
Volume 6, issue 3, September 2009
ISSN 1476-945X
ELSEVIER
Special Section: Fractal Modeling and Scaling in Natural Systems
ECOLOGICAL COMPLEXITY
An International Journal on
BIOCOMPLEXITY IN THE ENVIRONMENT AND THEORETICAL ECOLOGY
Editor-in-Chief:
Bai-Lian (Larry) Li

Ecological Complexity 6 (2009) 246–253

Contents lists available at ScienceDirect

Ecological Complexity

journal homepage: www.elsevier.com/locate/ecocom

On the fractal modelling of biomass distributions: An application to size class in fisheries

Carlos García-Gutiérrez [a], Miguel Ángel Martín [a], José-Manuel Rey [b,*]

[a] Department of Applied Mathematical for Agricultural Engineering, E.T.S.I. Agrónomos, Technical University of Madrid, 28040 Madrid, Spain
[b] Department of Economic Analysis, Universidad Complutense, 28223 Madrid, Spain

ARTICLE INFO

Article history:
Available online 26 June 2009

Keywords:
Biomass distribution
Fractal cascade
Balanced entropy

ABSTRACT

Size structure is a fundamental property in an ecosystem. We explore the potential of fractal models to describe biomass distribution within a community across a range of body-weight size. We propose a consistent method selecting a self-similar cascade generating a fractal distribution that replicates field data within a range of scales. Admissible cascades constructed from biomass distribution data are filtered out using the non-parametric Kolmogorov–Smirnov test. A specific analysis is then performed to detect the biomass spreading across a suitable size scales. This is done by computing a parameter (the balanced entropy index) that evaluates evenness in mass splitting. This method is successfully applied to replicate the distribution of fish biomass obtained from a fisheries dataset. The proposed fractal description of the community size structure of the sampled marine ecosystem may prove useful when analyzing key ecological processes which involve all individuals in a community regardless of their species.

1. Introduction

The statistical description and modelling of the size structure and biomass distributions of an ecosystem are fundamental issues in ecology. Energetic considerations provide evidence of their significant role (Peters, 1983; Lurié and Wagensberg, 1983), in particular in the celebrated allometric scaling of metabolic rate with respect to adult body mass (Kleiber, 1932). This result was later verified for a wide range of body sizes (Hemmingsen, 1960) and it was recently extended to biomass production and growth rates in plants (Nikkas and Enquist, 2001; Damuth, 2001). Damuth (1981, 1987) – considering a variety of species – showed that population abundance (N) changes allometrically with body mass M, i.e. $N \propto M^{-x}$. Further studies supported evidence of this behaviour across a wide body size range (Peters and Raelson, 1984; Marquet et al., 1990; Tokeshi, 1993).

Key properties and processes of an ecosystem, such as bioaccumulation, energy flow, or nutrient processing, depend on the size structure of the community, and in particular on its biomass distribution (Peters, 1983). It is consequently important to design models and methods describing and simulating biomass distributions.

There are many distributions in nature that display scaling features like those described above. In soil sciences, Turcotte's law claims that the number $N(R)$ of particles which have diameters larger than a characteristic diameter R, follows the scaling rule $N(R) \sim R^{-D}$ (Turcotte, 1986). In geography, Korcak (1938) found that the distribution of island areas in the Aegean sea verifies a similar power law, where $N(R)$ refers in their case to the number of islands of area greater than R. Also, the number of arthropods versus body length (Morse et al., 1985) or the number of mammals and other species versus their body mass (Damuth, 1981) verify similar power laws. All such distributions follow a general rule, essentially expressed by the fact that there is a huge profusion of elements small in size while large elements are relatively scarce. In fact, they are all versions of Pareto's law, originally found to rule income distributions (Pareto, 1897).

The examples above all refer to a law concerning mass distributions and their relationship to size. The words "mass" and "size" have different meanings in different contexts. In this paper, "size" refers to body weight and "mass" means biomass. The total biomass is spread among individuals with a characteristic body size. The sizes are arranged in increasing order so that the whole body weight range is represented by an interval $I = [a,b]$. Information about mass distribution is usually limited. It is obtained from field data as mass proportions $p_1, p_2, \ldots, p_n$ ($\sum_i p_i = 1$) on a set of size classes partitioning the size range, i.e. $I_1 = [a,a_1]$, $I_2 = [a_1,a_2], \ldots, I_n = [a_{n-1},b]$, with $a < a_1 < \ldots < a_{n-1} < b$, so that class I_i carries a fraction p_i of total mass.

---

* Corresponding author. Tel.: +34 91 394 2407; fax: +34 91 394 2561.
E-mail address: j-man@ccee.ucm.es (J.-M. Rey).

1476-945X/$ – see front matter © 2009 Elsevier B.V. All rights reserved.
doi:10.1016/j.ecocom.2009.05.006

These kinds of Pareto's law have been considered by Mandelbrot (1982) as hints of a scale invariance or a fractal behaviour that he claimed to be ubiquitous in nature. Enormous research effort in many scientific areas has been devoted to check the fractal nature of shapes, distributions, or processes, matching Mandelbrot's intuition.

Since the 1980s, fractal geometry has been used to describe and parameterize structures with scale invariant features which are assumed to be the result of a more or less explicit iteration process acting across a range of scales. The term fractal or self-similar is loosely attached to objects, distributions or processes when they have some scale invariant features which are commonly revealed by associated power laws. In some cases, the exponents of the power law correspond to fractal dimensions.

A principle of maximum entropy has been claimed to lie behind the typical scaling laws associated with fractal objects (Pastor-Satorras and Wagensberg, 1998). A form of entropy maximization was proposed as a modelling principle for the biomass distribution of an ecological population (Lurié and Wagensberg, 1983). Building on both ideas, we propose here a fractal model for the biomass distribution in an ecosystem. Specifically, we consider the distribution of biomass across body-weight sizes in a marine ecosystem. While biomass is a significant variable when representing size distribution, other measures are also used to describe size distribution in a marine community, e.g. mass, length, equivalent spherical diameter, etc. (Peters, 1983). The analysis of their relationships – notably allometric – is itself a research issue (see e.g. Safran, 1992).

