Generalized entropy indices to measure $\alpha-$ and $\beta-$diversities of macrophytes

H. B. A. Evangelista*, S. M. Thomaz*, R. S. Mendes†, and L. R. Evangelista†
*Departamento de Biologia (Nupélia) and †Departamento de Física and National Institute of Science and Technology for Complex Systems, Universidade Estadual de Maringá, Avenida Colombo, 5790 - 87020-900 Maringá, Paraná, Brazil

(Received on 17 February, 2009)

A family of entropy indices constructed in the framework of Tsallis entropy formalism is used to investigate ecological diversity. It represents a new perspective in ecology because a simple equation can incorporate all aspects of $\alpha-$diversity, from richness to dominance and can be also related to a measure of species rarity. In addition, a generalized Kullback-Leibler distance, constructed in the framework of a nonextensive formalism, is recalled and used as a measure of $\beta-$diversity between two systems. These tools are applied to data relative to the macrophytes collected from two not far apart arms of Itaipu Reservoir, in Paraná River basin.

Keywords: Diversity indices, Ecology, Generalized entropy.

1. INTRODUCTION

Biodiversity is a central issue in ecology and biological conservation and has attracted the attention of ecologists and conservationists especially over the last two decades [1, 2]. The number (richness) of species has been largely used as a measure of biological diversity [3]. However, several indices were created to measure diversity, such as the Shannon-Wiener index (H) [4, 5], derived from information theory, and the Simpson index (D) [6], derived from probability theory. Both indices have been widely used since the 50's and they are appealing because they summarize, in a single number, information on species richness (S) and species evenness (E). However, controversy on which index should be used exists in the literature, because diversity (and its broader meaning "biodiversity") can not be fully captured in a single number [7]. For these reasons, the introduction of a broader way to quantitatively measure diversity is tempting. In this direction, we agree with the idea that generalized diversity indices are superior to the traditional, one-dimensional ones, which are responsible for point descriptions of ecological assemblages, as discussed in Refs. [8–10]. Consequently, it is necessary to explore the potentialities of generalized entropy indices within the ecological contexts in more details.

In this paper we first review the most common families of entropy indices that have been used in ecology in the last decades, with particular emphasis on the parametric Patil and Taillies indices (known in statistical physics as Tsallis entropy). This discussion puts in evidence how Tsallis entropy highly employed in physics and other areas of inquiry can be useful in ecology. Along with this family of indices, we will use a recently proposed index developed in the framework of this family, to explore its characteristics measures of diversity in ecology with applications to aquatic macrophytes. The term "aquatic macrophytes" refers to a diverse group of aquatic photosynthetic organisms, all large enough to be seen with bare eyes [11]. Macrophytes are important in aquatic ecosystems, because they provide food and habitat for a variety of organisms, and interfere in the ecosystem functioning [12–14]. Reservoirs are usually suitable for macrophytes development and they can deteriorate multiple uses of these artificial environments (navigation, water sports and losses in the generation of electricity). For this reason, in the Itaipu Reservoir as well as in other similar systems, investigation of the diversity of macrophytes has to be correctly addressed.

The data collected in two arms of the Itaipu Reservoir are firstly analyzed to investigate $\alpha-$diversity in the framework of Tsallis entropy indices, as extensively discussed in Ref. [15]. A further development of these ideas is the exploration of the concept of $\beta-$diversity. To accomplish this task, we use the generalized Kullback-Leibler distance, as given by Patil and Taillie [16] and Borland et al. [17]. In some sense, this follows the program started by Gorelick to extend Shannon's and Simpson's indices to simultaneously account for species richness and relative abundance, but now we consider a continuous family of q values [18].

