

Available online at www.sciencedirect.com



Journal of Marine Systems 69 (2008) 310-327

JOURNAL OF MARINE SYSTEMS

www.elsevier.com/locate/jmarsys

Towards a seascape typology. I. Zipf versus Pareto laws

Laurent Seuront^{a,b,*}, James G. Mitchell^b

^a School of Biological Sciences, Flinders University, GPO Box 2100, Adelaide SA 5001, South Australia, Australia

^b Station Marine de Wimereux, CNRS UMR 8013 ELICO, Université des Sciences et Technologies de Lille, 28 avenue Foch, F-62930 Wimereux, France

Received 11 June 2005; received in revised form 22 December 2005; accepted 11 March 2006 Available online 22 February 2007

Abstract

Two data analysis methods, referred to as the Zipf and Pareto methods, initially introduced in economics and linguistics two centuries ago and subsequently used in a wide range of fields (word frequency in languages and literature, human demographics, finance, city formation, genomics and physics), are described and proposed here as a potential tool to classify space–time patterns in marine ecology. The aim of this paper is, first, to present the theoretical bases of Zipf and Pareto laws, and to demonstrate that they are strictly equivalent. In that way, we provide a one-to-one correspondence between their characteristic exponents and argue that the choice of technique is a matter of convenience. Second, we argue that the appeal of this technique is that it is assumption-free for the distribution of the data and regularity of sampling interval, as well as being extremely easy to implement. Finally, in order to allow marine ecologists to identify and classify any structure in their data sets, we provide a step by step overview of the characteristic shapes expected for Zipf's law for the cases of randomness, power law behavior, power law behavior contaminated by internal and external noise, and competing power laws illustrated on the basis of typical ecological situations such as mixing processes involving non-interacting and interacting species, phytoplankton growth processes and differential grazing by zooplankton.

© 2007 Published by Elsevier B.V.

Keywords: Marine ecosystem; Topology; Patchiness; Power law; Data analysis; Zipf; Pareto

1. Introduction

The investigation of the space-time structure of marine plankton distributions set the priority of the biological oceanographers in the early 20th century, so that by about 1930 many investigators had shown that planktonic organisms are neither uniformly nor randomly distributed in the ocean over a wide range of spatial and temporal scales (Hardy, 1926; Hardy and Gunther, 1935; Hardy, 1939). Despite an impressive body of literature on the subject (e.g. Legendre and Demers, 1984; Mackas et al., 1985; Daly and Smith, 1993), little is known about the space–time structure of biological processes on smaller scales (Seuront et al., 1999, 2002). Considering the increasing awareness of the heterogeneous nature of plankton distributions at decimeter and centimeter scales (Seymour et al., 2000; Waters and Mitchell, 2002; Waters et al., 2003) where the most ecologically relevant processes of viral infection, nutrient uptake, cell division and behavior occur, the ocean should be regarded as a "seascape" (Seuront et al., 2004), and should subsequently be classified by analogy with landscape ecology (e.g.

^{*} Corresponding author. School of Biological Sciences, Flinders University, GPO Box 2100, Adelaide SA 5001, South Australia, Australia. Tel./fax: +61 882015080.

E-mail address: laurent.seuront@flinders.edu.au (L. Seuront).

the complexity of the related data analysis techniques. Since the introduction of ship-board fluorometry (Platt, 1972), satellite data, (Feldman et al., 1984) and, more recently, microscale sensors providing up to 512 measurements per second (Mitchell, 2004; Wolk et al., 2004), marine scientists have had to deal with an ongoing data inundation that prevents an individual from conducting data preparation, reduction, analysis and interpretation. In addition, the more traditional, widely used mathematical descriptors, such as the variance-to-mean ratio (e.g. Taylor, 1961; Frontier, 1972; Downing et al., 1987) have been progressively replaced by numerical techniques more difficult to handle from the conceptual and practical points of view. Thus, spectral analysis is now a standard method in time series analysis since the seminal work of Platt in the early seventies (e.g. Platt and Denman, 1975; Chatfield, 1996) and can be found in most statistical software and programming books (e.g. Press et al., 1992). However, exotic and recently developed techniques such as nonlinear forecasting (Sugihara and May, 1990; Sugihara et al., 1990; Strutton et al., 1996, 1997; Hsieh et al., 2005), fractals (Seuront and Lagadeuc, 1997, 1998; Waters and Mitchell, 2002; Waters et al., 2003), multifractals (Pascual et al., 1995; Seuront et al., 1996a,b, 1999; Lovejoy et al., 2001 Seuront et al., 2002) and wavelets (Dremin et al., 2004; Fisher et al., 2004) are seldom used in marine ecology despite an increasing interest in describing patterns and processes (e.g. Seuront and Strutton, 2004).

There are several potential reasons for the limited applications of such techniques in marine ecology. First, the crelated formalisms, mainly developed and used in the fields of nonlinear dynamical systems and physical sciences, are not straightforward to use and require a large time investment to master. Second, unlike most of the numerical techniques used to analyze spatial data sets and time series, no software is commercially available yet for nonlinear forecasting, fractals, and multifractals. The wavelet packages now provided by some software company or as freeware on the internet require intensive training. Third, the numerical implementations of the related algorithms are challenging and time consuming. In addition, the implicit assumption of a Gaussian distribution is rarely tested and unless one is fortunate, the data set usually requires some massaging (e.g. despiking and detrending) before the analysis can

be performed. Exactly what happens in these processes and which options to choose (e.g. Tukey–Hamming vs. Bartlett's filtering method) requires careful consideration. The intrinsic intermittent properties of most real world patterns and processes are not compatible with the Gaussian hypothesis, and thus with the related techniques such as spectral analysis. In some instances, the most extreme events are the most ecologically meaningful (Seuront et al., 1999). Finally, some data sets are inherently non-stationary, and when there are less than 100 samples or the sampling interval is variable the above methods are unsuitable and require further processing or interpretation, or altogether different software programs and numerical techniques.

The methodological goals of this paper are to introduce the Zipf and Pareto laws (Pareto, 1896; Zipf, 1949) for use in marine ecology because they do not require any assumptions about the distribution of the data set, the regularity of the sampling interval, and are easy to implement. Their simplicity and ease appealed in the pre-computer, pre-electronic calculator era. Since that time the appeal has continued and the process has been widely used in areas such as human demographics, linguistic, genomic and physics, but surprisingly seldom in marine ecology (see Table 1). This paper is the first in a series treating the concept of seascape typology. The papers that follow (Mitchell and Seuront, this issue; Seuront et al., in press) illustrate the applicability of this concept to one- and two-dimensional data sets. The present paper provides a brief description of the theoretical forms of Pareto and Zipf laws, and their few applications to date in marine ecology. We next demonstrate that Pareto and Zipf laws are strictly equivalent, provide a one-to-one correspondence between their characteristic exponents, and argue that the choice of a technique is rather a matter of convenience. In order to allow marine ecologists to identify and classify readily any structure in their data sets, we finally provide a step by step overview of the characteristic shapes (i.e. typology) expected for both laws in case of (i) randomness, (ii) power law behavior, (iii) power law behavior contaminated by internal and external noise, and (iv) competing power laws.

2. Zipt and Pareto laws: theoretical background

2.1. The Zipf law

2.1.1. Background

The Zipf's law, named after the Harvard linguistic professor G.K. Zipf (1902–1950), is the observation that frequency of occurrence of any event, as a function

Table 1 Review of the systems studied using Pareto or Zipf laws in physical, biological and ecological sciences

System	Pareto/Zipf law	Reference
X-ray intensity from solar flares	Pareto	1
Ecosystem models dynamics	Pareto	2
Sand pile dynamics	Pareto	3
Volcanic acoustic emission	Pareto	4
Earthquakes dynamics	Pareto	5-6
Granular pile dynamics	Pareto	7
Himalayan avalanches	Pareto	8
Intensity of "starquakes"	Pareto	9
Evolution model dynamics	Pareto	10-11
Noncoding DNA sequences	Zipf	12
Landscape formation	Pareto	13-14
Sediment deposition in the ocean	Pareto	15
Coding/noncoding DNA sequences	Zipf	16
Word frequencies	Zipf	17
Word frequencies	Zipf	18
Formation of river networks	Pareto	19-20
Rice pile dynamics	Pareto	21
Noncoding DNA sequences	Zipf	22
Percolation process	Zipf	23
Linguistics	Zipf	24
Tropical rainforests dynamics	Pareto	25-27
City formation	Zipf	28
Bird population dynamics	Pareto	29
Aftershock series	Pareto	30
City distribution	Zipf	31
Discrete logistic systems	Pareto	32
Procaryotic protein expression	Zipf	33
Ion channels	Pareto	34
Dynamics of atmospheric flows	Pareto	35
US firm sizes	Pareto	36
Distribution of city populations	Pareto	37
Economics	Pareto	38
Microphytobenthos 2D distribution	Pareto	39
Marine species diversity	Zipf	40-44
Size spectra in aquatic ecology	Pareto	45
Phytoplankton distribution	Zipf	46

