic. Careful
scale by scale analyses of the anisotropy would be rewarding see e.g. Pflug, et al (1993), Lewis,
et al. (1999), Beaulieu et al. (2007), but is outside our present scope. Analysis of the individual
bands revealed a degree of isotropy similar to the indices shown in fig. 2.

The roundness of the $P$ contours justifies the use of the “isotropic” spectrum $E(k)$

obtained by angle integrating $P$:

$$E(k) = \int_1^{k'} P(k') dk'$$

where $k$ is the modulus of the wavevector (the notation indicates angle integration in fourier
space). In fig. 3a, we see $E$ on a log-log plot showing that with the exception of the single
lowest wavenumber $k=1$ that the $E$ is quite accurately of the power law form:

$$E(k) \propto k^{-\beta}$$
where $\beta$ is the “spectral exponent”. Note that sometimes angle averaging (rather than integrating) is performed; in 2D, the corresponding exponent is $\beta-1$. The advantage of using the present (turbulence based) definition is that if the process is isotropic, then $\beta$ is independent of the dimension of space (e.g. 1-D sections will have the same exponent). We can also see from the figure that the channel with strong artifacts (“banding” in channel 5, indicated by blue in fig. 3a) nevertheless has good scaling except at harmonics of the basic banding wavenumber. In table 2, we indicate the $\beta$’s (estimated for $k\geq2$), finding for all the bands $\beta\approx1.17\pm0.08$ where the uncertainty indicates the band to band spread in exponents.

The power law for eq. 4 is called “scaling” because the form of the spectrum is invariant under isotropic “zooms” ($x \rightarrow \lambda^{-1}x$; $k \rightarrow \lambda k$), the exponent $\beta$ is “scale invariant”. In order to better judge the quality of the scaling, we can “compensate” the spectrum by dividing by the mean behavior; this is shown in figs. 3b, c (where we have removed the single $k=1$ values in order to blow up the ordinate). Deviations from flatness indicate deviations from scaling with $\beta=1.17$. In this way, we can easily see that the extreme high wavenumber factor of roughly 2 in scale has extra energy, presumably because of noise (possibly introduced in the post processing). Also, the single scale surrogates are both more strongly affected at the highest wavenumbers, in addition their $\beta$’s are lower indicating that they also have stronger variability at the lower wavenumbers. Presumably the extra high wavenumber variability is because they are the result of a nonlinear operation at the single pixel scale, this operation breaks the scaling. Note that although the Fourier angle integration has reduced the variability, it is still present; this is quite normal for multifractal processes. Also note that the angle integration smoothes more effectively at the higher wavenumbers since there are many more small scale structures than large ones. Finally, note that the low wavenumber break in the scaling (for $k=1$) is probably caused by post-processing which attempts to correct for atmospheric attenuation. In effect, the algorithm works as a low pass filter ensuring that the overall intensity over the scene is roughly constant.
3. Multifractal analyses:

3.1 (Generalized) Structure functions:

As indicated in eq. 2 the spectrum is a second order statistic; if we assume statistical horizontal translational invariance (“statistical homogeneity”, this is the spatial counterpart of statistical “stationarity” which refers to statistical translational invariance in time), then the Wiener-Khintchin theorem shows that the spectrum is the Fourier transform of the autocorrelation function. Before the advent of multifractals, it was believed that many turbulent and turbulent like processes could be well-modeled by Gaussian processes in which the statistics are completely determined by $P(k)$ (isotropic Gaussian processes are completely determined by $E(k)$). However, in order to explain the extreme variability associated with the phenomenon of “intermittency”, cascade models were developed in which a simple mechanism repeats scale after scale in a multiplicative manner while “conserving” the mean of the process. The resulting statistical behavior is given by:

$$
\langle \varepsilon_{\lambda} \rangle = \lambda^{K_{\varepsilon}(\lambda)}; \quad \lambda = L/l
$$

where $\lambda$ indicates the scale ratio over which the process has been developed, $L$ is the “outer scale”, $l$ is the inner scale, and $K_{\varepsilon}(\lambda)$ is the (convex) moment scaling exponent. The symbol “$\varepsilon$” is used for the cascade process since the prototypical cascade quantity is the turbulent energy flux denoted $\varepsilon$ which is conserved from one scale to another in hydrodynamic turbulence. Note that we are discussing systems far from equilibrium and it is not the energy which is conserved but rather the ensemble mean flux of energy from large scales to smaller scales (in fact, $\varepsilon$ is only a true flux in fourier space). In terms of the moments in eq. 5, this conservation implies that the (appropriately normalized) process respects $\langle \varepsilon_{\lambda} \rangle = 1$ so that $K_{\varepsilon}(l)=0$. Actually, although this ensemble “canonical” conservation is the most general one, it is not the only one possible. Indeed, more restrictive “microcanonical” conservation principles are quite popular – in spite of
the fact that the variability of the corresponding multifractal processes is much lower and that such processes do not have simplifying universal behaviors (see Schertzer and Lovejoy, 1992, below).

The typical inner scales of turbulent processes in the atmosphere are the viscous dissipation scales which are typically of the order of millimeters or less; in remote sensing applications, the relevant resolution scale in eq. 5 is the resolution of the images which is typically much larger. In this case, the small sub-sensor scales average out most of the small-scale variability. However, in general, they will not completely smooth it out; this is the origin of the interesting strong variability of the extremes which goes variously under the names “multifractal butterfly effect”, “non-classical SOC”, “first order multifractal phase transitions”, “divergence of statistical moments”; this will be discussed in section 5 below (for a review of this and other multifractal properties, see [Schertzer, et al., 1997]).

