Biological communities as interacting compartments: thermodynamic properties and diversity indices

Fernando Meloni,[1] Gilberto M. Nakamura,[1] and Alexandre Souto Martinez[1]
Universidade de São Paulo, Ribeirão Preto 14040-901, Brazil

(Dated: today)

Diversity indices provide simple and powerful metrics for assessing biological communities. Based on entropy measures, the approach considers statistical and thermodynamic inferences to deduce ecological patterns. However, concerns exist regarding the accuracy of diversity indices. Because relative quantities depend on the sorting of organisms (*e.g.*, guilds and species) and their interactions, field observations carry inherent imprecision, thus leading to misinterpretation. Here, we present a framework that is able to appropriately achieve the thermodynamic properties in ecological systems and ensure the inference power. We demonstrate that effective abundances rather than raw abundances provide a trustful estimator of probabilities, which is evaluated through massive tests. We use empirical and synthetic data to show the advantages and reliability of this new framework under a broad range of conditions. The tests demonstrate that the replication principle is always optimized by the new estimator. Compared to other methods, this approach is simpler and reduces the importance of schemes used for sorting organisms. We highlight the robustness and the valor of effective abundances for ecological contexts: *i)* to assess and monitor the biodiversity, *ii)* to define the best sorting of organisms according to maximum entropy principles, and *iii)* to link local to regional diversity (α-, β-, and γ-diversity).

keywords: biodiversity assessment, entropy, extensive and additive properties, estimator of probabilities, interacting compartments, predictable patterns

I. INTRODUCTION

The assessment of biodiversity is a primary concern among ecologists. They are interested in monitoring species and ecosystems to explain how climate, soil type and several other environmental features affect organisms and their organization in nature[15,50]. Beyond basic knowledge, the motivation for this interest has increased over the past decades due to the increasing human impacts on climate and natural ecosystems that place species survival and ecosystem services at risk[2,26].

The best appraisal of patterns in biological communities considers details of several taxa, such as their biology, genetic variability, ecology, behavior, and so forth. However, this level of information is rarely available for practical contexts, and objective measures are considered along with external inferences to deduce the community patterns[54,55]. Taxonomic/functional compositions, the shape of the curves describing species-abundance distributions (SAD), and the ecological indices consist of distinct approaches for measuring and comparing biological communities[50]. Each approach addresses levels of information, and their use shows advantages and disadvantages for distinct contexts [41,55]. Among these metrics, ecological indices deserve special attention, and we explore them in this paper.

Also referred to as diversity indices, ecological indices are deduced from the Boltzmann-Shannon-Gibbs entropy, and they are proposed to measure order-disorder in biological communities. These indices enjoy advantages from well-supported concepts, interpretations, formulations, and inferences from information theory, physics, statistics, and thermodynamics, and they allow deducing further relationships from data[13,15,16,36,74,78]. Briefly, the calculation of diversity indices considers the number of categories and the number of organisms (abundances) as primary information [50]. The W categories are defined by any criteria of interest (*e.g.*, species, genera, behavior, guild, and so forth) and are used for sorting the organisms, while relative quantities provide a quantitative inference. This type of basic information can be obtained for a broad range of practical conditions and reflects the applicability of diversity indices. Furthermore, diversity indices provide reliable information about global patterns of biological organization even if information about organisms is limited.

Accordingly, once the W categories are determined *a priori*, the abundance A_k considers the number of organisms in the k-th category and determines the probability $p_k = A_k/A$, for $A = \sum_{k=1}^{W} A_k$ and $\sum_{k=1}^{W} p_k = 1$. These probabilities are used to calculate the classic Shannon diversity index H [50] as follows:

$$H = -\sum_{k=1}^{W} p_k \ln p_k, \tag{1}$$

which is equivalent to the Shannon entropy [74]. In practical contexts, the H values of biological communities living under distinct influences are compared by similar sampling effort. In such cases, the variation ΔH is inter-

preted as a metric that addresses the influence of a particular driver (*e.g.*, climate changes or human impacts) on the ecological framework [13,50,53].

In addition to the H index, there are several other ecological indices that provide complementary aspects of the biological communities. The notation

$$^{a}H = \frac{1}{1-a} \ln \sum_{k=1}^{W} p_k^a \tag{2}$$

unifies the three most important ecological indices[13,15,36]: the species richness W (number of categories), the H index (Shannon entropy), and the Ginni-Simpson index $D = \sum_{k=1}^{W} p_k^2$, where the last infers the curve asymmetry of ranked relative quantities. For $a = 0$ and $R \geqslant 2$, we have $^{0}H = \ln(W)$, which is the harmonic mean; for $a \to 1$, $^{1}H = H$, which is the geometric mean; and for $a = 2$, $^{2}H = 1/D$, which is the arithmetic mean. These cases are special cases of the Rényi generalized entropy and are important because they predict useful relationships. One expects to find correlations between the abundances of organisms and the values assumed by distinct diversity indices, such as $^{a}f(A) \equiv {}^{a}H$, which also implies that $g(^{0}H) \equiv {}^{1}H$ and so on [15,36]. These correlations match the concept of the replication principle in ecological contexts[13,15,38], and they are true only if the additive and extensive thermodynamic properties are ensured[13,34,78]. Assuming this circumstance, the entropy observed for small samples can be re-scaled to predict the entropy of the entire system. This type of inference would be used, e.g., to explain how the biological diversity is spatially partitioned (γ-, α- and β- diversities)[12,38]. Although H and other diversity indices provide an interesting approach to assess and interpret biological communities, tests using empirical data often show a different reality, and the topic remains open [10,12,14,24].

The inconsistency between theoretical predictions and empirical results is justified by the lack of a precise definition of appropriate classes for ecological systems[34]. A common approach for sorting organisms considers taxonomic arguments, where individuals are strictly grouped by their phylogeny. This method is intuitive but leads to granular schemes that poorly represent ecological constraints. In fact, taxonomic sorting reflects the evolutionary history of groups better than the immediate effect of environmental forces on niche occupancy [72], often resulting in probabilities p_k that fail in evaluating the diversity patterns [76]. In contrast to taxonomic schemes, the guild approaches employ functional traits and few taxonomic arguments, maximizing the niche context for sorting organisms. The individuals are grouped because they share similar traits and explore similar niches, a scheme that minimizes ecological redundancies. Consequently, guild methods tend to improve the assessment of environmental disturbances on biological communities [9,37,43,72]. Multiple arguments (such as phylogeny, morphological traits, genetic information, behavior, niche occupancy and *ad hoc* information) can be addressed together to elaborate constraints and produce distinct *a priori* hypotheses, and they are compared next in accordance with MaxEnt principles [13,34,35]. The arrangement that better achieves the thermodynamic properties defines the most plausible granular scheme. However, the approach also brings important concerns. For instance, knowledge of tens, hundreds, or sometimes thousands of distinct organisms is a prerequisite for accurate results using guilds. Beyond the necessity of specialists, this implies some subjectivity in selecting the traits and respective weights (if weights are addressed). Finally, guilds are considered for particular groups of organisms and under specific situations (*e.g.*, tree species in forests, arthropods in soil, fishes in lakes, and so on), and they cannot be readily inferred for wide proposes. Therefore, the opinion and context may affect the evaluation of probabilities p_k, the values assumed by the diversity indices, and the subsequent conclusions.

Notwithstanding the importance of the assortment of organisms, we emphasize here that the granular scheme is only one side of the problem for the accuracy for biodiversity assessment. We claim that ecologists should address further sources of data variability before defining the best assortment. Ecological systems admit complex spatial-temporal dynamics (many of which are poorly understood), which finally drives the presence and quantities of organisms. Beyond environmental influences, such dynamics derive from intricate food webs and interaction networks, which finally frame the biological organization in the ecosystems. The consequence is a mutual and non-trivial dependence of relative quantities. Because ecological information is generally acquired through field observations and samplings, fluctuations in relative abundances may accrue relevant "noise" in empirical data[45,51], leading to uncertain probabilities p_k for diversity indices. Consequently, the reproducibility of ecological experiments remains a topic of scientific concern[7].

Taking these arguments into consideration, we conjecture that insights taken from non-equilibrium systems [19,40,52,80] could be adapted to a static approach for reducing the effects of short-range fluctuations on observed quantities. In this paper, we explore the inherent inaccuracy of diversity indices as a consequence of a poor interpretation of raw abundances when used for estimate probabilities. We propose a new estimator that is able to separate data variations produced by entropic and non-entropic drivers, assuming samples as stationary (or as *quasi*-stationary) states. The approach is demonstrated to optimize the results obtained by diversity indices, making them reliable tools for assessing biological organization. The additive and extensive thermodynamic properties of biological communities are recovered from effective abundances, thus allowing a broad range of inferences about time, space, and external influences since the adequate context is considered. Extensive tests with empirical and synthetic data remove any doubt regarding the applicability and reliability of the new estimator of probabilities. This new estimator of probabilities rep-

resents a shift in our understanding about abundances, which is taken as the result of interacting compartments rather than independent ones. Our rationale, formulation and interpretations broadly match the framework presented by Würtz and Annila[80], with differences for the general interpretation and mainly for practical application. We discuss the phenomenological process behind the quantitative changes in the data structure, as well as the potential of this approach to answer classic problems concerning biodiversity patterns.

The remainder of this paper is organized as follows. In Sec. II, we present the rationale used to obtain an alternative estimator of probabilities, which enables optimizing the thermodynamic properties in ecological systems. In Sec. III, we use data of ground arthropods sampled in a successional gradient to show that organism sorting schemes are insufficient for assessing the entropy production in ecological systems, elucidating how effective abundances provide a distinct approach to improve the biodiversity assessment. In the same section, we test the framework in further empirical data, showing that thermodynamic properties are recovered in distinct communities, including plants, arthropods, mammals, fishes and others, that may used to infer regional diversity from few samples and is minimally affected by the shape of SAD. We finally discuss the implications of our findings and the phenomenological aspects concerning the topic in Sec. IV.

II. ENTROPY OF BIOLOGICAL COMMUNITIES AND ESTIMATORS OF PROBABILITIES

Contextualizing the problem, empirical data are generally obtained from limited samples, and the dynamics driving the spatio-temporal patterns of diversity under *in situ* conditions are hardly accessible. As distinct influences on relative abundances, on the one hand, we have the phenomena that alter the free energy in ecosystems and directly modify the ecological framework (*e.g.*, vegetation suppression, ecological succession, climate changes, and so forth). These phenomena disturb ecological systems and lead to a system reorganization, a consequence of the second law of thermodynamics. We name these phenomena entropic influences, or φ. On the other hand, predation, competition, seasonality, and so forth may also affect the quantities of organisms, but these are non-entropic influences, or θ. At homeostasis or any stationary state, such phenomena make abundances fluctuate (regularly or not) over time around a fix point. Despite being thermally neutral, these fluctuations address systematic sub- and/or overestimation of organisms in samples. Consequently, the non-entropic influences impart a high level of uncertainty to A_k values (primarily because of the sampling process) and hinder the assessment of entropy variation. Separating entropic from non-entropic effects on observed abundances is possibly the most important barrier for assessing the effects of environmental changes on biological diversity [11,17,25,32,42,44,71,73,79].