1: McHardy and Czerny (1987); 2: Bak et al. (1989); 3: Held et al. (1990); 4: Diodati et al. (1991); 5: Feder and Feder (1991); 6: Olami et al. (1992); 7: Jaeger and Nagel (1992); 8: Noever (1993); 9: Garcia-Pelayo and Morley (1993); 10: Bak and Sneppen (1993); 11: Paczuski et al. (1995); 12: Mantegna et al. (1994); 13: Somfai et al. (1994); 14: Somfai et al. (1995); 15: Rothman et al. (1994); 16: Mantegna et al. (1995); 17: Kanter and Kessler (1995); 18: Cizrók et al. (1995); 19: Rigon et al. (1994); 20: Rinaldo et al. (1996); 21: Frette et al. (1996); 22: Israeloff et al. (1995); 23: Watanabe (1996); 24: Perline (1996); 25: Solé and Manrubia (1995a); 26: Solé and Manrubia (1995b); 27: Manrubia and Solé (1996); 28: Makse et al. (1995); 29: Keitt and Marquet (1996); 30: Correig et al. (1997); 31: Marsili and Zhang (1998); 32: Biham et al. (1998); 33: Ramsden and Vohradsky (1998); 34: Mercik et al. (1999); 35: Joshi and Selvam (1999); 36: Axtell (2001); 37: Malacarne et al. (2002); 38: Burda et al. (2002); 39: Seuront and Spilmont (2002); 40: Margalef (1957); 41: Frontier and Bour (1976); 42: Frontier (1977); 43: Frontier (1985); 44: Frontier (1994); 45: Vidondo et al. (1997); 46: Mitchell (2004).

of the rank *r*, when the rank is determined by the above frequency of occurrence (i.e. from *n* events, the most and less frequent ones will then have ranks r=1 and r=1, respectively). More specifically, Zipf's law states that the frequency f_r of the *r*th largest occurrence of the event is inversely proportional to its rank *r* as:

$$f_r = \frac{f_1}{r} \tag{1}$$

where f_1 is the frequency of the most frequent event in the distribution. This is typically referred to as the Zipf's law. In log–log scales, the Zipfs' law gives a straight line with slope of -1. As in many phenomena large events are rare, but small ones quite common. For example, there are few large earthquakes and avalanches, but many small ones. There are a few words, such as "*the*", "*of*" and "*to*" that occur very frequently, but many which occur rarely, such as "*Zipf*". The generalized Zipf's law is subsequently defined as:

$$f_r = \frac{f_1}{r^{\alpha}} \tag{2}$$

where the log–log plot can be linear with any slope $\boldsymbol{\alpha}.$

Going back to Zipf's original work, Eqs. (1) and (2) can be more generally written as:

$$X_r \, \alpha \, r^{-\alpha} \tag{3}$$

where X_r is the value taken by any random variable relative to its rank r, and $\alpha = 1$ and $\alpha \neq 1$ for the Zipf's and the generalized Zipf's law, respectively. The concept related to X_r is very general and refers without distinction to frequency, length, surface, volume, mass or concentration. Discrete processes such as linguistic or genetic structures would nevertheless still require frequency computations, and thus refer to Eqs. (1) and (2). Alternatively, Eq. (3) can be thought as a more practical alternative that can be directly applied to continuous processes such as phytoplankton distributions (Mitchell, 2004).

2.1.2. Zipf's law in marine ecology

To our knowledge, Zipf's law has been only applied to the description of species diversity (Margalef, 1957; Frontier and Bour, 1976; Frontier, 1977, 1985, 1994), and very recently to the distribution of vertical phytoplankton distribution patterns (Mitchell, 2004).

More specifically, a modified version of the generalized Zipf law (see Eq. (2); Mandelbrot, 1953):

$$f_r = f_1 (r + \phi)^{-\alpha} \tag{4}$$

is useful for describing living communities in aquatic and terrestrial ecosystems (Frontier, 1985, 1994). Thus, in Eq. (4) f_r must be thought as the frequency of the rth species after ranking the species in decreasing order of their frequency. The two parameters α and ϕ are characterizing the species diversity and the evenness of a given community; the diversity is given by $H = -\sum_{i=1}^{N} f_i \log_2 f_i$ and the evenness by $R = H/\log_2 N$, where f_i is the relative frequency of the species i and N the number of species. The formulation of evenness shows that for the same number of species, the diversity is high when species have equivalent probability (high evenness), and low when few species are frequent and other are scarce (low evenness). More specifically, a low value of α means a slow decrease of the species abundance (i.e. a more even distribution of individuals among species), and a high value of α means a rapid decrease of species abundance (i.e. a more heterogeneous distribution). The former and the latter give less and more vertical rank-frequency distributions, hence high and low evenness and diversity. On the other hand, positive values of ϕ result in a greater evenness among the most frequent species, then a higher diversity index. Alternatively, a negative ϕ describes a community marked by the dominance of a few (even one) species and provides a low diversity index and a low evenness. In summary, ϕ and α act upon the diversity and evenness respectively through the niche diversity (i.e. the number of alternatives in each type of previous environmental condition) and through the predictability of ecosystem (probability of the appearance of a species when its environmental conditions are satisfied; Frontier, 1985). Despite appealing and meaningful properties, Eq. (4) has seldom been used in aquatic ecology. Margalef (1957) was the first to fit the Mandelbrot distribution to Mediterranean tintinnids with $\phi = 8.4$ and $\alpha = 4.5$, while Frontier and Bour (1976) and Frontier (1977) estimated $\alpha = 1$ and $\alpha = 2$ for chaetognaths and pteropods, respectively. More recently, Mitchell (2004) demonstrated the applicability of Zipf's law to ocean phytoplankton data sets and subsequently applied Eq. (3) to low and high resolution phytoplankton profiles. For instance, he showed that the Zipf exponents α , estimated for low resolution time series, range from $\alpha = 0.20$ to $\alpha = 0.65$ and exhibit a significant increase below a critical chlorophyll concentration. On the other hand, the exponents α estimated for high resolution vertical profiles of fluorescence have been shown to fluctuate from 0.04 to 0.15. Even if the ecological significance of such preliminary results still needs further investigations (see Mitchell and Seuront, this issue; Seuront et al., in press), it has nevertheless been suggested that rank sizing could be

regarded as a first step in detecting structure and encouraging a higher resolution examination of common phytoplankton distributions.

2.2. The Pareto distribution

2.2.1. Background

The Pareto's law was originally introduced in economics to describe the number of people whose personal incomes exceed a given value (Pareto, 1896). More generally, Pareto's law of any random variable X is described in terms of the cumulative distribution function (CDF):

$$P[X>x] \alpha x^{-\beta} \tag{5}$$

where *x* is a threshold value, and β is the slope of a log–log plot of P[X>x] vs. *x*. Note that Eq. (5) can be equivalently rewritten in terms of the probability density function (PDF) as (Faloutsos et al., 1999):

$$P[X=x]\,\alpha x^{-\gamma} \tag{6}$$

where γ ($\gamma = \beta + 1$) is the slope of a log-log plot of *P* [*X*>*x*] vs. *x*.

2.2.2. The Pareto distribution in marine ecology

Pareto's law has been connected to the biomass size spectrum (Vidondo et al., 1997) and has been used in a slightly different framework related to Self-Organized Criticality only once to characterize microscale patchiness of microphytobenthos biomass (Seuront and Spilmont, 2002). Below a critical biomass, Seuront and Spilmont (2002) found a power law fit ($r^2=0.99$) for microphytobenthos patch concentration, with $\beta=5.31$, i.e. $P[X \ge x] \propto x^{-5.31}$. The ecological relevance of such a finding will be explored thoroughly elsewhere (Mitchell and Seuront, this issue; Seuront et al., in press).

2.3. From Zipf to Pareto laws

Zipf and Pareto laws have often been described as separate power laws (e.g. Faloutsos et al., 1999), having been compared in a paper demonstrating that Zipf's law for the rank statistics is strictly equivalent to a power law distribution of frequencies (Troll and Graben, 1998). This comparison is unfortunately based on complicated mathematical analyses and does not provide any link between the Zipf and Pareto exponents α and β . Such a comparison is nevertheless a crucial prerequisite step to reconcile and compare results that could be obtained using one of these two methods. We demonstrate in a simple manner that Zipf and Pareto laws are strictly equivalent, and subsequently provide a one-to-one correspondence between the exponents α , β , and γ .