The behavior described by eq. 5 is called “multiscaling” because each statistical moment is scaling but with a different exponent. Continuing with the turbulence example, the famous Kolmogorov law for isotropic turbulence relates that energy flux to velocity gradients (Δv) as follows:

\[ \Delta v_\lambda = \varepsilon_\lambda^{a} \lambda^{-H}; \quad H = 1/3; \quad a = 1/3 \]  

(6)

We can see that the typical observables have an extra linear scaling term \( Hq \) and we have introduced the (generalized) structure function exponent \( \xi(q) \) (the usual structure function is for \( q=2 \), for a single realization – no ensemble averaging, the latter is called a “variogramme”). \( H \) thus characterizes the distance from the (conserved) pure multiplicative process \( \varepsilon \); it is the
degree of non (scale by scale) conservation of the process. In order to check this behavior on the MODIS data and surrogates, we need only estimate the gradients using:

$$\Delta v_\lambda = \left| v(x + \Delta x) - v(x) \right| ; \quad \lambda = L / |\Delta x|$$

(8)

where we have assumed translational invariance and isotropy. It is worth noting that this definition of the fluctuation $\Delta v_\lambda$ is sometimes called the “poor man’s wavelet”; other choices of definition are possible; wavelets provide a systematic framework for this (see e.g. Holschneider, 1995). However, in practice, the definition eq. 8 is usually adequate, the main restriction being that it is only appropriate when $0 \leq H \leq 1$, a condition which is usually (although not always) satisfied in geophysical applications (here we will see that $H \approx 0.18$). For example, when $H > 1$, one must measure fluctuations with respect to a local linear trend; this can be done either by fractionally differentiating the process (power law filtering, Schertzer and Lovejoy, 1987), using appropriate wavelets (Bacry, et al., 1989) or using the “Multifractal Detrended Fluctuation Analysis” technique (Kantelhart, et al., 2002). Note that using the Wiener-Khintchin theorem, we obtain a simple relation between the second order structure function exponent and the spectral exponent: $\beta = 1 + \xi(2)$; this is indeed approximately verified (see table 2 and fig. 5).

Using eq. 8 to estimate the fluctuation $\Delta v_\lambda$, we obtain the generalized structure functions shown in fig. 4a for bands 1, 2, 6, 7 and fig. 4b for vegetation index and soil moisture index. We can see that except for the smallest factor of 4-8, the (multiple) scaling is excellent. In order to quantify the differences in the scaling, in fig. 5a, b, we show the slopes which are our estimates of $\xi(q)$. We can see that (as expected) $\xi(q)$ is concave downwards. It is interesting to note that while the vegetation index has a $\xi(q)$ intermediate between it’s defining band $\xi(q)$’s, the soil moisture $\xi(q)$ is somewhat lower, outside the range. It is clear from the figures that a quantitative inter comparison of these continuous concave curves $\xi(q)$ is not possible; we first need to reduce the problem to a finite number of manageable parameters. Although we discuss this in much more detail below, we may already make a first quantification using the exponent $H$. If we assume that due to the scale by scale conservation of the underlying nonlinear cascade
that $K(1)=0$, we see from eq. 6 that $H=\xi(1)$ (actually, according to eq. 7, $H=\xi(1)+K(a)$, and in
general, we don’t know $a$; however the correction $K(a)$ is generally small and will be ignored
below). In table 2 we compare these estimates finding that over all 7 bands: $H=0.189\pm0.034$
whereas $H_{vi}=0.162$, $H_{sm}=0.144$ so that the surrogates have H’s which are a little lower than those
of the individual bands. For comparison, we have already mentioned that the classical
Kolmogorov turbulent result is $H=1/3$. Another classical theoretical $H$ is for passive scalars in
turbulence (the Corrsin-Obhukhov law also with $H=1/3$). Empirically, the topography has
$H=0.4 -0.7$ (oceans and continents respectively; Gagnon, et al., 2006) whereas many infra red
an visible signals from volcanic and other surfaces give $H$ quite close to those found here (e.g.
Laferrière and Gaonac’h, 1999, Harvey, et al., 2002, Gaonac’h, et al., 2003; see also Lovejoy
and Schertzer, 2006 for cloud radiances).

It remains to characterize the remaining nonlinear part, $K(q)$, although we do this more fully in
section 6 below, we can nevertheless go a step further by characterizing the slope of the $\xi(q)$
curve near the mean, i.e. $\xi'(1)=H-K'(1)$. Defining $C_1=K'(1)$ as the “codimension of the mean”
(see below for the justification of this terminology), we obtained the mean for all 7 bands: $C_1=
0.032\pm0.010$ whereas $C_{1vi}=0.033$, $C_{1sm}=0.010$. Although these values appear small, typical
values in turbulence are only a bit larger e.g. $C_1\approx0.07$ for the horizontal wind in the horizontal
(Schmitt, et al., 1992; Schmitt, et al., 1994), and $C_1\approx0.04$ for passive scalars in the horizontal
(Lilley, et al., 2004). The corresponding values in the vertical are about 9/5 times larger; this is
as predicted by the 23/9D model of scaling stratification, (Schertzer and Lovejoy, 1985a).
Finally, topography has $C_1\approx0.12$ (pretty much the same for both continents and oceans). For an
early review of these and other results, see Lovejoy and Schertzer (1995).
3.2 Multiscale versus Single scale surrogates:

We have already mentioned that the vegetation and soil moisture surrogates are defined in an ad hoc way using the (subjective, finest available) resolution. The fact that the surrogates have different scaling ($\xi(q)$’s) implies that if the surrogates are defined at different resolutions that their statistical properties will be different, in other words, the single scale surrogate can be (at most) correct at a single resolution. To see this more clearly, these “single scale surrogates” (sss, at scale $\lambda$, $\sigma_{\lambda}^{(s)}$) can be contrasted with the corresponding “multi scale surrogates” (mss, at scale $\lambda$, $\sigma_{\lambda}^{(m)}$). Mathematically, the difference can be expressed as:

$$\begin{align*}
\sigma_{\lambda}^{(s)} &= \left(\sigma_{\lambda}^{(s)}\right)_{\Lambda} = \left(\frac{I_{i,\Lambda} - I_{j,\Lambda}}{I_{i,\Lambda} + I_{j,\Lambda}}\right), \\
\sigma_{\lambda}^{(m)} &= \left(\frac{I_{i,\lambda} - I_{j,\lambda}}{I_{i,\lambda} + I_{j,\lambda}}\right), \\
\lambda &= \left(I_{i,\lambda}\right)_{\lambda}
\end{align*}$$

