10

Probabilistic Multifractals and Negative Dimensions

Ashvin B. Chhabra and K.R. Sreenivasan

Abstract

We propose that negative dimensions can be best understood using the concept of level-independent multiplier distributions and show that, by utilising them, one can extract the positive and negative parts of the $f(\alpha)$ function with exponentially less work than by using conventional boxcounting methods. When the underlying multiplicative structure is not known, both methods of computing negative dimensions can give spurious results at finite resolution. Applications to fully developed turbulence are discussed briefly.

1 Introduction

Fractal and multifractal concepts [1, 2, 3, 4, 5] are now widely used in the study of nonlinear systems. The specific idea of decomposing a singular measure into interwoven sets of singularity strengths α, with dimension $f(\alpha)$ [6, 7], has turned out to be a useful and compact way of describing the scaling properties of such a measure.

However, the $f(\alpha)$ analysis of measures constructed from experiments (e.g. the dissipation field of turbulence) yields two surprising results. First, the analysis of long data sets yields negative numbers for the $f(\alpha)$ of sets of extremal iso-α values [8, 9]. Second, although the data show unambiguous scaling (often over several decades), the actual exponent or dimension fluctuates from sample to sample by an amount greater than the (least square) error bars on any one sample [10, 11]. This seems to contradict the implicit assumption that unambiguous scaling indicates a well-defined and convergent $f(\alpha)$ function.

The purpose of this paper is two-fold: First, we wish to provide a simple way of understanding the above observations using the concept of level-independent multiplier distributions. This has been discussed briefly in [12, 8]. Second, we show that the positive and negative parts of the $f(\alpha)$ function can be computed by suitably manipulating these multiplier distributions. We will demonstrate that, for probabilistic multifractals (those generated

by random multiplicative processes), this method (hereafter referred to as the multiplier method) is significantly more accurate and requires exponentially less work compared to conventional boxcounting methods. Finally, for energy dissipation in the atmospheric surface layer, the $f(\alpha)$ deduced from the multiplier distribution is compared with previous measurements [8] by conventional boxcounting methods.

In the traditional multifractal formalism, one starts with a singular measure that may, for example, be the amount of dissipation in various regions of the dissipation field of fully developed turbulence [10, 13, 14], or the density of points in regions of phase space [7]. Cover the entire region with boxes of size ϵ and denote the integrated probability in the i^{th} box of size ϵ as $P_i(\epsilon)$. Then a singularity strength α_i can be defined by

$$P_i(\epsilon) \sim \epsilon^{\alpha_i}. \tag{1}$$

Denote $N(\alpha)$ as the number of boxes where P_i has singularity strength between α and $\alpha + d\alpha$. Then $f(\alpha)$ can be loosely defined [7] as the fractal dimension of the set of boxes with singularity strength α [4] and written as

$$N(\alpha) \sim \epsilon^{-f(\alpha)}. \tag{2}$$

This formalism leads [6, 7] to the description of a multifractal measure in terms of interwoven sets of Hausdorff dimension $f(\alpha)$ and singularity strength α. In practice, to compute $f(\alpha)$ directly, one must take into account the ϵ-dependent prefactors in Eq. 2. The canonical way of doing so is described in Ref. [15].

An equivalent way of computing $f[\alpha(q)]$ is to use the partition function and evaluate the average value of $< \tau(q) >$ via

$$Z(q) = \sum_i P_i^q(\epsilon) = \epsilon^{<\tau(q)>}, \tag{3}$$

and then use the relation

$$f(q) = q < \alpha(q) > - < \tau(q) > . \tag{4}$$

The refinement process of a fractal measure can be mapped onto a cascading process where each interval contracts into smaller pieces and thus divides the total measure among them according to some process. The scaling properties of measures created by simple models of deterministic cascading processes have been shown to provide good approximations to the scaling properties of the *low order moments* of various intermittent fields in turbulence [13]. It is believed that for cases like turbulence, a more realistic model for such cascading process would be probabilistic, where multipliers are picked randomly from some probability distribution reflecting the underlying physics. We will elaborate on this point to show that such models can incorporate sample to sample fluctuations as well as negative dimensions, both of which have been observed in turbulence.

2 Negative dimensions in random multiplicative processes

Consider the scaling properties of measures arising from deterministic systems e.g. period doubling. Every observed α value will by definition occur at least once and thus $f(\alpha)$, which is the logarithm of the number of times an α value occurs, will always be a non-negative function. In addition, to within some small numerical error, the $f(\alpha)$ function should be identical for two different realisations (of the measure) of the same size.

