Ecological Indicators 15 (2012) 198–207

Contents lists available at SciVerse ScienceDirect

Ecological Indicators

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

Underlying thermodynamic relations of a species diversity index: Freshwater crabs from Colombia

Diógenes Campos[a,*,1], Martha R. Campos[b,1]

[a] Faculty of Natural Sciences and Engineering, Universidad "Jorge Tadeo Lozano", Bogotá, Colombia
[b] Instituto de Ciencias Naturales, Universidad Nacional de Colombia, Bogotá, Colombia

ARTICLE INFO

Article history:
Received 5 March 2011
Received in revised form 14 August 2011
Accepted 19 September 2011

Keywords:
Probability distribution
Thermodynamic description
Freshwater crabs
Biological diversity
Biodiversity index

ABSTRACT

The goal of this paper is to apply the theoretical method described in Campos (2011) to the characterization of an ecological system represented by data corresponding to a collection of freshwater crabs consisting of 100 species and 6397 specimens recorded in Colombia between 1910 and 2010.

The registers in the collection are translated into a set of probabilities (relative abundances) $P = \{P_1, P_2, \ldots, P_{\mathcal{N}}\}$, where $\mathcal{N}$ is the species richness. The index for species diversity is calculated as the product of a function of the species richness times a function of the Helmholtz free energy $\mathcal{F}(P, q)$, where the entropic parameter q allows the system to be considered under various hypothetical environmental conditions.

1. Introduction

In this paper, a novel methodology for the characterization of an ecological system is proposed by combining a thermodynamic-like approach for the study of probabilistic systems with a geometrical index for measuring species diversity (Campos, 2011; Campos and Isaza, 2009). The usefulness of the method in ecology is illustrated by using a set of registers for a collection of freshwater crabs consisting of 6397 specimens from the Pseudothelphusidae (85 species) and Trichodactylidae (15 species) families, and transforming the data into probability distributions (species relative abundances), i.e., $P = \{P_1, P_2, \ldots, P_{\mathcal{N}}\}$, where $\mathcal{N}$ is the species richness. The method allows to treat incomplete sampling ($P_1 + P_2 + \cdots + P_{\mathcal{N}} \leq 1$), and also to simulate the effects of various hypothetical ecological conditions by adding a nonnegative real exponent ($q \geq 0$) to the probabilities, in the form $(P_1)^q + (P_2)^q + \cdots + (P_{\mathcal{N}})^q$. The geometrical index for measuring biological diversity distinguishes two contributions (see Eq. (12)), the first is associated with the species richness $\mathcal{N}$, and the second is due to the relative abundance of them (it is also called evenness, equitability or dominance). We are also able to get a visual and geometrical representation of biodiversity (index $\mathcal{B}_k$ or $\tilde{\mathcal{B}}_k$) by using cartesian bidimensional planes (see Eqs. (9) and (10)), either the *biodiversity plane* $r - \mathcal{B}_k$ or the *entropic plane* $q - \tilde{\mathcal{B}}_k$, where the effective radius $r(P, q)$ describes the relative abundance of species, $0 < r(P, q) \leq 1$.

The physical approach is enhanced with a biological perspective in which we estimate the biological diversity of the Colombian freshwater crabs by taking advantage of the geometrical index for measuring species diversity that was proposed in this journal by Campos and Isaza (2009). Since the freshwater collection represents an imperfect "photo" of the species diversity of the natural system, we can only roughly estimate the biological diversity of the real system, i.e., Colombia and its geographical regions. Despite this fact, the collection captures the most complete information available on the freshwater crabs from Colombia.

For the purposes of this work and hereafter, we will use the shorthand ICN *collection* when referring to the Crustacean collection of the Instituto de Ciencias Naturales (ICN), Universidad Nacional de Colombia, Bogotá. It includes specimens of Pseudothelphusidae (abbreviated as PS) and Trichodactylidae (TR) families.

