Received: 8 June 2020 | Revised: 9 January 2021 | Accepted: 14 January 2021

DOI: 10.1111/geb.13270

MACROECOLOGICAL METHODS

Global Ecology and Biogeography A Journal of Macroecology WILEY

From zero to infinity: Minimum to maximum diversity of the planet by spatio-parametric Rao's quadratic entropy

Duccio Rocchini[1,2] | Matteo Marcantonio[3] | Daniele Da Re[4] | Giovanni Bacaro[5] | Enrico Feoli[5] | Giles M. Foody[6] | Reinhard Furrer[7] | Ryan J. Harrigan[8] | David Kleijn[9] | Martina Iannacito[10] | Jonathan Lenoir[11] | Meixi Lin[12] | Marco Malavasi[2] | Elisa Marchetto[1] | Rachel S. Meyer[12] | Vítězslav Moudry[2] | Fabian D. Schneider[13] | Petra Šímová[2] | Andrew H. Thornhill[14,15] | Elisa Thouverai[1] | Saverio Vicario[16] | Robert K. Wayne[12] | Carlo Ricotta[17]

[1]BIOME Lab, Department of Biological, Geological and Environmental Sciences, Alma Mater Studiorum University of Bologna, Bologna, Italy

[2]Department of Spatial Sciences, Faculty of Environmental Sciences, Czech University of Life Sciences Prague, Praha - Suchdol, Czech Republic

[3]Department of Pathology, Microbiology, and Immunology, School of Veterinary Medicine, University of California, Davis, California, USA

[4]Georges Lemaître Center for Earth and Climate Research, Earth and Life Institute, UCLouvain, Louvain-la-Neuve, Belgium

[5]Department of Life Sciences, University of Trieste, Trieste, Italy

[6]School of Geography, University of Nottingham, Nottingham, UK

[7]Department of Mathematics and Department of Computational Science, University of Zurich, Zurich, Switzerland

[8]Center for Tropical Research Institute of the Environment and Sustainability, University of California-Los Angeles, Los Angeles, California, USA

[9]Plant Ecology and Nature Conservation Group, Wageningen University, Wageningen, The Netherlands

[10]Inria Bordeaux - Sud-Ouest, Talence, France

[11]UR "Ecologie et Dynamique des Systèmes Anthropisés" (EDYSAN, UMR 7058 CNRS-UPJV), Université de Picardie Jules Verne, Amiens, France

[12]Department of Ecology and Evolutionary Biology, University of California-Los Angeles, Los Angeles, California, USA

[13]Jet Propulsion Laboratory, California Institute of Technology, Pasadena, California, USA

[14]Environment Institute, The University of Adelaide, Adelaide, South Australia, Australia

[15]Department for Environment and Water, State Herbarium of South Australia, Botanic Gardens and State Herbarium, Adelaide, South Australia, Australia

[16]Institute of Atmospheric Pollution Research–Italian National Research Council C/O Department of Physics, University of Bari, Italy

[17]Department of Environmental Biology, University of Rome "La Sapienza", Rome, Italy

Correspondence
Duccio Rocchini, Alma Mater Studiorum University of Bologna, Department of Biological, Geological and Environmental Sciences, via Irnerio 42, 40126, Bologna, Italy.
Email: duccio.rocchini@unibo.it

Funding information
H2020 project SHOWCASE; H2020 COST Action CA17134 'Optical synergies for spatiotemporal sensing of scalable ecophysiological traits (SENSECO); Swiss National Science Foundation, Grant/Award Number: SNSF 175529

Editor: Thomas Gillespie

Abstract

Aim: The majority of work done to gather information on the Earth's biodiversity has been carried out using in-situ data, with known issues related to epistemology (e.g., species determination and taxonomy), spatial uncertainty, logistics (time and costs), among others. An alternative way to gather information about spatial ecosystem variability is the use of satellite remote sensing. It works as a powerful tool for attaining rapid and standardized information. Several metrics used to calculate remotely sensed diversity of ecosystems are based on Shannon's information theory, namely on the differences in relative abundance of pixel reflectances in a certain area. Additional metrics like the Rao's quadratic entropy allow the use of spectral distance beside abundance, but they are point descriptors of diversity, that is they

---

Duccio Rocchini and Matteo Marcantonio contributed equally to the paper.

can account only for a part of the whole diversity continuum. The aim of this paper is thus to generalize the Rao's quadratic entropy by proposing its parameterization for the first time.