On the other hand, consider measures created by finite realisations (samples) of a random multiplicative process (hereafter referred to as RMP [4]). The $f(\alpha)$ curve will fluctuate from sample to sample depending on the particular collection of multipliers picked from the probability distribution. One now has two distinct possibilities regarding the averaging procedure. First, one can average the exponents $f(\alpha)$ and α from each sample (quenched averaging). Since in any single sample, every *observed* value of α will occur at least once and hence have a positive dimension, the resultant $f(\alpha)$ curve will always be non-negative. The second procedure is to average the partition function (annealed averaging). One adopts the view that $f(\alpha)$ is the logarithm of a histogram, and defines the dimension [16, 17] by

$$f(\alpha) \sim -\frac{\log < N(\alpha) >}{\log(\epsilon)} \tag{5}$$

where the averages are arithmetic (not Boltzmann), taken over various samples. This procedure has been called supersampling [18]. Now consider values of α that occur rarely e.g. less often than one per typical sample. The dimension assigned to these values by Eq. 5 would be negative. One physical example would be randomly oriented one-dimensional cuts through three-dimensional turbulent dissipation fields [10, 13, 14]. Events occurring in three-dimensional space with low probability would be missed in any given one-dimensional cut. However, they can be recovered by sampling over many such cuts and using Eq. 5. This enables one to gather information about higher-dimensional spaces from low-dimensional measurements. Mandelbrot [18] thus stressed that the analysis of single samples suppresses valuable information about these rare events.

On the other hand Cates and Witten [17] recognised that, in principle, the negative $f(\alpha)$ should be computable from a single sample. To this end they suggested breaking up the measure into smaller pieces, normalising each such sample and then supersampling or averaging the new (sub)partition functions (with the correspondingly reduced scaling range). This reasoning is motivated by the presumed self-similarity of the partition function. Thus, both ways of observing negative dimensions involve supersampling of either different realisations or parts of the same realisation. One would, however, like to understand the occurrence of negative dimensions

from a microscopic point of view i.e., from the underlying multiplicative or refinement process, just as we do for the positive dimensions [7].

The method to be described is based on the view that scaling properties reflecting self-similarity in the measure can be described by a distribution of multipliers that are level-independent [4, 7, 18]. These multipliers define how the measure in a given piece will rearrange into smaller pieces. For stochastic or random multiplicative processes (including randomly oriented lower-dimensional cuts of deterministic processes) all the scaling properties of the measure, including $f(\alpha)$ and sample to sample fluctuations can be understood in terms of the properties of the probability distribution of level-independent multipliers. The multiplier distribution in stochastic systems is the natural analog of the scaling function [19] in deterministic systems. For the latter, the scaling function contains information about the level to level contraction ratios (multipliers) and, in addition, organises them correctly in time. For RMP, there is no natural ordering in time, but the multipliers are characterised by their value as well as by the probability with which they occur.

To understand the relationship between the probability distribution of the multipliers $P(M)$ and $f(\alpha)$, consider a binary RMP where at every level of refinement an interval breaks up into two equal pieces, but the measure is distributed in the ratio M and $1-M$ where M is either 0.7 or $1-0.7=0.3$ [13]. If we assign the larger ratio to any one piece randomly, then we have two rules $[0.7, 0.3]$ and $[0.3, 0.7]$, which are applied with equal probability. If the process proceeds to n levels, the redistributed measure will consist of 2^n pieces and one can compute its $f(\alpha)$ curve. Clearly the left extreme of the $f(\alpha)$ curve ($\alpha = \frac{\log(0.7)}{\log(0.5)} = 0.514...$) will be given by the box containing the string of multipliers $0.7, 0.7, 0.7....n$ times. Since, in any sample, such a string will occur exactly once (the probability of such a piece is 2^{-n} and the number of pieces is 2^n), the dimension of this iso-α set will be $log(1) = 0$. Similarly the iso-α set corresponding to the string of multipliers $0.3, 0.3, 0.3...$ will also occur exactly once and have a dimension of zero. All the other strings will occur more often and thus the entire $f(\alpha)$ curve will be positive. Consider a simple generalisation [20] of the binomial measure, where now we have four sets of multipliers $[0.7, 0.3]$, $[0.3, 0.7]$, $[0.8, 0.2]$, and $[0.2, 0.8]$, which are chosen with equal probability. Then the smallest singularity strength ($\alpha = \frac{\log(0.8)}{\log(0.5)} = 0.321...$) will correspond to the box containing the multipliers $0.8, 0.8, 0.8....n$ times. However, such a box will occur with probability $(\frac{1}{4})^n$. Since there are 2^n boxes per sample, one expects such a singularity strength to be observed only once every $(2^n 4^{-n})^{-1} = 2^n$ samples. Using Eq. 5, one finds that the dimension of the singular set ($\alpha = 0.321...$) is $\frac{\log(2^{-n})}{\log(2^n)} = -1$. The $f(\alpha)$ function corresponding to this probability distribution now ranges from -1 to 1, with the number $f(\alpha)$ quantifying the relative frequency of observing a singularity strength α in a given number of samples of finite size [18].