This paper is organized as follows. In Sec. 2, we present in a summarized way the most common families of entropy indices that are also known in ecological applications, even if they have not been applied until now as intensively as they should be. The emphasis lies on the family constructed in the framework of Tsallis entropy formalism. In Sec. 3, we recall the definition of an alternative index, also based on Tsallis entropy, relevant to diversity and evenness. These indices are applied to a dataset (presented in Sec. 3.1) to quantitatively investigate the $\alpha-$diversity of macrophytes. In Sec. 4, we review the derivation of the generalized Kullback-Leibler information gain and discuss its relevance for the analysis of $\beta-$diversity. The mathematical tools built in this framework are applied to a dataset obtained not far apart in the Itaipu Reservoir as a measure of the dissimilarities among samples. Finally, some concluding remarks are drawn in Sec. 5.

2. GENERALIZED ENTROPY INDICES IN ECOLOGY

The use of families of indices have a long history in ecology [8, 19]. The first family was proposed by Rényi (1961), who extended the concept of Shannon's entropy by defining the entropy of order α, originating a family of $\alpha-$diversity indices [19], in the form

$$R(\alpha) = \frac{1}{1-\alpha} \ln \sum_{i=1}^{W} p_i^{\alpha}, \quad \text{for} \quad \alpha \geq 0 \quad \text{and} \quad \alpha \neq 1, \tag{1}$$

where p_i is the probability of the state i and W is the number of states. The limit $\alpha \to 1$ recovers the Shannon's entropy. Hill (1973) proposed a family of diversity indices that may be interpreted as mean numbers of species [10, 21]. These numbers can be written in an entropic form as

$$N_\alpha = \left(\sum_{i=1}^{W} p_i^\alpha\right)^{1/(1-\alpha)}, \quad \text{for} \quad \alpha \geq 0 \quad \text{and} \quad \alpha \neq 1. \qquad (2)$$

In the same direction, Daróczy (1970) and Aczél and Daróczy (1975) also proposed an entropy family of type α as [22, 23]:

$$H^\alpha = \frac{1}{2^{1-\alpha}-1}\left(\sum_{i=1}^{W} p_i^\alpha - 1\right), \quad \text{for} \quad \alpha \geq 0 \quad \text{and} \quad \alpha \neq 1. \qquad (3)$$

It is easy to show that the Shannon's entropy is a limiting function of H^α when $\alpha \rightarrow 1$. Patil and Taillie (1979) proposed a parametric diversity index family β, in the form [24]

$$\Delta_\beta = \frac{1}{\beta}\left(1 - \sum_{i=1}^{W} p_i^{\beta+1}\right), \quad \text{for} \quad \beta \neq 0 \quad \text{and} \quad \beta \geq -1. \qquad (4)$$

The Patil and Taillie's indices have been intensely studied recently in the context of statistical physics [25–27] and are usually written in the form:

$$S_q = \frac{1 - \sum_{i=1}^{W} p_i^q}{q-1}, \qquad (5)$$

known as Tsallis entropy, where $q = \beta + 1$ is a real parameter which is considered non-negative to ensure that S_q is concave. Motivated by these studies, Keylock (2005) explored these families of indices in an ecological context. A critical discussion of these indices was made by Jost (2006) [10]. Recently, a new index, S_{q*}, was introduced as a unified way to measure ecological diversity and species rarity [15]. It is based on Patil and Tallies's indices and the corresponding evenness. This family of indices (based on Tsallis entropy) captures multiple aspects of biodiversity and provides a better perspective that goes beyond the indices currently used in ecology. From this family, special diversity and evenness indices that balance commonness and rareness, a practice still unemployed by ecologists, was proposed [15].

As a family of diversity indices, Eq. (5) interpolates the well known Simpson ($q = 2$),

$$S_2 = 1 - \sum_{i=1}^{W} p_i^2, \qquad (6)$$

and Shannon- Wiener indices ($q \rightarrow 1$),

$$S_1 = -\sum_{i=1}^{W} p_i \ln p_i. \qquad (7)$$

In general, each specific application of the entropy S_q requires the determination of a particular value of q. This is not an easy task, especially when dealing with statistical mechanical systems. On the other hand, a desirable measure of diversity has to take all the relevant aspects that characterize ecological systems into account, from richness to species dominance. Along these lines, when the possible values of q are considered, S_q becomes a family of diversity indices because it embodies and accounts for the fundamental properties of the usual diversity indices in a simple and unified way. For instance, besides incorporating H and D, the Tsallis entropy can be used as a measure of richness because when $q = 0$, $S_1 = S - 1$, with $S = W$ and $p_i^0 = 1$, for $p_i \neq 0$.