Fishery data has been chosen in order to test whether the proposed model can replicate biomass distributional data. There are some reasons for this choice. First, hauls are naturally expected to produce a more faithful random sample of the marine community—obviously within a biomass range constrained by, e.g. the size of fishing meshes. Notice, however, that a full ecosystem is not surveyed by the dataset; fisheries data is not representative of all species in a community sampled by a selective fishing gear. Secondly, the biomass dataset analyzed here is the same as the one tested by Lurié and Wagensberg (1983) and thus there is a chance that a fractal model might fit the distribution, as we claim in this paper.

Our starting point is that biomass distributions can be explained as the result of an asymmetrical biomass spreading operating within a range of scales. In the absence of any other information, it is natural to claim that biomass is recursively distributed within smaller scales of individual body-weight sizes while matching the distribution of biomass at larger scales. This is consistent with the idea of scale invariance. This approach was first considered in soil sciences to model soil particle size distributions (Martín and Taguas, 1998). The assumption that disparity in mass spreading occurs systematically within a range of scales is the main feature of a self-similar distribution intensively studied in fractal geometry (Falconer, 1990).

Trinomial cascades are used in this paper to construct self-similar distributions. These are mass fragmentation processes that repeatedly subdivide the (bio)mass carried by a size class interval into three mass fractions supported by size subclasses partitioning the size interval. Both the rule of mass spreading at different scales and the partitioning scheme are preserved across scales. A so-called self-similar distribution is obtained in the limit process (Falconer, 1990).

Since biomass range has enormous variations, the biomass on a size interval and the size range covered by the interval may typically be far from being proportional. This feature can also be observed in models of biomass distribution (Lurié and Wagensberg, 1983). As a result, natural partitions of the size domain used to report biomass fractions are composed of intervals of different lengths. In turn, useful biomass fragmentation schemes consist of spreading comparable mass fractions on unequal size intervals, in contrast to the most popular fractal constructions (e.g. Cantor measures or binomial cascades) spreading unequal masses on intervals of the same length. The trinomial cascades proposed here for biomass-size distribution are defined on asymmetric size partitions.

The choice of ternary partitions keeps the need of data to a minimum. However, it will produce a consistent replication of biomass distribution data. It is a remarkable fact that a complex distribution, such as the biomass-size distribution, can be replicated from the (bio)mass fractions supported by three size classes only. The model also supplies synthetic distributional data which when unavailable must be valid when the model fits the field distribution.

In order to select the best model, we simulate self-similar cascades that match biomass data at a very crude level, i.e. agreeing on the biomass carried by three basic size classes. The Kolmogorov–Smirnov (KS) statistical test is used to compare sample and synthetic distributions. The KS test discards many simulated distributions. Mass splitting is then analyzed across the partitioning schemes defined by the surviving cascades. This is done using the balanced entropy (*BE*) index introduced in Martín et al. (2005). The *BE* analysis determines the self-similar cascade that best replicates the process of biomass splitting. The *BE* index is defined to deal with non-uniform size partitions. While *BE* analysis has been applied to test the continuity of field soil particle size distributions (Rey et al., 2006), its potential use to test self-similarity has not been considered so far.

2. Materials and methods

2.1. Self-similar cascades

A self-similar cascade is an iterative process that involves partitioning intervals into subintervals according to a fixed rule and correspondingly splitting the (probability) mass supported by the intervals following a different (maybe random) rule. The process typically defines a limiting (multi)fractal distribution, i.e. such that the mass concentration (or Hölder) exponent varies continuously through the mass support interval (Falconer, 1990). The limit object is a highly intricate distribution that is generated in a very simple manner.

More specifically, a self-similar (random) cascade model can be defined as follows. Let $I = [a,b]$ be whole size interval partitioned in, say, three subintervals $I_1 = [a,s_1]$, $I_2 = [s_1,s_2]$, $I_3 = [s_2,b]$, $a < s_1 < s_2 < b$. These are the basic intervals of the cascade. Let $p_1 = P(I_1)$, $p_2 = P(I_2)$ and $p_3 = P(I_3)$ be the mass proportions of the intervals I_1, I_2 and I_3, respectively, where $p_1 + p_2 + p_3 = 1$. The size fragmentation is defined by the linear mappings φ_i (similarities) that transform I into I_i for $i = 1, 2, 3$; that is, $I_i = \varphi_i(I)$. These three transformations are applied to interval I to get the first generation of size intervals, I_i, then applied to those three to obtain the second generation of 9 size intervals, $I_{i_1 i_2} \subset I_{i_1}$ where $I_{i_1 i_2} = \varphi_{i_1}(\varphi_{i_2}(I)),\ i_1, i_2 = 1, 2, 3$ and so on. After k repetitions, the kth generation of size intervals $I_{i_1 i_2 \ldots i_k}$ is obtained. As k increases the kth generation of size intervals gives an increasingly finer partition of I.

The mass is iteratively distributed at each stage among new sized intervals according to the following scheme: the mass supported by an interval $I_{i_1 i_2 \ldots i_k, j}$ is obtained by

$$P(I_{i_1 i_2 \ldots i_k, j}) = P(I_{i_1 i_2 \ldots i_k}) \cdot V_j,$$

where V_j are independent random variables with mean p_j (for instance, following a normal distribution of mean p_j and variance σ_j^2). If $\sigma_j = 0$, the cascade is deterministic and the limiting distribution – self-similar – is the prototype model of a multifractal measure (Falconer, 1990). In this paper, deterministic cascades are considered as initial models for biomass distributions. They are shown to replicate biomass data. In the deterministic case, the mass of an interval $I_{i_1 i_2 \ldots i_k}$ is given by $P(I_{i_1 i_2 \ldots i_k}) = p_{i_1}\ p_{i_2}\ \ldots\ p_{i_k}$. This produces a (pre-fractal) probability distribution at each stage of the construction.