Consider now a more general process [21, 4] where the multipliers are randomly picked from a given distribution $P(M)$. To derive a general relation between the $f(\alpha)$ function defined by Eq. 5 and the distribution of multipliers, average the partition function over K samples (supersampling) of equal size so that

$$< \tau(q) >= \frac{\log < Z(q) >}{\log(\epsilon)} = \frac{\log[(\frac{1}{K}) \sum_{j=1}^{K} \sum_{i=1}^{a^n} P_{ij}^q(\epsilon)]}{\log(\epsilon)} \tag{6}$$

where $\epsilon = a^{-n}$ (2^{-n} for binary cascades) and P_{ij} is the measure in the i^{th} box of the j^{th} sample. But, due to the self-similarity of the measure, this relation should hold at any n so *we put* $n = 1$. We can do this as we are dealing with a model with no level to level correlations. Thus we have a collection of K sets of boxes, where the measure in any one box is simply a multiplier picked randomly from $P(M)$ (subject to the constraint of conservation of the measure). Denoting the multiplier by M, and remembering that for $n = 1$ the measure in the i^{th} box is $P_{i1} = M_i$, we can write

$$< \tau(q) >= -D_0 - \frac{\log(< M^q >)}{\log(a)}, \tag{7}$$

and correspondingly

$$< \alpha(q) >=< \frac{\partial \tau(q)}{\partial q} >= -\frac{< M^q \log(M) >}{< M^q > \log(a)} \tag{8}$$

with $f(q)$ given by Eq. 4. Note that the averages are over the distribution of the multipliers $P(M)$.

The explanation for the occurrence of negative dimensions is a little more complicated here than for the binomial measure. Now we have multipliers picked randomly from a distribution and multiplied together to create an effective value of α. In the binomial measure the strings containing an equal number of 0.7 and 0.3's would behave like a string of average multiplier value of $\sqrt{0.7 * 0.3}$. The logarithm of the number of ways that such strings can occur is the dimension of that iso-α set. Such an α value and the dimension of the corresponding iso-α set can also be calculated by using the parameter q. This is done by evaluating the partition function according to Eq. 3 at some fixed q to compute $\tau(q)$, taking the derivative of the partition function with respect to q to compute $\alpha(q)$ and, finally, by computing $f(q)$ using Eq. 4. Similarly, the various random values of multipliers from $P(M)$ also multiply to produce different α values. The number of different ways they can do so depends on the multiplier distribution, and the logarithm of this number suitably normalised is the dimension of that iso-α set. Once again $\alpha(q)$ and $f(q)$ can be evaluated by using Eqs. 7, 8 and 4.

These equations provide a recipe for relating the distribution of multipliers with the scaling properties of the measure. The problem of computing

the positive and negative parts of the $f(\alpha)$ function is thus reduced to the problem of computing $P(M)$. We will now demonstrate that the $f(\alpha)$ curve computed using the multiplier method converges exponentially faster to the asymptotic $f(\alpha)$ curve than that computed from conventional boxcounting methods using Eqs. 3, 4 and 5.

3 Advantages of the multiplier method over conventional boxcounting

For convenience we will elucidate our arguments with a binary RMP. The simplest way of computing $P(M)$ for such a process is to cover a measure at the n^{th} stage of refinement, with boxes of size $2^{-(n-1)}$, compute P_i $(i = 1, 2, 3, ...2^{(n-1)})$, then subdivide each of these boxes in two pieces and compute the ratios of the measures in the original box to any one of the two subdivided boxes. Each subdivided box will give a value for M, and using the entire measure one can compute $P(M)$. Clearly the computation of $P(M)$ is helped by considering samples at more levels of refinement (increasing n), for at the n^{th} level, we have 2^n realisations of M with which to construct $P(M)$. One gets better statistics by averaging $P(M)$ over different levels in addition to averaging over different samples (i.e. supersampling).

Now, in the conventional boxcounting method described by Eq. 6, given a single sample at the n^{th} level (consisting of 2^n pieces for a binary process), one would see only those α values that had a probability greater than or equal to 2^{-n}. However, this α value comes from a string of n multipliers, each of which must (on the average) be picked with a probability of at least $1/2$. (In general, for a process that subdivides a piece into a smaller pieces at each level of refinement, one would only see α values corresponding to string of n multipliers, each of which (on the average) is picked with a probability of $1/a$.) To see α values consisting of strings of multipliers of lower probability, the currently used procedure is to supersample [18, 8]. In this procedure, to observe α values which occur with a probability of, say, 4^{-n} one would need $\frac{4^n}{2^n} = 2^n$ samples; that is, as one refines the measure more and more, i.e increases n, one needs an *exponentially increasing* number of samples to see the same α value in that ensemble of samples. Thus one understands the statement [18] that any increase in the level of the cascade must be accompanied by an exponential increase in the number of samples to achieve the same supersampling effect. A mathematically transparent way of understanding this statement is to notice from Eq. 2 that the number of occurrences of an iso-α set with a negative dimension will decrease as $\epsilon \to 0$. Setting $\epsilon = a^{-n}$, one notices that the number of such α values will decrease exponentially with the level of refinement of the measure. This statement, although correct, is paradoxical. If one is interested in describing a measure, it stands to reason that refining it should lead to better information about its scaling properties. However, following the supersampling procedure one does increasingly worse as the level of refinement increases. The attempt to resolve this apparent paradox is what has led us to emphasize the multiplier method.