To end this section, it is necessary to emphasize, as we did before, that even if the idea of a family of indices and the indices themselves are known in ecology, they surely were scarcely applied in the last decades.

3. THE ALTERNATIVE INDEX S_q*

As underlined above, S_q represents a parametric family of indices labeled by q, with some limiting values representing well-known indices that measure biological diversity. Similarly to what happened with S_q, it is possible to introduce a family of evenness indices also labeled by q [9]. This family is defined as

$$E_q = \frac{S_q}{S_q^{\max}}, \qquad (8)$$

where

$$S_q^{\max} = \frac{1 - W^{1-q}}{q-1}$$

represents the maximum value of S_q when the constraint $\sum_{i=1}^{W} p_i = 1$ is imposed. As before, the limiting cases of evenness can be obtained by considering particular values of q [15]. These indices will be used in this section to analyze part of the data described below.

3.1. The dataset

The Itaipu Reservoir, a major impoundment of the Upper Paraná River located on the Brazil-Paraguai border, is colonized by a rich assemblage of aquatic plants. From January 2001 to July 2007, two arms located along the reservoir were studied (São João and São Vicente). Sixty stands (30 per arm) were surveyed from a boat, at constant and low velocity. In each stand, two independent samplers spent 10 min observing or collecting aquatic macrophytes for identification. To represent the time employed in the analysis presented in this paper, we fixed the time for the first sample (made in both arms in January 2001) as $t_1 = 0$ month. After that, 13 other samplings were made (described in Table I), just to give an idea of the oscillation between summer and winter in the collects. The relative abundance (p_i) of each species was measured as:

$$p_i = \frac{n_i}{\sum_{i=1}^{S} n_i},$$

where n_i is the number of stands in which the species i was recorded.

In Fig. 1, the evenness introduced in Eq. (8) is shown as a function of the parameter q for illustrative samplings collected along three consecutive years in the São Vicente arm

TABLE I: Periods in which the samples were collected in two arms of the Itaipu Reservoir: São Vicente River and São João River. The time for the first sample was fixed, for reference, as $t_1 = 0$ and the others are given in months on the right column.

Sampling	Time (months)
January, 2001	$t_1 = 0$
June, 2001	$t_2 = 5$
January, 2002	$t_3 = 11$
August, 2001	$t_4 = 18$
February, 2003	$t_5 = 24$
July, 2003	$t_6 = 29$
January, 2004	$t_7 = 35$
July, 2004	$t_8 = 41$
January, 2005	$t_9 = 47$
July, 2005	$t_{10} = 53$
February, 2006	$t_{11} = 60$
July, 2006	$t_{12} = 65$
January, 2007	$t_{13} = 71$
July, 2007	$t_{14} = 77$

(see Sec. 3.1). The trend of the curve is the expected one, showing the existence of a minimum that defines $q*$. Each sample, corresponding to a given arm, is associated to a minimum. The temporal variation of $q*$ is depicted in Fig. 2 for all samples shown in Table I. It is remarkable that $q*$ remained below the value $q = 1$ in every month, exhibiting oscillations (minima in summer and maxima in winter). Notice that this point $q = q*$ defines the maximum deviation of the perfect equitability ($E_q = 1.0$) of a given sample. For this reason and because the negative relationship between $q*$ and S was demonstrated, the index E_{q*}, defined for $q = q*$, was interpreted as a parameter associated with species rarity[15].