Eq. (3), shows that there are kr variables X_r (where k is a constant) greater than or equal to $r^{-\alpha}$. This leads to rewrite Eq. (5) as:

$$P[X > kr^{-\alpha}] \propto r \tag{7}$$

and

$$P[X > X_r] \, \alpha \, X_r^{-1/\alpha} \tag{8}$$

From Eqs. (3), (6) and (8), the relationship between the exponents α , β and γ is given by:

$$\begin{cases} \alpha = \frac{1}{\beta} \\ \gamma = 1 + \frac{1}{\alpha} \end{cases}$$
(9)

As a consequence, the Zipf and Pareto laws can be regarded as equivalent. Specifically, the *x*-axis of the Zipf law is conceptually identical to the *y*-axis of the Pareto law (Eqs. (7) and (8)). The use of one or the other distribution is simply a matter of convenience.

Because of the correspondence between Zipf and Pareto distributions we will, from this point on refer to Zipf's law. In particular, we provide an overview of the characteristic shapes expected for Zipf's law in the cases of (i) pure randomness, (ii) power law behavior, (iii) power law behavior contaminated by internal and external noise, and (iv) competing power laws. In addition, we provide some practical keys for the marine ecologists to avoid misinterpreting Zipf graphs and to conduct a successful analysis.

3. Zipt and Pareto laws: the bridge to applications

3.1. Random processes

Fig. 1 shows the characteristic signatures of five simulated random processes (i.e. white noise) with 100, 500, 1000, 5000 and 10,000 data points in linear and logarithmic plots of Zipf distributions. In linear plots (Fig. 1A), the Zipf law for random noise appears as linear. On log-log plots (Fig. 1B), the simulated random noises do not produce any power law behavior as expected from Eq. (3), but instead a continuous roll off from a horizontal line (i.e. $\alpha \rightarrow \infty$). This is representative of the fact that no



Fig. 1. Linear and log–log Zipf plots of random processes with 100, 500, 1000, 50,000 and 10,000 data points (from left to right).

value is more likely to be more common than any other value.

The previous observations can be extended by considering a specific class of random processes, referred to as fractional Brownian motions (fBm; Fig. 2A, B). Because fBm have the desirable property of exhibiting antipersistent (i.e. an increase in the value of the random variable is expected to be followed by a decrease) and persistent (i.e. an increase in the value of the random variable is expected to be followed by a decrease) behaviors (see e.g. Feller, 1971), they explore a certain range of values before moving off more or less gradually to another range of values. These properties lead to a weaker version of randomness in the Zipf framework (Fig. 2C–F). For antipersistent fBm (Fig. 2A), the Zipf plots do not exhibit any clear linear behavior (Fig. 2C), mainly



Fig. 2. Antipersistent (A) and persistent (B) fractional Brownian motions (fBm), shown together with their characteristic signatures in linear (C, E) and $\log -\log (D, F)$ plots. In antipersistent and persistent processes, an increase in the value of a random variable is expected to be followed by a decrease and an increase, respectively. The resulting Zipf plot exhibit different deviations from randomness. The dashed lines in (B, D) indicate a range of values explored by the fBM before moving off more or less gradually to another range of values. The same colors have been used for the different fBm (A, B) and their related Zipf plots (C–F); the darker colors characterize the more antipersistent/persistent fBm.

because of the upward and downward roll off observed for low and high rank values, respectively. This is, however, simply the result of an under sampling of the highest and lowest values that can be regarded as an implicit consequence of antipersistence. The distributions are characterized by a weak evenness for high and low values, the distribution being dominated by a few (ultimately one) high and low values. On the other hand, for persistent fBm (Fig. 2B), the step shape of the Zipf plot (Fig. 2D, F) reflects the property of persistent processes to visit one particular range of values and then to change to another range sharply. This step function becomes clearer when the fBm exhibit more persistence (Fig. 2D, F). The main difference between antipersistent and persistent Zipf plots then relies in the quantity of values taken by the fBm between transitions that will be more gradual in the antipersistent case, and thus contain more points than in the persistent case. Because the scale expansion related to log-log plots may hide, at least partially, the specific structural features of Zipf plots when compared to noise (see Fig. 2C-F), we strongly recommend the user to use both linear and logarithmic plots. Finally, we stress that any step in Zipf plots indicates structural discontinuities within the data set.

3.2. Deterministic processes

Deterministic patterns and processes are well know in time series analysis and referred to as monotonic and periodic trends. Gradients and sine waves are examples. Periodicity is common in marine ecology. Here, we simulate an increased linear trend and a sine wave trend. Both of them have been subsequently contaminated by observational white noise (Fig. 3A, D). The Zipf plots of the increasing trends exhibit the characteristic signature of white noise (Fig. 3B, C). In contrast, the Zipf plots of the sine waves exhibit distinct features (Fig. 3E, F). The sine wave has the characteristic Zipf shape of a pattern oversampled for its higher and lower values, while the noisy sine wave converges towards a random Zipf signature.

3.3. Pure power laws

Patterns and/or processes characterized by a power law function (e.g. Eq. (3)) will appear as a straight line in log–log plots (Fig. 4). However, this theoretical case is rare in nature, and we will study hereafter a series of situations where power laws may be hidden by a wide range of contaminating processes.

3.4. Contaminated power laws

Before focusing on the processes susceptible to modify the characteristic exponents of Zipf power laws, we will consider here the potential effects of external and internal noise on the extent of the power laws. In the first approach, the variability of a given descriptor is driven by "news" events, which represent exogenous variables (exogenous in the sense that they are not a part of an internal mechanism which drives the descriptor fluctuations), for instance, the motion of dinoflagellate cells induced by vertical turbulent eddy diffusivity. On the other hand, internal noise refers to the existence of an engine within the cells (i.e. endogenous) which generate motion by mechanisms of feedback of the motion of the cells upon themselves.

3.4.1. External (white) noise

If varying amounts of noise are added to the power function $X_r \propto r^{-\alpha}$ as:

$$Y_r = (r^{-\alpha} + \varepsilon) \tag{10}$$

where ε is a white noise whose amplitude is defined as being a given percentage of the maximum value of X_r , then the noise causes a rightward departure from the straight line at a rank proportional to the amount of noise added (Fig. 4). Measuring the point of departure from a power law for a variety of noise levels (here 0.01, 0.1, 1 and 10%) recovers the original function for Zipf plots. Such a graphical approach could be very valuable to estimate the extent to which noise contaminates or contributes to the measured signal.

Consider two situations where a simulated power law function $(X_r \propto r^{-\alpha})$, with $\alpha = 0.18$) is mixed with a random noise ε_i , vertically offset so as not to overlap, i.e. $\varepsilon_1 \in [\max_{X_c}, \chi_1]$ and $\varepsilon_2 \in [\chi_2, \min_{X_r}]$, as $Y_r = X_r + \varepsilon_i$. This could illustrate the expected outcome of a transect crossing a hydrodynamic front separating two distinct structural entities and/or a vertical profile crossing a strong thermocline (Fig. 5A, B). In both cases, the subsequent Zipf plots exhibit a clear step function indicative of a structural discontinuity (Fig. 6A, B) between the characteristic behaviors expected in case of randomness and power law. However, while we used the same power law in both cases, the exponents and the goodness of the power law fits are different (Fig. 6).

This result could lead to misinterpretation of Zipf plots. The widely acknowledged assumption that any range of values with the same extent (e.g. 10 to 100, or 10000 to 10090) on the *x*-axis produce the same range of values of the *y*-axis is no longer valid in the nonlinear framework of power laws. Thus, different ranges of rank, *r*, values, 225 to 450 (Fig. 6A) and 1 to 450 (Fig. 6B), return different ranges on the *y*-axis, and thus different laws. As a consequence, to conduct Zipf analyses successfully and for the results to be meaningful, we recommend here to analyze separately



Fig. 3. Simulated increasing and periodic trends before (grey curve) and after (open diamonds) being contaminated by observational white noise (A, D), and their subsequent Zipf signatures in linear (B, C) and log–log (D, E) plots. In both cases, the symbol θ represents space or time in case of time series or transect studies, respectively.

the different ranges of values identified in a preliminary global analysis as being separated by a step function. Fig. 7 thus illustrates how the simulated power law $X_r \propto r^{-\alpha}$ is recovered by separately analyzing the values characterized by ranks ranging from 225 to 450 for ε_1 (see Figs. 5A and 6A).