2.2. Fishery data

Data from the ATLOR VI expedition (Manríquez and Rucabado, 1976) has been processed to test the hypothesis proposed in this paper, namely that fractal cascades may explain biomass data. The same raw data considered by Lurié and Wagensberg (1983) has been processed here. The ATLOR VI report consists of raw data from 53 hauls in an upwelling zone at the northwest coast of Africa obtained during 24 days on the basis of several catches a day. We processed the ATLOR expedition data to obtain weight data of fishes of approximately 280,000 individuals from 131 different species, which are basically pelagic and demersal. While the whole ecosystem may have not been sampled, it is plausible to argue that the ATLOR data provides a fair picture of the fish community within a scale range. The dataset contains 28 different species, each one contributing at least 1% to the total biomass. Table 1 shows the taxonomic description of this sub-sample together with the biomass contributions by species. This sub-sample accounts for about 85% of the total biomass, so that 15% of the biomass data is covered by the majority of species (103), each one representing less than 1% of total biomass. Since we are interested in describing biomass distribution structure within a range of body mass sizes regardless of species, we do not discard those 103 species from the target fish biomass distribution.

Raw data entries are obtained for the different body weight measurements (sizes) annotated in the fishery report, ranging from 0 kg to 12.4 kg. The range of fish body weights considered in this paper is in between 0 kg and 0.2 kg. This choice is suitable for comparisons with Lurié and Wagensberg (1983) who limited their study to the 0–0.085 kg size range. It turns out that only 2.5% of the sample population have weights outside the 0–0.2 kg range. Instead, 9% of the population is left out of the Lurié and Wagensberg 0–0.085 kg range. These weight measurements are not equally spaced, as can be expected from this type of mass-size distributions. Indeed, if the pelagic community biomass is binned into logarithmic size classes the amount of biomass in each is claimed to be approximately constant over a size ranging from bacteria to whales (Sheldon et al., 1972, 1977). The body size classes are naturally defined by their weight measurement data figures. The size partition so obtained is made up by 161 size classes: $J_1 = [0{,}0.01]$, $J_2 = [0.01{,}0.02], \ldots, J_{158} = [0.194{,}0.197]$, $J_{159} = [0.197{,}0.198]$, $J_{160} = [0.198{,}0.199]$, $J_{161} = [0.199{,}0.2]$. Because of asymmetry in the measurement data, the associated size partition is not uniform. The total biomass contributed by each class J_k is computed by adding the body masses of individuals whose weights fall inside class J_k—regardless of their species. Finally, biomass class contents are normalized in order to define a probabilistic distribution $m(J_k)$, $k = 1, \ldots, 161$, referred to as the ATLOR distribution for the sequel. When necessary, it will be assumed that biomass is uniformly distributed inside the size intervals J_k.

2.3. Model testing

In order to test whether the self-similar cascade model can replicate fishery data, we construct all possible trinomial cascades available from the initial biomass data. A cascade is constructed by setting $\{\varphi_1, \varphi_2, \varphi_3;\ p_1, p_2, p_3\}$, as explained above. In fact, selecting the size limits s_1, s_2 for the first generation of size intervals defines not only the similarities but also the probabilities of any admissible cascade since it is defined to coincide with the biomasses of the three basic size intervals. Any admissible cascade is thus generated according to the following scheme:

Table 1
Biomass-species description of the sample distribution processed in the article from the ATLORVI dataset (only species contributing at least 1% of the biomass are listed).

Species	Average body-mass (kg)	Biomass percentage
Argyrosomus regius	6.728	1.124
Aspitrigla obscura	0.099	2.062
Buglossidium luteum	0.005	0.966
Capros aper	0.029	1.643
Conger conger	3.573	1.892
Dentex macrophthalmus	0.032	1.402
Diplodus annularis	0.023	8.498
Llepidotrigla cavillone	0.016	2.133
Loligo forbesi	0.179	3.577
Loligo vulgaris	0.059	1.354
Macroramphosus gracilis	0.009	1.015
Octopus vulgaris	0.917	5.819
Pagellus coupei	0.024	1.045
Pagellus erythinus	0.044	3.047
Plectorhinchus mediterraneus	0.217	2.142
Pomadasys bennetti	0.109	1.861
Raja naevus	1.029	1.873
Raja undulada	1.463	1.101
Sardina pilchardus	0.026	10.176
Scyliorhinus canícula	0.382	3.422
Sepia officinalis	0.168	1.144
Soleidae miscellanea atl-6	0.079	1.294
Spondyliosoma cantharus	0.008	1.707
Trachinus draco	0.094	1.002
Trachurus trachurus	0.048	15.335
Zeus faber	0.855	3.098

1. Two size limits $a < s_1 < s_2 < b$ are selected among the 162 available weight data points. This choice defines the three basic size intervals $I_1 = [a{,}s_1]$, $I_2 = [s_1{,}s_2]$, $I_3 = [s_2{,}b]$ partitioning the whole size range $[a{,}b]$. In the ATLOR distribution, we have $a = 0$ and $b = 0.2$.
2. The (bio)mass contents $p_1 = m(I_1)$, $p_2 = m(I_2)$ and $p_3 = m(I_3)$ of the basic intervals are obtained from the ATLOR data by aggregating the biomass contents of the data size intervals J_k contained in each basic interval I_i. That is, $p_i = \sum_{J_k \subset I_i} m(J_k)$, for $i = 1, 2, 3$.
3. The size fragmentation mapping $\varphi_i(x)$ is defined by $\varphi_i(a) = s_{i-1}$, $\varphi_i(b) = s_i$, $i = 1, 2, 3$ (in the understanding that $a = s_0$, $b = s_0$). Notice that the contraction ratio of φ_i is given by $(s_i - s_{i-1})/(b - a)$.
4. A self-similar distribution is defined at any required size scale from the mass-replicating formula, $p(I_{i_1 i_2 \ldots i_k}) = p_{i_1}\ p_{i_2}\ \ldots\ p_{i_k}$.