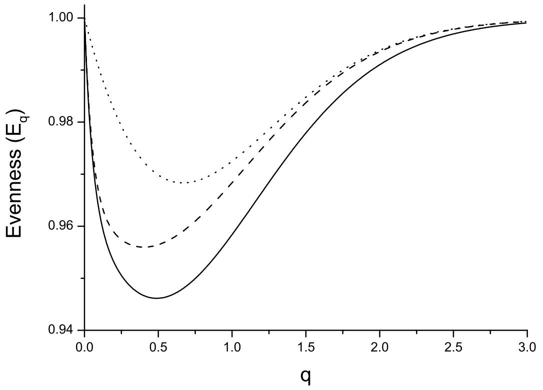

FIG. 1: Evenness indices E_q versus q for the data described in Sec. 3.1: January, 2001 (dotted line), January, 2002 (solid line), and February, 2003 (dashed line). The minima in these curves correspond to $q = q*$.

For comparative purposes, in Fig. 3 E_{q*} and the usual evenness E (i.e., the evenness related to the Shannon's index) are shown for São Vicente River along time.

The absolute values of these indices are clearly different, as expected, because they correspond to different values of q, but their values were significantly correlated (Pearson = 0.66). However, the new index E_{q*} enhanced the temporal variations characterizing the sample. This behavior reinforces the conclusion that E_{q*} can be helpful as a new index, to be added to the classical ones, in order to form a family of indices and to achieve a better description of the diversity.

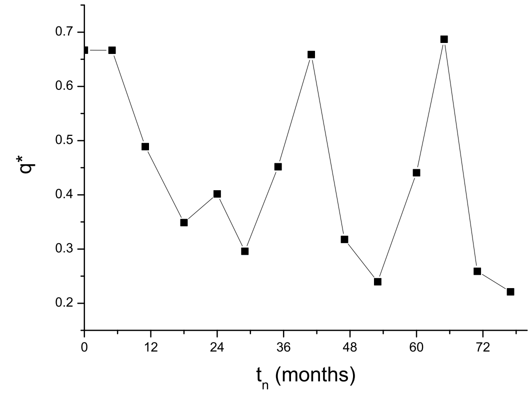

FIG. 2: $q*$ versus the t_n (in months) corresponding to the samples described in Sec. 3.1. A continuous line in this and subsequent graphs was drawn to visualize trends.

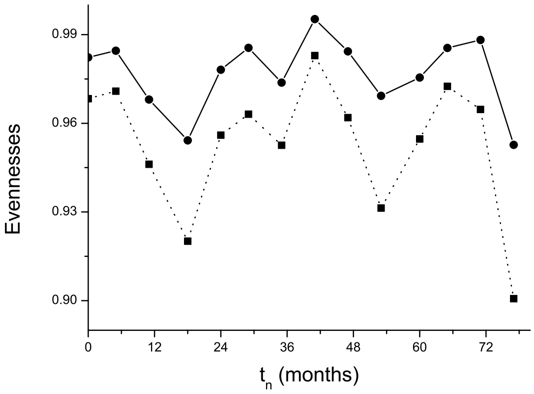

FIG. 3: Evennesses indices E_{q*} (dotted line) and E_1 (solid line) versus t_n for the dataset described in Table I.

In Fig. 4, the diversity indices corresponding to the data of Table I are shown. The new index, S_{q*} (Fig. 4b) is shown along with the classical ones of Shannon and Simpson (Fig. 4a). Again, it is remarkable that S_{q*} is the one that best evidences the oscillations and the marked variations along time. In fact, the Simpson index (S_2) seems to remain essentially constant ($CV = 0.064$) whereas the Shannon index (S_1) oscillates with low amplitude ($CV = 0.339$) as compared with S_{q*} ($CV = 8.99$).

4. THE GENERALIZED KULLBACK INFORMATION GAIN

As stated before, the Shannon's entropy can be obtained as a particular case of Eq. (5) when $q \to 1$, thus yielding Eq. (7), which can be rewritten as