where the maximum scale ratio (satellite image scale/single pixel scale) $L/l=\Lambda$ and the notation $(I_{i,\lambda})_{\lambda}$ denotes averaging from this finest resolution up to the intermediate resolution $\lambda<\Lambda$. The mss is the surrogate that would be obtained by applying an identical algorithm (eq. 1) on satellite data at the lower resolution $\lambda$. In fig. 4c, we show that resulting $\xi_{\text{mss}}(q)$, obtained over the same range of scales as the sss $\xi$ (fig. 4b) is quite different (significantly larger); we quantify this below. To partially quantify this difference, we see from table 2, that $H_{\text{mssvi}}=0.231$, $H_{\text{msssm}}=0.213$ which is significantly larger than the sss estimates of the previous subsection; $H_{v}=0.162$, $H_{s}=0.144$. Similarly, determining the corresponding $C_1$ values, we find $C_{1\text{mssvi}}=0.049$, $C_{1\text{mssvi}}=0.059$ (c.f. the values 0.033, 0.010 for the sss $C_1$’s which are substantially smaller, especially the soil moisture values). These statistical differences imply that if the indices are calculated at the MODIS resolution $\Lambda$ and then degraded to an intermediate resolution $\lambda$, that the results would be quite different than if the same algorithms were used but directly on data from a lower resolution satellite. In addition, the problem gets worse as the range of scales increases.
because the scaling exponents are different. We therefore conclude that due to the scale
dependence of the algorithm, that the surrogate indices cannot be valid over a very significant
range of scales. This underscores the need to develop scale invariant algorithms based on the
scale invariant exponents; see Lovejoy, et al. (2001) for a discussion.

4. Trace Moment Analysis:

We have seen that the generic statistical properties of processes which are repeated scale
after scale are characterized by a nonlinear exponent $K(q)$, and that the observables will
generally have an extra linear scaling term $qH$. Since at least for low $q$; the linear term $qH$ is
often larger than the nonlinear $K(q)$, in analyses, it can mask the latter, it is therefore
advantageous to first estimate the conserved flux $\varepsilon$ from the observed $v$, and estimate $K(q)$
directly. From eq. 6, we see that in principle, this can be done by removing the $\lambda^{-H}$ scaling. To
do this, note that if we start with a field $\varepsilon^a$ and fractionally integrate it by $H$; (a power law filter
$k^{-H}$), that the resulting field will have the fluctuation statistics indicated by eq. 6 (see Marsan, et
al., 1996). This suggests that to obtain a flux from $v$ (i.e. a conserved field with $H=0$), that it
suffices to invert the power law filter, i.e. to fractionally differentiate by an order $H$. It turns out
that a finite difference approximation to this followed by an integer order $H'>H$ and then taking
the absolute value of the result is sufficient (the absolute value is necessary since the
multiplicative cascade $\varepsilon>0$; see however complex and vector cascades (Scherzer and Lovejoy,
1995). Since usually, $0\leq H\leq 1$, a first order finite difference is sufficient. One simply takes the
absolute differences at the finest available resolution $\varepsilon_\Lambda$ and degrades them:

$$\varepsilon_\Lambda = (\varepsilon_\Lambda)_\lambda; \quad \varepsilon_\Lambda = |\Delta v_\Lambda|; \quad 0 < H < 1$$

(10)

where again the notation $(\varepsilon_\Lambda)_\lambda$ indicates the average of the finest resolution data $\varepsilon_\Lambda$ over the
intermediate resolution $\lambda$. In 1D (series), one can simply use $\Delta v_\Lambda(x) = v(x+l) - v(x)$ where
the inner scale is $l=L/\Lambda$ where $L$ is again the outer scale. In higher dimensions on data on
regular grids (as here) – if the data is fairly isotropic – then we can simply treat each line separately and then use the 1D method. However, there are more possibilities, for example we may use finite difference moduli of gradient vectors, or finite difference Laplacians (the latter being an isotropic order 2 derivative). In 2D, this yields respectively:

\[
\begin{align*}
|\Delta v_\lambda (x,y)| &\approx \left[ (v(x+l,y) - v(x,y))^2 + (v(x,y) - v(x,y+l))^2 \right]^{1/2} ; \quad l = L / \Lambda \\
|\Delta v_\lambda (x,y)| &\approx |v(x,y) - (v(x+1,y) + v(x-1,y) + v(x,y+1) + v(x,y-1))/4|
\end{align*}
\]

(l is the inter-pixel distance). These and other variants have been extensively tested on numerical simulations and on data, see for example Tessier et al. (1993) and Lavallée, et al. (1993); there is generally not much difference between the various choices (or between direct fractional differentiation using Fourier techniques).

In fig. 6a, b we show the results on the $\text{sss}$ indices. The plot shows various statistical moments of $\varepsilon_\lambda$ at various intermediate resolutions obtained by spatially averaging $\varepsilon_\lambda$ obtained from the modulus of the gradient vector over larger scales (smaller scale ratios) $\lambda$. Since the intermediate scale $\Delta x = L / \lambda$, the smallest scales are on the right, the largest scales are on the left.

As a first remark, we see that the overall range of variability is quite a bit smaller than for the corresponding structure functions; this is because for the moments shown ($q<2$), the $qH$ term is larger than the nonlinear $K(q)$ term. Notice that for $q>1$ the effect of spatial averaging is to decrease the values ($K>0$) whereas for $q<1$, it increases them ($K<0$). Next, we can see that due to the nonlinear operations at the smallest scales (the definition in terms of the various bands, the modulus of the gradient operator), the scaling is not so good at the finest resolutions; the exponents $K(q)$ (the slopes on the log-log figure) were estimated over the intermediate range shown. At the very largest scales the scaling is also poor partly because of poor statistics (there aren’t very many large scale structures), but also because of the low wavenumber atmospheric corrections mentioned earlier.
To quantify the scale by scale statistical differences between the bands and the indices, we can use the scaling of the moments (the slopes in fig. 6) to estimate $K(q)$ and compare the various moment scaling exponents; this is done in fig. 7a,b,c. Again, we can quantify the differences by the behavior near $q=1$; numerically evaluating the slope $C_1=K'(1)$, we obtain a second series of estimates shown in table 2, for the bands, these are very close to the previous (for the seven bands: $C_1=0.0367 \pm 0.001$, c.f. the previous value $0.032 \pm 0.010$), but for the surrogates, the values are somewhat different: $C_{i,v}=0.064$, $C_{i,s}=0.053$ compared to the structure function estimates $0.033$, $0.010$ respectively. The fact that the two methods agree so well for the bands but not so well for the surrogates is due to the poorer scaling of the latter, itself due to their scale dependent definitions.