Innovation: The parametric Rao's quadratic entropy, coded in R, (a) allows the representation of the whole continuum of potential diversity indices in one formula, and (b) starting from the Rao's quadratic entropy, allows the explicit use of distances among pixel reflectance values, together with relative abundances.

Main conclusions: The proposed unifying measure is an integration between abundance- and distance-based algorithms to map the continuum of diversity given a satellite image at any spatial scale. Being part of the `rasterdiv` R package, the proposed method is expected to ensure high robustness and reproducibility.

KEYWORDS

biodiversity, ecological informatics, modelling, remote sensing, satellite imagery

1 | INTRODUCTION

Since Alexander von Humboldt (1769–1859), the spatial component of nature has played a relevant role in natural science. In the development of theoretical and empirical models in ecology, spatial structure represents a key concept to allow scientists to link ecological patterns to the generating processes and to the functional networking among organisms (Borcard & Legendre, 2002). The majority of the work done to gather information about Earth diversity has been carried out using in-situ data, with known issues related to epistemology (e.g., species determination and taxonomy), spatial uncertainty, logistics (time and costs), among others (Rocchini et al., 2011).

Using satellite remote sensing can at least help to attain rapid and standardized information about Earth diversity (Gillespie, 2005; Rocchini et al., 2005, 2018). Furthermore, remote sensing can also be used to monitor some ecosystem functions and parameters such as temperatures, precipitation, photosynthesis, and vegetation biomass production (Schimel et al., 2019; Zellweger et al., 2019) that can be useful to define the different niches of in-situ species, following Goodall's (1970) ideas, who envisaged future diversity measures as those based on niche theory (Hutchinson, 1959). The free access to remote sensing data (see Zellweger et al., 2019) has opened new ways to study ecosystem diversity and biodiversity issues (Aerts et al., 2017). The spectral data related to pixels, as operational geographical units, are descriptions of pieces of land that allow us to define a new kind of Earth 'diversity', which may complement in-situ biodiversity measurement (Hernandez-Stefanoni et al., 2012).

Diversity varies with area, thus investigating multiple spatial grains, until wide extents, is important to effectively monitor spatial diversity change in space and time (MacArthur et al., 1966). This is especially true in macroecology, where the primary aim is to model large-scale spatial patterns to infer the ecological processes that generated them, particularly considering the recent effect of global changes worldwide (Hobohm et al., 2019). In order to determine the horizontal distribution of diversity within a satellite image (i.e., which areas within the image are more diverse than others), diversity indices are usually spatially referenced by calculating the index within a moving window.

Several metrics that measure diversity from satellites rely on the Shannon's theory of entropy (Shannon, 1948), with diversity being measured as $H = -\sum_{i=1}^{N} p_i \log p_i$, where p_i is the proportion of the ith pixel value (e.g., digital number, DN) found within a moving window containing N pixels. Shannon's H basically summarizes the partition of abundances (sensu Whittaker, 1965) by taking into account both relative abundance and richness of DNs (Figure 1).

However, Shannon's entropy is a point descriptor of (remotely sensed) diversity. As such, it shows only one part of the whole potential diversity spectrum at a glance. The use of generalized entropies has been advocated to address this problem. In this case, one single formula represents a parameterized version of a diversity index, thus providing a continuum of potential diversity indices. In the context of the measurement of diversity, the Rényi (1970) parametric entropy

$$H_\alpha = \frac{1}{1-\alpha} \log \sum_{i=1}^{N} p_i^\alpha \tag{1}$$

with $0 \le \alpha \le \infty$ represents a powerful tool to account for the continuum of diversity (Figure 1).

One particularly convenient property of H_α is that by varying the parameter α there is a continuum of possible diversity measures, which differ in their sensitivity to rare and abundant DNs, becoming increasingly dominated by the most common DNs for increasing values of α. Note that for $\alpha \to 1$, H_1 equals the Shannon's entropy. A similar formulation was then proposed by Hill (1973) who expressed parametric diversity as the 'numbers equivalent' of Rényi generalized entropy (Supporting Information Appendix S1).