$$S = \sum_{i=1}^{W} p_i I_i, \tag{9}$$

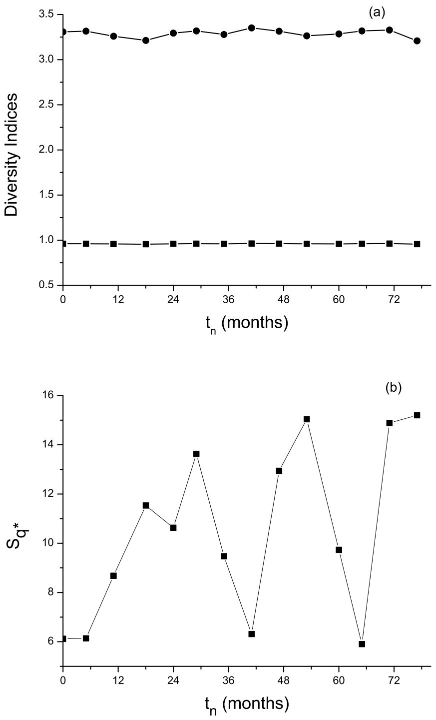

FIG. 4: Diversity indices versus t_n for the data of Table I: (a) the classical indices $S_1 = \mathtt{H}$ (Shannon, circles) and $S_2 = \mathtt{D}$ (Simpson, squares) and (b) the new index, S_{q*}.

where $I_i = -\ln p_i$, which is the information content of outcome i. The change of information, ΔI_i, between two sets of measurements can be defined as

$$\Delta I_i = -\left(\ln p'_i - \ln p_i\right), \tag{10}$$

where p'_i denote the first set of measurements and p_i the new set of measurements. The so-called Kullback-Leibler information gain or *relative entropy* is defined as [17]:

$$K(p,p') = \sum_{i=1}^{W} p_i \Delta I_i = \sum_{i=1}^{W} p_i \ln \frac{p_i}{p'_i}. \tag{11}$$

Similarly, it is possible to show that a generalized Kullback-Leibler measure follows naturally from the standard deviation of the Kullback entropy by employing the nonextensive formalism. One obtains [17]:

$$\begin{aligned} K_q(p,p') &= \sum_{i=1}^{W} \frac{p_i^q}{1-q}\left(p_i^{1-q} - {p'}_i^{1-q}\right) \\ &= \sum_{i=1}^{W} \frac{p_i}{1-q}\left[1 - \left(\frac{p_i}{{p'}_i}\right)^{q-1}\right]. \end{aligned} \tag{12}$$

When $q \to 1$, we get Eq. (11); for $q = 0$, $K_q(p) = 0$. It is possible also to introduce a distance connected with Simpson's measure ($q = 2$):

$$K_2(p,p') = 1 - \sum_{i=1}^{W} \frac{p_i^2}{{p'}_i}. \tag{13}$$

In ecology, $\alpha-$diversity has been defined as the species diversity within community plots [28]. Thus, the analysis presented in previous section concerns specifically with $\alpha-$diversity. The $\beta-$diversity is, however, defined as the amount of turnover in species composition from one location to another [29]. To investigate $\beta-$diversity some indices have been proposed. Among them, we mention the qualitative Sorensen's index, defined as

$$C_s = \frac{2N}{S_A + S_B}, \tag{14}$$

where N is the number of species found in both sites and S_A and S_B are the number of species found in sites A and B, respectively. The other index is the quantitative Morita-Horn index:

$$C_{mH} = \frac{2\sum_i N_{iA} N_{iB}}{(\sum_i N_{iA}^2/N_A^2 + \sum_i N_{iB}^2/N_B^2) N_A N_B}, \tag{15}$$

where N_A and N_B are the total number of individuals and N_{iA}, N_{iB} are the number of individuals of the ith species in sites A and B, respectively. These are similarity indices, i.e., they can be used to quantify community changes due to natural succession or environmental perturbations [30].

In another direction, dissimilarity measures between communities have been proposed [29]. In particular, Tóthmérész suggested summarizing $\beta-$diversity using the distribution of plot-to-plot dissimilarities within a vegetation sample instead of scalars [31]. Some years before, Wilson and Mohler [32] proposed to measure compositional change along gradients with data on species relative abundance, by means of the "gradient rescaling method". By expanding on an idea of Tóthmérész [31], Ricotta and Avena [29] proposed to characterize $\beta-$diversity for data on species relative abundances based on the distribution of Kullback's information-theoretical distance of single plots from the pooled sample.