The multiplier method described by Eqs. 7, 8 and 4 assumes that the scaling properties of the measure arise from the repeated composition of multipliers from the same distribution. This means that, for each value of α, there exists a value of the multiplier M^* which, when composed n times, would produce the same α value. That is,

$$< \alpha >= \lim_{n\to\infty} \sum_{j=1}^{n} \frac{\log(M_j)}{\log(2^{-n})} = -\frac{\log(M^*)}{\log(2)}. \tag{9}$$

The probability of choosing the multiplier M^* is related to the dimension of the iso-α set by

$$< f(M^*) >= -\lim_{\epsilon\to 0} \frac{\log < N_\epsilon(M^*) >}{\log(\epsilon)} = 1 + \frac{\log(P(M^*))}{\log(2)}, \tag{10}$$

where $N_\epsilon(M^*)$ is the number of times a string with the average multiplier M^* occurs at resolution ϵ. Note that $P(M^*)$ is a scale-invariant multiplier distribution that is derived from $P(M)$ but different from it.

In order to compute α and $f(\alpha)$ from the multiplier distribution, one makes use of the parameter q (see Eqs. 7 and 8). As q in Eq. 7 moves from $-\infty$ to ∞, different multipliers ranging from M_{min} to M_{max} get accentuated, thus reproducing the entire $f(\alpha)$ curve. From Eq. 10 one can see that $< f(M^*) >$ is negative for a binary RMP if $P(M^*) < 1/2$. Now if one increases the number of levels n (i.e. sample size), then one can better approximate $P(M)$. In particular, with this method we will be able to detect any multiplier with a probability of more than 2^{-n}. Thus as we increase the number of levels in the cascade, the multiplier method gets better by computing $P(M)$ to a precision of 2^{-n}, in contrast to conventional boxcounting which needs exponentially larger number samples to maintain its precision. Even at a fixed level n, a single sample of boxcounting will see only α values corresponding to $P(M^*) > 1/2$, while the method of multipliers will pick up α values corresponding to $P(M) > (1/2)^n$. Thus, we expect the multiplier method to be correspondingly more accurate for computations of positive dimensions as well. It is because of these two improvements that the multiplier method requires exponentially less work and is commensurately more accurate.

To fix these ideas, consider a binary RMP, where the multipliers are chosen from a uniform distribution [18] (hereafter referred to as a uniform RMP). One can analytically compute its α and $f(\alpha)$ as

$$\alpha(q) = \frac{1}{(q+1)\log(2)} \tag{11}$$

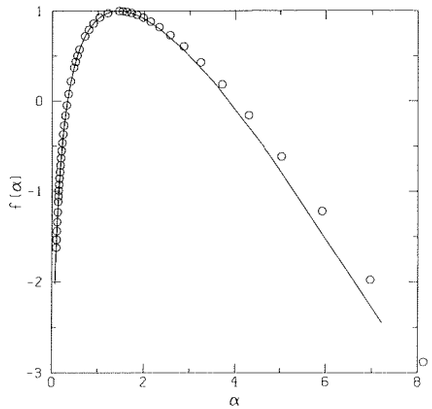

FIGURE 1. Shows that one can extract both the positive and negative parts of the $f(\alpha)$ curve from a single sample. The sample is generated from a binary cascade (of 15 levels) where multipliers were chosen randomly from a uniform distribution (uniform RMP). The solid line is the exact $f(\alpha)$ curve. The circles are from the multiplier method with $-1 < q < 15$.

and

$$f(q) = 1 + \frac{q}{(q+1)\log(2)} - \frac{\log(q+1)}{\log(2)} \tag{12}$$

i.e, the $f(\alpha)$ curve goes all the way down to $-\infty$ with α ranging from 0 to ∞ [22]. We apply the multiplier method on a single sample (of the uniform RMP) at $n = 15$. Fig. 1 shows that it is indeed possible to compute both the positive and the negative parts of the $f(\alpha)$ curve from a single sample. Fig. 2 shows the convergence of the exponent $f(q = 5)$ to its exact value as one refines the box size. Since the box size can only shrink to a value where each box would contain just one point, to improve accuracy one should increase the number of levels in the cascade. Also, to decrease local fluctuations, one can calculate the exponent by averaging multiplier distributions from several different box sizes.

The most stringent way of comparing supersampling using boxcounting and the multiplier method is to fix both the level of the cascade in each sample (n) and the number of samples in the ensemble. Let $n = 12$ and consider 32 samples i.e. 32,768 points in all. Fig. 3 compares these two methods with the known theoretical $f(\alpha)$ function for the uniform RMP. Clearly, the multiplier method is capable of yielding far smaller dimensions with greater accuracy. If one now increased the level of refinement of the measure to, say, $n = 15$ but kept the number of samples (32) the same, for reasons mentioned earlier, we would expect the results of conventional boxcounting to get worse. This is demonstrated in Fig. 4, where the two arrows demarcate the minimum dimensions found by the boxcounting method (using $-1 < q < 15$). (The multiplier method does not suffer from this defect, and improves as the number of levels is increased: as more statistics are gathered, exponents from higher and higher q values become more reliable as demonstrated in Fig. 2.)