From the 161 weight data points that are available, 12,720 possible trinomial cascades can be defined. All of them were implemented as described above. By #2 above we see by construction that the biomass fractions of the basic size intervals I_i for each simulated distribution coincide with those provided by actual data. The remaining synthetic biomass data, defined for any arbitrary size class, is used to compare the simulated distributions with the ATLOR distribution.

The Kolmogorov–Smirnov (KS) test gives an important statistical criterion for comparing any two one-dimensional probability distributions. The two-sample version of the KS test (DeGroot and Schervish, 2002) is used here to test whether the ATLOR distribution differs from each simulated cascade. The two-sample KS test is a general non-parametric method widely used in comparisons of any two samples. The KS statistic is defined by

$D_n = \sup_x |M(x) - P_n(x)|$, where $M(x)$, $0 \leq x \leq 0.2$, is the ATLOR cumulative distribution defined out of the 162 size data points of the registered sample, and $P_n(x)$ is a self-similar cumulative distribution defined from $n = 3^k$ intermediate size point obtained from running the cascade up to the kth stage. The null hypothesis that the two underlying distributions are the same is rejected at level α if $D_n > \left(\sqrt{(162+n)/162n}\right)k_\alpha = \underline{k}_\alpha$, where k_α is found from the Kolmogorov distribution satisfying $Pr(K \leq k_\alpha) = 1 - \alpha$.

2.4. The BE index

While the KS test is a critical test to decide whether two distributions statistically differ, it does not supply information on the process of mass spreading across sizes, which is characteristic of self-similar cascades. The balanced entropy (*BE*) index, that was introduced by Martín et al. (2005) as a generalization of Shannon's entropy (1948) is a suitable parameter to evaluate uniformity in mass fragmentation when partitions are refined (Rey, 2006). The *BE* index is defined out of an underlying mass distribution *p*, and for a given partition $\prod = \{I_j\}_j$ of the size interval *I*, by

$$BE(p, \Pi) = \frac{\sum_j p(I_j) \log p(I_j)}{\sum_j p(I_j) \log r_j},$$

where the r_j's are the relative lengths of intervals in the partition, i.e. $r_j = length(I_j)/length(I)$. Noticed that r_j is given by the contraction ratio of φ_i in the case when the I_i's are the basic intervals of a self-similar cascade. Whenever evident we omit the term "*p*" in the notation above. For the ATLOR data, $length(I) = 0.2$. Notice that, for any partition, $\{r_j\}_j$ forms a probability distribution. The *BE* index offers significant information about mass spreading both at a given scale (defined by a partition) and across different scales. Also, the *BE* index is adapted to deal with non-uniform size partitions.

For a given partition $\prod$, *BE* can be seen as a distance of the mass distribution *p* to the uniform distribution, characterized by $p(I_j) = r_j$. Values of *BE* near 0 mean that *p* is very far from being uniform, whereas values of *BE* near 1 correspond to distributions that are close to satisfying the uniform distribution property $p(I_j) = r_j$.

When two different partitions $\prod$, $\prod'$ are considered such that $\prod'$ is finer than $\prod$ (i.e. each interval in $\prod'$ is either an interval of $\prod$ or a subinterval of one of $\prod$), the difference $BE(\prod) - BE(\prod')$ gives information about how uniform the global mass spreading is when splitting its mass from the intervals in $\prod$ into those defined by $\prod'$. A rise of *BE* is consistent with a more uniform redistribution of mass across $\prod'$ whereas a lowering of *BE* is compatible with a less uniform redistribution. Self-similar mass fragmentation processes produce a constant value of $BE(\prod_j)$, for any partition $\prod_j$ composed of intervals $I_{i_1 i_2 \ldots i_j}$ at the *j*th level of the cascade. The theory concerning the static properties and inter-scale working principles of balanced entropy can be seen in (Martín et al., 2005; Rey, 2006).

2.5. Testing for self-similarity

BE analysis offers a complementary procedure yielding information about the process structuring the mass spreading. In particular it tells us whether the output distribution of a fractal cascade is compatible with the ATLOR distribution within a range of scales. Our purpose is to learn which one of the biomass splittings defined for the self-similar cascades passing the KS test is closer to the ATLOR data.

The method is implemented as follows. For each one of the simulation cascades that passed the KS test, the *BE* indices of the ATLOR distribution are computed for the partitions defined by the cascade up to the $k = 8$ level. The sequence of these *BE* values is compared with that of the *BE* values of the self-similar cascade, which are all given by the same number (in turn coinciding with the fractal (entropy) dimension of the limit distribution (Deliu et al., 1991)). The cascade for which the sum of the absolute values of the differences of the eight *BE* values is minimal is then chosen as the best replication of the ATLOR data. The procedure thus systematically explores the ATLOR mass distribution across each one of the partitioning schemes associated with the admissible cascades choosing the best scheme. This is specified below.

1. Let s_1, s_2, define one of the admissible cascades *P* (that passed the KS test at a level *k* given). Let $\prod_1 = \{I_1, I_2, I_3\}$ be the basic partition associated with *P* and $\prod_j$ ($j = 1, 2, \ldots, k$) be the sequence of partitions generated by the cascade (i.e. $\prod_j$ is made up by the *j*th generation of size intervals $I_{i_1 i_2 \ldots i_j}$).
2. For the ATLOR *m* mass distribution and for each partition $\prod_j$, $j = 1, 2, \ldots, k$, compute the *BE* index to obtain the sequence $BE(j) = BE(m, \prod_j)$, $j = 1, 2, \ldots, k$.
3. Calculate the *BE* index for the self-similar cascade, defined by

$$BE(P) = \frac{\sum_{i=1}^{3} P(I_i) \log P(I_i)}{\sum_{i=1}^{3} P(I_i) \log r_i},$$

and verifying $BE(P, \prod_j) = BE(P)$ for any *j*. Compute the differences $BE(j) - BE(P)$ for $j = 1, 2, \ldots, k$.