The Rényi (and Hill) parametric functions summarize diversity by taking into account the pixel values of a satellite image and their relative abundances. However, they do not allow explicit consideration of the differences among these values. As an example, two arrays of

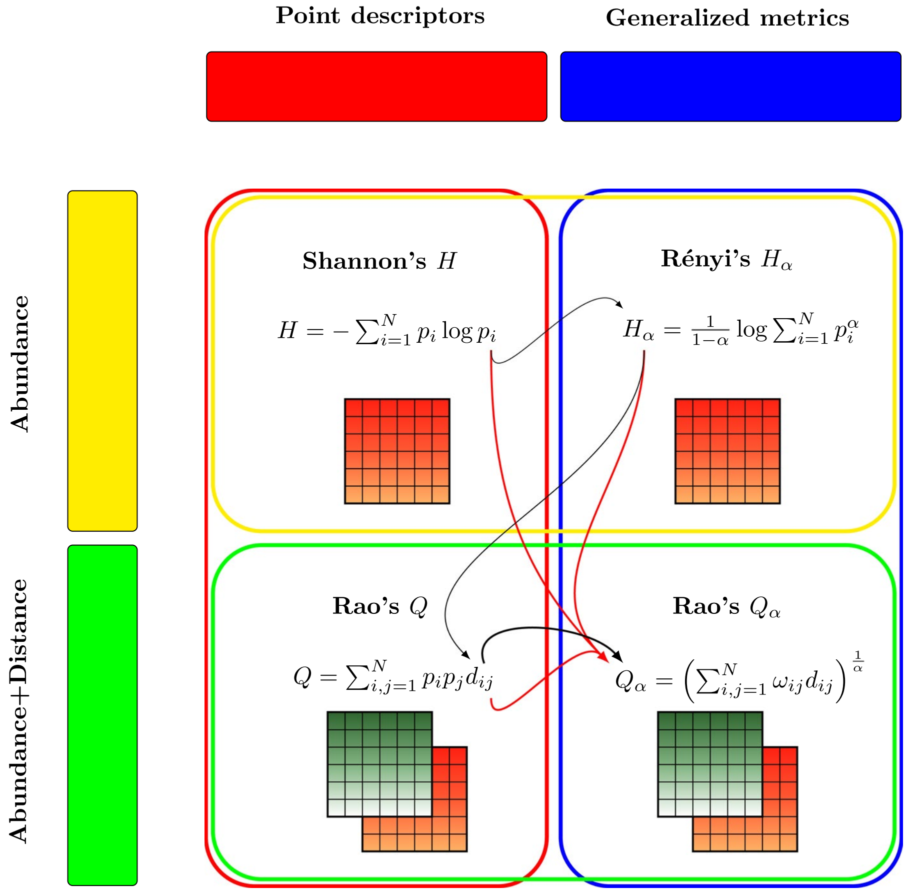

FIGURE 1 Grounding theory of this paper. Diversity measures can encompass abundance-based as well as abundance-distance-based metrics (yellow and green boxes, respectively). Abundance-distance-based metrics allow multiple layers to be used. The black arrows represent the theoretical flow of this paper, with the thickness representing the complexity of each index, starting from Shannon's information theory (point descriptor) to Rényi's H_α (generalized entropy), which do not make use of distance. Distance enters the Rao's Q formula, but this is still a point descriptor of diversity. Finally, parametric Rao's Q_α comprises the use of distances and the generalized entropy concept. The red arrows represent the properties of the Rao's Q_α: (a) it is grounded in information theory starting from Shannon's H, (b) it is a generalized entropy like the Rényi H_α, and (c) it makes use of distances like the Rao's Q [Colour figure can be viewed at wileyonlinelibrary.com]

nine pixels with maximum richness and evenness (i.e., both containing nine different DNs with relative abundances $p_i = \frac{1}{9}$) but differing in their values will attain the same Shannon diversity irrespective of the values of the DNs in both arrays.

By introducing a distance parameter d_{ij} among each pair of values i and j, Rao's quadratic entropy (Rao, 1982)

$$Q = \sum_{i,j=1}^{N} p_i p_j d_{ij} \tag{2}$$

explicitly considers the differences among the pixel values in the calculation of diversity (Figure 1). Hence, two different pixels with values [2,3] will attain a lower diversity with respect to two pixels with values [0,100]. For instance, to make an ecological parallel, this is somewhat similar to the phylogenetic distance between two species: the values [2,3] would be equivalent to two sister species closely related on the tree of life while [1,100] would be equivalent to two very distant species on the tree of life.