In their theoretical study, Würtz and Annila[80] considered the paths to solve this question. The authors considered ecosystems as systems out of equilibrium, and they interpreted the ecological interactions as chemical potentials. Using ecological succession as context, the authors explicitly deduced the time and matched the thermodynamic entropy to ecological succession. However, the most important insight was the estimator of probabilities, which is assumed as the result of a product. Unfortunately, this topic has remained poorly explored until now. We develop an independent reasoning that widely converges to a similar mathematical framework. Because the concept was previously proven, we spend less time with justifications in order to focus on the applicability and implications. Our approach also has some differences because we address each sample as a stationary state, a static approach that does not demand that time be explicitly considered [20,21]. In fact, the stationarity addressed here is a condition imposed to the sample, and it is not a necessary condition for the system. The approach strictly focuses on variability in abundances, which for methodological and practical concerns enables wider inference power from ΔH because time, space or any other driver can be deduced from appropriate experimental designs. With some contextualization, we next present a step by step description of the alternative probability estimator q_k, which is a function of A_k and φ but not necessarily proportional to them.

Following Haegeman and Loreau[33], the Shannon entropy [Eq. (1)] describes the organization of biological communities in a given ecological system. The main aspect that is necessary to obtain a reliable value of H is the proper definition of groups according to which the organisms are classified. As previously mentioned, there are several ways to group organisms in compartments, and we start with the taxonomic view. At a basic granular level, species define the categories, and the abundance A_k represents the number of organisms in the k-th species of the system, while the total abundance is $A = \sum_{k=1}^{W} A_k$ and W is the species richness. The ratio

$$p_k = A_k/A \tag{3}$$

estimates the occurrence probability of finding a given element of the k-th compartment. Naturally, this definition ensures that $\sum_{k=1}^{W} p_k = 1$.

Despite the intuitive formulation of p_k, applying Eq. 3 to Eq. 1 raises several experimental questions. For instance, small changes in abundance $\Delta A_k = \varepsilon_k A$ with $\varepsilon_k \ll 1$, either due to sampling errors or measurements in different ecosystems, produce corrections $\Delta p_k = \varepsilon_k - p_k(\sum_{\ell=1}^{W} \varepsilon_\ell) + o(\varepsilon^2)$. Since ε_k are assumed to be small, corrections to the estimators themselves are also very small. However, they may produce large corrections

to the entropy,

$$\Delta H = -\sum_{k=1}^{W} \varepsilon_k \ln p_k - H \sum_{k=1}^{W} \varepsilon_k + o(\varepsilon^2). \tag{4}$$

For clarity, let us consider the particular case of a single variation in abundance, i.e., $\varepsilon_k \equiv \varepsilon$. Under this circumstance, $\Delta H = -\varepsilon \sum_k \ln p_k - \varepsilon W H$. Note that in the first parcel, $\varepsilon \sum_k \ln p_k$, compartments with small probability $p_k \ll 1$ provide considerably larger contributions to ΔH than compartments with a larger presence in the ecosystem. This result should be compared against compartments with equal probability $p_k = 1/W$, $\Delta H = \varepsilon W H - \varepsilon W H = 0$. Therefore, large deviations in H arise due to the estimator Eq. (3), making the index sensitive to population dynamics.

Another pressing concern regarding the reliability of Eq. (3) as an ecological index revolves around the organism interactions and their sorting in categories. These interactions are expected to shape the various levels of the biological organization either directly or indirectly. In thermodynamics, even though interactions are known *a priori*, the sheer amount of components in the system prohibits the exact computation of properties. The chemical potential μ overcomes this problem by introducing an effective energy change caused by the addition of another particle to the system, including temperature and pressure. In this way, the chemical potential encodes all resulting effects of underlying interactions in the system, with Gibbs free energy as $G = \mu N$, for N identical particles. For ecological problems, the translation is straightforward: we associate a chemical potential $\mu_k \equiv \mu_k(\varphi, A_1, \ldots, A_W)$ to the k-th compartment such that the total Gibbs free energy is

$$G = \sum_{k=1}^{W} \mu_k A_k - G_0(\varphi), \tag{5}$$

where $G_0(\varphi)$ depends only on entropic variable φ. From Eq. (5), it is clear that the chemical potentials μ_k represent weights to the compartment occupation A_k, from which we infer the estimator

$$q_k = \frac{\mu_k}{\sum_{n=1}^{W} \mu_n} \,. \tag{6}$$

One reobtains the estimator p_k when $\mu_k \propto A_k$, *i.e.*, the effort required to add another organism to the k-th compartment increases linearly with A_k.

For more complex environments, we should define or estimate the values μ_k in some other way. Indeed, from the Maxwell relation $\beta\mu_k = -(\partial H/\partial A_k)$ and $H = -\sum_{k=1}^{W} p_k \ln p_k$ (Eq. (1)) with estimators $p_k = A_k/A$ ($A = \sum_{k=1}^{W} A_k$), one obtains

$$\beta\mu_k^{\text{est}} = \frac{\ln A_k - \ln A}{A} + \frac{H}{A}. \tag{7}$$

It is easy to show that the term $\ln(A_k) - \ln(A)$ in Eq. (6) refers to the H_{max} (maximum entropy possible) only if all relative abundances are the same (Suppl. 1). For ecological systems, further considerations are necessary. The ratio $H/A \ll 1$ for biological communities since $H \sim o(1)$, whereas $A \geqslant o(10^3)$. Moreover, $A^{-1} \ln A$ is a constant for fixed A and only shifts the free energy. From these considerations, we have

$$\mu_k \equiv \frac{1}{\beta A} \ln\left(1 + A_k\right), \tag{8}$$

which in turn produces our estimator corresponding to the k-th compartment

$$q_k \equiv \frac{1}{Z} \ln(1 + A_k), \tag{9}$$

with the partition function $Z = \sum_{j=1}^{W} \ln(1 + A_j)$ guaranteeing that $\sum_{k=1}^{W} q_k = 1$. Comparing distinct estimators and their thermodynamic interpretation, the Gibbs partition function is

$$p_k = \frac{A_k}{A} \equiv \frac{e^{-\beta\varepsilon_k}}{\sum_{n=1}^{W} e^{-\beta\varepsilon_n}} \tag{10}$$

while q_k is

$$q_k = \frac{\ln A_k + 1}{\sum_{n=1}^{W} \ln A_n + 1} \equiv \frac{-\beta\varepsilon_k}{\sum_{n=1}^{W} -\beta\varepsilon_n}. \tag{11}$$

The estimator q_k vanishes for vanishing A_k similar to p_k. However, q_k exhibits notable differences from p_k in terms of the relative variations $\Delta A_j = \bar{\varepsilon}_j(1 + A_j)$ with $\bar{\varepsilon}_j \ll 1$,[1]

$$\Delta q_k = \frac{1}{Z}\left(\bar{\varepsilon}_k - q_k \sum_{\ell}^{W} \bar{\varepsilon}_\ell\right) + o(\bar{\varepsilon}^2). \tag{12}$$

A brief comparison between Eq. (12) and Δp_k reveals an interesting fact. Let us assume that both estimators q_k and p_k share similar numerical values, $p_k \approx q_k$, with constant $\bar{\varepsilon}_k \approx \varepsilon_k = \varepsilon$. These particular constraints are unlikely to naturally occur in uncontrolled environments; they serve only to highlight general properties. For instance, despite the numerical equivalence between the estimators in this hypothetical scenario, the variations Δq_k are reduced by a constant factor $Z \sim o(W) > 1$, which mitigates fluctuations of abundances A_k. For more realistic systems, we consider the fluctuations $\Delta\varphi$ due to entropic influences:

$$\Delta p_k = \Delta\varphi \left[\frac{1}{A}\frac{\partial A_k}{\partial \varphi} - p_k \sum_{\ell}^{W} \frac{1}{A}\frac{\partial A_\ell}{\partial \varphi}\right], \tag{13a}$$

$$\Delta q_k = \frac{\Delta\varphi}{Z}\left[\frac{\partial}{\partial\varphi}\ln(1 + A_k) - q_k \sum_{\ell}^{W} \frac{\partial}{\partial\varphi}\ln(1 + A_\ell)\right] \tag{13b}$$

[1] Note the difference between $\varepsilon_k = \Delta A_k/A$ and $\bar{\varepsilon}_k = \Delta A_k/(1 + A_k)$, i.e., $\varepsilon_k/\bar{\varepsilon}_k < 1$

Other than the factor Z^{-1}, Δq_k and Δp_k differ mostly in the way the variations behave: Δq_k depends only on relative variations, whereas Δp_k requires the variation of A_k with φ in addition to the total value $A = A(\varphi)$. Consequently, q_k may produce a better separation of effects from the different compartments in the system when compared to p_k for the same compartmentalization scheme. In other words, the estimator q_k reduces the interdependences among the various compartments.

From the relation $q_k \leqslant A_k/Z$, one obtains the inequality $q_k/p_k \leqslant (A/Z)$. Therefore, the positive index

$$Q = \left(\frac{1}{W}\frac{Z}{A}\right)\sum_{k=1}^{W}\frac{q_k}{p_k} \leqslant 1 \tag{14}$$

which measures the deviation between the probability estimator q_k from the more common estimator p_k. This assumption shifts the concept of units in ecological systems, meaning that only exponential variations in relative abundances are properly inferred in terms of entropy variations. Note that our rationale matches the general MaxEnt framework for thermodynamic variational principles [19–21], whereas our partition function Z is similar to that assumed in Würtz and Annila[80]. In fact, Eq. 11 demands deduction and inferences, and we present them in terms of abundances. Eq. 5 indicates that abundances are the result of distinct terms. To clarify this topic, we consider $F_k(\varphi)$ to be the entropic influence on the abundance of the k-th compartment, and θ_k is the non-entropic influence of the W compartments on quantities found in the k-th compartment.Then, any entropy production is related to $\Delta F(\varphi)$, while non-entropic effects result from the internal dependence among categories (biological interactions, population dynamics, and so forth).