Fig. 4. Log-log plot of a power law (diamonds) with different percentage of additive noise (0.01, 0.1, 1 and 10%, from bottom to top).

The combination of randomized values of the power law $X_r \propto r^{-\alpha}$ and the non-overlapping noise, ε_i , (Fig. 5A–C), leads to results similar to those in Fig. 6. Zipf analysis is then revealed to be extremely powerful and valuable in the identification and the quantification of hidden structural properties of any data sets.

3.4.2. Internal (process) noise

Instead of considering an external process (i.e. observational or instrumental noise), the power law itself can be contaminated by internal variability. In such cases, the power law function $X_r \propto r^{-\alpha}$ (with $\alpha = 0.18$) is rewritten as:

$$Y_r = \left(r \pm r \times \varepsilon\right)^{-\alpha} \tag{11}$$

where ε is still a white noise term whose amplitude is defined as being a given percentage of the maximum value of X_r , here randomly chosen as being positive or negative. Whatever the amount of noise considered (here between 5 and 100%), the exponents α estimated from Eq. (3) cannot be statistically regarded as being different from the expected values of 0.18 (p < 0.01).

3.4.3. Competing power laws

In this section, we will investigate examples that could be encountered in the ocean, and that result in a modification of α .

3.4.3.1. Case study 1: mixing non-interacting species. Consider two theoretical phytoplankton populations separated by a sharp hydrological gradient. One is composed of diatoms which can reasonably be thought as following a power law form, $X_r \propto r^{-\alpha}$ (here $\alpha = 0.18$), with respect to their large size and their aggregative properties. The



Fig. 5. Power law distribution $X_r \propto r^{-0.18}$ combined with a white noise distribution ε_i as $Y_r = X_r + \varepsilon_i$, as a caricature of two populations separated by a sharp hydrological gradient (A, B) or mixed (C). The two populations, characterized by a power law X_r and a random distribution ε_i , have been considered as fully separated (A, B), with $\varepsilon_i = \varepsilon_1$ ($\varepsilon_1 \in [\max_X, \chi_1]$) and $\varepsilon_i = \varepsilon_2$ ($\varepsilon_2 \in [\chi_2, \min_X]$), and fully mixed (C), with $\varepsilon_i = \varepsilon_1$. The symbol θ represents space or time in case of time series or transect studies, respectively.

Fig. 6. Zipf plots of the two theoretical situations illustrated in Fig. 5A, B. Note that while the same power law have been used in both situations, the original power law $X_r \propto r^{-0.18}$ is recovered only when $X_r > \varepsilon_i$ (B); when $X_r < \varepsilon_i$ the power law fit to the power law population is not significant (A). The dotted circles indicate the step function behavior of the Zipf plot that should be regarded as being indicative of structural changes within the data set.

other one is composed of dinoflagellates that, because of their smaller size, high concentration and motility, are more homogeneously distributed and are then simply represented here as a background concentration k_i . The resulting pattern can then be thought of as the combination $Y_r = X_r + k_i$. It is emphasized here that any change in the background concentration k_i does not affect the power law nature of the original data set X_r . However, the exponent α' of the resulting power laws $Y_r \propto r^{-\alpha'}$ decreases with increasing values of k_i . The addition of an increasing background concentration thus

Fig. 7. Illustration of the "density–dependence" of Zipf plot. The Zipf behavior of the power law population X_r characterized by $X_r < \varepsilon_i$ (black squares; Fig. 6A) is recovered (open squares) when the range of values identified as being separated by a step function (see Fig. 6A, B) have been analyzed separately.

200

r

250

300

350

400

450

smoothed out the differences between ranks. The observation of such a decrease in empirical power law exponents from field data sampled at the same point before and after the disruption of a hydrological gradient, or at different period of the seasonal cycle, would



Fig. 8. Log–log plot signature of the Zipf behavior resulting from mixing two theoretical populations characterized by two distinct power laws and overlapping ranges of concentrations. The range of values corresponding to the overlapping of the two power laws presents an intermediate power law behavior with a characteristic exponent α' defined as $\alpha_1 < \alpha' < \alpha_2$ and $\alpha' = k\alpha_1 + (1-k)\alpha_2$.



30

25

20

15

10

5

0

100

0

50

100

 $Y_r = (X_r + \varepsilon_1)$

 $X_r = 1395r^{-0.8}$

 $r^2 = 0.83$

 $X_{*} = 27.98i$

 $r^2 = 0.99$

150

strongly indicate a structural change in the relative organization of the studied biological communities.

Now, consider a situation where two spatially separated phytoplankton populations are mixed, e.g. two monospecific diatom populations, characterized by overlapping ranges of concentrations and distinct power law forms, $X_{1r} \in [2.99, 13.79]$ and $X_{1r} \propto r^{-\alpha_1}$ with $\alpha_1 = 0.18$, and $X_{2r} \in [3.19, 31.37]$ and $X_{2r} \propto r^{-\alpha_2}$ with $\alpha_1 = 0.24$, respectively. Evenly mixing these two populations without considering any interactions will result in the Zipf structures shown in Fig. 8. The range of values corresponding to the overlapping of the two power laws presents an intermediate power law behavior with a characteristic exponent $\alpha' = 0.196$ (Fig. 8). More generally, the values of α' are implicitly bounded between $\alpha_1 < \alpha' < \alpha_2$, where α_1 and α_2 are the Zipf exponents of the original power laws, and depend on the proportion of values from each original power law, following $\alpha' = k\alpha_1 + (1-k)\alpha_2$. Finally, as stated above, a separate analysis of the values greater than the critical concentration (i.e. 13.79) associated with the step function shown in Fig. 8 is necessary to recover the original exponents $\alpha_2 = 0.24$.

3.4.3.2. Case study 2: mixing interacting species.

Here, we consider one of the previous phytoplankton populations whose concentration X_r is characterized by a power law form $X_r \propto r^{-\alpha}$, with $\alpha = 0.24$. We will now investigate the effects of processes capable of locally decreasing (i.e. mortality related to inter- and intraspecific competition, or grazing) or increasing (phytoplankton growth or coagulation processes) phytoplankton concentration on the Zipf signature of the population $X_r \propto r^{-0.24}$.

3.4.3.2.1. Decrease in local phytoplankton concentration. First, under the assumption of evenly distributed grazers, the grazing impact of copepods can be estimated as a percentage or a Michaelis– Menten function of the local phytoplankton concentration. Assuming that the ingestion of phytoplankton cells by copepods is a percentage or a random function of food availability, the resulting food distributions can be described by:

$$Y_{1r} = X_r - kX_r \tag{12}$$

and

$$Y_{2r} = X_r - \varepsilon X_r \tag{13}$$

where k is a constant, $0 \le k \le 1$, and ε is a random noise process, i.e. $\varepsilon \in [0,1]$. For increasing values of k,



Fig. 9. Log–log plot signature of the Zipf behavior expected in case of a power law X_r (open diamonds) competing with a random mortality component ($Y_{2r}=X_r-\varepsilon X_r$), where $\varepsilon = 0.05$, 0.25, 0.50 and 0.75 (from top to bottom).

the function Y_{1r} is simply shifted downward on a loglog Zipf plot (not shown), indicating a decrease in the background concentration of the population. A similar trend can be identified in the variable Y_{2r} for an increasing amount of noise, but with a characteristic 'noise roll off' for low rank values (Fig. 9).

Alternatively, following laboratory data on the feeding of copepods suggesting that ingestion rate can be fairly represented by a Michaelis–Menten function (e.g. Mullin et al., 1975), Eqs. (12) and (13) are modified as:

$$Y_{3r} = X_r - I_{\max} X_r / (k_s + X_r)$$
(14)

where I_{max} is the maximum ingestion rate, k_s is the halfsaturation constant for feeding, and X_r the concentration of food. Fig. 10 shows the Zipf structure of the resulting phytoplankton concentration Y_{3r} , for different values of the half-saturation constant k_s and the maximum ingestion rate I_{max} . It clearly appears that the effect of grazing is mainly perceptible for low values of Y_{3r} , a direct consequence of the convex form of the Michaelis–Menten function (see Eq. (14)), and leads to a significant divergence from a power law when I_{max} is high and k_s is low (compare Fig. 10A, B and C).