The aim of this paper is thus to propose, for the first time, a parameterization of Rao's quadratic entropy in order to provide a generalized entropy that accounts for both relative abundances and distances among pixel values. The proposed approach is now part of the `rasterdiv` R package, a package dedicated to diversity measures of spatial matrices, increasing its capability to discern among different diversity measures by a single formula.

2 | SPATIO-PARAMETRIC RAO'S QUADRATIC ENTROPY

Inter-pixel spectral distances are directly related to landscape heterogeneity and they are capable of describing species habitats, starting with a satellite image (Rocchini et al., 2005). A satellite image can be viewed as a matrix of numbers describing Earth reflectance in different dimensions stored as pixels. A sensor per light wavelength records the reflectance values of a certain object in that wavelength, which are stored as numbers in a certain range (e.g., digital numbers in eight bits, ranging from 0 to 255). In general, the higher the variability in the spectral space defined by the pixel reflectance values, the higher the diversity of the ecosystem under study.

Consider a window of N pixels moving across the whole image to calculate a diversity index. Let i and j be two pixels randomly chosen with repetition within the moving window. Let d_{ij} be a symmetric measure of the (multi)spectral distance between i and j such that $d_{ij} = d_{ji}$ and $d_{ii} = 0$. Rao's Q (Rao, 1982) is defined as:

$$Q = \sum_{i,j=1}^{N} p_i p_j d_{ij} = \sum_{i,j=1}^{N} \frac{1}{N} \times \frac{1}{N} d_{ij} \tag{3}$$

Therefore, Q measures the expected (i.e., mean) distance between two randomly chosen pixels and $\frac{1}{N}$ is the probability of extracting each pixel. Note that, unlike H_α the calculation of Rao's quadratic entropy is not limited to single bands but can be extended to multispectral systems of any dimension. For the connection between quadratic entropy and variance, see Rocchini et al. (2019).

A more direct approach for developing a parametric version of quadratic entropy stems from the work of Guiasu and Guiasu (2011). Let $\omega_{ij} = \frac{1}{N} \times \frac{1}{N}$ be the combined probability of selecting pixels i and j in this order. Guiasu and Guiasu (2011) noted that Rao's Q can be expressed as a linear function of the combined probabilities of all pairs of pixels:

$$Q = \sum_{i,j=1}^{N} \omega_{ij} d_{ij} = \sum_{i,j=1}^{N} \frac{1}{N} \times \frac{1}{N} d_{ij} = \sum_{i,j=1}^{N} \frac{1}{N^2} d_{ij} \tag{4}$$

In practice, Rao's Q is the arithmetic mean of the distances d_{ij} between all pairs of pixels i and j. Hence, in order to implement a parametric version of Rao's Q, it seems natural to substitute the arithmetic mean in Equation 3 with a generalized mean (Hardy et al., 1952):

$$Q_\alpha = \left(\sum_{i,j=1}^{N} \omega_{ij} d_{ij}^{\alpha} \right)^{\frac{1}{\alpha}} = \left(\sum_{i,j=1}^{N} \frac{1}{N^2} d_{ij}^{\alpha} \right)^{\frac{1}{\alpha}} \tag{5}$$

This operation connects Q_α with other diversity metrics that are expressed as generalized means, such as Hill's (Hill, 1973) or Jost's (Jost, 2006) numbers (Supporting Information Appendix S1) equivalents (see also Leinster & Cobbold, 2012).

The Rao's Q, viewed as an arithmetic mean, is one of all the possible means in its generalized form Q_α:

$$Q_\alpha = \begin{cases} \alpha \to 0, Q_0 = \sqrt[N^2]{\prod_{i,j=1}^{N} d_{ij}} \leftarrow \textit{geometric} \\ \alpha = 1, Q_1 = Q = \sum_{i,j=1}^{N} \frac{1}{N^2} d_{ij} \leftarrow \textit{arithmetic} \\ \alpha = 2, Q_2 = \sqrt{\sum_{i,j=1}^{N} \frac{1}{N^2} d_{ij}^2} \leftarrow \textit{quadratic} \\ \alpha = 3, Q_3 = \sqrt[3]{\sum_{i,j=1}^{N} \frac{1}{N^2} d_{ij}^3} \leftarrow \textit{cubic} \\ \alpha \to \infty, Q_{\alpha \to \infty} = \max d_{ij} \leftarrow \textit{max}_d \end{cases} \tag{6}$$

The mathematical proof that (a) for $\alpha \to 0$, Q_0 corresponds to the geometric mean, and (b) for $\alpha \to \infty$, Q_∞ corresponds to the maximum distance between pixel values pairs is provided in Supporting Information Appendix S1.