The structure of the paper is the following. Section 2 summarizes the method proposed by Campos (2011) for the study of probabilistic systems. Section 3 describes the ecological system of our interest: namely, the ICN collection of freshwater crabs. In Section 4, some intermediate calculations are explained so that the emphasis is on the ecological meaning of the involved quantities. In Section 5, the freshwater crabs diversity in Colombia and their geographical regions is evaluated. Section 6 includes some conclusions and a short discussion.

* Corresponding author. Tel.: +57 1 6194054.
E-mail addresses: dcamposr@utadeo.edu.co, dcamposr@cable.net.co (D. Campos), mhrochad@unal.edu.co (M.R. Campos).
[1] Members of the Colombian Academy of Sciences.

1470-160X/$ – see front matter © 2011 Elsevier Ltd. All rights reserved.
doi:10.1016/j.ecolind.2011.09.020

2. Macro-description of a probabilistic system

In this section, we summarize the thermodynamic-like method described by Campos (2011). As a general guide for understanding the relations between the quantities used in this paper, we refer the reader to Fig. 1. The starting point is a probability distribution for $\mathcal{N}$ species, $P = \{P_1, P_2, \ldots, P_{\mathcal{N}}\}$, where the n-th species is represented by the probability P_n (relative abundance). The degree of completeness of P is given by $\sum_{n=1}^{\mathcal{N}} P_n = \omega_{\mathcal{N}}(P)$, i.e., P can be complete ($\omega_{\mathcal{N}}(P) = 1$) or incomplete ($\omega_{\mathcal{N}}(P) < 1$).

2.1. Three basic quantities

Associated with P, let us define three basic quantities (Fig. 1):

(a) The set $\mathcal{E}(P) := \{\mathcal{E}_1(P), \mathcal{E}_2(P), \ldots, \mathcal{E}_{\mathcal{N}}(P)\}$, where $\mathcal{E}_n(P) := -\ln P_n$ is the *Hartley information* provides by the observation of the n-th species, and ln is natural logarithm (units of information in nats). We recall that the logarithm of a number depends on the base b of logarithm used ($\ln = \ln_b$): the logarithm to base $b = 2$ is called bit (a contraction of binary digit) and it is the amount of information stored by a system that exists in one of two possible distinct states, and the logarithm to base $b = e \approx 2.71828$ is called nat (a contraction of natural digit); i.e., one nat corresponds to about $\ln_2 e \approx 1.4427$ bits. The quantity $\mathcal{E}_n(P)$ is also called *surprise* or *pseudo-energy*: the lower the probability P_n, the higher the surprise when we learn of the occurrence of species n.

(b) The *partition function*

$$Z_{\mathcal{N}}(P, q) := \sum_{n=1}^{\mathcal{N}} (P_n)^q = \sum_{n=1}^{\mathcal{N}} \exp(-q\mathcal{E}_n(P)), \tag{1}$$

where the *entropic parameter* q is a nonnegative real number, $q \geq 0$ (Tsallis, 2009). Eq. (1) has been used in the developments of several diversity indices, see, e.g., Mathai and Rathie (1975), Rényi (1960), and Tsallis (2009).

(c) The *escort probability* set of P, $p(P, q) := \{p_1(P, q), p_2(P, q), \ldots, p_{\mathcal{N}}(P, q)\}$, that is obtained by rewriting Eq. (1) in the form

$$\sum_{n=1}^{\mathcal{N}} p_n(P, q) = 1, \quad \text{with} \quad p_n(P, q) := \frac{(P_n)^q}{Z_{\mathcal{N}}(P, q)}. \tag{2}$$

Notwithstanding that P can be incomplete or complete, the set $p(P, q)$ is complete for all $q \geq 0$. The escort probability $p_n(P, q)$ is an effective probability for the n-th species that has the real probability P_n. This concept has its origin in the non-extensive statistical mechanics (Tsallis, 2009).