5. The Multiscaling of the probabilities and the Probability Distribution Multiple Scaling Technique (PDMS):

We have concentrated on the statistical moments as a simple method of characterizing scale invariant fields. However, the moments are integrals of the probability densities, and the multiscaling of the former implies the multiscaling of the latter. In addition, under fairly general conditions, the relationship can be inverted so that knowledge of the scaling of the moments determines the scaling of the probabilities. In particular, we obtain:

$$\Pr(\varepsilon_\lambda > \lambda^\gamma) \approx \lambda^{-c(\gamma)}$$  \hspace{1cm} (12)

which states that the probability (“Pr”) that a (resolution $\lambda$) value $\varepsilon_\lambda$ exceeds the threshold $\lambda^\gamma$ is a power of the resolution with exponent $-c(\gamma)$ (Schertzer and Lovejoy, 1987). The equality sign “$\approx$” is to within constant or slowly varying factors (e.g. including “sub-exponential” factors such as $\log \lambda$). Also, note that although in the above we use the expression “probability distribution”, it is not strictly appropriate since it represents the integral of the probability density from $\lambda^\gamma$ to $\infty$ rather than the usual $-\infty$ to $\lambda^\gamma$ (the latter corresponds to a true “cumulative distribution function”). $\gamma$ is called a “singularity” since it quantifies the rate at which the field values $\varepsilon_\lambda$
diverge in the small scale limit $\lambda \to \infty$ (if $\gamma<0$, it is in fact a “regularity”). The function $c(\gamma)$ defined via eq. 12 is called the (statistical) “codimension function” (Schertzer and Lovejoy, 1987) or sometimes the “Cramer function” (Mandelbrot, 1989). If it is low enough ($c(\gamma)<d$, where $d$ is the dimension of the embedding space, $d=2$ for single images), then $d(\gamma)=d-c(\gamma)$ is the (geometrical) fractal dimension of the set of points (at resolution $\lambda$) with $\varepsilon_\lambda > \lambda^\gamma$. While the geometrical $d(\gamma)$ satisfies $0 \leq d(\gamma) \leq d$, the only restriction of $c(\gamma)$ is that it is nonnegative.

The use of $c$ in place of $d$ is necessary in stochastic processes since events with $c>d$ (i.e. those which would have an impossible negative geometrical dimension $d(\gamma)<0$) correspond to events which are almost surely absent in a $d$ dimensional space but on the contrary are almost surely present in a large enough sample. Indeed, if we have $N_s$ independent realizations of the process, each over a range scales $\lambda$, then the “effective” or “sampling dimension” of the sample is $D_s=\log N_s/\log \lambda$, and one will expect to find all levels of activity up to extreme events such that $c(\gamma_s)=d+D_s$ where $\gamma_s$ is the “sampling singularity” (Schertzer and Lovejoy, 1989). In particular, there will be a largest moment $q_s=c'(\gamma_s)$ which can be estimated from a finite sample of $N_s$ realizations, for $q>q_s$ the $K(q)$ becomes (spuriously) linear following the tangent $K'(q_s)$.

As mentioned above, knowledge of $K(q)$ is equivalent to knowledge of $c(\gamma)$, and visa versa, indeed, the two are conveniently related via Legendre transformations (Parisi and Frisch, 1985):

$$K(q) = \min_{\gamma} (q\gamma - c(\gamma))$$

$$c(\gamma) = \min_{q} (q\gamma - K(q))$$

(13)

This establishes a one to one correspondence between orders of singularity $\gamma$ and statistical moments $q$: $\gamma = K'(q)$, $q = c'(\gamma)$, we can now see the justification for the term $C_1=K'(1)$ introduced earlier; in a precise sense, it corresponds to the singularity which gives the dominant contribution to the mean, in addition since $K(1)=0$, we find $C_1=c(C_1)$ is also the codimension of the $\gamma=C_1$ singularities.
Before using eq. 11 to directly estimate \( c(\gamma) \) we recall that the “≈” sign in eq. 11 means that using the approximation 

\[ c(\gamma) = \frac{-\log P_r}{\log \gamma} \]

is not so accurate. It is therefore better to calculate the histogrammes at a series of resolutions \( \lambda \) and then, for a fixed \( \gamma = \log \varepsilon_\lambda / \log \lambda \), regress \( \log P_r \) against \( \log \lambda \); the slope is \(-c(\gamma)\) (this is the Probability Distribution Multiple Scaling Technique (PDMS) (Lavallée, et al., 1991). We show the result in fig. 8a, b, c. As can be seen at the larger \( \gamma \) (the extremes) the statistics become poor so that even though in principle there is information up until \( c = d = 2 \), it is very noisy. Indeed, if seek to study the extremes, then determining the histogrammes directly at the high resolution is quite revealing. This is especially true since in general, as mentioned in section 3 in general “canonical” cascade processes, there is a fundamental distinction that must be made between the “bare cascade”, quantities \( \varepsilon_{\lambda b} \), i.e. that developed from the outer scale \( L \) down to a smaller (but finite) scale \( l \) and the “dressed” process \( \varepsilon_{\lambda d} \) obtained by averaging (“dressing”) a bare process continued down to the small scale limit. While the former has all its positive moments converge (the corresponding bare \( K_b(q) \) is always finite), the dressed process is more variable, it generally has diverging moments for all moments \( q > q_D \) where \( q_D \) is critical exponent dependent on the dimension \( D \) over which the process is averaged set (Mandelbrot, 1974, Schertzer and Lovejoy, 1987); \( K_d(q) = K_b(q) \) for \( q < q_c \). In general:

\[
\left\{ \Pr(\varepsilon_{\lambda, d} > s) \approx s^{-q_0}; \ s \gg 1 \right\} \leftrightarrow \left\{ \langle \varepsilon_{\lambda, d}^q \rangle \rightarrow \infty; \ q > q_D \right\}
\]