However, the two previous approaches are implicitly based on the hypothesis of a homogeneous phytoplankton distribution, which is now recognized as an oversimplified hypothesis (e.g. Seuront et al., 1996a, 1999; Waters and Mitchell, 2002; Waters et al., 2003), and did not take into account potential behavior adaptation of grazers to varying food concentrations (e.g. Tiselius, 1992). If one considers that the remote



Fig. 10. Log-log plot signature of the Zipf behavior expected in case of a power law X_r (open diamonds) competing with a Michaelis– Menten grazing component $(Y_{3r}=X_r-I_{\max}X_r/(k_s+X_r))$. For a given maximum ingestion rate I_{\max} , the effect is stronger for high values of the half-saturation constant k_s .

sensing abilities (e.g. Doall et al., 1998) of copepods can lead to aggregation of grazers in areas of high phytoplankton concentrations as investigated both empirically and numerically (e.g. Tiselius, 1992; Saiz et al., 1993; Leising and Franks, 2000; Seuront et al., 2001), Eq. (13) can be modified as:

$$Y_{4r} = X_r - 10^{(X_r/k)} \tag{15}$$

where k is a constant and the ingestion function $I(X_r)=10^{(X_r/k)}$ represents an increased predation impact on higher phytoplankton concentration. The advantage of the function $I(X_r)$ is that it can be regarded as a representation of both aggregation of copepods with constant ingestion rates and evenly distributed copepods with increasing ingestions rates in high density phytoplankton patches. Decreasing values of the constant k increases the grazing impact on high density patches (Eq. (15)). The grazed phytoplankton population then diverges from a power law form for high values of Y_{4r} , but asymptotically converges to the original power law for the smallest values of Y_{4r} , i.e. $Y_{4r} \propto r^{-\alpha}$ for $r \rightarrow r_{min}$ (Fig. 11).

While the previous examples have been based on zooplankton grazing on phytoplankton, we nevertheless stress the generality of our approach as similar results could have been obtained considering two phytoplankton populations competing for the same nutrient resource using Michaelis–Menten or Droop functions.

3.4.3.2.2. Increase in local phytoplankton concentration. For the sake of simplicity, we consider that phytoplankton growth (in response to physical coagulation or nutrient uptake) could be represented as a percentage or a random function of the actual phytoplankton concentration X_r . Eqs. (12) and (13) are then respectively rewritten as:

$$Y_{5r} = X_r + kX_r \tag{16}$$

and

$$Y_{6r} = X_r + \varepsilon X_r \tag{17}$$

where k is a constant, $0 \le k \le 1$, and ε is a random noise process, i.e. $\varepsilon \in [0,1]$. For increasing values of k, the function Y_{5r} is, in full agreement with what have been concluded from Eq. (12), shifted upward on a log–log Zipf plot (not shown), indicating an increase in the background concentration of the population. Using different values of k in Eq. (16) has no effect on the shape of the related Zipf plots and exponents α'



Fig. 11. Log–log plot signature of the Zipf behavior expected in case of a power law X_r (open diamonds) competing with a preferential grazing component for high phytoplankton concentrations $(Y_{4r}=X_r-10^{(Xr/k)})$. The grazed phytoplankton population diverges from a power law form for high concentrations, but asymptotically converges to the original power law for the smallest values. The extent of the observed divergence is controlled by increasing grazing pressure *k* (from top to bottom).

 $(Y_{5r} \propto r^{-\alpha'})$ that remain equal to the original power law, i.e. $Y_{5r} \propto r^{-0.24}$. Slightly different conclusions can be drawn from the behavior of the Zipf plots of the function Y_{6r} (Fig. 12). First, increases in the amount of noise ε (ranging from 25 to 100%) lead to a vertical offset of the function Y_{6r} when compared to the original power law. The resulting functions exhibit the characteristic downward roll-off signature related to randomness, but might also locally show increasing trends that are intrinsically caused by the random component in Eq. (17). They could be misleading, especially when they occurred over the highest rank range (see arrows in Fig. 12), and must not be related to break points indicative of structural discontinuities (see above) that would erroneously lead to a separate analysis of different subsections of the original data set. Finally, even if the exponents α' fluctuate around the original value, they are never significantly different (p < 0.05).

3.4.3.2.3. Increase vs. decrease in local phytoplankton concentration. Because the two previous situations are unlikely to be found individually in the ocean, but should also occur concomitantly, we combine Eqs. (12) and (16), and Eqs. (13) and (17) as:

$$Y_{7r} = X_r + (k_1 - k_2)X_r \tag{18}$$

and

$$Y_{8r} = X_r + (\varepsilon_1 - \varepsilon_2)X_r \tag{19}$$

where k_1 and k_2 are constants $(0 \le k_1 \le 1, \text{ and } 0 \le 1)$ $k_2 \leq 1$), and ε_1 and ε_2 are random noise processes, i.e. $\varepsilon_1 \in [0,1]$ and $\varepsilon_2 \in [0,1]$. The resulting functions Y_{5r} and Y_{6r} , exhibit intermediate behaviors between what have been observed from Eqs. (12) and (16), and Eqs. (13) and (17). For $k_1 = k_2$, the original power law, $Y_{7r} \propto r^{-0.24}$, is recovered, the growth component compensates for the death component. In contrast, when $k_1 < k_2$ and $k_1 > k_2$ the resulting function Y_{7r} is shifted downward and upward on a log-log plot as previously observed from Eqs. (12) and (16). While the overall structure is preserved, the latter and the former cases lead to decreases and increases in the background concentration of the population. The Zipf plot of the function Y_{8r} , shown in a log-log plot, exhibits the different characteristic features previously identified: a power law behavior not significantly different from the original one (i.e. $\alpha' \approx \alpha = 0.24$) followed by a roll off towards low Y_{8r} values (Fig. 13). As stated above, successive positive random fluctuations might lead to local increasing trends slightly diverging from a power law (grey arrow), but should not be associated to a step function.



Fig. 12. Log–log plot signature of the Zipf behavior expected in case of a power law (open diamonds) competing with a random growth component ($Y_{6r}=X_r+\varepsilon X_r$) where ε =0.25, 0.50, 0.75 and 1.00 (from bottom to top). The arrows indicate the minimum of a range of Y_{6r} values locally diverging from a power law because of successively increasing random increments. The dashed lines indicate show the power law behavior of the initial values X_r .



Fig. 13. Log–log plot signature of the Zipf behavior expected n case of a power law X_r (open diamonds) competing with random growth and mortality components ($Y_{8r}=X_r+(\varepsilon_1-\varepsilon_2)X_r$), where the random processes ε_1 and ε_2 have been chosen as $\varepsilon_1=0.75$ and $\varepsilon_2=0.25$, $\varepsilon_1=0.50$ and $\varepsilon_2=0.50$, and $\varepsilon_1=0.25$ and $\varepsilon_2=0.75$ (from bottom to top). The dashed lines indicate show the power law behavior of the initial values X_r .

Next, we consider a situation where the positive and negative fluctuations are purely randomly driven as:

$$Y_{9r} = X_r \pm kX_r \tag{20}$$

and

$$Y_{10r} = X_r \pm \varepsilon X_r \tag{21}$$

where k is a constant $(0 \le k \le 1)$ and ε is a random noise process, i.e. $\varepsilon \in [0,1]$, whose amplitude is defined as being a given percentage of the maximum value of X_r ; k and ε are randomly chosen as being positive or negative. The resulting Zipf signatures of the functions Y_{9r} and Y_{10r} are shown in Fig. 14 as log-log plots (Fig. 14). The positive and negative components of Eq. (20) clearly appeared as separated by a step function (Fig. 14A). The positive components lead to power laws that are not significantly different from the original. A separate analysis of the range of values separated from the power laws by step functions (arrows; Fig. 14A) did not show any power law behavior. Alternatively, the effects of Eq. (21) on the initial power law behavior are the characteristic downward roll-off signature related to randomness and the fluctuations around a power law behavior that is not significantly different from the original (Fig. 14B). To ensure the relevance of Zipf analysis-as introduced by Eqs. (2)–(3)–the few data points diverging upward, or flattening, towards the first rank values must not be included in the regression analysis aimed at estimating α . Indeed, the former case describes a distribution marked by the dominance of a few (ultimately one) 'hotspot' that is likely to be chronically undersampled (Seuront et al., 1999), while



Fig. 14. Log–log plot signature of the Zipf behavior expected in case of a power law X_r (open diamonds) competing with combined constant random and growth components ($Y_{9r}=X_r\pm kX_r$; a) and combined random growth and mortality components ($Y_{10r}=X_r\pm \epsilon X_r$; b), where the constant k and the noise ϵ have been chosen as k=0.25, 0.50 and 0.75, and $\epsilon=0.25$, 0.50, 0.75 and 100 (from bottom to top). The grey arrows indicate a step function (A), while the black one (B) indicates the beginning of a local departure from a pure power law due to successively increasing random increments. The dashed lines indicate show the power law behavior of the initial values X_r .