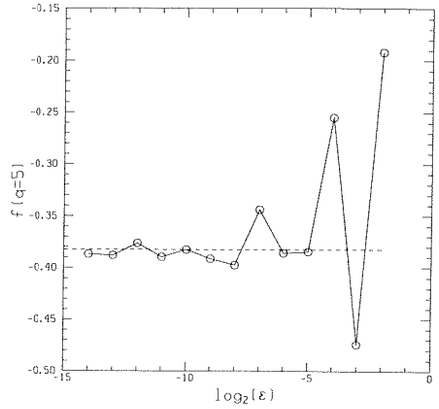

FIGURE 2. The quantity $f(q = 5)$ using the method of multipliers and varying the box size for the same multiplicative process (uniform RMP) as in Fig. 1. The dashed line is the exact value.

Let us discuss the case where one does not understand how to partition the measure correctly and covers the measure arbitrarily with boxes of uniform size. Here one has a trade-off between the number of errors caused by improper boxing (which increases as the number of boxes) and the magnitude of the finite-size errors caused (which decreases as the box size becomes smaller). In the multiplier method with uniform partitioning, the obvious course is to go to more levels of refinement. In doing so the errors due to improper boxing decrease. One also has the advantage that the increased number of boxes results in correspondingly improved statistics for computing multiplier averages. On the other hand, since the number of boxes increases so do the number of errors. Thus there is a tradeoff and one may have to go to very fine resolution (a large number of levels) to produce accurate results. This may not always be possible, especially if one is dealing with experimental data. Fig. 5a shows the results of using the multiplier method on the uniform RMP (for a single sample at $n = 15$) with multipliers computed by comparing measures from boxes one third their size. Since the underlying structure of the cascade process is binary, one expects convergence problems and spurious results. Comparing this with Fig. 1 one notes, however, that the fit is surprisingly good. There is some error in the negative q region (right-half of the curve corresponding to large α values), but the positive q region (left-half of the curve corresponding to small α values) seems, if anything, better at finding the negative dimensions. This fortuitous result comes from improper boxing, as demonstrated in Fig. 5b,

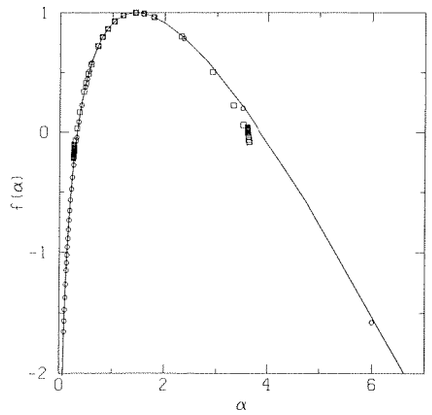

FIGURE 3. Comparison between boxcounting (squares) and the multiplier method (circles) at $n = 12$ for the uniform RMP described in Fig 1. Both methods have been supersampled (averaged) over 32 different realisations.

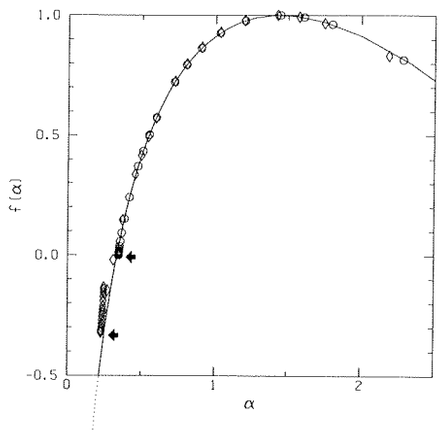

FIGURE 4. Illustrates the degradation boxcounting results with increasing refinement level. Diamonds represent data from cascades with $n = 10$, while the circles are for $n = 15$ (for the uniform RMP described in Fig. 1). The data have been averaged over 32 samples in both cases.

where we have performed the same improper boxing (with base 3) on the generalised binomial measure discussed earlier in the text. From theoretical arguments we know that the minimum dimension for this measure is -1. Note in Fig. 5b both the existence of spurious dimensions (which are less than -1) and the increased error for the negative q region. One reason for the increase in error is that the uniform RMP arises from a smooth distribution of multipliers so improper boxing computes a slightly shifted distribution with largely similar scaling properties. This is not true for the generalised binomial measure whose multiplier distribution consists of four delta functions. Improper boxing here replaces the true distribution with a poorly overlapping smooth one. The spurious negative dimensions arise from the scaling properties of the non-overlapping region which also contaminates the negative q portion (large α portion). (One however might expect that multiplier distributions arising in nature would have smooth multiplier distributions.) Thus the results of improper boxing vary depending on the distribution as well as the level under consideration. It follows that, while attempting to compute negative dimensions for experimental data with an unknown multiplicative structure, it is highly advisable to use different bases for the computation of the multiplier distribution and check the convergence of $f(\alpha)$ with the number of levels in each sample. In addition, one might also average data over several samples.