(14)

In practice, the moments of a finite data set are always finite, nevertheless, the divergence of moments implies that the \( K(q) \) estimated for \( q > q_D \) will be dominated by a single (largest) singularity, the scaling will be spurious, there will be discontinuity in the slope of \( K(q) \) at \( q = q_D \). Since there is a mathematical analogy between classical thermodynamics and multifractals, the discontinuity is \( K'(q) \) at \( q = q_D \) is called a “first order multifractal phase transition”. Another term for the phenomenon is the “multifractal butterfly effect” (Lovejoy and Schertzer, 1998), a term which is justified because – just as for the classical deterministic chaos “butterfly effect” - the large scale moments for \( q > q_D \) are in fact determined by the small scale
details. Finally, the divergence of moments occurs in combination with scaling, fractal
structures (indeed, it is the result of the build up of such strong variability from large to small
scales that the spatial averaging over finite sets cannot sufficiently “calm” the higher order
statistical moments). Elsewhere, in statistical physics, the combination of fractal structures with
algebraic extreme probabilities has been called “Self-Organized Criticality” (SOC; Bak, et al.,
1987), hence, we see that multifractals provide a “non-classical” route to SOC (Schertzer and
Lovejoy, 1994). Actually, this non-classical SOC may be more generally physically relevant
since it based on a (quasi) constant flux whereas the classical SOC is in fact only strictly valid in
the (unrealistic) “zero flux” limit.

In order to check for the divergence of moments, we therefore calculated the probability
distributions of $\varepsilon_\lambda$; see fig. 9a,b. Due to the slight excess smoothing at the highest factor of two
in scale (see, the spectra, section 2), we averaged the (gradient estimated) $\varepsilon_\lambda$ over 2X2 pixels.
We can see that the forms of the extreme probability tails are roughly the same for the indices
and the bands from which they were derived, but it is not obvious that any asymptotic linear
behavior is followed. The problem is that it is notoriously difficult to estimate the exponent of a
tail, especially when theoretically; we know neither the value of the exponent, nor where the
asymptotic regime begins. In order to quantify the tail behavior, we estimated the tail slope for
the greatest factor of 2 in scale (see table 2); for the bands 1,2, 6, 7 we obtained a mean $q_D \approx 6.64$
(the mean of all the bands is $6.4 \pm 1.4$) this reference slope is placed on the distributions in figs.
9a, b. We see that the agreement is quite suggestive (especially when we consider that various
instrumental (smoothing) effects could be responsible for artificially reducing the values of the
extreme gradients used to estimate $\varepsilon$). The evidence for algebraic tails is apparently strongest for
the soil moisture surrogate and the corresponding bands 6, 7. Alternatively, the absolute
logarithmic slope in the tails in fig. 9 yield the maximum moments that can be accurately
measured with the available sample. It may be that this is an estimate of $q_s$ and that $q_s < q_D$ so
that the data set is simply too small to observe $q_D$ (recall that as $N_i$ increases, so does $q_D$).
6. A complete, manageable and physically motivated characterization of the processes:

**Universal multifractals:**

We have argued that a general feature of processes whose mechanism repeats over a wide range of scales is that they are scaling, characterized by convex moment functions \((K(q), c(\gamma))\). If this was all that we could deduce, then multifractals would be quite unmanageable, theoretically it would mean that an infinite number of parameters were important (e.g. each value of \(K(q)\)), or empirically, it would require the determination of an infinite number of coefficients (e.g. regression slopes). However, in physics it has generally been found that when processes are iterated sufficiently or if there are a large enough number of interactions, that only a few of their characteristics are important in the limit of many interactions/iterations. This is the idea of “universality”. Perhaps the oldest and most familiar example of universality is the central limit theorem in probability theory which is routinely invoked to justify assumptions about gaussianity of measurement noises. Although the Gaussian example is well known, the full result – the “Generalized Central Limit theorem” (Levy, 1925) for sums of independent identically distributed (i.i.d.) random variables (r.v.’s) is less known. It states that if one appropriately centers and normalizes sums of a sufficiently large number of i.i.d. r.v.’s with finite variance that the result is indeed a Gaussian. However, if the finite variance requirement is dropped and one allows for algebraic distributions with exponents \(\alpha<2\), then one obtains Levy distributions. This result is relevant to cascade processes because if one considers the “generator” of the process, \(\log \epsilon\), then this is the sum of the logarithms of the random cascade factors, hence one expects the generalized central limit theorem to apply to the logs. This would lead to log Gaussian and log Levy multifractals. Although this basic idea turns out to be correct, the nontrivial small scale limit of the cascade process has obscured the issue even leading to strong statements that there are no universal multifractal properties ([Mandelbrot, 1989]). Indeed, to obtain universal properties, one must consider the universality issue on cascades developed only over a finite range of scales, and only then - after central limit convergence has
been achieved - to consider the small scale limit (see Schertzer and Lovejoy (1997) for more details and the debate). The resulting bare universal multifractals have the following exponent functions:

\[ K(q) = \frac{C_1}{\alpha - 1} (q^{\alpha} - q); \quad q \geq 0 \]

\[ c(\gamma) = C_1 \left( \frac{\gamma}{C_1 \alpha'} + \frac{1}{\alpha'} \right)^{\alpha'}; \quad \frac{1}{\alpha'} + \frac{1}{\alpha} = 1 \]  