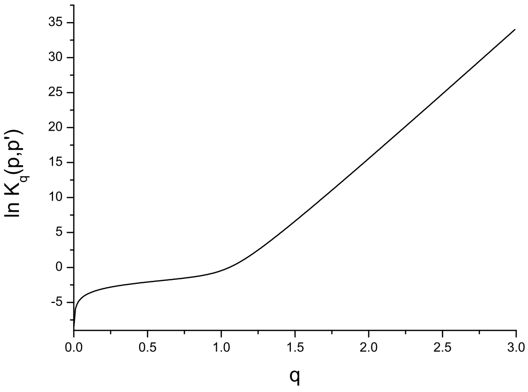

FIG. 5: Logarithm of the Kullback-Leibler distance vs. q. The two sites are arms of the Itapiu Reservoir: São Vicente and São João.

As a natural extension of this idea, in this paper we propose to quantitatively investigate some aspects of the $\beta-$diversity by using the generalized Kullback-Leibler distance, Eq. (12), introduced above. To illustrate the usefulness of this approach, we applied this measure to the dataset presented in Sec. 3.1. We used two close arms of the Itaipu Reservoir just to emphasize the power of K_q to enhance the dissimilarities among these two (in principle) similar communities (as shown in Fig. 5). In Fig. 6, the logarithm of the Kullback-Leibler distance correspondent to the Shannon and Simpson indices is shown.

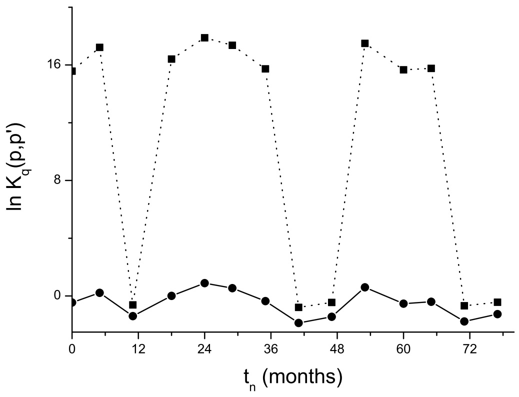

FIG. 6: Logarithm of the Kullback-Leibler distance vs. time for Shannon $q \to 1$ (circles) and Simpson $q = 2$ (squares) limits.

Both indices have the same overall trend. However, the distance connected with $q = 2$ (Simpson) enhanced the differences between the two systems. On the other hand, it is clear that the similarity between the two sites is low for January 2002, January 2005, and January 2007. This is the summer period, in which the two systems present similar diversity. However, in the winter, represented by the months June 2001, July 2003, and July 2005, the values of $K_2(p,p')$ are high, indicating that the species found in these arms are different. These preliminary results confirm the potentiality of a generalized distance to be associated with the $\beta-$ diversity investigations in ecology.

5. CONCLUDING REMARKS

As we have emphasized above, the idea of a family of diversity indices is not novice in ecology, but it has not been explored in all its potentialities. After reviewing the main families proposed in this endeavor, we used a family of indices constructed in the framework of the Tsallis's entropy formalism to investigate diversity of aquatic macrophytes in two arms of the Itaipu Reservoir. To accomplish this task, we used the new indices linked with this formalism. In particular, we used an index connected with the special value of $q = q*$, the parameter characterizing this family of entropy indices, whose meaning still deserves investigation. We found that the indices associated to $q*$, i.e., the value for which the evenness presented its maximum deviation from the perfect equitability, can be particularly useful as an additional information to explore $\alpha-$diversity in ecological samples. In addition, we also discussed the possible role of a Kullback-Leibler distance in investigations of $\beta-$diversity. In this sense, this paper represents a step further to carry on a detailed investigation of a broad aspect of diversity, at least for two fundamental reasons: i) it demonstrates the applicability of an unified tool to describe $\alpha-$diversity by means of a set of unified parameters, embodying diversity from richness to dominance and species rarity; ii) it permits us to face aspects of $\beta-$diversity in the same framework and with the same mathematical tools used for $\alpha-$diversity. We are then convinced that this approach, based on a nonextensive formalism, can be relevant not only in the framework of statistical physics but will also find broad applications in ecological systems.