4. Determine P^* such that $D(P) = \sum_{j=1}^{k} |BE(j) - BE(P)|$ is minimum.

3. Results

Fig. 1 displays in a log–log plot the number of individuals $N(s)$ – in the full ATLOR dataset – with body weights greater than *s*. The sequence follows a remarkable power-law behaviour, namely $N(s) \sim s^{-12}$ within the whole range of scales. This is the kind of scale invariance claimed to be typically associated with an underlying fractal behaviour.

Species rank–size plotting is a common method to look at underlying mass-size structure of ecological communities (Magurran, 2004). The biomass contributions by each species are arranged according to their rank (from the most to the least important contributions) to obtain the plot in Fig. 2—only the list of ATLOR species contributing at least 0.1% of the total biomass are included in the count (these are 71 species accounting for 98.5% of the total ATLOR biomass data).

The amount of biomass by each species turns out to be an inverse power of its rank. This is a type of Pareto's law which, as explained above, holds for a number of different phenomena and essentially remains unexplained.

All 12,720 possible trinomial cascades that can be defined by choosing two intermediate ATLOR weight data points $s_1 < s_2$ (with $0 = a < s_1 < s_2 < b = 0.2$) have been simulated up to stage $k = 8$. This amounts to generating 6561 intermediate size points within the full size interval, thereby inducing a discretization of the simulated distribution which is much finer than the original data.

For level $\alpha = 0.05$ and stage $k = 8$, the KS test rejected 11,941 simulations. Thus 779 self-similar cascades (6.12% of the total number) passed the KS test at a 95% confidence level. Among those 779 cascades a best self-similar one P^* is chosen according to its performance in terms of the *BE* criteria explained in the preceding section, i.e. P^* minimizes $D(P) = \sum_{j=1}^{k} |BE(j) - BE(P)|$ over the class of cascades that passed the KS test.

Table 2 shows the best choice P^* for $k = 8$. The distribution P^* is defined for the $s_1 = 0.008$, $s_2 = 0.026$ size limits. This corresponds to a distribution of total biomass that assigns 5.53% of it to the weight interval [0,0.008], 23.43% to the interval [0.008,0.026] and 71.04% to the [0.026,0.2] interval. In Fig. 3 we plot the ATLOR and the P^* distributions.

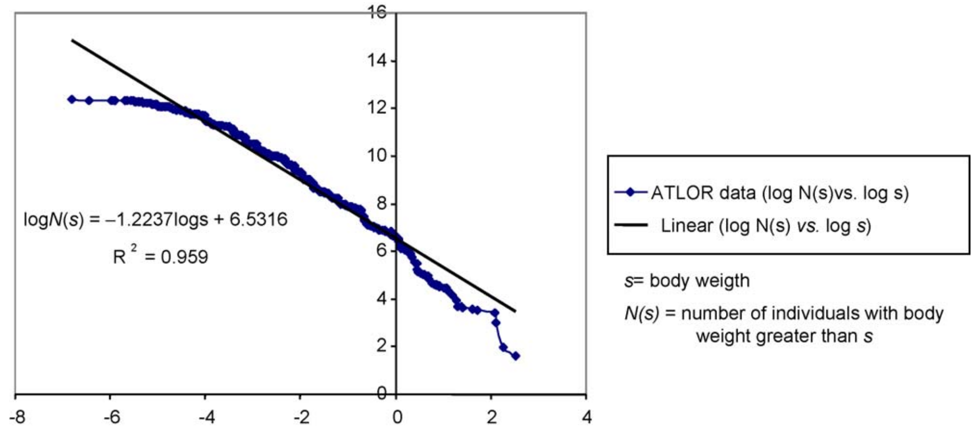

Fig. 1. A Pareto-type law for ATLOR data. Using the full ATLOR dataset, the number of individuals $N(s)$ with body weight greater than s is plotted against s in a log–log plot together with the best linear fit, with shows an excelent coefficient of determination. A Pareto law seems to holds for $N(s)$ which may be seen as a consequence of scale invariance.

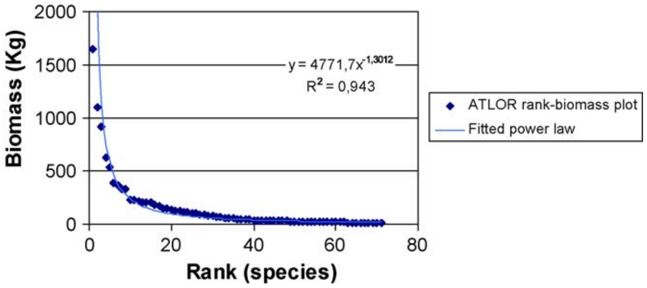

Fig. 2. Rank/biomass plot for ATLOR data. The total biomasses contributed by each species in the ATLOR data are represented from larger to smaller contributions (for species with biomass percentages >0.1%). A remarkable inverse power law dependence (Pareto-type law) is obtained for the species rank–biomass disribution.

4. Discussion

It is remarkable that a preliminary description of the data size structure (see Fig. 1) renders features which are consistent with a fractal genesis of the distribution. Also, Fig. 2 suggests strong evidence in favour of Pareto's law for the ATLOR biomass data, which is not an uncommon structure for the biomass species distribution (Magurran, 2004). This is an interesting result, but not directly related to our approach, since size classes hold together individuals of different species.

At a 95% confidence level, 6.12% of the total number of cascades produces distributions that do not differ statistically from the ATLOR distribution (at the stage $k = 8$ of the construction, i.e. using 6561 intermediate points). It is a significant qualitative finding that synthetic distributions generated from just three size data points (s_1, s_2 and 0.2) can reproduce the ATLOR distribution, which originates from 162 points.

By minimizing $D(P)$, not only does P^* resemble the ATLOR data at the size resolution scale defined by $k = 8$ but it also supplies a self-similar biomass spreading scheme across scales (those defined from $k = 1$ to $k = 7$) that is consistent with that of the ATLOR distribution.

In order to get a fair picture of the minimality property satisfied by P^*, we show in Fig. 4 the differences $BE(j) - BE(P)$ for $j = 1, 2, \ldots, 8$ for the best case ($P = P^*$) and for the worst case, within the family of cascades that passed KS test.