Applying these terms to abundances, the classic approach assumes that

$$A = \sum_{k=1}^{W} A_k = \sum_{k=1}^{W} F_k(\varphi) \times \theta_k, \tag{15}$$

where A is the result of an arithmetic relation among the W compartments, a framework that makes little sense in terms of dependent compartments. For instance, consider the effects produced by the introduction of an exotic species to a random ecosystem. One does not expect that the addition of n predators will reduce exactly n prey, as assumed by the arithmetic ratio. In fact, a predator consumes several prey, affecting the growth rate of prey because the number of prey also affects the growth rate of predators. Therefore, this mechanism produces a cascade of influences on relative abundances, which are better expressed by multiple dependences[80]. Considering A as the result of a product rather than a sum, Eq. 15 assumes that

$$A = \prod_{k=1}^{W} F_k(\varphi) \times \theta_k. \tag{16}$$

The continuous observation during the stationary state would imply that θ_k is a function able to describe how A_k fluctuates around the fixed point determined by $F_k(\varphi)$. However, the sampling process makes θ_k assume discrete values indicating whether A_k deviates from the value produced by $F_k(\varphi)$ at the right moment of the sample. Then, $\theta_k = [\lambda_k F(\varphi)]/[F(\varphi)]$, where $\lambda_k, \forall k$ is a Lagrange multiplier, and

$$A_k = F_k(\varphi) \times \frac{\lambda_k F_k(\varphi)}{F_k(\varphi)}. \tag{17}$$

For $\theta_k = 1$, the influence of the further compartments on the kth is null, whereas for $\theta_k \neq 1$, some influence is detectable.

As categories interact, the quantities in one category are expected to increase at the expense of a decrease in another during the stationary state, maintaining the thermodynamic balance according to the second law. Consequently, all values assumed by θ alternate around 1 through the W compartments. Concerning Eq. 16, entropic and non-entropic influences on abundances A_k can be isolated at the logarithm scale as

$$\ln(A) = \sum_{k=1}^{W} \ln F_k(\varphi) + \sum_{k=1}^{W} \ln \theta_k. \tag{18}$$

For the consequence of $\theta_k \to 1$, $\sum_1^W \ln \theta_k \to 0$, and therefore, we have the "effective" abundances $\hat{A}_k$ as

$$\hat{A} = \sum_{k}^{W} F_k(\varphi) = \sum_{k}^{W} ln[A_k + 1], \tag{19}$$

where unit summed to A_k guarantees the relation even when one or more categories are not represented in the sample. This framework states that $\hat{A}$ is a function of "observed" raw abundances A, and it exclusively holds the entropic influences on A_k. Therefore, it is expected that

$$\Delta\hat{A} \equiv \Delta\varphi. \tag{20}$$

Importantly, as $\theta_k \to 1, \forall k$, the average value $\bar{\theta} \approx 1$, and the expression $\sum_{k=1}^{W} \ln \theta_k$ can be rewritten as $W[ln(W^{-1}]$, making Eq. 18 assume that

$$\tau_W \equiv ln[A(\varphi)] - W[\ln(W^{-1}))] \equiv \sum_{i}^{W} \ln(A_k + 1)]^*, \tag{21}$$

where τ_W is the optimized entropy for the granular scheme that produced W. The errors associated with Eqs. 19 - 21 reflect whether the assumption $\theta_k \to 1$ deviates from the true distribution produced by θ_1,...,θ_W, inferring whether the assumed granular scheme deviates from the actual maximum entropy possible for the system. Accordingly, the multiplicative approach and consequent probabilities calculated as q_k offer a robust approach for comparing granular schemes and for assessing

the entropy production in ecological systems. Because time is not necessary and thermodynamic properties are recovered, further inferences can be explored, such as the space and consequently the partitioning diversity. In the following, we shall test the plausibility of this rationale using empirical and synthetic data.

III. TESTING THE FRAMEWORK IN EMPIRICAL DATA

Entropy production in soil systems

In the first test with empirical data, we verify whether the aforementioned framework is applicable for practical contexts. For this purpose, we use TGA, which is a data set that contains information of plant biomass and counts of ground arthropods in a successional gradient of tropical forests. Concurrent to the granular scheme, we aim to evaluate how diversity indices of ground arthropods respond to the entropy production when the classic probabilities, p_k, and the new one, q_k, are committed.

TGA covers 13 forests in distinct stages of development through a gradient constituted by 10 reforestations and 3 forest remnants. All sites are located in Southeast Brazil and bear a high diversity of tree species (Atlantic forests). Data of plant biomass are removed to infer the entropic influences on the diversity of ground arthropods. Then, samples of tree densities ($ind./m^2$), area of trunks (m^2), and litter biomass (kg/m^2) are converted to only one variable, φ, by using principal component analysis[45,59]. φ represents a general quantitative descriptor of forest development, and its non-dimensional values stand for the linear coefficients of the eigenvector of the largest eigenvalue (see Supplement 2 for details). Concerning ground arthropods, the input data are counts of ground arthropods sampled in soil and litter during winter and summer, totaling 40 samples per site (see Supplement 2 for details). Part of these data were previously analyzed to explain the colonization of ground arthropods during the development of the forest [57,60]. The organisms are sorted by species (occasionally morpho-species), and data of distinct samples are joined in only one large sample per site. With data of fauna in hand, we consider the species sorting to calculate the number of categories W (species richness), the abundance A and the classic Shannon diversity H (Eq. 1) for the 13 sites of TGA. Next, we plot respective values against φ quantities, and we observe the correlation between them.

Figures 1A-C depict that the variations in the three metrics, namely, W, A and H, show no correlation with the values of φ. Consequently, two possible explanations may be assumed. The first one admits no cause-effect relationship between the forest development and the biological organization of ground arthropods. This explanation would imply that the successional process does not produce entropy in litter and soil layers or that ground arthropods are not sensitive to such a process. This hypothesis appears to be implausible due to strong experimental evidence attesting to the contrary[1,4,17,18,27,57,60,64]. The alternative explanation assumes that the cause-effect relationship exists, but W, A and H are unable to properly assess entropy production in soil communities. We assume that the second explanation is correct because it is corroborated with experimental evidence.

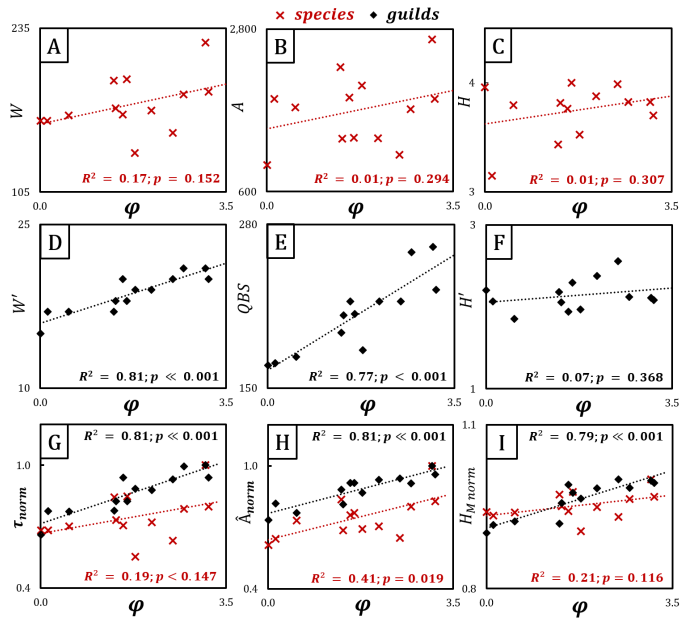

Figure 1. Relationship between successional process and ecological indices of ground arthropods considering a gradient of thirteen sites ($n = 13$). φ: indicator of forest development, obtained from a PCA conducted using values of tree dominance, tree density and litter biomass. For ground arthropods, R: species richness; A: total abundance; H: Shannon diversity index using species as compartments; W: number of guilds according to Parisi *et al.*[67]; QBS: Qualitat Biológica del Suolo, an indicator of soil health based on ground arthropods[67]; H': Shannon index obtained for guilds as compartments; $\tau_W = \ln(A) - W\ln(W^{-1})$, the maximum entropy; $\hat{A} = \sum_{k=1}^{W} \ln A_k$, the effective abundance; and H_M: Shannon entropy achieved by probabilities $q_k = \ln(A_K)/\hat{A}$. A-C show that R, A and H of ground arthropods are poorly correlated with the forest development. The use of guilds increases the correlation between the number of categories W and the successional stage, as does QBS. However, the use of guilds as compartments produces values of entropy H' that are poorly correlated with the forest development. Finally, the use of q_k probabilities, derived from the effective abundances approach, show that the forest development affects the organization of ground arthropods, mainly when q_k is considered together with guilds.

There is a clever way to avoid the concerns around H and related indices: sorting organisms with categories other than taxonomy, in which relevant biological and ecological constraints are explicitly addressed. In fact, conceptual and empirical evidence suggest that guild ap-

proaches may improve the extraction of ecological information because they fairly reflect the niche occupancy [1,9,61,67]. Ideally, functional approaches would pursue new categories that alter H values and improve the thermodynamic properties. Consequently, a better inference about entropy production for the TGA case is also expected, and we test this hypothesis.

Thus, we sort the arthropods of TGA in guilds following the criteria described in Parisi *et al.*[67]. Tailored from ground arthropod data, this methodology collects informative content about the biological quality of the soil and produces a reliable indicator of soil health, the QBS (*Qualità Biological del Suolo*) [5,30,47,61,67]. Accordingly, arthropods are sorted by high-level taxa and morphological traits into taxonomic-functional categories (≈ 28). Each category receives an *ad hoc* score, the EMI (eco-morphological index), that provides a quantitative inference of how adapted its organisms are to soil niches (Supplement 3). For a random sample, the EMI scores of all represented categories are summed to produce the QBS value. Note that relative abundances show limited importance for the calculation of QBS, and single or multiple occurrences at the same k-th category result in equal contributions. Concerning the TGA example, one expects that the number of EMI categories and QBS values increase over the forest development as a consequence of soil maturation[17]. We then conjecture that the EMI categories could also improve the assessment of entropy production in TGA if abundances are used rather than the *ad hoc* values.

In the following, we refer to the EMI categories as guilds, and we add the prime symbol to differentiate its use (primed) from the use of species (unprimed). For arthropods sorted according to the EMI categories, W' corresponds to the number of guilds, and QBS is the respective index calculated following the approach of Parisi *et al.*[67]. Using guilds as a granular scheme, we also calculate H', as $p'_k = A'_k/A$, where A'_k is the relative abundance of the k-th guild. Then, we plot W', QBS, and H' against the φ values, and we compare these results with those obtained for species.

Figures 1D-E depict how the number of guilds W' and the QBS index vary as functions of φ. The linear fits indicate that both quantities correlate with the forest development, inferring that the successional process affects the organization of soil communities. In fact, guilds join several organisms in a few categories, which naturally reduces the total variance. However, the same result does not hold for H' (Figure 1F). In comparison to H (species), H' reveals that guilds produce negligible improvements in the assessment of entropy production. Furthermore, the H' behavior is in clear contrast to that of QBS, although both metrics address the same granular scheme. How is this result possible? Because QBS computes *ad hoc* quantities and H' computes abundances, it is clear that the estimator p_k makes H' inefficient. Therefore, the granular schemes alone cannot assess the entropy production in systems such as TGA, making the use of an appropriate estimator of probabilities necessary.

Then, we follow the framework presented in Sec. II, considering abundances as the result of interacting categories. Employing q_k probabilities, we calculate the Shannon index (Eq. 1) for two granular schemes: species and guilds. We use the notations H_M (species) and H'_M (guilds) to differentiate the indices calculated by q_k from those calculated by p_k. The indices τ_W and $\hat{A}$, predicted by Eq. 19 and Eq. 21, are also calculated, and all them are then plotted against φ.