Acknowledgments

Many thanks are due to the Brazilian agencies, CNPq and Capes, for partial financial help. H. B. A. Evangelista thanks PDTA/PTI Parque Tecnológico de Itaipu for a fellowship.

---

[1] E. O. Wilson and F. M. Peter (eds), *Biodiversity* (National Academic Press, Washington DC, 1988).
[2] G. Moorel and E. O. Wilson, Nature **405**, 254-256 (2000).
[3] A. E. Magurran, *Ecological diversity and its measurement* (Princeton University Press, Princeton, 1988).
[4] C. E. Shannon, Bell Syst. Technol. J. **27**, 379 (1948).
[5] C. E. Shannon and E. W. Weaver, *The mathematical theory of communication* (University of Illinois, 1949).
[6] E. H. Simpson, Nature **163**, 688 (1949).
[7] A. Purvis and A. Hector, Nature **405**, 212-219 (2000).
[8] G. Lövei, Community Ecology **6**, 245, 2005.
[9] C. J. Keylock, Oikos **109**, 203 (2005).
[10] L. Jost, Oikos 113, 363 (2006).
[11] P. A. Chambers, P. Lacoul, K. J. Murphy, and S. M. Thomaz, *Hydrobiologia* **595**, 9 (2008).
[12] R. G. Wetzel, *Limnology* (Saunders College Publishing House, Philadelphia, 1983).
[13] C. M. Duarte, D. Planas, and J. Pañuelas, *Macrophytes: taking control of an ancestral home*, in R. Margalef, (ed). *Limnology now: a paradigm of planetary problems* (Elsevier Science, Amsterdam, 1994).
[14] F. Esteves, *Fundamentos de Limnologia* (Interciência, FINEP, Rio de Janeiro, Brasil, 1998).
[15] R. S. Mendes, L. R. Evangelista, A. A. Agostinho, S. M. Thomaz, and L. C. Gomes, Ecography **31**, 450 (2008).
[16] G. P. Patil and C. Taillie, J. Am. Stat. Soc. **77**, 548 (1982).
[17] L. Borland, A. R. Plastino, and C. Tsallis, J. Math. Phys. **39**, 6490 (1998).
[18] R. Gorelick, Ecography **29**, 525 (2006).
[19] B. Tóthmérész, J. Veg. Sci. **6**, 283 (1995).
[20] A. Rényi, *On measures of entropy and information*. In: *4^{th} Berkeley Symposium on Mathematical Statistics and Probability* (Berkeley, USA, 1961) edited by J. Neyman.
[21] M. O. Hill, Ecology **54**, 427 (1973).

[22] Z. Daróczy, Inform. Control. **16**, 36 (1970).

[23] J. Aczél and Z. Daróczy, *On Measures of Information and their Characterizations* (Academic Press, New York, 1975).

[24] G. P. Patil and C. Taillie, *An overview of diversity*. In: *Ecological Diversity in Theory and Practice* (International Cooperative Publishing House, Fairland, MD, 1979), edited by J. F. Grassle, G. P. Patil, W. Smith, and C. Taillie.

[25] C. Tsallis, J. Stat. Phys. **52**, 479 (1988).

[26] C. Tsallis, R. S. Mendes, and A. R. Plastino, Physica A **261**, 534 (1998).

[27] C. Tsallis, Braz. J. Phys. **39**, 337 (2009).

[28] R. H. Whittaker, Ecological Monographs **30**, 279 (1960).

[29] C. Ricotta and G. Avena, Plant Biosystems **137**, 57 (2003).

[30] A. Ludovisi and M. I. Taticchi, Ecological Modelling **192**, 299 (2006).

[31] B. Tóthmérész, Abstracta Botanica **22**, 149 (1998).

[32] M. W. Wilson and C. L. Mohler, Vegetario **54**, 129 (1983).