What about the effects of improper boxing in the supersampling method? Here one runs into an even more serious problem. As one increases the number of levels (to reduce errors due to improper boxing) one must supersample over an exponentially larger number of samples to observe the same negative dimensions. However, this has the consequence of increasing the number of errors exponentially (due to improper boxing), which in turn can easily be mistaken for real events. Let us note, generally, that the idea of negative dimensions is to describe rarely occurring events. Boxcounting with boxes of arbitrary size creates errors, which in turn can be mistaken for precisely the events we are seeking to quantify. The reason that these spurious values (errors) scale with box size (just like the real exponents), is that their number is proportional to that of the boxes, which in turn increases exponentially with increasing levels of refinement. Thus the spurious values also increase exponentially with some smaller exponent. The occurrence of such spurious dimensions for boxcounting is shown in Fig. 6a and Fig. 6b, where the 10,000 samples of the generalised binomial measure (at $n = 5$) have been boxcounted using several different box sizes. The canonical method for computing $f(\alpha)$ and α directly [15] has been employed, and the entropy and energy of the measure have been plotted with increasing box size. The slopes yield $f(q = 25)$ and $\alpha(q = 25)$ respectively. In both figures the lower solid line is the correct exponent (to within a percent) obtained from the least squares fit using box sizes of the form 2^{-n}. The upper solid line is the least square fit obtained from box sizes other that 2^{-n}. These data from incorrect box sizes also scale but with spurious

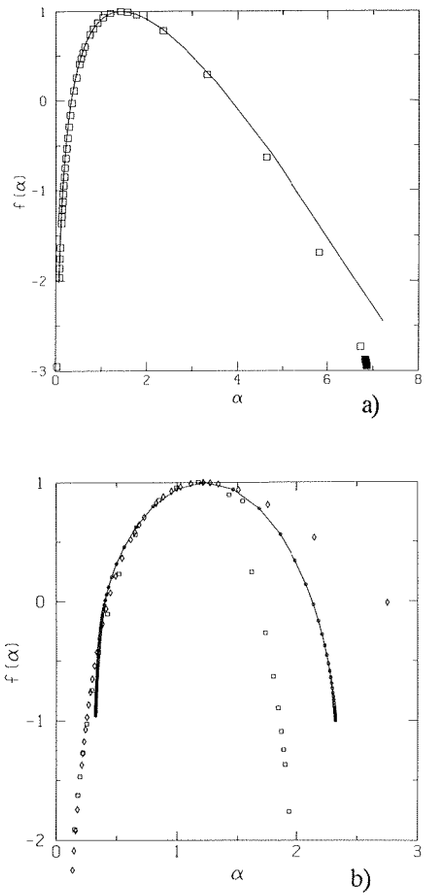

FIGURE 5. (a) Demonstrates the effect of using the wrong base for the multiplier method on a single sample of the uniform RMP at $n = 15$. The multipliers were computed by computing the ratio of the measure in boxes with those in boxes a third of the original size (base 3). The figure demonstrates that the computed $f(\alpha)$ is still reasonable. The solid line shows the analytical solution.(b) Demonstrates the effect of using the wrong base (base 3) for the multiplier method on a single sample of the generalised binomial measure at $n = 10$ (squares) and $n = 15$ (diamonds). Notice the spurious dimensions ($f < -1$) and the large error in the negative q region. The solid line is the theoretical curve.

exponents. Here improper boxing yields $f(q = 25) = -2.24$ which is clearly incorrect. The dashed line in the two figures is the least square fit to all the data (from proper and improper boxing) and is incorrect for both f and α.

4 Applications to turbulence

We will now apply the multiplier method for the determination of the negative dimensions [23] for the distribution of energy dissipation in the atmospheric surface layer several meters above the ground [10, 8].

To obtain negative dimensions in the atmosphere using the supersampling method, one needs an enormous amount of data. For example, to be able to observe a dimension of -2 in the atmospheric dissipation field one needs to distinguish multipliers that have a probability of occurrence $1/8$. Following Ref. [8] we estimate an integral length scale to consist of $\sim 10^4$ data points (sampled at about 6000 Hz). Thus assuming a binary cascade we can estimate from the appropriate Reynolds number that $n \sim 12$ and solve $(1/8)^{12} * (\text{ number of samples }) * (2^{12}) \sim 1$ which gives us an estimate of about ten million samples. Remembering that each sample is about 10^4 points we arrive at an estimate that one needs roughly 10^{11} data points which at the sampling rate of 6000 Hz. would require several years of data acquisition alone. So, we conclude that supersampling ill-suited for the purpose of measuring negative dimensions in atmospheric (or other high Reynolds number) flows.

On the other hand, the measurement of their scaling properties is rather important in order to be able to make statements about universality. In addition, the examination of the multiplier distribution itself may be quite useful in understanding the underlying fractal structure of turbulence.

Figure 7a shows the probability distribution of the multipliers for the energy dissipation at several levels of the cascade process in the inertial range. This was obtained by assuming that a random multiplicative binary cascade in one dimension models the scaling properties of the one dimensional signal. In addition it was assumed that box-averaging the energy dissipation duplicated the splitting process in reverse. To compute the multiplier distribution, the data (a component of the energy dissipation [10, 8]) were divided into bins of m points each. Each bin was then subdivided into two bins of $m/2$ points each and the ratio of the the dissipation contained in the smaller bin to that of the larger bin was computed. Since a conservative binary cascade was assumed, the multipliers from the two bins must add to unity and hence the resulting multiplier distribution is symmetrical. Extension to cascade of bases other than binary is trivial and has been done, but will not be disussed here. The multiplier distributions shown here were obtained from a record length of 409,600 points of the atmospheric dissipation field obtained by hot-wire measurements. In spite of the scatter in the data, it appears that there is rough self-similarity in the cascade process.