Each generalized mean always lies between the smallest and largest of its values. Increasing the parameter α will increase the weight of the highest values of d_{ij}, thus providing a continuum of potential diversity indices (Figure 1).

3 | THE ALGORITHM

3.1 | How does it work?

Starting from a satellite image, a spatial moving window might be used to make the calculation on predefined extents of analysis. The grain (sensu Dungan et al., 2002) will be the resolution of the image while the extent of analysis will be the size of the moving window (see Figure S1 in Supporting Information Appendix S2). The calculation is based on a distance matrix of type:

$$M_d = \begin{pmatrix} d_{\lambda_1,\lambda_1} & d_{\lambda_1,\lambda_2} & d_{\lambda_1,\lambda_3} & \cdots & d_{\lambda_1,\lambda_n} \\ d_{\lambda_2,\lambda_1} & d_{\lambda_2,\lambda_2} & d_{\lambda_2,\lambda_3} & \cdots & d_{\lambda_2,\lambda_n} \\ d_{\lambda_3,\lambda_1} & d_{\lambda_3,\lambda_2} & d_{\lambda_3,\lambda_3} & \cdots & d_{\lambda_3,\lambda_n} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ d_{\lambda_n,\lambda_1} & d_{\lambda_n,\lambda_2} & d_{\lambda_n,\lambda_3} & \cdots & d_{\lambda_n,\lambda_n} \end{pmatrix} \tag{7}$$

among all the potential pairs of pixels inside the moving window. The diagonal terms of the matrix (which equal zero) will have no effect for $\alpha > 0$ (Equation (5)), since they would enter the $\sum$ term. On the contrary, for $\alpha \to 0$, they would enter the $\prod$ term by nullifying Q_0.

We coded the proposed parameterization of Rao's quadratic entropy as an R function (R Core Team, 2021), implementing the previously developed `rasterdiv` package (Marcantonio et al., 2020; https://CRAN.R-project.org/package=rasterdiv). The calculation of different Q_α by automatically changing the range of potential α values is done by the function `paRao`, as:

```
> paRao(x, alpha=c(0:4,Inf), method="classic",
 dist_m="euclidean", window=9, na.tolerance=0.5, simplify=3,
 np=8, cluster.type="SOCK", diag=TRUE)
```

where x is the input data set, which can be a `RasterLayer` or a matrix class object, and `alpha` is the α parameter of Equation 4, which can be a single value or a vector of integers. In the example above, α is a vector of integers ranging from 0 to 4, plus Inf, which in the R language is a reserved word representing positive infinity ($\alpha \to \infty$). The option `method` decides if `paRao` is calculated with one single layer (`classic`) or with more than one layer (`multidimension`). With `method="multidimension"` then x must be a list of objects. The argument `dist_m` is the type of distance considered in the calculation of the index, and can be set to any distance class implemented in the R package proxy, such as "`euclidean`", "`canberra`"