Fig. 15. Log–log plot signature of the Zipf behavior expected in case of a power law X_r (open diamonds) competing with a persistent fractional Brownian motion.

the latter case refers to a distribution that has been systematically oversampled. These issues have been specifically detailed elsewhere in the framework of information theory (Mandelbrot, 1953), and led to the modified version of the generalized Zipf law presented above; see Eq. (4).

It is now known that the distributions of nutrients, phytoplankton and zooplankton exhibit different levels of persistence (Tsuda, 1995; Seuront et al., 1996a,b, 1999; Seuront and Lagadeuc, 2001; Seuront et al., 2002). In addition, the interplay between the biotic properties of individuals and populations and abiotic processes produce space-time structures characterized by long-range correlation (i.e. persistence, Kendall et al., 2000). We consider, finally, a situation where the local concentration of a phytoplankton population initially driven by a power law $X_r \propto r^{-0.24}$ could be influenced by a fractional Brownian motion resulting from the combined effects of local biological (nutrient uptake, inter- and intraspecific competition, grazing pressure, infection) and physical (advection, diffusion, turbulence) processes following:

$$Y_{11r} = X_r \pm f \operatorname{Bm} X_r \tag{22}$$

where fBm is a persistent fractional Brownian motion (see Fig. 2b) whose amplitude is defined as being a given percentage of the maximum value of X_r , and randomly chosen as being positive or negative. The resulting Zipf signature (Fig. 15) exhibits downward roll-off charac-

teristic of randomness for high rank values, and longrange correlations around a power law behavior $(Y_{11r} \propto r^{-\alpha'})$, with $\alpha' = 0.240 \pm 0.005$; $\overline{x} \pm \text{SD}$) that is significantly different from the original power law $(X_r \propto r^{-0.24}, p > 0.05)$. These long range correlations exist whatever the values of *r*, but are more clearly visible for the low values of *r*, i.e. high values of X_r in Eq. (22).

4. Discussion and conclusions

We have demonstrated the potential for the seldom used Zipf power law to be a powerful tool in the analysis and the classification of marine ecosystems in the presence of randomness, monotonic and periodic trends, internal and external noise, mixing and environmental forcing. In particular, we show that Zipf analysis can be directly and easily applied to any data set without intensive computational, mathematical or statistical analysis, and with a minimum amount of calculation. It can be conducted in a few minutes with most standard software packages even for a data set of several thousands data points. We nevertheless stress that Zipf analysis should not be used without a preliminary visual inspections (an absolute prerequisite in data analysis that is often neglected), as they could erroneously be used as a direct index of patchiness. Indeed, a distribution characterized by a patch of ten high density data points, ten randomly or regularly spaced 'hotspots', or ten ranked 'hotspots' will return exactly the same Zipf shape. Zipf analysis cannot be then substituted for spatially and temporally informative techniques such as spatial autocorrelation analysis and spectral analysis. In addition, the one-to-one correspondence between Zipf and Pareto distributions analytically derived here could be regarded as a way to reconcile previous and future works using one or the other technique. The strength of the Zipf framework also has the desirable properties of not requiring any assumptions about the statistical distribution, regularity of sampling intervals and stationarity of the data set that are sometimes absolute prerequisites to some statistical data analysis techniques.

Finally, we stress that the Zipf framework can be conveniently used as a tool to identify and classify structures in marine ecosystems, but also to infer the underlying processes that generate the observed patterns. The characteristic shapes introduced here, and most importantly their potential changes in time and/or space, make it possible to hypothesize the origin and the ecological implications of such modifications, as well as providing useful insights on what further analysis to conduct and how to design of sampling schemes. For instance, a phytoplankton transect study providing a step function in a Zipf plot (cf. Fig. 2D) will indicate different levels of organization within the same populations, and maybe different sub-populations that would require separate analysis and/or additional sampling. For phytoplankton distribution, it is easy to imagine that mixing, changing nutrient and/or zooplankton concentrations will alter distribution and intensity to the extent that the characteristic exponents α and β for a set of data varies due to natural processes. Thus, a phytoplankton population exhibiting a single power law behavior before wind stratification and investigated temporally from a fixed point might exhibit successive changes (e.g. Figs. 10 and 11), and the identification and the classification of the Zipf shapes will then allow one to infer the nature of the observed changes. In turn, a study mainly focusing on phytoplankton distributions which results in transitions such as those shown in Figs. 6 and 8 in spring and autumn respectively may well be modified and adapted to investigate the potential differences in the grazers community. More generally, if phytoplankton properties such as growth or distribution follow a power law, then mortality processes such as grazing and lysis may well follow a similar but competing power law as hypothesized and illustrated above. Thus, if such power law behavior can be shown in phytoplankton, the removal of the first ranks (large values) could be interpreted as an indication of predation. However, the ubiquity of power laws is not an absolute requirement as many nonpower law processes could be involved in the modification of the pure power law behavior, as well as the removal of the last ranks (low values). The establishment of a comprehensive typology of marine ecosystems structures and functions requires checking and experimenting, but would undoubtedly be a useful model in that it could work spatially as well as temporally, adding another approach to studying marine ecosystem dynamics. This assumes that effects from mixing, infection and nutrient starvation can be teased out from those of grazing. Alternatively, perhaps a more unified approach would provide a better predictor of overall dynamics. The general applicability, as well as the subsequent ecological interpretations, of these two related techniques are fully illustrated elsewhere (Mitchell and Seuront, this issue; Seuront et al., in press).

Acknowledgements

Justin Seymour and Raechel Waters are acknowledged for their constructive and inspiring discussions. We thank two anonymous reviewers for their useful comments. This work was supported by the Australian Research Council and Flinders University of South Australia. The Flinders Faculty of Science and Engineering provided a visiting scholarship to LS.

References

- Axtell, R.L., 2001. Zipf distribution of U.S. firm sizes. Science 293, 1818–1820.
- Bak, P., Sneppen, K., 1993. Punctuated equilibrium and criticality in a simple model of evolution. Phys. Rev. Lett. 24, 4083–4086.
- Bak, P., Chen, K., Creutz, M., 1989. Selforganized criticality in the game of life. Nature 342, 780–782.
- Biham, O., Malcai, O., Levy, M., Solomon, S., 1998. Generic emergence of power law distributions and Lévy-stable intermittent fluctuations in discrete logistic systems. Phys. Rev., E Stat. Phys. Plasmas Fluids Relat. Interdiscip. Topics 58, 1352–1358.
- Burda, Z., Johnston, D., Jurkiewicz, J., Kaminski, M., Nowak, M.A., Papp, G., Zahed, I., 2002. Wealth condensation in Pareto macroeconomies. Phys. Rev., E Stat. Phys. Plasmas Fluids Relat. Interdiscip. Topics 65, 1–4.
- Chatfield, C., 1996. The Analysis of Time Series. Chapman and Hall, New York.
- Cizrók, A., Mantegna, R.N., Havlin, S., Stanley, E., 1995. Correlations in binary sequences and a generalized Zipf analysis. Phys. Rev., E Stat. Phys. Plasmas Fluids Relat. Interdiscip. Topics 52, 446–450.
- Correig, A.M., Urquizú, M., Vila, J., 1997. Aftershock series of event February 18, 1996: an interpretation in terms of self-organized criticality. J. Geophys. Res. 102, 27404–27420.
- Daly, K.L., Smith, W.O., 1993. Physical-biological interactions influencing marine plankton production. Ann. Rev. Ecolog. Syst. 24, 555–585.
- Diodati, P., Marchesoni, F., Piazza, S., 1991. Acoustic emission from volcanic rocks: an example of self-organized criticality. Phys. Rev. Lett. 67, 2239–2242.
- Doall, M.H., Colin, S.P., Strickler, J.R., Yen, J., 1998. Locating a mate in 3D: the case of *Temora longicornis*. Philos. Trans. R. Soc. Lond., B 353, 681–689.
- Downing, J.A., Pérusse, M., Frenette, Y., 1987. Effect of interreplicate variance on zooplankton sampling design and data analysis. Limnol. Oceanogr. 32, 673–680.
- Dremin, I.M., Ivanov, O.V., Nechitailo, V.A., 2004. An introduction to wavelets. In: Seuront, L., Strutton, P.G. (Eds.), Handbook of Scaling Methods in Aquatic Ecology: Measurement, Analysis, Simulation. CRC Press, Boca Raton, pp. 279–296.
- Faloutsos, M., Faloutsos, P., Faloutsos, C., 1999. On power-law relationships of the internet topology. Comput. Commun. Rev. 29, 251–262.
- Feder, H.J.S., Feder, J., 1991. Experiments and simulations modeling earthquakes. In: Riste, T., Sherrington, D. (Eds.), Spontaneous Formation of Space-time Structures and Criticality. Kluwer Academic Publishers, Dordrecht, pp. 107–111.
- Feldman, G.C., Clark, D., Halpern, D., 1984. Satellite color observations of the phytoplankton distribution in the eastern Equatorial Pacific during the 1982–83 El Nino. Science 226, 1069–1071.
- Feller, W., 1971. An Introduction to Probability Theory and its Application. Wiley, New York.
- Fisher, K.E., Wiebe, P.W., Malamud, B.D., 2004. Fractal characterization of spatial distributions of plankton on Georges Bank. In: Seuront, L., Strutton, P.G. (Eds.), Handbook of Scaling Methods in Aquatic Ecology: Measurement, Analysis, Simulation. CRC Press, Boca Raton, pp. 297–320.