2.2. Three thermodynamic potentials

Now, by using the probability set $p(P, q)$ we define the expectation values (Campos, 2011), see Fig. 1:

- The average value of the Hartley information set (or average surprise),

$$\mathcal{U}(P, q) := \sum_{n=1}^{\mathcal{N}} p_n(P, q)\, \mathcal{E}_n(P). \tag{3}$$

- The Shannon entropy for the escort probability set,

$$S(P, q) := -k_{\mathrm{B}} \sum_{n=1}^{\mathcal{N}} p_n(P, q) \ln p_n(P, q), \tag{4}$$

  where k_{B} is a positive constant (choosing an appropriate unit we can set it to one). $S(P, q)$ measures the amount of information which is missing due to the probabilistic nature of the available information (relative abundances of $\mathcal{N}$ species).

- The Helmholtz *free energy* defined by

$$\mathcal{F}(P, q) := \mathcal{U}(P, q) - \frac{1}{k_{\mathrm{B}} q} S(P, q) = -\frac{1}{q} \ln Z_{\mathcal{N}}(P, q), \quad q > 0. \tag{5}$$

As far as the biodiversity indices considered in Section 5.1 are concerned, $\mathcal{F}(P, q)$ alone is the important function, because $\mathcal{U}(P, q)$ and $S(P, q)$ are only intermediate functions for the definition of the free energy.

3. Collection of freshwater crabs from Colombia

3.1. Geographical regions of Colombia

Colombia is a country located in Northern South America, bordering the Caribbean Sea (~1600 km) and the Pacific Ocean (~1300 km). There are three chains of high mountains (cordilleras) that run from south to northeast, and divide the country into five geographic regions (Fig. 2, left): Andean (~305,000 km^2), Pacific (~83,170 km^2), Caribbean (~132,218 km^2), Amazonian (~403,348 km^2, predominantly jungle) and Orinoquian (~310,000 km^2, predominantly plains). The Caribbean region is lowland, but includes also the Sierra Nevada de Santa Marta, an isolated mountain system with peaks reaching heights over 5700 m.

Freshwater crabs are represented in the neotropics by two endemic families: the Pseudothelphusidae (PS) with approximately 260 known species, and the Trichodactylidae (TR) with nearly 50 known species. From these two families, 85 (PS) and 15 (TR) species have been recorded in Colombia.

Since the Pacific region merges into and is connected with the Andean region without having a definite border, we make no difference between the data from the two regions and merge them into the data for the Andean region (see Table 1). Fig. 2 (right) shows the number of specimens accumulated in the ICN collection: 100 species and 6397 specimens recorded in Colombia between 1910 and December 2010.

The study of freshwater crabs of Colombia is important from a biological, social, taxonomic, and geographical points of view. Freshwater crabs perform an essential function in aquatic ecosystems since they constitute an important element in the trophic chain. They contribute to the acceleration of the decomposition of organic material, and they are also bio-indicators of non-contaminated water. In many regions of Colombia freshwater crabs are an important component of the local diet (Campos, 2005).

3.2. Probability distributions

As a facility for avoiding taxonomic names, the species in the ICN collection are enumerated from 1 to 100: #1 to #85 correspond to Pseudothelphusidae species, and #86 to #100 to the Trichodactylidae species (see Table A.5). For this study, the data is a set of registers, where each register includes the following information: number assigned to the species (#), name of the species, geographical region, numbers of collected specimens, and the year in which the specimens were collected (see Appendix A).

The registers in the collection are translated into five probability distributions of the form $P = \{P_1, P_2, \ldots, P_{\mathcal{N}}\}$, where $\mathcal{N}$ is the species richness, and neither of the $\mathcal{N}$ elements are null. The probability $P_n = N_n/N_0$ is the proportional abundance of the nth species, where N_n is the number of specimens for the n-th species and N_0 is the total number of specimens in the sample. Different sample spaces can be chosen: PS species alone (P1), TR species alone (P2), whole