Figure 1G shows that correlations between τ_W or τ'_W with φ are compatible with the respective correlations observed for W and W', which confirms the gain produced by guilds in comparison to the species scheme (Figures 1A,D,F). Furthermore, the tests indicate that the q_k probabilities always attain better inference about φ than the p_k (Figures 1H-I). For both species and guilds, the effective abundances $\hat{A}$ are correlated with the values of φ, with even the granular schemes still being important. Once TGA addresses successional context, guilds naturally produce stronger inference power about entropy production than species. When q_k is employed together with guilds, abundances finally reveal the entropy production in the TGA data set, with $\hat{A}'$ achieving even better results than QBS (Figure 1E,H). As predicted by Eq. 19, H'_M also correlates with φ, which is evidence that effective abundances allied to guild compartments recover the thermodynamic properties in the TGA case. Such optimization is predicted by Eq. 21, and since it is confirmed, it ensures the replication principle. Because the replication principle is of great importance for thermodynamic inference, we apply further tests to explicitly evaluate this hypothesis.

Recovering the replication principle in soil systems

The replication principle states that diversity indices are functions of abundances[13,15], which requires that extensive and additive thermodynamic properties be ensured. Concerning the extensive property, it is expected that an increase of abundance leads to a monotonic increase in the number of compartments and in entropy values. For the additive property, it is expected that the entropies of samples considered separately are equivalent to the entropies of samples considered together, implying the independence of categories. Subsequently, we evaluate the reliability of such relationships in the TGA example.

Objectively, we test whether q_k can actually optimize the extensive property in the TGA case by evaluating whether the increase of $\hat{A}$ leads to a monotonic increase in W, H_M and τ_W. Such correlations are expected only when the replication principle is ensured. We also test the influence of granular schemes by evaluating the curves obtained for species and guilds separately. As predicted in Sec. II, Figure 2 depicts that the values of H_M and H'_M

linearly correlate with respective effective abundances $\hat{A}$ and $\hat{A}'$, the number of categories W and W', and maximum entropy τ_W and τ'_W. These results explain the better inferences about entropy production depicted in Figure 1, suggesting that the q_k probabilities clearly achieve the extensive properties in the TGA case. Furthermore, the results clearly indicate that the optimization of the extensive property does not necessarily depend on the granular scheme, which also confirms the predictions of Sec. II.

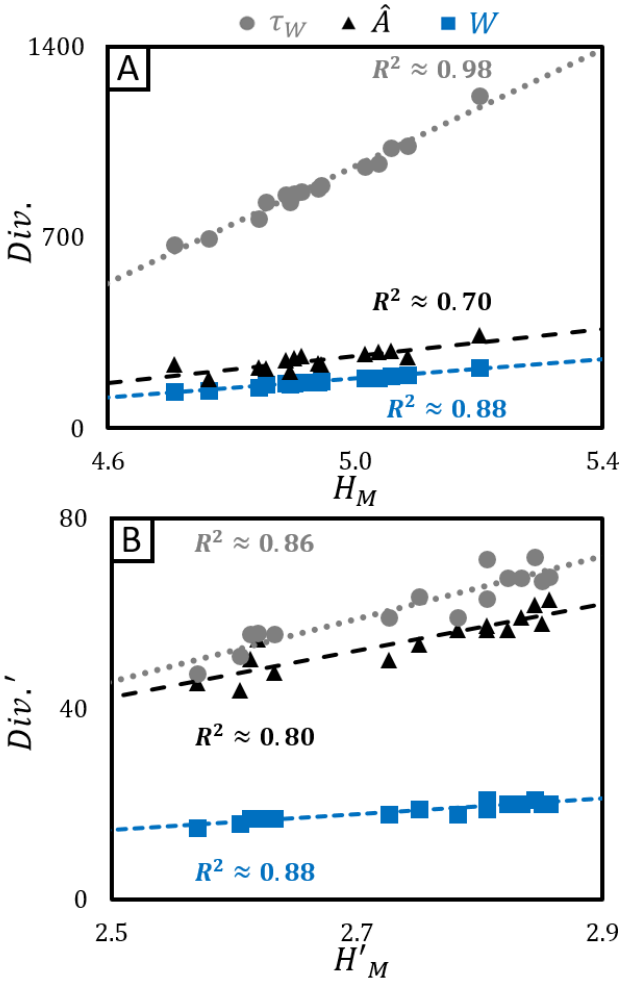

Figure 2. Thermodynamic relations for ecological indices of ground arthropods sampled in secondary tropical forests when using probabilities q_k. **A**: arthropods sorted as species; **B**: arthropods sorted as guilds[67]. *Div.*: diversity indices; W: number of categories; $\hat{A}$: effective abundances; τ_W: maximum entropy for W compartments and A. The figure shows that probabilities q_k can recover the extensive and additive properties of ground arthropod communities. Under such circumstances, the thermodynamic relationships are guaranteed, thus allowing the prediction of how changes in the number of organisms are related to the number of compartments, as well as the system organization.

Next, we explicitly evaluate whether the additive property is also recovered by q_k in TGA. The additive property states that $HA + HB \equiv HAB$, where A and B are samples, while AB is their union. This would be final evidence that q_k ensures the replication principle (thermodynamic properties) in ecological systems. To avoid any influence of the entropy production on TGA data, which address correlations among samples, we develop a strategic test to properly evaluate the additivity in this data set.

The test consists of a shuffling routine followed by systematic comparisons between the variances (σ^2) of sample data separately with the variances of the same sample data together.The sites of TGA ($m = 13$) are labeled as U_a, U_b,...,U_m and are used to create c clusters $C_1, ..., C_c$. Each cluster combines arthropod counts of four random sites, such as $C_1 = (U_a \cup U_b \cup U_c \cup U_d)$, $C_2 = (U_a \cup U_c \cup U_d \cup U_e)$, and so on. Next, a second set of x clusters is created, now combining the clusters $C_1, ..., C_m$ two by two, such as $C_{(1,2)}, ..., C_{(x-1,x)}$, where $C_{i,j} = (C_i \cup C_j)$. With this shuffled structure in hand, we calculate the variances (σ^2) of respective probabilities, such as $\sigma^2_{C_1}$, $\sigma^2_{C_2}$, $\sigma^2_{C_3}$, and so on, for the first group of clusters, and $\sigma^2_{C_{(1,2)}}$, $\sigma^2_{C_{(1,3)}}$, $\sigma^2_{C_{(2,3)}}$, and so forth for the second one. Because the additive property explicitly concerns the independence of compartments, the granular scheme is of great importance. Then, we propose four distinct treatments to evaluate the subject:

- Treatment 1: W for categories and p_k for probabilities;
- Treatment 2: W for categories and q_k for probabilities;
- Treatment 3: W' for categories and p_k for probabilities;
- Treatment 4: W' for categories and q_k for probabilities.

Briefly, each treatment evaluates the linear correlation between $\sigma^2_{C_i} + \sigma^2_{C_j} \equiv \sigma^2_{C_{(i,j)}}$. A good linear fit indicates that the treatment allows achieving the additive property. Conversely, the lack of a correlation suggests that the thermodynamic property is poorly achieved. Figures 3A-D depict the results.

Figure 3A indicates that, for Treatment 1, the sum of variances is not compatible with the variance of distinct communities together. Therefore, the species scheme and p_k probabilities cannot ensure the additive property. Figures 3B and 3C indicate that Treatments 2 and 3 partially recover the additive property, implying that the granular scheme and the estimator of probabilities consist of distinct approaches to reduce data variability in ecological systems. Finally, Treatment 4 (Figure 3D) shows that the guild scheme and q_k probabilities employed together produce an excellent linear fit between $\sigma^2_{C_i} + \sigma^2_{C_j} \equiv \sigma^2_{C_i \cup C_j}$, indicating that the additive property is completely recovered.

Once more, the tests support the results depicted in Figure 1 and the predictions of Sec. II. Therefore, the

tests confirm the independent effect of q_k probabilities on the optimization of thermodynamic properties of soil systems. Although the q_k probabilities ensure the replication principle in the TGA case, an inference about general ecological systems requires further tests. Thus, we subsequently evaluate the reliability of the approach in further contexts.

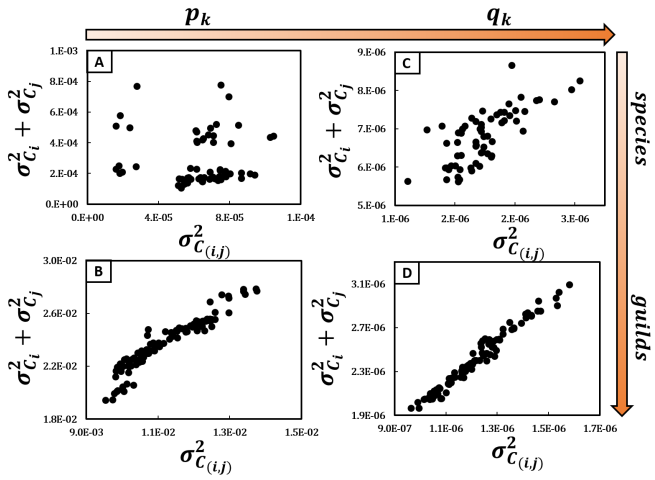

Figure 3. Comparing the variances (σ^2) of probabilities obtained using distinct approaches for quantities of ground arthropods found in secondary Atlantic forests. Thirteen sites showing different levels of development are used to create multiple arrangements (n=63). **A**: species compartments and p_k; the sum of variances shows a poor linear correlation with the variance of sums ($R^2 < 0.01; p = 0.593$), indicating that additive and extensive properties of the systems are not achieved. **B**: guilds[67] and p_k ($R^2 \approx 0.90; p << 0.001$); **C**: species and q_k ($R^2 \approx 0.55; p < 0.001$). In both cases, the correlation between the sum of variances and the variance of sums was improved in comparison to A, suggesting that the thermodynamic properties are better assessed. **D**: guilds and q_k; the excellent linear fit ($R^2 \approx 0.98, p << 0.001$) indicates that the additive and extensive properties of the ecological system are recovered by addressing both approaches together.

Beyond soil systems: the generality of the result obtained by q_k

Notwithstanding the predictions of Sec. II and evidence from soil systems, we hypothesize that the optimization of thermodynamic properties by q_k cannot hold for broader ecological contexts. Hence, we evaluate the replication principle in general situations, taking ten data sets available on-line and calculating the entropy produced by the estimators p_k and q_k. These data sets concern a broad range of biological communities in worldwide ecosystems, including mammals, perennial plants and ants from US deserts[3,62]; fishes and benthic invertebrates from temperate lakes[48,49]; ground arthropods sampled in Italy's organic and conventional crops[61]; butterflies living in open areas from Colorado, US[63]; plant species growing in steppes from Oklahoma[56]; understory plant communities of temperate forests from Canada[66]; and tree communities of tropical forests from India[69]. The experimental conditions also greatly vary among studies. There are times series, simple samples in spatially explicit conditions, counting of individuals, biomass estimation, frequencies, and so forth. We provide a list of the main characteristics of these data sets in Supplement 3, but we encourage a search in the original papers for more details (all of them are available as open source data).