(15)

the above Legendre transformation pairs (Schertzer and Lovejoy, 1987) are valid for \(0 \leq \alpha < 1\), \(1 < \alpha \leq 2\) (when \(\alpha = 1\), \(K(q) = C_1 q \log q\)), and \(0 \leq C_1 \leq d\) (if \(C_1 > d\), the process cannot be normalized on the subspace). The restriction to \(\theta \geq 0\) is necessary (unless \(\alpha = 2\), the “log normal” multifractal) since generally the \(q < 0\) moments diverge; this is why we do not plot them in figs. 5, 7. Once again, the results for the dressed cascades apply so that these formulae are only valid for \(q < q_0\) and \(\gamma < \gamma_0 = K'(q_0)\). Because of this the term “log-Levy” and “log-Gaussian” are misnomers; the probability tails are actually somewhat stronger than those terms would imply. It could be mentioned that in the literature, a weaker form of universality has been proposed leading to log-Poisson multifractals (She and Levesque, 1994). While log-Poisson multifractals share the infinite divisibility properties of the universal (Levy-generator) multifractals (necessary for continuous in scale cascades), the corresponding processes are neither stable nor attractive so that it would be surprising to for them to emerge from a complex real world process.

The universal form eq. 15 shows that only two parameters are needed to specify the conserved flux of universal multifractals; for the “observables” \((\nu)\), we have a third parameter \((H)\). We have seen in section 3 that two of these parameters \((H, C_1)\) can be estimated from \(\xi(1)\), \(\xi'(1)\); the above shows that only a single additional parameter \(\alpha\) is needed to characterize the entire process. This Levy index \(\alpha\) can be estimated either from the radius of curvature of \(\xi(q)\) (or \(K(q)\)) at \(q=1\) (equivalently, determined from the second derivatives), or – in practice more accurately - by considering \(K(q)\) or \(\xi(q)\) over a wider range of \(q\) values and exploiting the dependence on \(q^\alpha\). We test one of these methods ([Schmitt, et al., 1995]) on the MODIS data
(for another popular method, “Double Trace Moments”, see [Lavallée, et al., 1993]). As long as \( \nu (0) \) is finite, we may remove the linear part of \( \xi(q) \) by considering the “residue” \( r(q) \):

\[
r(q) = q \xi'(0) - \xi(q)
\]

(16)

Applied to universal multifractals, as long as \( \nu > 1 \), we have:

\[
r(q) = \frac{C_1}{\alpha - 1} q^\alpha
\]

(17)

Hence we can estimate \( \nu \) by a linear regression of \( \log r(q) \) versus \( \log q \). We show this in fig. 10a, b using \( q \) values from 0.1 to 3 at intervals of 0.1 ( \( \nu (0) \) is estimated from a local quadratic fit to \( \xi(q) \) with \( q=0, 0.1, 0.2 \)). As can be seen, the residue is accurately a power law; for the bands we find (table 2): \( \nu = 1.91 \pm 0.03 \), while for the indices, we find \( \alpha_{vi} = 2.02 \), \( \alpha_{sm} = 2.23 \) both of which are slightly outside the theoretically allowed range for universal processes (it can thus only be an approximate empirical characterization). This is presumably a further consequence of the poor scaling of the indices (but not the radiances). Finally, we compare these results to those of a nonlinear regression on the trace moment estimate \( K(q) \) curve. Due to the presence of the linear term in the universal \( K(q) \) (eq. 15), the nonlinear regression is not as accurate. We might also note that a potential problem with it is that one must only make a fit for \( q < \min(q_s, q_D) \) (recall that for \( q > q_s \), or \( q > q_D \), \( K_d(q) \) becomes linear, it is spurious), and a priori, the appropriate values of \( q_s, q_D \) are unknown. Here we have seen that the extreme \( q \) which can be used is of the order \( q \approx 6 \) or greater and we made our nonlinear regressions using the relatively well estimated moments \( q \leq 2 \).

Using these universal multifractal parameters, we can make multifractal simulations with the same statistics/resolution dependencies (Schertzer and Lovejoy, 1987, Wilson, et al., 1991, Pecknold, et al., 1993), see fig. 11a, b for examples with roughly the same universal multifractal parameters as the radiances. For more simulations, showing the effect of varying not only the \( H, \ C_1, \ \alpha \) parameters, but also various scaling anisotropies, see Lovejoy and Schertze, (2007) and also the Multifractal Explorer site: http://www.physics.mcgill.ca/~gang/multifrac/index.htm.
7. Conclusions:

Due to extreme variability over huge ranges of scale, it is often difficult or impossible to obtain adequate quantities of in situ data; remotely sensed surrogates are attractive alternatives. The classical approach combines remote data from judiciously chosen radiance channels in semi-empirical algorithms which are calibrated at the best available resolutions. The problem is that this is predicated on the classical geostatistical assumption that the fields of interest are sufficiently regular, smooth so that they have no significant resolution dependencies, in effect by postulating \textit{a priori} that the subpixel variability is not serious. In this paper, we argue that on the contrary, the fields are multifractal having strong resolution dependencies (they are singular with respect to Lebesgue measures). This implies that at best the surrogate fields derived in this manner can be correctly calibrated at a single resolution, they will be incorrect at other resolutions. This shows that new, resolution independent algorithms must be developed.

In this paper, we illustrate these ideas using soil moisture and vegetation indices and surrogates obtained from red, and far red MODIS satellite imagery. We systematically review the key notions and analysis methods which have emerged after over twenty years of study of cascades and other multifractal processes. First, using the classical technique of fourier analysis, we showed that the basic isotropic scaling was reasonably well obeyed over most of the available range of scales (here, a factor of 512). We then considered the statistics of fluctuations which can be systematically analyzed as functions of scale and intensity (by varying the order, \( q \)) using generalized structure functions \( \xi(q) \). This showed that the basic radiance bands had good multiscaling (i.e. scaling with concave \( \xi(q) \)), motivating the introduction of two different surrogates: the first, the classical “single scale surrogate”, (\( sss \)) is defined by a nonlinear combination of radiances at the finest resolution; the second, the “multiple scale surrogate” (\( mss \)) is based on the same algorithm except that it is derived by successively lower resolution satellite radiances. Although the basic algorithm is identical (the surrogates are equal to the difference in radiances at two bands divided by their sums), the results as functions of resolution
were found to be different. This is a consequence of the scale dependence of the radiances and the nonlinear operation defining the surrogates. The fact that the scale by scale properties of the sss and mss surrogates were different (a fact that we quantified further with parameters $H, C_1$) shows that unless the finest resolution of the satellite data just happens to coincide with the inner scale of the soil moisture and vegetation indices, that the surrogates cannot be more than rough approximation to the indices, valid only near the resolutions at which they were calibrated.