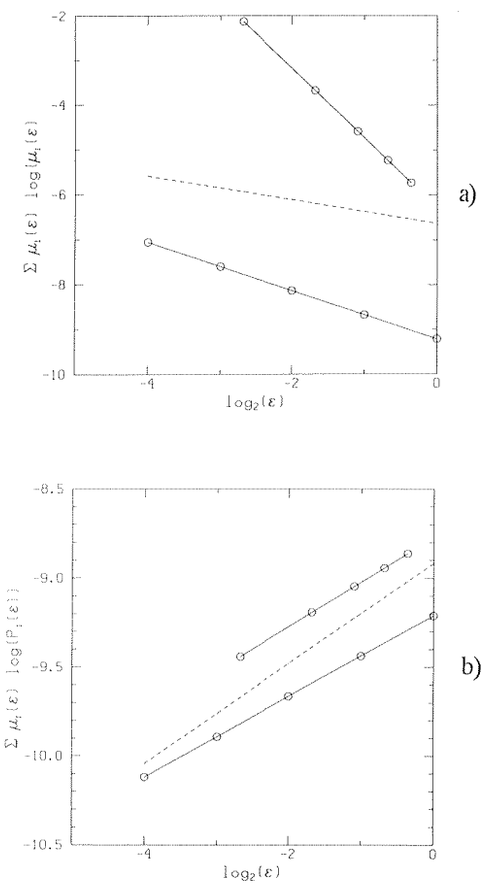

FIGURE 6. (a) Spurious scaling of negative dimensions that should not exist, arising from improper boxcounting of a generalised binomial measure at $n = 5$ averaged over 10^5 samples. Plotted is the entropy of the measure, whose slope with respect to different box sizes yields $f(q = 25)$. The lower solid line is the least square fit to data obtained from using boxes of 2^{-n} and is the correct value within a percent. The upper solid line is a least square fit obtained from box sizes other that 2^{-n} (boxes consisted of $5, 10, 15, 20, 25$ points each). The data from these box sizes also scale but with incorrect exponents. The dashed line is the least square fit to all the data (from proper and improper boxing) and also gives incorrect results. (b) The same as Fig. 6a but plots the internal energy of the measure to compute $\alpha(q = 25)$.

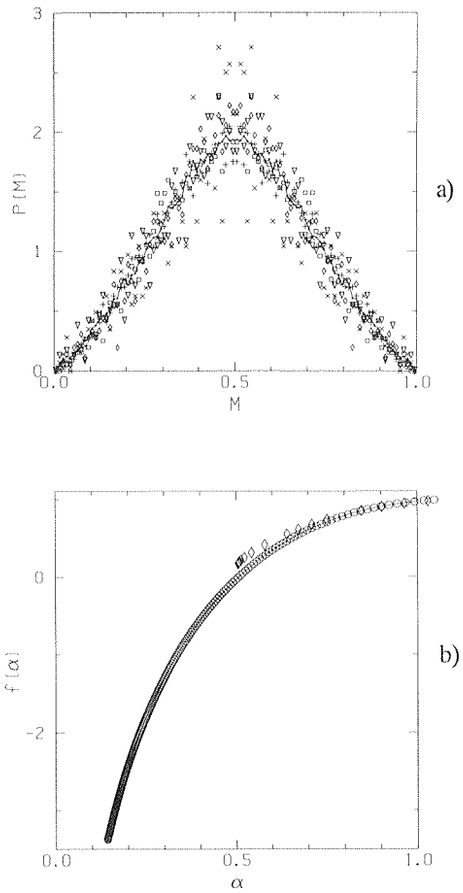

FIGURE 7. (a) The multiplier distribution $P(M)$ obtained by assuming a binary cascading process on the data (409,600 points) from the dissipation field of an atmospheric surface layer. Different symbols correspond to different levels of coarse-graining of the data. The symbol + represents multipliers determined by comparing the total dissipation in boxes of 50 points each with those containing 100 data points. The other symbols are as follows $(\square)100 : 200, (\diamond)200 : 400, (\nabla)300 : 600, (\times)500 : 1000$. The joined solid circles are the mean distribution. The symmetry about 0.5 is a consequence of assuming a conservative binary cascade. (b) Comparison between conventional boxcounting (combined with supersampling) (diamonds) and the multiplier method (joined circles) for the dissipation field of fully developed turbulence in the atmospheric surface layer using $409,600$ data points.