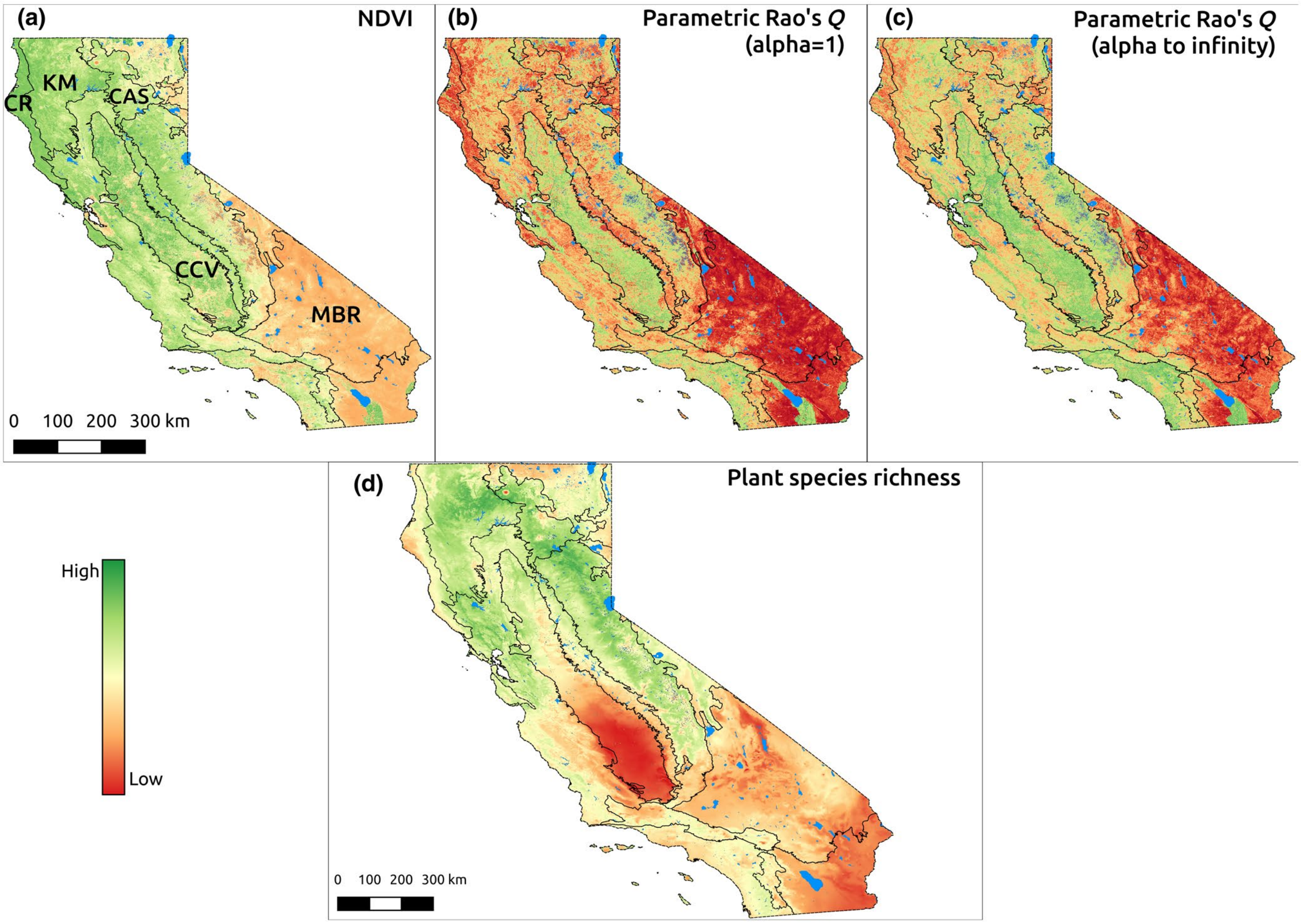

FIGURE 2 Maps showing normalized difference vegetation index (NDVI), the Rao's *Q* index and native plant species richness for the ecoregions of California. The NDVI values shown in the top-left box (a) (100 m resolution) were derived from the European Space Agency Copernicus Sentinel-2 data set, processed with Google Earth Engine and range between −0.26 (red) and 0.99 (dark green). The Rao's *Q* indices in the top-centre and top-right boxes were calculated from the NDVI map using $\alpha = 1$ and $\alpha \to \infty$ ((b) and (c)) and a moving window of 9 × 9 pixels. Higher Rao's *Q* values represent pixels whose surrounding NDVI values are more 'diverse' than pixels with lower Rao's *Q* values. The bottom map (d), reporting the pixels' potential native plant species richness (resolution: 810 m), was derived by summing the binary potential distribution ranges of 5,222 native plant species modelled by Thornhill et al. (2017). Species richness ranges between 134 (red) and 1,029 (green) species per pixel (1 km^2). Five ecoregions are labelled on the NDVI map and discussed in the main text: Coast Range (CR), Klamath Mountains (KM), Cascades (CAS), California Central Valley (CCV) and Mohave Basin and Range (MBR). A mask with water bodies was added to the maps. Refer to the main text for additional information [Colour figure can be viewed at wileyonlinelibrary.com]

or "`manhattan`". Moreover, `dist_m` can also be a user-defined matrix of distances. However, if `method` is set to "`classic`" (unidimensional `paRao`) all distance types reduce to the Euclidean distance. The argument `window` is the side length in cells of the moving window (in this case set to 9), whereas `na.tolerance` is the proportion (0–1) of NA cell allowed in a moving window: if the proportion of NA cells in a moving window exceeds `na.tolerance` then the value of the moving window central pixel will be NA. The option `simplify` allows the number of decimal places to be reduced to ease the calculation by reducing the number of numerical categories, that is, if `simplify = 3` only the first three digits of data will be considered for the calculation of the index. The argument `np` is the number of parallel processes used in the calculation. If `np` > 1 then the `doParallel` package will be called for parallel calculation, and `cluster.type` will indicate the type of cluster to be opened (default is "`SOCK`", "`MPI`" and "`FORK`" are the alternatives). The `diag` argument refers to the diagonal term of Equation (6). It will have no effect on the function for $\alpha > 0$, while it will nullify the value of Q_α if set to `TRUE`, as previously explained in Equation (6).