We then went on to make a more complete scale by scale statistical characterization of the data and surrogates, showing how to obtain and analyze the underlying conserved fluxes using a technique called “Trace Moments”, as well as a scale invariant characterization of the probability distributions called the “PDMS technique”. The analysis techniques were originally developed in order to quantify multifractal turbulent processes and we reviewed the relevant theory while explaining the various techniques. At this level, the inter-comparison of the scaling involves two exponent functions, one for the moments, $(K(q))$, one for the probabilities $(c(\gamma))$.

Although these are related by a Legendre transform – so that they contain essentially the same information – they still effectively involve an infinite number of parameters (e.g. a different exponent for each $K$ value). In order to make the problem manageable, we argued that in the real world, multifractal processes are likely to be in the basin of attraction of the three parameter “universal” multifractal processes so that analysis requires the estimation of just three parameters; in addition to the $H, C_1$ mentioned above, the Levy index $\alpha$ which characterizes the degree of multifractality. We then estimated the remaining $\alpha$ index. Overall (table 2), we found the scaling parameters $(H, C_1, \alpha)$ were not too different for the different bands, but were significantly different for the surrogates (which also had scaling which was less accurately followed, a consequence of the nonlinear transformation at the smallest scale). Although the multifractal exponents $H, C_1$ we find are not so large, nevertheless, precisely because they act over wide ranges of scale, they lead to large effects. In contrast, there is another multifractal approach to remote sensing (Lévy-Vehel and Mignot, 1994, Lévy-Vehel, 1995) which is agnostic
about the existence of wide range scaling; in practice only considers very small ranges of scale (e.g. factors of 2-4) so that it leads to applications which are not too different from classical remote sensing techniques.

The similitude of the scaling properties of the different radiance channels does not mean that overall they have the same statistics; indeed, all our analysis techniques were essentially isotropic so that even though by spectral analysis we showed that the anisotropy is not so large – it can still lead to significant differences which can build up scale by scale (indeed, the techniques we used generally wash out any scale by scale direction dependence of statistics/structures). This is perhaps most graphically shown in the simulations (fig. 11). However systematic analysis of the scaling anisotropy is not easy and is outside our present scope. Finally although we criticized classical scale bound remote sensing algorithms, we did not propose a specific alternative. The general problem is to obtain a scale invariant characterization of the statistical inter-relation between the various radiances/ fields. The basic approach is to consider a “state vector” with each relevant field represented by a different component. We are thus lead to scale invariant vectors (“Lie cascades”, Schertzer and Lovejoy, 1995), an approach which has been discussed to some extent in Lovejoy et al., (2001). However, very little is known about vector multifractals, nor about the appropriate analysis techniques.

References.


Harvey, D. A., et al. (2002), Multifractal characterization of remotely sensed volcanic features: a case study from Kilauea volcano, Hawaii, Fractals, 10, 265-274.


Schertzer, D., and S. Lovejoy (1985b), Generalised scale invariance in turbulent phenomena, Physico- 
Chemical Hydrodynamics Journal, 6, 623-635.

Schertzer, D., and S. Lovejoy (1987), Physical modeling and Analysis of Rain and Clouds by 

Schertzer, D., and S. Lovejoy (1989), Nonlinear variability in geophysics: multifractal analysis and 
simulation, in Fractals: Physical Origin and Consequences, edited by L. Pietronero, p. 49, Plenum, 
New-York.


Schertzer, D., and S. Lovejoy (1994), Multifractal Generation of Self-Organized Criticality, in Fractals In 

Schertzer, D., and S. Lovejoy (1995), From scalar cascades to Lie cascades: joint multifractal analysis of 
rain and cloud processes, in Space/time Variability and Interdependence for various hydrological 


Schertzer, D., et al. (1997), Multifractal cascade dynamics and turbulent intermittency, Fractals, 5, 427- 
471.


Schmitt, F., et al. (1994), Empirical study of multifractal phase transitions in atmospheric turbulence, 
Nonlinear Processes in Geophysics, 1, 95-104.


Tessier, Y., et al. (1993), Universal Multifractals: theory and observations for rain and clouds, Journal of 
Applied Meteorology, 32, 223-250.

Vereeckena, H., R. Kasteela, J. Vanderborght and T. Harterb.2007. Upscaling Hydraulic Properties and 

Viña, A., A.A. Gitelson, Donald C. Rundquist, Galina Keydan, Bryan Leavitt, and James Schepers. 
Table 1: Characteristics of the MODIS data, taken over Guadalajara (Central Spain) over 250x250 Km², 29/7/2006.

<table>
<thead>
<tr>
<th>Name</th>
<th>Wavelengths</th>
</tr>
</thead>
<tbody>
<tr>
<td>Band 1</td>
<td>620 – 670 nm</td>
</tr>
<tr>
<td>Band 2</td>
<td>841 – 876 nm</td>
</tr>
<tr>
<td>Band 3</td>
<td>459 – 479 nm</td>
</tr>
<tr>
<td>Band 4</td>
<td>545 – 565 nm</td>
</tr>
<tr>
<td>Band 5</td>
<td>1230 – 1250 nm</td>
</tr>
<tr>
<td>Band 6</td>
<td>1628 – 1652 nm</td>
</tr>
<tr>
<td>Band 7</td>
<td>2105 – 2155 nm</td>
</tr>
</tbody>
</table>
Table 2: Fits are over highest range down to factor 4 from lowest, except TM which also
excludes small scale factor 2.