In addition Fig. 7a shows the mean multiplier distribution (continuous line). From this mean distribution, we obtain the $f(\alpha)$ curve as discussed in the earlier sections. For the present, we show in Fig. 7b only the singular part of the curve ($\alpha < 1$). Box-counting methods using the same length of data yields only positive dimensions (diamonds), whereas both the positive and negative dimensions (circles) are easily obtained using the multiplier method. Furthermore, the negative dimensions obtained for this atmospheric flow are in good agreement with the negative dimensions obtained by supersampling data from laboratory flows at lower Reynolds numbers [8]. (Note that supersampling becomes a feasible method at lower Reynolds numbers; even so, Meneveau & Sreenivasan [8] had to use record lengths containing ten million points to obtain reliable results.) We conclude that the multiplier method is much more economical for computing negative dimensions, and that the concept of universality in the multiplicative process is a reasonable one for turbulence.

5 Conclusions

We have discussed a simple way of computing the scaling properties of fractal measures arising from probabilistic processes. The only assumption concerns the existence of a scale-invariant probability distribution of multipliers, which in any case is always necessary for the existence of self-similarity. In cases where the underlying multiplicative process is understood, the multiplier method requires exponentially less work than the boxcounting method. Where one does not know how to partition the measure, the reader is strongly warned about the pitfalls of quantifying such events. However, if there exists other information that allows one to check against spurious scaling, the accuracy of the multiplier method will improve as the refinement gets finer. On the other hand, conventional boxcounting methods are inherently doomed in the search for such rare events.

In this paper, we have not touched upon the important case where level to level correlations exist. One then has to compute these correlations and incorporate them into the relevant equations for calculating dimensions. Alternatively one can increase the base of the process (thus effectively increasing the amount of coarse-graining per level) to decrease these correlations.

We hope that this paper will motivate examination of such scale invariant multiplier distributions whenever one observes sample to sample fluctuations in the dimension of objects or measures - as indeed one does in turbulence and growth models like DLA. The application of these ideas to turbulence has been discussed briefly under the assumption that a binary multiplicative process occurs. More details will be discussed elsewhere.

Acknowledgments

We thank C.J. Evertsz, R.V. Jensen, P.W. Jones, B.B. Mandelbrot and N. Read for useful discussions.

References

[1] B.B. Mandelbrot. *The Fractal Geometry of Nature.* W. H. Freeman & Co., New York, 1982.

[2] J. P. Eckmann and D. Ruelle. *Rev. Mod. Phys.* **57**, 3–617 (1985).

[3] G. Paladin and A. Vulpiani. *Physics Reports*, **156**, 147 (1987).

[4] B.B. Mandelbrot. *J. Fluid Mech.*, **62**, 331 (1974).

[5] H.G.E. Hentschel and I. Procaccia. *Physica*, **8D**, 435 (1983).

[6] U. Frisch and G. Parisi. In *Turbulence and Predictability of Geophysical Fluid Dynamics and Climate Dynamics.* (Eds. M. Ghil, R. Benzi, and G. Parisi). North-Holland, New York, 1985, page 84.

[7] T.C. Halsey, M.H. Jensen, L.P. Kadanoff, I. Procaccia, and B.I. Shraiman. *Phys. Rev. A*, **33**, 1141 (1986).

[8] C. Meneveau and K.R. Sreenivasan. *To appear in the Journal of Fluid Mechanics*, 1990.

[9] The measurements of negative dimensions reported in [8] were motivated by conversation with Benoit Mandelbrot.

[10] C. Meneveau and K.R. Sreenivasan. *Nucl. Phys. B (Proc. Suppl.)*, **2**, 49 (1987).

[11] A.B. Chabra and K.R. Sreenivasan. *In preparation*, 1990.

[12] B.B. Mandelbrot. *STATPHYS 17: Proc. of the Rio de Janeiro meeting.* North-Holland, Amsterdam, 1989.

[13] C. Meneveau and K.R. Sreenivasan. *Phys. Rev. Lett.*, **59**, 797 (1987).

[14] R.R. Prasad, C. Meneveau, and K.R. Sreenivasan. *Phys. Rev. Lett.*, **61**, 74 (1988).

[15] A.B. Chabra and R.V. Jensen. *Phys. Rev. Lett.*, **62**, 1327 (1989).

[16] B.B. Mandelbrot. *J. Stat. Phys.*, **34**, 895 (1984).

[17] M.E. Cates and T.A. Witten. *Phys. Rev. A*, **35**, 1809 (1987).

[18] B.B. Mandelbrot. *Fractals: Proc. of the Erice meeting.* (Ed. L. Piettronero). Plenum, New York, 1989.

[19] M.J. Feigenbaum. *J. Stat. Phys.*, **25**, 669 (1978).

[20] C.J. Evertsz. *Laplacian Fractals* [PhD. Thesis]. The Cheese Press, Edam, The Netherlands, 1989. We thank C.J. Evertsz for bringing to our attention this analytically tractable example.

[21] E.A. Novikov. *P.M.M.*, **35**, 266 (1971).

[22] In this example $\alpha \to \infty$ as $q \to -1$. Thus the entire $f(\alpha)$ curve is reproduced by letting q vary from ∞ to -1.

[23] One should in principle allow for the possibility of completely deterministic processes in turbulence, in which case the need for negative dimensions disappears; so far, however, nobody has been able to determine one such. The closest attempt, which yields results in good agreement with the positive part of the $f(\alpha)$ curve, is given in [13].