Frette, V., Christensen, K., Malthe-Sørenssen, A., Feder, J., Jøssang, T., Meakin, P., 1996. Dynamics in a pile of rice. Nature 379, 49–51.

Frontier, S., 1972. Calcul de l'erreur sur un comptage de zooplancton. J. Exp. Mar. Biol. Ecol. 8, 121–132.

- Frontier, S., 1977. Réflexions pour une théorie des écosystèmes. Bull. Ecol. 8, 445–464.
- Frontier, S., 1985. Diversity and structure in aquatic ecosystems. Oceanogr. Mar. Biol. Ann. Rev. 23, 253–312.
- Frontier, S., 1994. Species diversity as a fractal properties of biomass. In: Nowak, M.M. (ed.), Fractals in the Natural and Applied Sciences (A-41). Elsevier, North-Holland, pp. 119–127.
- Frontier, S., Bour, W., 1976. Note sur une collection de Chaetognathes récoltée au dessus du talus continental près de Nosy Be (Madagascar) Cah. ORSTOM, Sér. Océanogr. 14, 267–272.
- Garcia-Pelayo, R., Morley, P.D., 1993. Scaling law for pulsar glitches. Europhys. Lett. 23, 185–188.
- Hardy, A.C., 1926. The Discovery expedition: a new method of plankton research. Nature 118, 630–632.
- Hardy, A.C., 1939. Ecological investigation with Continuous Plankton Recorder: object, plan and methods. Hull Bull. Mar. Ecol. 1, 1–57.
- Hardy, A.C., Gunther, E.R., 1935. The plankton of the South Georgia whaling grounds and adjacent waters, 1926–1927. Discov. Rep. 11, 1–456.
- Held, G.A., Solina, D.H., Keane, D.T., Haag, W.J., Horn, P.M., Grinstein, G., 1990. Experimental study of critical mass fluctuations in an evolving sandpile. Phys. Rev. Lett. 1120–1132.

Hsieh, C.H., Glaser, S.M., Lucas, A.J., Sugihara, G., 2005. Distinguishing random environmental fluctuations from ecological catastrophes for the North Pacific Ocean. Nature 435, 336–340.

- Israeloff, N.E., Kagalenko, M., Chan, K., 1995. Can Zipf distinguish language from noise in noncoding DNA? Phys. Rev. Lett. 76, 1976–1979.
- Jaeger, M.H., Nagel, S.R., 1992. Physics of the granular state. Science 255, 1523–1526.
- Joshi, R.R., Selvam, A.M., 1999. Identification of self-organized criticality in atmospheric low frequency variability. Fractals 7, 421–425.
- Kanter, I., Kessler, D.A., 1995. Markov processes: linguistics and Zipf's law. Phys. Rev. Lett. 74, 4559–4562.
- Keitt, T.H., Marquet, P.A., 1996. The introduced Hawaiian avifauna reconsidered: evidence for self-organized criticality? J. Theor. Biol. 182, 161–167.
- Kendall, B.E., Bjørnstad, O.N., Bascompte, J., Keitt, T.H., Fagan, W.F., 2000. Dispersal, environmental correlation, and spatial synchrony in population dynamics. Am. Nat. 155, 628–636.
- Kolasa, J., Pickett, S.T.A., 1991. Ecological Heterogeneity. Springer Verlag, New York.
- Legendre, L., Demers, S., 1984. Towards dynamic biological oceanography and limnology. Can. J. Fish. Aquat. Sci. 41, 2–19.
- Leising, A.W., Franks, P.J.S., 2000. Copepod vertical distribution within a spatially variable food source: a simple foraging-strategy model. J. Plankton Res. 22, 999–1024.
- Lovejoy, S., Currie, W.J.S., Tessier, Y., Claereboudt, M.R., Bourget, E., Roff, J.C., Schertzer, D., 2001. Universal multifractal and ocean patchiness: phytoplankton, physical fields and coastal heterogeneity. J. Plankton Res. 23, 117–141.
- Mackas, D.L., Denman, K.L., Abbott, M.R., 1985. Plankton patchiness: biology in the physical vernacular. Bull. Mar. Sci. 37, 652–674.
- Makse, H.A., Havlin, S., Stanley, H.E., 1995. Modelling urban growth patterns. Nature 377, 608–612.
- Malacarne, L.C., Mendes, R.S., Lenzi, E.K., 2002. q-exponential distribution in urban agglomeration. Phys. Rev., E Stat. Phys. Plasmas Fluids Relat. Interdiscip. Topics 65 (doi:017106).

- Mandelbrot, B.B., 1953. Contribution à la théorie mathématique des communications. Thèse Univ. Paris Publ. Inst. Stat. Univ. Paris, 2:1–121.
- Manrubia, S.C., Solé, R.V., 1996. Self-organized criticality in rainforest dynamics. Chaos, Solitons Fractals 7, 523–541.
- Mantegna, R.N., Buldyrev, S.V., Goldberger, A.L., Havlin, S., Peng, C.K., Simons, M., Stanley, H.E., 1994. Linguistic features of noncoding DNA sequences. Phys. Rev. Lett. 73, 3169–3172.
- Mantegna, R.N., Buldyrev, S.V., Goldberger, A.L., Havlin, S., 1995. Systematic analysis of coding and noncoding DNA sequences using methods of statistical linguistics. Phys. Rev., E Stat. Phys. Plasmas Fluids Relat. Interdiscip. Topics 52, 2939–2941.
- Margalef, R., 1957. La teoria de la informacion en ecologia. Mem. R. Acad. Cienc. Artes Barc. 32, 373–449.
- Marsili, M., Zhang, Y.C., 1998. Intercating individuals leading to Zipf's law. Phys. Rev. Lett. 80, 2741–2744.
- McHardy, I., Czerny, B., 1987. Fractal X-ray time variability and spectral invariance of the Seyfert galaxy NGC5506. Nature 325, 696–698.
- Mercik, S., Weron, K., Siwy, Z., 1999. Statistical analysis of ionic current fluctuations in membrane channels. Phys. Rev., E Stat. Phys. Plasmas Fluids Relat. Interdiscip. Topics 60, 7343–7345.
- Mitchell, J.G., 2004. Rank-size analysis and vertical phytoplankton distribution patterns. In: Seuront, L., Strutton, P.G. (Eds.), Handbook of Scaling Methods in Aquatic Ecology: Measurement, Analysis, Simulation. CRC Press, Boca Raton, pp. 257–278.
- Mitchell, J.G., Seuront, L. Towards a seascape typology. II. Onedimensional patterns. J. Mar. Syst. (revised).
- Mullin, M.M., Stewart, E.F., Fuglister, F.J., 1975. Ingestion by planktonic grazers as a function of concentration of food. Limnol. Oceanogr. 20, 259–262.
- Noever, D.A., 1993. Himalayan sandpiles. Phys. Rev., E Stat. Phys. Plasmas Fluids Relat. Interdiscip. Topics 47, 724–744.
- Olami, Z., Feder, J., Christensen, K., 1992. Self-organized criticality in a continuous, nonconservative cellular automaton modeling earthquakes. Phys. Rev. Lett. 68, 197–200.
- Paczuski, M., Maslov, S., Bak, P., 1995. Avalanche dynamics in evolution, growth and depinning models. Phys. Rev., E Stat. Phys. Plasmas Fluids Relat. Interdiscip. Topics 53, 414–418.
- Pareto, V., 1896. Oeuvres complètes. Droz, Geneva.
- Pascual, M., Ascioti, F.A., Caswell, H., 1995. Intermittency in the plankton: a multifractal analysis of zooplankton biomass variability. J. Plankton Res. 17, 1209–1232.
- Perline, R., 1996. Zipf's law, the central limit theorem, and the random division of the unit interval. Phys. Rev., E Stat. Phys. Plasmas Fluids Relat. Interdiscip. Topics 54, 220–223.
- Pickett, S.T.A., Cadenasso, M.L., 1995. Landscape ecology: spatial heterogeneity in ecological systems. Science 269, 331–334.
- Platt, T., 1972. Local phytoplankton abundance and turbulence. Deep-Sea Res. 19, 183–187.
- Platt, T., Denman, K.L., 1975. Spectral analysis in ecology. Ann. Rev. Ecolog. Syst. 6, 189–210.
- Press, W.H., Teukolsky, S.A., Vetterling, W.T., Flannery, B.P., 1992. Numerical Recipes in Frotran. Cambridge University Press, Cambridge.
- Ramsden, J.J., Vohradsky, J., 1998. Zipf-like behavior in procaryotic protein expression. Phys. Rev., E Stat. Phys. Plasmas Fluids Relat. Interdiscip. Topics 58, 7777–7780.
- Rigon, R., Rinaldo, A., Rodriguez-Iturbe, I., 1994. On landscape selforganization. J. Geophys. Res. 99, 11971–11987.
- Rinaldo, A., Maritan, A., Colaiori, F., Flammini, A., Rigon, R., Ignacio, I., Rodriguez-Iturbe, I., Banavan, J.R., 1996. Thermodynamics of fractal river networks. Phys. Rev. Lett. 76, 3364–3367.