This broad range of experimental conditions is suitable for an important propose: assessing and comparing the replication principle and the thermodynamic inference produced by the classic and the new estimators of probabilities. Note that we are not attempting to replicate the original experiments using a new tool, or evaluating the reliability of the previous findings. In fact, several details about the original experimental designs and inferences about external influences are not addressed here, because they are pointless for our propose. Accordingly, climate, human manipulation, entropy production, granular schemes, and other aspects described in these data sets are here considered as only sources of entropy variation. Our only intention is to check the consistency of the estimator q_k concerning the replication principle, observing how the diversity indices fit variations in organism quantities. Based on the consistency of thermodynamic laws, it is expected that the data variability and the distinct conditions considered by these data sources minimally affect the general patterns and that the abundance variability present in each data set would provide sufficient information to observe the thermodynamic properties. Finally, we aim to evaluate the applicability of the framework described in Sec. II as a general tool for assessing biological diversity in ecological systems.

Accordingly, for a data set S containing n samples, we simply calculate the diversity indices of each sample, and we analyze the correlation among distinct indices under the conditions found in S. Because we intend to prove the optimization of the replication principle, we evaluate the correlation between the same diversity indices considered by Hill numbers [13,15,36], although not necessarily as Eq. 2. For p_k, we observe the values of W, H and D as functions of A, whereas for q_k, we observe W, H_M and D_M as functions of $\hat{A}$. Furthermore, we assess the correlation between W and H and H_M. In contrast to the tests with TGA, only one granular scheme is assessed per data set. Based on MaxEnt principles, the best estimator is considered to be the one capable of producing index values that optimize the inference produced by relative quantities. The results are depicted in Figure 4 and Figure 5.

Concerning the ten data sets, the use of p_k almost always produces values for the diversity indices (W, H, and D) that poorly correlate with abundances A. Even the correlation between W and H is important only for one case (Figure 5[56]). Therefore, the test corroborates the poor inference provided by the classic estimator of prob-

abilities, which assumes *a priori* the biological categories as independent compartments. Conversely, probabilities q_k always achieve strong correlations between effective abundances and the ecological indices or between W and H_M (Figures 4 and 5). Note that probabilities q_k are appropriate even for the case in which p_k probabilities are satisfactory.

Therefore, the results corroborate the evidence provided by the TGA example, as well as the prediction of Sec. II. Concerning the broad range of conditions regarded by this test, the results depicted in Figures 4 and 5 stress the valor of the framework presented here for the assessment of diversity patterns in ecological systems. These new and strong evidence reinforce that effective abundance, as well as the entire conceptual framework behind its assumption, consists of a trustful manner to recover the thermodynamic entropy of ecological systems, which is a substantial gain in the inference power produced by diversity indices based on Boltzmann-Gibb-Shannon entropy. In fact, the approach is so consistent in all experimental situations that we further proceed to find additional inferences, as follows.

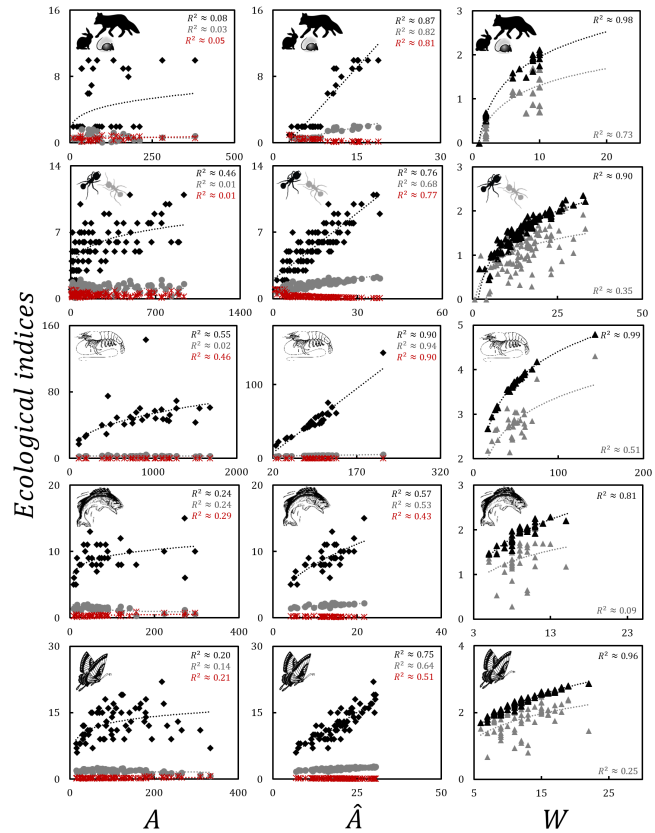

Figure 4. Relationship between quantities of organisms and entropy measures according to distinct estimators of probabilities. A: abundances; $\hat{A}$: effective abundance; W: number of compartments (species, genera, guilds, and so forth); black diamonds represent the number of compartments W; gray diamonds represent the Shannon entropy; and dark red asterisk represents the Gini-Simpson asymmetry D. On the left, the indices are calculated for probabilities taken as $p_k = A_k/A$, and in the middle, by probabilities as $q_k = \ln A_k / \sum_{k=1}^{W} \ln A_k$. On the right, black triangle represents the H obtained by q_k, while gray triangle represents the H obtained by p_k. Accordingly, probabilities q_k improve the maximum entropy for distinct biological communities, revealing their additive and extensive properties. In lines: 1st: Mammals from Utah, US[3]; 2nd: Ants from Chihuahuan Desert, Arizona, taken by baits[62], $n = 135$, $W = 20$; 3rd: Benthic invertebrates from North Lake region[48], $n = 30$, $W = 19$; 4th: fishes from temperate lakes of North America taken by electro fishing[49], $n = 35$, $W = 27$; 5th: Butterflies from open areas of Colorado, US,[63] $n = 66$, $W = 58$. **Art:** *oogazone.com.*

Probabilities q_k as a tool to link the partitioning of diversity

The examples presented above provide strong evidence supporting the use of q_k probabilities to achieve the thermodynamic properties in ecological systems. In turn, we conjecture that q_k probabilities can also be employed to solve an old question in ecology: the partitioning of biodiversity[24]. Briefly, the problem consists of whether the diversity of local samples represent the entire diversity found in the region. The local diversity is named α-diversity, the regional diversity is named γ-diversity, and their relationship is molded by the β-diversity. This important subject has generated debate among ecologists for decades and still lacks a consensus, with dozens of frameworks already proposed[24]. Because probabilities q_k optimize the extensive properties in general ecological systems, we conjecture that an appropriate experimental design could reveal how diversity patterns are affected by the sampling area. If the correlation between sampling area and H_M is confirmed, then the q_k probabilities would provide the basis for definitively solving the dilemma, at least under the thermodynamic perspective.

Hence, we use a data set that contains counts of vascular plants, where the samples are taken in an experimental design with spatial correlation[65]. This data set considers a large parcel of $256m \times 256m$, partitioned into 256 samples of $16m \times 16m$ arranged side by side, as in a square lattice. We calculate A, $\hat{A}$, W, H_{max}, H and H_M for one parcel alone, then for two parcels together, and then for three parcels together, and so on, until all 256 of the parcels are addressed together. We plot the respective values against the respective area (m^2) represented by the data. For extensive systems, a monotonic increase of all indices is expected. Figure 6 shows that the abundances, effective abundances and the number of species monotonically increase as a function of the sampling area, as expected for extensive systems. The same holds for H_{max} and for H_M, but not for H. In fact, the entropy calculated by p_k produces an anomalous behavior, suggesting that entropy would decrease despite the increase in sampling area. Therefore, no doubts exist that q_k can actually recover the extensive properties of ecological systems and can be used to link γ-diversity to

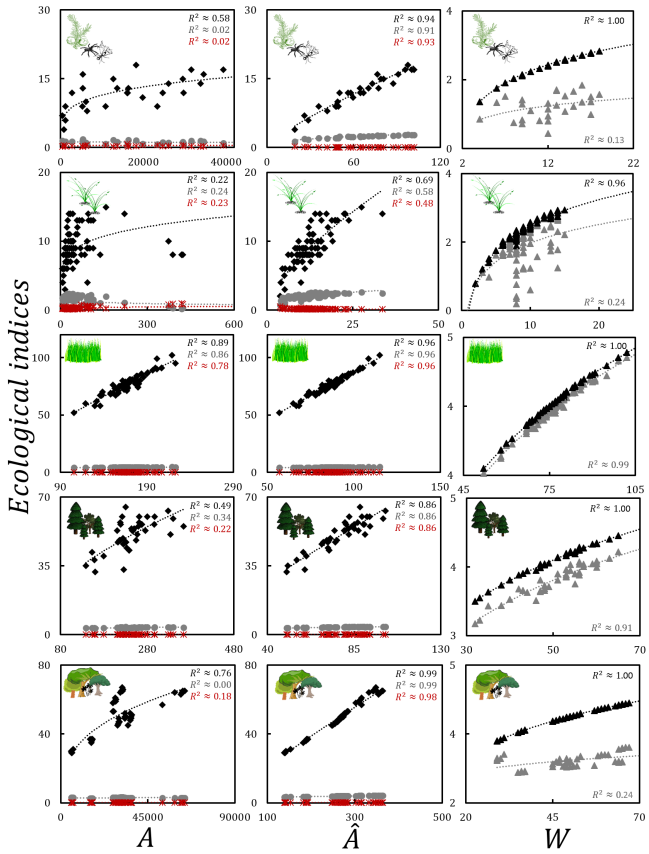

Figure 5. Relationship between quantities of organisms and entropy measures according to distinct estimators of probabilities. A: abundances; $\hat{A}$: effective abundance; W: number of compartments (species, genera, guilds, and so forth); black diamonds represent the number of compartments W; gray diamonds represent the Shannon entropy; and dark red asterisk represents the Gini-Simpson asymmetry D. On the left, the indices are calculated for probabilities taken as $p_k = A_k/A$, and in the middle by probabilities as $q_k = \ln A_k / \sum_{k=1}^{W} \ln A_k$. On the right, black triangle represents the H obtained by q_k, while gray triangle represents the H obtained by p_k. Accordingly, probabilities q_k improve the maximum entropy for distinct biological communities, revealing their additive and extensive properties. In lines: 1st: Abundance of soil arthropod groups found in cultivated soils from Italy $n = 28$, $W = 28$[61]; 2nd: Perennial plants from Chihuahuan Desert, Arizona, $n = 42$, $W = 59$[62]; 3rd: Vegetation cover of grass and shrub species in prairie steppes from Oklahoma, US[56], $n = 60$, $W = 320$; 4th: Semi-quantitative crown cover from trees in the Lac Croche temperate forest, Quebec, Canada, $n = 40$, $W = 72$[66]; 5th: Basal area of trees in tropical forests from India, $n = 46$, $W = 100$[69]. **Art:** *oogazone.com.*

α-diversity only using a scale correction, while p_k is not a proper metric. Naturally, β-diversity is a function that links local and regional diversity, and it can unequivocally predict how much of the regional diversity is represented in a set of samples.