Also plotted in Fig. 4 are the mean, maximum and minimum differences $BE(j) - BE(P)$ among all P's that passed the KS test, along with the difference values for the best P. Notice that the choice of P^* implies that the area between the (◆) line (corresponding to P^*) and the x-axis, is minimal.

The distribution P^* is plotted together with the ATLOR distribution in Fig. 3. The closeness appears particularly fitting for the size range 0–0.85 kg – considered in Lurié and Wagensberg (1983) – probably because the fishery data are richer within this range. By construction, the distribution P^* and the ATLOR distribution assign the same biomass fraction to each the basic intervals [0,0.008], [0.008,0.026], and [0.026,0.2], respectively. The self-similar modelling permits the scanning of the biomass distribution lying inside those intervals at any desired size resolution. Notice that this is achieved at the price of just two generating parameters, namely $s_1 = 0.008$ and $s_2 = 0.026$ (plus the accompanying biomass proportions).

The obtained index value $BE(P^*) = 0.88$ – not too far from unity – indicates that the cascade spreads the mass not too asymmetrically but not uniformly either. In fact, in the limit, the construction generates a self-similar distribution whose fractal dimension is 0.88. This is a mathematical property of the BE index: for a self-similar cascade: $BE(P)$ coincides with the entropy dimension and

Table 2
Best self-similar cascade replicating the ATLOR distribution.

Stage (intermediate weight data points)	Best P^* parameters				KS test (95%)		*BE* analysis	
					$\underline{k}_{0.05}$	*K*-value	$BE(P)$	$D(P)$
$k = 8$ ($n = 6561$)	∏ (kg) Mass	$I_1 = [0,0.008]$ 0.055285	$I_2 = [0.008,0.026]$ 0.234294	$I_3 = [0.026,0.2]$ 0.710421	0.1085	0.1015	0.8833	0.0260

The parameters defining the best cascade model are presented along with the values of the KS statistic and BE analysis. The best self-similar replication P^* of the ATLOR data is defined in two steps. First, self-similar cascades passing the Kolmogorov–Smirnov test (95%) are filtered out among the class of admissible trinomial self-similar cascades (at stage $k = 8$ of their construction). Then P^* is selected within that subclass as the minimizer of $D(P) = \sum_{j=1}^{k} |BE(j) - BE(P)|$ which is a measure of closeness in the way the mass is split in size classes across scales ($BE(j)$ is the balanced entropy at stage j of ATLOR data while $BE(P)$ is the balanced entropy – constant at each stage – of a self-similar cascade P). The winner cascade P^* corresponds to the size limits $s_1 = 0.008$, $s_2 = 0.026$. This amounts to split the total biomass by assigning 5.53% of biomass within weight interval [0,0.008], 23.43% within [0.008,0.026] and 71.04% within [0.026,0.2].

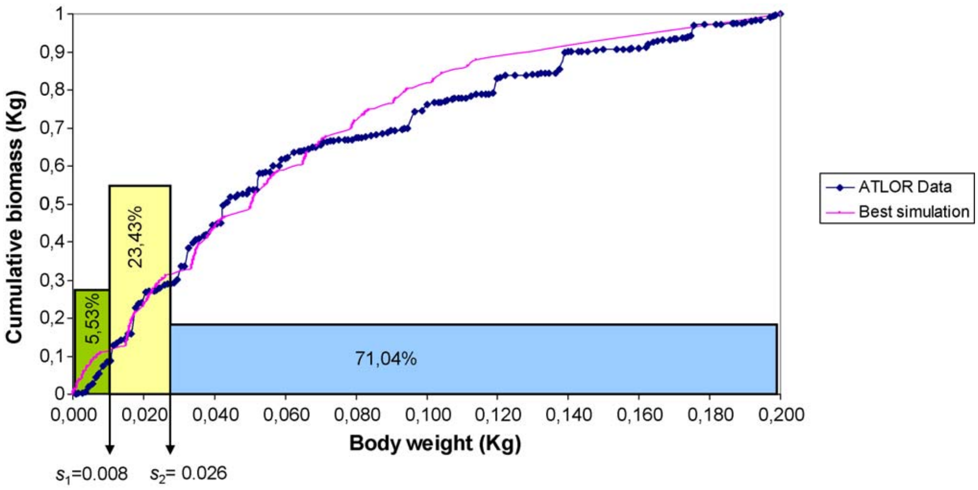

Fig. 3. Cumulative distributions of ATLOR data and of the best simulation P^*. While ATLOR cumulative is obtained from 161 intermediate size data points, the cumulative of P^* is constructed from 6561 size point generated by the cascade (at stage $k = 8$ of the construction). The biomass histogram of P^* on the initial size partition [0,0.008], [0.008,0.026], [0.026,0.2] is also displayed. By construction, this histogram is the same for the ATLOR distribution.