- Rothman, D.H., Grotzinger, J.P., Flemings, P., 1994. Scaling in turbidite deposition. J. Sediment. Res., Sect. A Sediment. Pet. Proc. 64, 355–365.
- Saiz, E., Tiselius, P., Jonsson, P.R., Verity, P., Paffenhöfer, G.A., 1993. Experimental records of the effects of food patchiness and predation on egg production of *Acartia tonsa*. Limnol. Oceanogr. 38, 280–289.
- Seuront, L., Lagadeuc, Y., 1997. Characterisation of space-time variability in stratified and mixed coastal waters (Baie des Chaleurs, Québec, Canada): application of fractal theory. Mar. Ecol., Prog. Ser. 159, 81–95.
- Seuront, L., Lagadeuc, Y., 1998. Spatio-temporal structure of tidally mixed coastal waters: variability and heterogeneity. J. Plankton Res. 20, 1387–1401.
- Seuront, L., Lagadeuc, Y., 2001. Multiscale patchiness of the calanoid copepod *Temora longicornis* in a turbulent coastal sea. J. Plankton Res. 23, 1137–1145.
- Seuront, L., Spilmont, N., 2002. Self-organized criticality in intertidal microphytobenthos patch patterns. Physica, A 313, 513–539.
- Seuront, L., Strutton, P.G. (Eds.), 2004. Handbook of scaling methods in aquatic ecology. Measurement, Analysis, Simulation. CRC Press, Boca Raton.
- Seuront, L., Schmitt, F., Lagadeuc, Y., Schertzer, D., Lovejoy, S., Frontier, S., 1996a. Multifractal analysis of phytoplankton biomass and temperature in the ocean. Geophys. Res. Lett. 23, 3591–3594.
- Seuront, L., Schmitt, F., Schertzer, D., Lagadeuc, Y., Lovejoy, S., 1996b. Multifractal intermittency of Eulerian and Lagrangian turbulence of ocean temperature and plankton fields. Nonlinear Process. Geophys. 3, 236–246.
- Seuront, L., Schmitt, F., Lagadeuc, Y., Schertzer, D., Lovejoy, S., 1999. Multifractal analysis as a tool to characterize multiscale inhomogeneous patterns. Example of phytoplankton distribution in turbulent coastal waters. J. Plankton Res. 21, 877–922.
- Seuront, L., Schmitt, F., Lagadeuc, Y., 2001. Turbulence intermittency, small-scale phytoplankton patchiness and encounter rates in plankton: where do we go from here? Deep-Sea Res., Part 1, Oceanogr. Res. Pap. 48, 1199–1215.
- Seuront, L., Gentilhomme, V., Lagadeuc, Y., 2002. Small-scale nutrient patches in tidally mixed coastal waters. Mar. Ecol., Prog. Ser. 232, 29–44.
- Seuront, L., Brewer, M., Strickler, J.R., 2004. Quantifying zooplankton swimming behavior: the question of scale. In: Seuront, L., Strutton, P.G. (Eds.), Handbook of scaling methods in aquatic ecology: measurement, analysis, simulation. CRC Press, Boca Raton, pp. 333–359.
- Seuront, L., Mitchell, J.G., Seymour, J.R., Waters, R.L., Leterme, C., Strutton, P.G., Towards a seascape typology. III. Two-dimensional patterns. J. Mar. Syst. (revised).
- Seymour, J.R., Mitchell, J.G., Pearson, L., Waters, R.L., 2000. Heterogeneity in bacterioplankton abundance from 4.5 millimetre resolution sampling. Aquat. Microb. Ecol. 22, 143–153.
- Solé, R.V., Manrubia, S.C., 1995a. Are rainforest self-organized in a critical state? J. Theor. Biol. 172, 31–40.

- Solé, R.V., Manrubia, S.C., 1995b. Self-similarity in rainforests: evidence for a critical state. Phys. Rev., E Stat. Phys. Plasmas Fluids Relat. Interdiscip. Topics 51, 6250–6253.
- Somfai, E., Czirok, A., Vicsek, T., 1994. Self-affine roughening in a model experiment on erosion in geomorphology. J. Phys. A 205, 355–366.
- Somfai, E., Czirok, A., Vicsek, T., 1995. Power-law distribution of landslides in an experiment on the erosion of a granular pile. J. Phys. A 27, 757–760.
- Strutton, P.G., Mitchell, J.G., Parslow, J.S., 1996. Non-linear analysis of chlorophyll a transects as a method of quantifying spatial structure. J. Plankton Res. 18, 1717–1726.
- Strutton, P.G., Mitchell, J.G., Parslow, J.S., 1997. Using non-linear analysis to compare the spatial structure of chlorophyll with passive tracers. J. Plankton Res. 19, 1553–1564.
- Sugihara, G., May, R.M., 1990. Nonlinear forecasting as a way of distinguishing chaos from measurement error in time series. Nature 344, 734–741.
- Sugihara, G., Grenfell, B., May, R.M., 1990. Distinguishing error from chaos in ecological time series. Philos. Trans. R. Soc. Lond., B 330, 235–251.
- Taylor, L.R., 1961. Aggregation, variance and the mean. Nature 189, 732–735.
- Tiselius, P., 1992. Behavior of *Acartia tonsa* in patchy food environments. Limnol. Oceanogr. 37, 1640–1651.
- Troll, G., Graben, P.B., 1998. Zipf's law is not a consequence of the central limit theorem. Phys. Rev., E Stat. Phys. Plasmas Fluids Relat. Interdiscip. Topics 57, 1347–1355.
- Tsuda, A., 1995. Fractal distribution of the oceanic copepod Neocalanus cristatus in the subarctic Pacific. J. Oceanogr. 51, 261–266.
- Vidondo, B., Prairie, Y.T., Blanco, J.M., Duarte, C.M., 1997. Some aspects of the analysis of size spectra in aquatic ecology. Limnol. Oceanogr. 42, 184–192.
- Watanabe, M.S., 1996. Zipf's law in percolation. Phys. Rev., E Stat. Phys. Plasmas Fluids Relat. Interdiscip. Topics 53, 4187–4190.
- Waters, R.L., Mitchell, J.G., 2002. Centimeter-scale spatial structure of estuarine in vivo fluorescence profiles. Mar. Ecol., Prog. Ser. 237, 51–63.
- Waters, R.L., Mitchell, J.G., Seymour, J.R., 2003. Characterization of two-dimensional phytoplankton patchiness sampled at millimetre and centimetre scales. Mar. Ecol., Prog. Ser. 251, 49–58.
- Wickham, J.D., Norton, D.J., 1994. Mapping and analyzing landscape patterns. Landsc. Ecol. 9, 7–23.
- Wolk, F., Seuront, L., Yamazaki, H., Leterme, S., 2004. Comparison of biological scale resolution from CTD and microstructure measurements. In: Seuront, L., Strutton, P.G. (Eds.), Handbook of Scaling Methods in Aquatic Ecology: Measurement, Analysis, Simulation. CRC Press, Boca Raton, pp. 3–15.
- Zipf, G.K., 1949. Human Behavior and the Principle of Least Effort. Hafner, New York.