Underlying ecological mechanism

The last doubt regards a possible influence of underlying ecological mechanisms on the inference produced by q_k. In fact, the forces driving the shape of SADs are poorly known, and the method presented here makes no *a priori* inferences about them. In fact, distinct mechanisms can produce SADs with different shapes, which could carry consequences for the assessment of entropy in biological communities[68]. Consequently, we test whether q_k probabilities are sensitive to the SAD shape using synthetic data.

We employ computational algorithms to generate data distributed as four probability density functions (PDFs) [31,70] for our treatments, including *a)* Poisson, *b)* exponential, *c)* lognormal, and *d)* power law. In each case, we produce $n = 50$ data vectors $(\vec{V}_1, ..., \vec{V}_{50})$, where $\vec{V}_j$ represents a "pseudo" biological community found in the jth site. The length of $\vec{V}_j$ denotes W_j, while its data enters the relative abundances $A_1, A_2, ..., A_{W_j}$. In all treatments, we intentionally produce "pseudo" communities, where respective species richness W_j varies according to a Poisson distribution ($\Lambda = 35$). Furthermore, we monotonically increase the parameter values of generator functions, guaranteeing that all the treatments contain some entropy variability (see Supplement 4 for details). Similar to the tests with empirical data, we compare the correlation between diversity indices produced by p_k and q_k in accordance with thermodynamical predictions.

Briefly, the results obtained for synthetic data (Figures S1-S4) show that probabilities calculated as p_k achieve extensive properties for systems whose SADs are distributed as a Poisson or exponential PDF. However, p_k was hardly efficient for SAD distributed as lognormal or power law functions. Conversely, probabilities calculated as q_k optimized the thermodynamic properties in all treatments. Therefore, the use of q_k probabilities is demonstrated to be robust even when data variability is large, as in lognormal and power law cases, and the optimization of entropy is hardly sensitive to the underlying mechanisms driving the SADs.

IV. DISCUSSION

We present an important advancement for the assessment and monitoring of biodiversity. In fact, our framework does not develop a new entropy or create a new diversity index. Rather, one simply employs a more appropriate estimate of abundances to well-known concepts and formulations. Then, the novelty concerns which quantities can actually recover the thermodynamic properties of ecological systems. The several tests presented here indicate that our approach maximizes the replication principle, providing a more precise measure of order-disorder patterns of biological diversity. The potential update in the ecological discussion is substantial because several themes showing conflicts between theoretical pre-

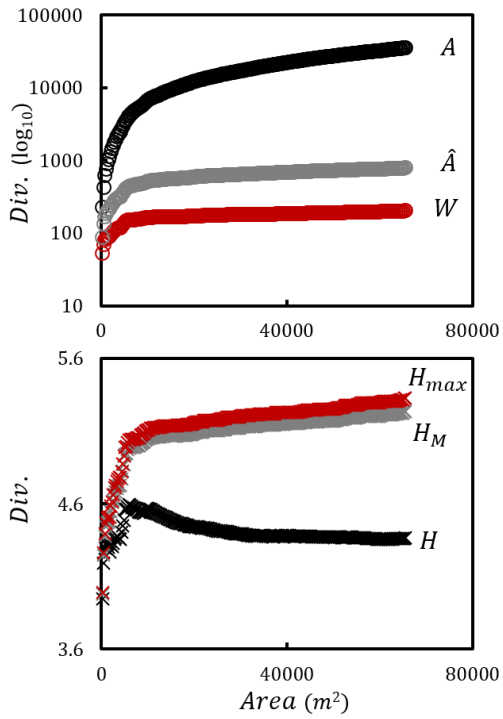

Figure 6. Influence of sampling area (m^2) on the diversity of plants according to different estimators of probabilities. *Div.*: diversity metrics; A: abundances; $\hat{A}$: effective abundance; W: species richness; H_{max}: maximum entropy possible for W compartments; H: Shannon entropy calculated by $p_k = A_k/A$; H_M: Shannon entropy calculated by $q_k = \ln(A_k + 1)/\sum_1^W \ln(A_k+1)$. The analyses consider the data of Palmer *et al.*[65]. On the top, the increase in sampling area led to the monotonic increase of abundances, effective abundances and species richness. The bottom shows that H_{max} and the entropy produced by q_k (H_M) monotonically increase with the area, whereas the entropy produced by p_k (H) presents an unpredicted pattern. Therefore, only H_M achieves the extensive properties of this system.

dictions and empirical data can now be reviewed without further samplings. Next, we discuss some particular implications.

The most important aspect of our findings concerns the estimator of probabilities. The approach proposed here is highly logical in biological terms. The quantitative dependence among categories is a well-known aspect of any ecological system that has remained poorly understood until now. Then, the major conceptual shift occurs for the calculation of probabilities and quantitative interpretations. The dependence among compartments assumes multiplicative processes rather than additive ones and yields important consequences for diversity indices and their inference power. For instance, only exponential increases in the number of organisms (or biomass) produce detectable variations in terms of entropy. Second, the emergence of a new category in ecological systems (*e.g.*, a new guild) depends on the geometric increase of the entire system. Such deductions appear to explain the general interpretation of biomass pyramids[77], for which distinct trophic levels show a geometric dependence. Further evidence comes from the effects produced by the inclusion/exclusion of species in natural ecosystems, such as in biological invasions. Such events can affect the entire community in non-trivial manners, which cannot be explained by the simple additive processes[22,46].

Our tests demonstrate that the classic estimator p_k that relies on additive processes yields weak results when assessing diversity in empirical data. The inefficacy of p_k is not new in ecological discussions, but the consequences of its inaccuracy have not received suitable attention. We wonder how many of the conflicting aspects concerning the biological diversity are in fact consequences of misinterpretations produced by this problematic estimator. In our tests, we include some important topics that exemplify our concern: the thermodynamic inference provided by diversity indices, the maximum entropy produced by distinct granular schemes, the assessment of entropy production in changing ecosystems, and the explicit relationship between local and regional diversities, which leads to the partitioning of diversity. All these topics represent a large part of the long-term discussion in ecology, and the estimator q_k appears to offer a potential solution for several of them. Therefore, we claim that the estimator p_k should be definitively left aside in favor of the adoption of q_k and that the topics mentioned here be reassessed under this new approach.

The mathematical formulation proposed here follows well-established concepts and methods used to assess non-equilibrium systems[19–21], which are adapted for a static approach. In fact, the probabilities and interpretations presented here match those proposed for ecological contexts in Ref.[80], which surprisingly has remained poorly explored until now. In comparison with the previous findings, in which time is an inherent aspect[80], the static approach here takes into account that each sample is being interpreted as a stationary (or *quasi*-stationary) state. Note that the hypothesis of stationarity is fragile under a more rigorous interpretation but that this poorly affects the accuracy of the calculation or the practical implications of the results. In this sense, our framework consists of an additional step in the direction of thermodynamic inferences, concepts, and views for wider ecological scopes. Once the diversity indices can produce trustworthy metrics, time, space or any other variable of interest can be experimentally contextualized and its phenomenological effects deduced from the ΔH obtained from samples.

Therefore, the importance of q_k is mostly practical and corroborated by empirical data. The framework is simple and is in accordance with thermodynamic rules, and it does not require additional theoretical assumptions and new formulations. The tests with empirical data revel how resilient the approach is to variations of granular schemes and is applicable to a broad range of ecological conditions. In fact, our set of tests could not find bounds

for the use of q_k, which is another highlight. However, the most clear advantage concerns the inference power and the prediction ability produced by the new estimator, which matches the replication principle. Because H_M uses a formulation similar to that used by the classic H, the approach shows another advantage in comparison to other recently proposed methodologies[8,16]: as previously mentioned, H_M can readily be employed to review previous studies without further field work. In fact, the reproducibility of results in ecological experiments is currently in question[7], and some studies with biodiversity showing conflicting results can now be reevaluated without further data samplings.

As demonstrated, the thermo-statistical interpretation achieved by q_k in ecological systems provides the basis for deep advancements in the knowledge about biodiversity[12,14,24,38]. As we demonstrated here, α-, $\gamma-$, and β-diversities are perfectly inferred by the thermodynamic inferences. Because entropy and the number of categories scale with effective abundances, few samples can be used to estimate the regional pool of species, as well as the total and relative abundances expected for the entire region. The same is applicable to studies searching for adequate schemes for sorting organisms, as we show for the TGA example (species $\times$ guilds). Our tests are not extensive, and we argue that several other tools and methods addressing the MaxEnt principles[33–35] can take advantage of the use of q_k. This potential deserves particular attention in future studies.

Phenomenological process behind the data transformation

Concerning the phenomenological implications of the new estimator, an important question arises: why does the logarithm of abundances better achieve the thermodynamic properties of biological communities? To answer this question, we need to interpret ecological patterns under a statistical view. The asymmetry of SAD curves is considered to be a universal pattern of ecological systems. Although the underlying mechanisms that produce SAD are still debatable, they are certainly related to the unbalanced fitness of species and the consequent number of respective decedents. As the logarithm function asymmetrically affects relative quantities, it is natural that log-transformed proportions are less uneven than raw proportions. This principle is well known from signal analysis, where mel frequencies[28,75] are used to reduce the signal-to-noise ratio and provide more clear human interpretation of sound waves. In fact, humans can only distinguish great variations in sound waves, while small oscillations are imperceptible. This concept is also used in statistics, where the Box-Cox techniques[6] are used to manipulate relative frequencies in random variables and generate Gaussian PDFs from data distributed according to another function. Similar to the TGA case, the entropic contribution for the formation of polymer chains is separated from non-entropic influences using a similar framework[29]. Therefore, it is clear that the geometric symmetry for relative proportions is not producing arbitrary results and only reflects the quantitative effects of interacting compartments on relative abundances. Therefore, the mechanism proposed here is known in other disciplines, and it is able to reduce data noise and produce a new probability distribution for a distinct physical system. In this sense, we also show its applicability for ecological systems. Compared with generalized entropy[78], the assumptions of the use of q_k do not require a new granular scheme[39] because only the quantitative signal inferred by the observables are considered.