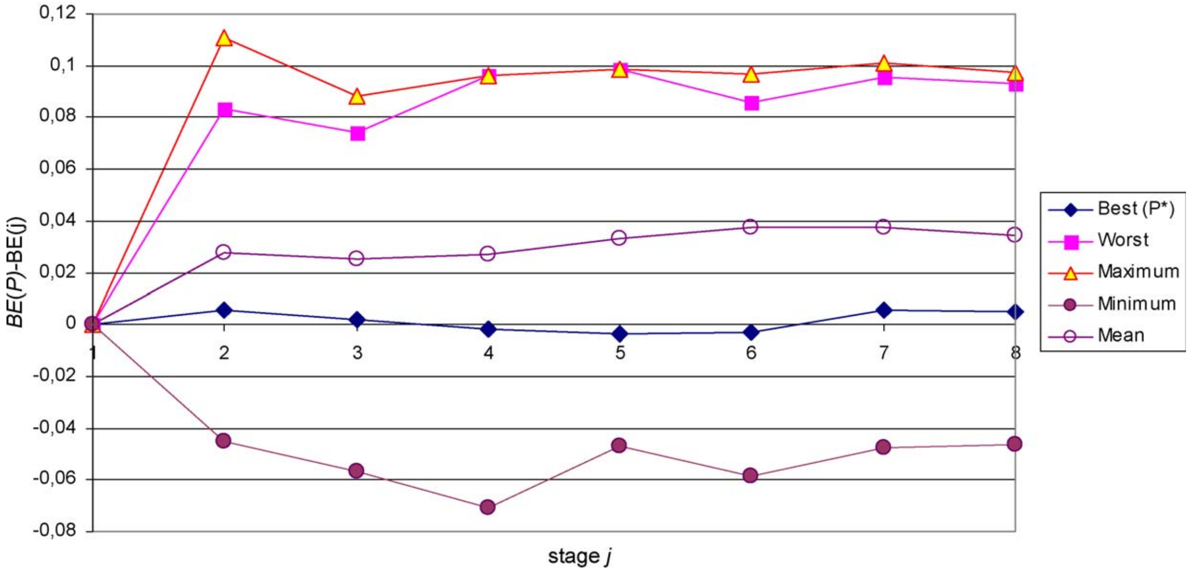

Fig. 4. Balanced entropy (*BE*) analysis for ATLOR distribution (weight size fish biomass). The best self-similar cascade was selected to minimize the sum of absolute values of the differences of *BE* indices of the cascade *BE*(P) and the ATLOR distribution *BE*(j) for $j = 1$–8 among the family of cascades that passed the KS test at stage $j = 8$ (6.12% of possible cascades). The maximum (▲), minimum (●), mean (○), and worst case (■) of the differences $BE(P) - BE(j)$ for each stage from $j = 1$–8 for all cascades P passing the KS test are plotted along with the sequence $BE(P^*) - BE(j)$ (◆) corresponding to the minimizing cascade P^*.

the Hausdorff dimension of the limit distribution P (Deliu et al., 1991). Thus, P^* supplies a pre-fractal distribution within any range of scales (in particular, up to stage $k = 8$) and serves as a true fractal model for the biomass ATLOR distribution.

It seems natural that the obtained self-similar modelling should not depend critically on their constituent parameters, i.e. s_1, s_2 plus the fractions of biomass supported by the size intervals, $p_1^* = m([0, s_1])$, $p_2^* = m([s_1, s_2])$, and $p_3^* = m([s_2, 0.2])$ (m stands for the ATLOR biomass distribution.) Thus, small changes in the parameters should produce models that are still valid, i.e. passing the KS test at a 95% significance and yielding values of $D(P)$ that remain close to $D(P^*)$. Since the *BE* index – in turn $D(P)$ – varies continuously with the masses $P(I_i)$ and sizes r_i, it is clear that small variations in the parameters s_i and/or p_i^* produce similar values of $D(P)$ and thus self-similar cascades splitting the biomass approximately in the same way across scales. This contributes to solve the issue of sensitivity with respect to the *BE* analysis. A similar reasoning may apply with respect to the KS test.

Notice that only the class of deterministic cascades has been considered so far as an initial model for biomass data. Since a deterministic model seems to do the job it appeared unnecessary to look for random model. However, if the deterministic model is to be considered plausible, small random perturbations of the distribution of biomasses defined by P^* should also do the job. Using random cascades like those described above, we have tested whether noisy versions of the deterministic cascade P^* still produce valid simulations. Specifically, at each level of the construction, the random variables V_j in the random scheme are defined to be Gaussian with mean p_i^* and standard deviation φ_j given by a fraction α of p_i^*, for $i = 1, 2$, so that $p_3 = 1 - p_1 - p_2$. We show in Table 3 the results of the KS test for different values of α obtained from an ensemble of 100 realizations of the random cascade for each considered value of α.

As expected, for small values of the perturbation of the biomass fractions ($\alpha = 0.05$), all of the noisy cascades pass the KS test at significance level of 95%. As the perturbation in the biomass

Table 3
Robustness of the ATLOR data self-similar modelling with respect to noisy cascades.

		Deterministic cascade	Noisy cascades			
		$\alpha = 0$	$\alpha = 0.05$	$\alpha = 0.1$	$\alpha = 0.5$	$\alpha = 1$
KS statistic	Mean	0.10147	0.08699	0.10040	0.2281	0.4479
	St. dev.		0.0062079	0.0160266	0.1220566	0.2467389
% of realizations passing KS (95%)			100	78	2	8

Random cascades are considered to check the robustness of the self-similar modelling of ATLOR distribution. The random cascades are noisy versions of the selected deterministic cascade P, defined by perturbing the self-similar mass splitting defined by (p_1^*, p_2^*, p_3^*) at each stage of the construction. Biomass is distributed according to a random triplet (p_1, p_2, p_3), where p_i $(i = 1, 2)$ are Gaussian variables with mean p_i^* and standard deviation given by αp_i^* $(i = 1, 2)$. An ensemble of 100 realizations for each cascade is considered for each value of α. The table shows mean and standard deviation of KS-statistic values and the percentage passing the KS test. For $\alpha = 0.05$, all realizations pass the KS test. As the perturbation increases (given by α) so does the percentage of cascades that are rejected by the KS test.

fractions p_i^* increases, less and less simulations pass the test. For $\alpha = 0.5$ practically none of the realizations pass the test. The results of this exercise confirm the robustness of the method.

The method combines a non-parametric (KS) statistical test, comparing sample and synthetic distributions at a chosen scale, with the (*BE*) analysis of biomass spread across scales given by the partitions defined by the cascades. Notice the difference between the KS test and the *BE* analysis. Whereas the former is concerned with a static property of the generated output distribution and looks at goodness of fit in a distributional sense, the latter allows us to look for the generating process underneath by scanning mass fragmentation within many partition sequences, eventually choosing the best in self-similarity terms.

5. Conclusions

In this paper we address the possibility of determining a fractal genesis for the distribution of biomass across body sizes in an ecological community. A fractal distribution model for biomass distributional data is proposed. The model distribution is generated by a deterministic self-similar cascade which is determined from only two intermediate size data points.