<table>
<thead>
<tr>
<th></th>
<th>$\beta$</th>
<th>Struc H</th>
<th>TMH</th>
<th>Struc C</th>
<th>TM C</th>
<th>$\alpha$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1.15</td>
<td>0.186</td>
<td>0.109</td>
<td>0.031</td>
<td>0.036</td>
<td>1.86</td>
</tr>
<tr>
<td>2</td>
<td>1.09</td>
<td>0.144</td>
<td>0.078</td>
<td>0.024</td>
<td>0.034</td>
<td>1.89</td>
</tr>
<tr>
<td>3</td>
<td>1.24</td>
<td>0.226</td>
<td>0.160</td>
<td>0.031</td>
<td>0.039</td>
<td>1.88</td>
</tr>
<tr>
<td>4</td>
<td>1.26</td>
<td>0.218</td>
<td>0.166</td>
<td>0.030</td>
<td>0.037</td>
<td>1.92</td>
</tr>
<tr>
<td>5</td>
<td>1.05</td>
<td>0.141</td>
<td>0.064</td>
<td>0.054</td>
<td>0.039</td>
<td>1.95</td>
</tr>
<tr>
<td>6</td>
<td>1.25</td>
<td>0.207</td>
<td>0.159</td>
<td>0.026</td>
<td>0.035</td>
<td>1.93</td>
</tr>
<tr>
<td>7</td>
<td>1.18</td>
<td>0.199</td>
<td>0.130</td>
<td>0.026</td>
<td>0.037</td>
<td>1.95</td>
</tr>
<tr>
<td>Vi sss</td>
<td>1.03</td>
<td>0.162</td>
<td>0.074</td>
<td>0.033</td>
<td>0.064</td>
<td>2.02</td>
</tr>
<tr>
<td>Vi multiscale surrogate</td>
<td></td>
<td>0.231</td>
<td></td>
<td></td>
<td>0.049</td>
<td></td>
</tr>
<tr>
<td>Sm sss</td>
<td>1.03</td>
<td>0.144</td>
<td>0.073</td>
<td>0.010</td>
<td>0.053</td>
<td>2.23</td>
</tr>
<tr>
<td>Sm multi scale surrogate</td>
<td></td>
<td>0.213</td>
<td></td>
<td></td>
<td>0.059</td>
<td></td>
</tr>
<tr>
<td>mean</td>
<td>1.17±0.08</td>
<td>0.189±0.034</td>
<td>0.124±0.04</td>
<td>0.032±0.010</td>
<td>0.037±0.001</td>
<td>1.91±0.03</td>
</tr>
</tbody>
</table>
Figures Caption

**Fig. 1.** Images of Guadalajara (Central Spain), which correspond to an area of 250x250 km. Left column: vegetation index (top), below, bands 1, 2, right soil moisture index, below, bands 6, 7.

**Fig. 2.** $\text{LogP}$ of: a) vegetation Index. b) soil moisture.

**Fig. 3.** a) Spectrum of all 5 channels, Hanning window. Orange=1, green 2, 3, cyan=4, blue spike 5, purple=6, magenta=7. b) Spectrum compensated by $k^{1.17}$ for compensated by 1.17 vegetation index (red) with bands 1 (orange), 2 (green) so that flat curves indicate $k^{1.17}$ spectra. Note the greatly expanded scale, the fluctuations are small. c) Spectrum compensated by $k^{1.17}$ soil moisture (bottom, blue) with bands 6, 7.

**Fig. 4.** a) Structure functions for bands 1, 2 6, 7. Moments of order 0.1, 0.3, ..1.9 are shown top to bottom, along with fits over the scaling range (8 pixels up to 256). $\Delta x$ is the number of pixels of the “lag” and $M = \langle \Delta v^q \rangle$. b) Vegetation index left, soil moisture index right, single scale surrogate. c) Vegetation index left, soil moisture index right, scale by scale surrogate structure functions.

**Fig. 5.** a) Structure function exponent, vegetation index (blue) with bands 1, 2. b) Soil moisture (bottom), with bands 6, 7. c) The multiscale $\xi(q)$ for vegetation index (blue), soil moisture (red).

**Fig. 6.** Trace moments using twenty values of $\lambda$ per order of magnitude in scale: a) vegetation index, b) soil moisture index.
Fig. 7. Trace moment determination of $K(q)$ for: a) vegetation index (purple) and soil moisture index (green). b) vegetation index (purple) and bands 1, 2 (orange, green respectively). c) soil moisture index (green) and blue, purple, bands 6, 7.

Fig. 8. The codimenison function for: a) the 7 bands. b) the vegetation index (red) and bands 1, 3 (yellow, green respectively). c) the soil moisture index (bottom) and channels 6, 7 (purple, magenta respectively). The reference line has slope 6.6.

Fig. 9. a) Vegetation index (blue), bands 1, 2 (orange, green). b) Soil moisture (bottom), bands 6, 7.

Fig. 10. Residues $r(q)$ for: a) the vegetation index (blue) and bands 1, 2 (orange, green). b) soil moisture (bottom) with bands 6, 7 (nearly superposed, top).

Fig. 11. A multifractal simulation with the parameters: a) $H=0.18$, $C_1=0.05$, $\alpha=1.9$. This simulation illustrates both the type of variability expected, but also the effect of a small amount of anistotropy; here the same at all scales. b) $H=0.18$, $C_1=0.05$, $\alpha=1.9$. This simulation illustrates the effect of a small amount of differential anistotropy; with the direction and elongation of structures slowly changing with scale.
Figure 1.
Figure 2.
Figure 3
Figure 4.
Figure 5.
Figure 6.
Figure 7.
Figure 8.
Figure 9.
Figure 10.

(a) Graph showing $\log_{10} q$ vs. $\log_{10} (q\xi'(0) - \xi(q))$.

(b) Graph showing $\log_{10} q$ vs. $\log_{10} (q\xi'(0) - \xi(q))$. 
Figure 11.