One important aspect addressed here remained underexplored: the underlying mechanism driving the shape of the SAD curves. The shape of SAD curves and its relation with the central limit theorem[23] can be explored to deduce the ecological dynamics. The theorem states that the additive processes would result in SAD distributed as Gaussian distributions or exponential distributions (if constrains are addressed), while multiplicative processes would result in lognormal distributions. Obviously, the theorem predicts such patterns as opposite limit cases when realizations are massive, but large spectra can be observed for limited sampling effort. Accordingly, the empirical evidence clearly shows that SAD curves are almost always distributed as a lognormal PDF, occasionally as a power-law PDF, but almost never as PDFs such as Gaussian or exponential. For the lognormal cases, we have a clear indication of multiplicative processes during the pattern formation. However, distinct processes related to sampling effort can produce a power law from data that are actually lognormal distributed[58,81]. Therefore, even for when the power-law hypothesis is the best approach to describe the SAD curves, the evidence point in the direction of multiplicative processes molding abundance, which should be explored in favor of defining the underlying mechanisms driving the biological diversity.

REFERENCES

[1] Andersen, A. N. 1997. Using ants as bioindicators: multiscale issues in ant community ecology. Conservation Ecology **1**:1–8. URL `http://www.consecol.org/vol1/iss1/art8/`.

[2] Aronson, J., S. J. Milton, and J. N. Blignaut. 2007. Restoring natural capital: science, business and practice. Island Press.

[3] Bartel, R. A., F. F. Knowlton, and L. C. Stoddart. 2005. Mammal Abundance Indices in the Northern Portion of the Great Basin, 1962-1993. Ecology **86**:3130–3130. URL `http://dx.doi.org/10.1890/04-1607`.

[4] Birkhofer, K., C. Dietrich, K. John, Q. Schorpp, A. S. Zaitsev, and V. Wolters. 2016. Regional Conditions and Land-Use Alter the Potential Contribution of Soil Arthropods to Ecosystem Services in Grasslands. Frontiers in Ecology and Evolution **3**:150. URL `http://journal.frontiersin.org/article/10.3389/fevo.2015.00150`.

[5] Blasi, S., C. Menta, L. Balducci, F. D. Conti, E. Petrini, and G. Piovesan. 2013. Soil microarthropod communities from Mediterranean forest ecosystems in Central Italy under different disturbances. Environmental Monitoring and As-

sessment **185**:1637–1655. URL http://dx.doi.org/10.1007/s10661-012-2657-2.

[6]Box, G., and D. Cox. 1964. An analysis of transformations. Journal of the Royal Statistical Society, Series B **26**:211–252. URL http://asa.scitation.org/doi/10.1121/1.1915893.

[7]Bruna, E., R. Chazdon, T. Errington, and B. Nosek. 2017. Reproducibility & repeatability in tropical biology: a call to repeat foundational studies. OSF Preprints pages 1–12. URL file:///C:/Users/PC/Downloads/Bruna_etal_RPTB_Commentary_8may2017%20(1).pdf.

[8]Butturi-Gomes, D., M. P. Junior, H. C. Giacomini, and P. D. M. Junior. 2014. Computer intensive methods for controlling bias in a generalized species diversity index. Ecological Indicators **37**:90–98. URL http://dx.doi.org/10.1016/j.ecolind.2013.10.004.

[9]Cadotte, M. W., K. Carscadden, and N. Mirotchnick. 2011. Beyond species: functional diversity and the maintenance of ecological processes and services. Journal of Applied Ecology **48**:1079–1087. URL http://dx.doi.org/10.1111/j.1365-2664.2011.02048.x.

[10]Cernansky, R. 2017. The biodversity revolution. Nature **546**:22–24. URL https://www.nature.com/polopoly_fs/1.22079!/menu/main/topColumns/topLeftColumn/pdf/546022a.pdf.

[11]Chalmandrier, L., T. Munkemuller, V. Devictor, S. Lavergne, and W. Thuiller. 2015. Decomposing changes in phylogenetic and functional diversity over space and time. Methods in Ecology and Evolution **6**:109–118. URL http://dx.doi.org/10.1111/2041-210X.12297.

[12]Chao, A., C.-H. Chiu, and T. Hsieh. 2012. Proposing a resolution to debates on diversity partitioning. Ecology **93**:2037–2051. URL http://dx.doi.org/10.1890/11-1817.1.

[13]Chao, A., C.-H. Chiu, and L. Jost. 2014. Unifying Species Diversity, Phylogenetic Diversity, Functional Diversity, and Related Similarity and Differentiation Measures Through Hill Numbers. Annual Review of Ecology, Evolution, and Systematics **45**:297–324. URL https://doi.org/10.1146/annurev-ecolsys-120213-091540.

[14]Chao, A., C.-H. Chiu, and L. Jost. 2016. Statistical challenges of evaluating diversity patterns across environmental gradients in mega-diverse communities. Journal of Vegetation Science **27**:437–438. URL http://dx.doi.org/10.1111/jvs.12420.

[15]Chao, A., N. J. Gotelli, T. C. Hsieh, E. L. Sander, K. H. Ma, R. K. Colwell, and A. M. Ellison. 2014. Rarefaction and extrapolation with Hill numbers: a framework for sampling and estimation in species diversity studies. Ecological Monographs **84**:45–67. URL http://dx.doi.org/10.1890/13-0133.1.

[16]Chao, A., L. Jost, T. C. Hsieh, K. H. Ma, W. B. Sherwin, and L. A. Rollins. 2015. Expected Shannon Entropy and Shannon Differentiation between Subpopulations for Neutral Genes under the Finite Island Model. PLOS ONE **10**:1–24. URL https://doi.org/10.1371/journal.pone.0125471.

[17]Coleman, D. C., , D. Crossley, , and P. F. Hendrix. 2004. Fundamentals of Soil Ecology (Second Edition). Second edition edition. Academic Press, Burlington.

[18]de Vries, F. T., E. Thebault, M. Liiri, K. Birkhofer, M. A. Tsiafouli, L. Bjornlund, H. Bracht Jorgensen, M. V. Brady, S. Christensen, P. C. de Ruiter, T. d Hertefeldt, J. Frouz, K. Hedlund, L. Hemerik, W. H. G. Hol, S. Hotes, S. R. Mortimer, H. Setala, S. P. Sgardelis, K. Uteseny, W. H. van der Putten, V. Wolters, and R. D. Bardgett. 2013. Soil food web properties explain ecosystem services across European land use systems. Proceedings of the National Academy of Sciences **110**:14296–14301. URL http://www.pnas.org/content/110/35/14296.abstract.

[19]Dewar, R. 2003. Information theory explanation of the fluctuation theorem, maximum entropy production and self-organized criticality in non-equilibrium stationary states. Journal of Physics A: Mathematical and General **36**:631 – 641.

[20]Dewar, R., and A. Maritan, 2014. A Theoretical Basis for Maximum Entropy Production. Pages 141–167 in Beyond the Second Law: Entropy Production and Non-equilibrium Systems. Springer.

[21]Dewar, R. C., 2014. A general maximum entropy framework for thermodynamic variational principles. Pages 49–72 in AIP Conference Proceedings, volume 1636. URL http://aip.scitation.org/doi/abs/10.1063/1.4903723.

[22]Didham, R. K., J. M. Tylianakis, N. J. Gemmell, T. A. Rand, and R. M. Ewers. 2007. Interactive effects of habitat modification and species invasion on native species decline. Trends in Ecology & Evolution **22**:489 – 496. URL http://www.sciencedirect.com/science/article/pii/S0169534707002054.

[23]Durrett, R., 2013. Probability: Theory and Examples.

[24]Ellison, A. M. 2010. Partitioning diversity1. Ecology **91**:1962–1963. URL http://dx.doi.org/10.1890/09-1692.1.

[25]Ettema, C. H., and D. A. Wardle. 2002. Spatial soil ecology. Trends in Ecology and Evolution **17**:177–183.

[26]Falk, D., M. Palmer, and B. Zedler. 2006. Foundations of Restoration Ecology. Island Press.

[27]Fitter, A. H., C. A. Gilligan, K. Hollingworth, A. Kleczkowski, R. M. Twyman, J. W. Pitchford, and T. M. of the Nerc Soil Biodiversity Programme. 2005. Biodiversity and ecosystem function in soil. Functional Ecology **19**:369–377. URL http://dx.doi.org/10.1111/j.0269-8463.2005.00969.x.

[28]Fletcher, H. 1938. Loudness, Masking and Their Relation to the Hearing Process and the Problem of Noise Measurement. The Journal of the Acoustical Society of America **9**:275–293. URL http://asa.scitation.org/doi/10.1121/1.1915893.

[29]Flory, P. J., 1993. Spatial configuration of macromolecular chains. Pages 156–177 in T. Frängsmyr and S. Forsén, editors. Nobel Lecture, Chemistry 1971-1980. World Scientific Publishing Co. URL https://www.nobelprize.org/nobel_prizes/chemistry/laureates/1974/flory-lecture.pdf.

[30]Galli, L., M. Capurro, C. Menta, and I. Rellini. 2014. Is the QBS-ar index a good tool to detect the soil quality in Mediterranean areas? A cork tree Quercus suber L. (Fagaceae) wood as a case of study. Italian Journal of Zoology **81**:126–135.

[31]Gillespie, C. S. 2015. Fitting Heavy Tailed Distributions: The poweRlaw Package. Journal of Statistical Software **64**:1–16. URL http://www.jstatsoft.org/v64/i02/.

[32]Gotelli, N. J., G. R. Graves, and C. Rahbek. 2010. Macroecological signals of species interactions in the Danish avifauna. Proceedings of the National Academy of Sciences **107**:5030–5035. URL http://www.pnas.org/content/107/11/5030.abstract.

[33]Haegeman, B., and M. Loreau. 2008. Limitations of entropy maximization in ecology. Oikos **117**:1700–1710. URL http://dx.doi.org/10.1111/j.1600-0706.2008.16539.x.

[34]Harte, J. 2011. Maximum Entropy and Ecology. Oxford Unversity Press.

[35]Harte, J., and E. A. Newman. 2014. Maximum information entropy: a foundation for ecological theory. Trends in Ecology & Evolution **29**:384 – 389. URL http://www.sciencedirect.com/science/article/pii/S0169534714001037.

[36]Hill, M. O. 1973. Diversity and Evenness: A Unifying Notation and Its Consequences. Ecology **54**:427–432. URL http://www.jstor.org/stable/1934352.

[37]HUBBELL, S. P. 2005. Neutral theory in community ecology and the hypothesis of functional equivalence. Functional Ecology **19**:166–172. URL http://dx.doi.org/10.1111/j.0269-8463.2005.00965.x.

[38]Jost, L. 2006. Entropy and diversity. Oikos **113**:363–375. URL http://dx.doi.org/10.1111/j.2006.0030-1299.14714.x.

[39]Keylock, C. J. 2005. Simpson diversity and the Shannon–Wiener index as special cases of a generalized entropy. Oikos **109**:203–207. URL http://dx.doi.org/10.1111/j.0030-1299.2005.13735.x.

[40]Kleidon, A., Y. Malhi, and P. M. Cox. 2010. Maximum entropy production in environmental and ecological systems. Philosophical Transactions of the Royal Society of London B: Biological Sciences **365**:1297–1302.