The method is applied to fishery data reporting fish body weight distribution from a large sample of catches in the northwest coast of Africa. It permits us to single out a cascade which mimics the fishery distribution in two different ways, both statistically and constructively. On one hand, the cascade cannot be distinguished statistically from data at a fine size resolution (defined by the cascade partition at a given level of construction). On the other hand, the field data splits the biomass approximately in a self-similar way across a range of scales (also defined by the cascade partitions).

Strictly speaking our proposed method cannot claim to reproduce the distribution of the whole community of a sampled ecosystem. This is mainly due to limitations imposed by data records. On one side, there are obvious constraints in the wide-spread of data size range because of the fishing gear and related sampling procedures. Also, the construction of a cascade requires the specification of a body size interval in which the synthetic distribution is supported. This truncation of the true distribution implies that the processed biomass distribution lacks biomass supported by tiny and huge body sizes and in turn that biomass self-similarity cannot be claimed to hold for the full community. It is remarkable, however, that a cascade can be defined that distributes biomass in a self-similar way across a range of scales in a way that cannot be statistically differentiated from biomass data at a fine scale of body mass sizes.

A main qualitative finding is that fractal distributions may serve as underlying models for biomass-size distributions. It structures biomass diversity within an ecosystem. Once a fractal model has been consistently tested to replicate data, it can be quantitatively exploited to reproduce intermediate distributional values that are not accessible from field observations. In the case of self-similar cascades it is remarkable that the distribution can be scanned at arbitrary intermediate values just from the knowledge of three size intervals along with their biomass content.

Acknowledgements

This work has partially been supported by Plan Nacional de Investigación Científica, Desarrollo e Innovación Tecnológica (I + D + I) (Spain), Ref. AGL2007-28609E and DGUI (Comunidad de Madrid), Technical University of Madrid (UPM), under Ref. M0700204135.

References

Damuth, J., 1981. Population density and body size in mammals. Nature 290, 699–700.

Damuth, J., 1987. Interspecific allometry of population density in mammals and other animals: the independence of body mass and population energy use. Biol. J. Linn. Soc. 31, 193–246.

Damuth, J., 2001. Scaling of growth: plants and animals are not so different. Proc. Natl. Acad. Sci. U.S.A. 98, 2113–2114.

Deliu, A., Geronimo, J.S., Shonkwiler, R., Hardin, D., 1991. Dimensions associated with recurrent self-similar sets. Math. Proc. Cambridge Philos. Soc. 110, 327–336.

DeGroot, M.H., Schervish, M.L., 2002. Probability and Statistics, 3rd ed. Addison–Wesley, New York.

Falconer, K.J., 1990. Fractal Geometry—Mathematical Foundations and Applications. Wiley, New York.

Hemmingsen, A.M., 1960. Energy metabolism as related to body-mass and respiratory surfaces and its evolution. Rep. Steno. Mem. Hosp. 9, 1–110.

Kleiber, M., 1932. Body size and metabolism. Hilgardia 6, 315–353.

Korcak, J., 1938. Deux types fondamentaux de distribution statistique. Bull. de l'Institut International de Statistique III 295–299.

Lurié, D., Wagensberg, J., 1983. On biomass diversity in ecology. Bull. Math. Biol. 45 (2), 287–293.

Magurran, A., 2004. Measuring Biological Diversity. Blackwell.

Mandelbrot, B.B., 1982. The Fractal Geometry of Nature. W.H. Freeman, New York.

Manríquez, M., Rucabado, J., 1976. Datos Informativos No. 1 del Afloramiento N.O: de Africa. Instituto de Investigaciones Pesqueras, Barcelona.

Marquet, P.A., Navarrete, S.A., Castilla, J.C., 1990. Scaling population density to body size in rocky intertidal communities. Science 250, 1125–1127.

Martín, M.A., Taguas, F.J., 1998. Fractal modelling, characterization, and simulation of particle-size distributions in soils. Proc. R. Soc. London Ser. A 454, 1457–1468.

Martín, M.A., Rey, J.-M., Taguas, F.J., 2005. An entropy-based heterogeneity index for mass-distributions in Earth Sciences. Ecol. Model. 182, 221–228.

Morse, D.R., Lawton, J.H., Dodson, M.M., Williamson, M.H., 1985. Fractal dimension of vegetation and the distribution of arthopod body lengths. Nature 314, 731–733.

Nikkas, K.J., Enquist, B.J., 2001. Invariant scaling relationships for interspecific plant biomass production rates and body sizes. Proc. Natl. Acad. Sci. U.S.A. 98, 2922–2927.

Pareto, V., 1897. Cours d'économique politique. McMillan, London.

Pastor-Satorras, R., Wagensberg, J., 1998. The maximun entropy principle and the nature of fractals. Physica A 251, 291–302.

Peters, R.H., 1983. The Ecological Implications of Body Size. Cambridge Univ. Press, Cambridge, UK.

Peters, R.H., Raelson, J.V., 1984. Relations between individual size and mammalian population density. Am. Nat. 124, 498–517.

Rey, J.-M., Montero, E., Martín, M.A., 2006. Inter-scale behavior of balanced entropy for soil texture. Geoderma 134, 415–427.

Rey, J.–M., 2006. Basic quantization properties of balanced entropy. Geoderma 134, 424–427.

Safran, P., 1992. Theoretical analysis of the weight–length relationship in fish juveniles. Mar. Biol. 112, 545–551.

Shannon, C.E., 1948. A mathematical theory of communication. Bell Syst. Technol. J. 27, 379–423 and 623–656.

Sheldon, R.W., Prakash, A., Sutcliffe Jr., H., 1972. The size distribution of particles in the ocean. Limnol. Oceanogr. 17, 327–340.

Sheldon, R.W., Sutcliffe Jr., H., Paranjape, M.A., 1977. Structure of pelagic food chain and relationship between plankton and fish production. J. Fish. Res. Board Can. 34, 2344–2353.

Tokeshi, M., 1993. Species abundance patterns and community structure. Advances in Ecological Research, 24. Academic Press, pp. 111–186.

Turcotte, D.L., 1986. Fractals and fragmentation. J. Geophys. Res. 91, 1921–1926.