[41]Laughlin, D. C., R. T. Strahan, M. M. Moore, P. Z. Fulé, D. W. Huffman, and W. W. Covington. 2017. The hierarchy of predictability in ecological restoration: are vegetation struc-

ture and functional diversity more predictable than community composition? Journal of Applied Ecology pages n/a–n/a. URL http://dx.doi.org/10.1111/1365-2664.12935.

[42]Lavelle, P., T. Decaens, M. Aubert, S. Barot, M. Blouin, F. Bureau, P. Margerie, P. Mora, and J.-P. Rossi. 2006. Soil invertebrates and ecosystem services. European Journal of Soil Biology **42**:S3 – S15. URL http://www.sciencedirect.com/science/article/pii/S1164556306001038.

[43]Lavender, J. T., K. A. Dafforn, M. J. Bishop, and E. L. Johnston. 2017. An empirical examination of consumer effects across twenty degrees of latitude. Ecology pages n/a–n/a. URL http://dx.doi.org/10.1002/ecy.1926.

[44]Lebrija-Trejos, E., E. A. Pérez-García, J. A. Meave, F. Bongers, and L. Poorter. 2010. Functional traits and environmental filtering drive community assembly in a species-rich tropical system. Ecology **91**:386–398. URL http://dx.doi.org/10.1890/08-1449.1.

[45]Legendre, P., and L. Legendre. 2012. Numerical Ecology, Volume 24, 3th edition. Elsevier.

[46]Levine, J. M. 2000. Species Diversity and Biological Invasions: Relating Local Process to Community Pattern. Science **288**:852–854. URL http://science.sciencemag.org/content/288/5467/852.

[47]Madej, G., G. Barczyk, and M. Gdawiec. 2011. Evaluation of Soil Biological Qualisty Index (QBS-ar): Its Sensivity and Usefulness in the Post-Mining Chronosequence - Preliminary Research. Polish Journal of Environmental Studies **20**:1367–1372.

[48]Magnuson, J., S. Carpenter, and E. Stanley, 2016. North Temperate Lakes LTER: Benthic Macroinvertebrates 1981 - current, Data set ID:11. URL https://lter.limnology.wisc.edu/dataset/.

[49]Magnuson, J., S. Carpenter, and E. Stanley, 2016. North Temperate Lakes LTER: Fish Abundance 1981 - current, Data set ID:7. URL https://lter.limnology.wisc.edu/dataset/north-temperate-lakes-lter-fish-abundance-1981-current.

[50]Magurran, A. 2013. Measuring Biological Diversity. Wiley.

[51]Maldonado, C., C. I. Molina, A. Zizka, C. Persson, C. M. Taylor, J. Alban, E. Chilquillo, N. Ronsted, and A. Antonelli. 2015. Estimating species diversity and distribution in the era of Big Data: to what extent can we trust public databases? Global Ecology and Biogeography **24**:973–984. URL http://dx.doi.org/10.1111/geb.12326.

[52]Martyushev, L., and V. Seleznev. 2006. Maximum entropy production principle in physics, chemistry and biology. Physics Reports **426**:1–45.

[53]Mason, N. W., F. de Bello, D. Mouillot, S. Pavoine, and S. Dray. 2013. A guide for using functional diversity indices to reveal changes in assembly processes along ecological gradients. Journal of Vegetation Science **24**:794–806. URL http://dx.doi.org/10.1111/jvs.12013.

[54]Matthews, T. J., and R. J. Whittaker. 2015. REVIEW: On the species abundance distribution in applied ecology and biodiversity management. Journal of Applied Ecology **52**:443–454. URL http://dx.doi.org/10.1111/1365-2664.12380.

[55]McGill, B. J., R. S. Etienne, J. S. Gray, D. Alonso, M. J. Anderson, H. K. Benecha, M. Dornelas, B. J. Enquist, J. L. Green, F. He, A. H. Hurlbert, A. E. Magurran, P. A. Marquet, B. A. Maurer, A. Ostling, C. U. Soykan, K. I. Ugland, and E. P. White. 2007. Species abundance distributions: moving beyond single prediction theories to integration within an ecological framework. Ecology Letters **10**:995–1015. URL http://dx.doi.org/10.1111/j.1461-0248.2007.01094.x.

[56]McGlinn, D. J., P. G. Earls, and M. W. Palmer. 2010. A 12-year study on the scaling of vascular plant composition in an Oklahoma tallgrass prairie. Ecology **91**:1872–1872. URL http://dx.doi.org/10.1890/09-2017.1.

[57]Meloni, F., 2012. Influência do desenvolvimento florestal sobre a comunidade edáfico-epígea de Arthropoda e a mirmecofauna: bases para a bioindicação do processo sucessional na restauração ecológica. Ph.D. thesis, FFCLRP - Universidade de São Paulo. URL http://www.teses.usp.br/teses/disponiveis/59/59131/tde-02012013-120543/pt-br.php.

[58]Meloni, F., C. R. F. Granzotti, S. Bautista, and A. S. Martinez. 2017. Scale dependence and patch size distribution: clarifying patch patterns in Mediterranean drylands. Ecosphere **8**:e01690–n/a. URL http://dx.doi.org/10.1002/ecs2.1690.

[59]Meloni, F., N. P. Lopes, and E. M. Varanda. 2012. The relationship between leaf nitrogen, nitrogen metabolites and herbivory in two species of Nyctaginaceae from the Brazilian Cerrado. Environmental and Experimental Botany **75**:268–276.

[60]Meloni, F., and E. M. Varanda. 2015. Litter and soil arthropod colonization in reforested semi-deciduous seasonal Atlantic forests. Restoration Ecology **23**:690–697. URL http://dx.doi.org/10.1111/rec.12236.

[61]Menta, C., A. Tagliapietra, C. G., Z. A., and P. S. 2015. Ibs-Bf and Qbs-Ar Comparison: Two Quantitative Indices Based on Soil Fauna Community. EC Agriculture **2**:427–439.

[62]Morgan Ernest, S. K., G. M. Yenni, G. Allington, E. M. Christensen, K. Geluso, J. R. Goheen, M. R. Schutzenhofer, S. R. Supp, K. M. Thibault, J. H. Brown, and T. J. Valone. 2016. Long-term monitoring and experimental manipulation of a Chihuahuan desert ecosystem near Portal, Arizona (1977–2013). Ecology **97**:1082–1082. URL http://dx.doi.org/10.1890/15-2115.1.

[63]Oliver, J., K. Prudic, and S. Collinge. 2006. Boulder County Open Space butterfly diversity and abundance. Ecology **87**.

[64]Ottermanns, R., P. W. Hopp, M. Guschal, G. P. dos Santos, S. Meyer, and M. Rob-Nickoll. 2011. Causal relationship between leaf litter beetle communities and regeneration patterns of vegetation in the Atlantic rainforest of Southern Brazil (Mata Atlântica). Ecological Complexity **8**:299 – 309. URL http://www.sciencedirect.com/science/article/pii/S1476945X11000444.

[65]Palmer, M. M., R. K. Peet, R. A. Reed, W. Xi, and P. S. White. 2007. A multiscale study of vascular plants in a North Carolina Piedmont forest. Ecology **88**:162–D1. URL http://esapubs.org/Archive/ecol/E088/162/default.htm.

[66]Paquette, A., E. Laliberté, A. Bouchard, S. de Blois, P. Legendre, and J. Brisson. 2007. LAC CROCHE UNDERSTORY VEGETATION DATA SET (1998–2006). Ecology **88**:3209–3209. URL http://dx.doi.org/10.1890/07-0513.1.

[67]Parisi, V., C. Menta, C. Gardi, C. Jacomini, and E. Mozzanica. 2005. Microarthropod communities as a tool to assess soil quality and biodiversity: a new approach in Italy. Agriculture, Ecosystems & Environment **105**:323 – 333. URL http://www.sciencedirect.com/science/article/pii/S0167880904000970.

[68]Pathria, R., and P. Beale. 2011. Statistical Mechanics. Elsevier.

[69]Pelissier, R., J.-P. Pascal, N. Ayyappan, B. R. Ramesh, S. Aravajy, and S. R. Ramalingam. 2011. Tree demography in an undisturbed Dipterocarp permanent sample plot at Uppangala, Western Ghats of India. Ecology **92**:1376–1376. URL http://dx.doi.org/10.1890/10-1991.1.

[70]R Development Core Team, 2017. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria. URL http://www.R-project.org.

[71]Ritchie, M. E. 2009. Scale, Heterogeneity, and the Structure and Diversity of Ecological Communities. Princeton University Press.

[72]Rosenfeld, J. S. 2002. Functional redundancy in ecology and conservation. Oikos **98**:156–162. URL http://dx.doi.org/10.1034/j.1600-0706.2002.980116.x.

[73]Schipper, A. M., H. W. M. Hendriks, M. J. Kauffman, A. J. Hendriks, and M. A. J. Huijbregts. 2013. Modelling interactions of toxicants and density dependence in wildlife populations. Journal of Applied Ecology **50**:1469–1478. URL http://dx.doi.org/10.1111/1365-2664.12142.

[74]Shannon, C. E. 1948. A Mathematical Theory of Communication. Bell System Technical Journal **27**:379–423. URL http://dx.doi.org/10.1002/j.1538-7305.1948.tb01338.x.

[75]Stevens, S., and J. Vokmann. 1937. A scale for the measurement of the psycological magnitude pitch. The Journal

of the Acoustical Society of America **8**:185–190. URL http://asa.scitation.org/doi/10.1121/1.1915893.

[76]Thompson, R., and B. M. Starzomski. 2007. What does biodiversity actually do? A review for managers and policy makers. Biodiversity and Conservation **16**:1359–1378. URL http://dx.doi.org/10.1007/s10531-005-6232-9.

[77]Trebilco, R., J. K. Baum, A. K. Salomon, and N. K. Dulvy. 2013. Ecosystem ecology: size-based constraints on the pyramids of life. Trends in Ecology & Evolution **28**:423 – 431. URL http://www.sciencedirect.com/science/article/pii/S0169534713000888.

[78]Tsallis, C. 2009. Introduction to Nonextensive Statistical Mechanics. Springer.

[79]Valencia, E., F. T. Maestre, Y. Le Bagousse-Pinguet, J. L. Quero, R. Tamme, L. Borger, M. Garcia-Gomez, and N. Gross. 2015. Functional diversity enhances the resistance of ecosystem multifunctionality to aridity in Mediterranean drylands. New Phytologist **206**:660–671. URL http://dx.doi.org/10.1111/nph.13268.

[80]Wurtz, P., and A. Annila. 2010. Ecological succession as an energy dispersal process. Biosystems **100**:70 – 78. URL http://www.sciencedirect.com/science/article/pii/S0303264710000055.

[81]Zhao, K., M. Musolesi, P. Hui, W. Rao, and S. Tarkoma. 2015. Explaining the power-law distribution of human mobility through transportation modality decomposition **5**. URL http://dx.doi.org/10.1038/srep09136.