In the next section, we provide an ecological example of the application of the spatio-parametric Rao's Q, done at regional scale to allow a comparison with in-situ data on species diversity. Supporting Information Appendix S2 provides an example at worldwide spatial scale.

3.2 | Case study: the diversity of vegetation greenness and the ecoregions of California

A comparison between in-situ and remotely sensed diversity at worldwide scale might be difficult due to known biases in, for example, sampling effort, taxonomies, spatial uncertainty (Rocchini et al., 2017). Hence, we decided to calculate the parametric Rao's *Q* index on a normalized difference vegetation index (NDVI) raster layer of California (USA) to be compared with data in the field on native plant species diversity provided in Thornhill et al. (2017) from Baldwin et al. (2017). We chose California as a case study due to its high ecological diversity as well as to the availability of plant species field-data for this region.

In practice, we aimed at visualizing and describing differences in both diversity and structure of vegetation for the state of California, USA. First, an NDVI raster layer was derived from Copernicus Sentinel-2 data (European Space Agency, reference period: January 2017 to July 2018) and processed through Google Earth Engine to filter out cloud cover, select the greenest pixel of the time series and resample at 100 m pixel resolution. Then, the `paRao` R function was used to derive the Rao's *Q* index, considering both the original formulation of the Rao's *Q* ($\alpha = 1$, Equation (5)) and the formulation with $\alpha \to \infty$ maximizing β-diversity (Supporting Information Appendix S2), with a moving window of 9×9 pixels.

A map of plant species richness was derived relying on the potential distribution range of 5,222 native California vascular plants modelled by Thornhill et al. (2017) and used to report alpha-diversity (plant species richness per pixel), whereas a vector map of the ecoregions of California (level III; from the United States Environmental Protection Agency) was used to identify beta-diversity. In Figure 2, we show NDVI, the Rao's *Q* indices with $\alpha = 1$ and $\alpha \to \infty$ and native plant species richness, reporting the boundaries of the different ecoregions for California. This comparison revealed macro-ecological and biogeographical patterns that can be better interpreted considering the information condensed in the Rao's *Q* index (Figure 3). We focused on five important ecoregions, which are used in Figure 2 and in the boxplots of Figure 3 to discuss the most evident ecological patterns.

Overall, the Rao's *Q* index matched native plant species diversity in areas characterized by a lower human impact, such as in the north-west and south-east portions of the state. The 'Cascades' (CAS) and 'Klamath Mountains' (KM) ecoregions are characterized by highly dissected ridges, foothills, and valleys and host a very diverse flora, rich in endemic and relic species (Griffith et al., 2016). Rao's *Q* was the highest in these two areas (apart for the agricultural and urbanized areas described later), despite moderate NDVI values (Figures 2 and 3). The low mountains covered by highly productive, rain-drenched evergreen forests composing the 'Coast Range' (CR) ecoregion showed lower plant diversity and Rao's *Q*, despite very high NDVI values. By contrast, the dry and warm 'Mohave basin and range' (MBR) is characterized by broad basins and scattered mountains and showed low NDVI, Rao's *Q* and plant species richness.

On the contrary, Rao's *Q* performed poorly when compared with native plant species richness in agricultural and developed lands with high productivity (i.e., high NDVI) and heterogeneity. This is the case of the 'California Central Valley' (CCV) region, which is composed of flat, urbanized and intensively farmed plains. The extensive presence of irrigated crops intersected with urbanized areas caused medium to high NDVI values and an apparently high Rao's *Q* diversity, but a low native species richness, especially in the drier southern portion of the valley.

Passing from the pure Rao's *Q* index ($\alpha = 1$) to its parameterization with $\alpha \to \infty$ helped to increase the discrimination among areas, due to the fact that when $\alpha \to \infty$ the Rao's *Q* corresponds to the maximum distance (β-diversity) among pixel values in a site. Very

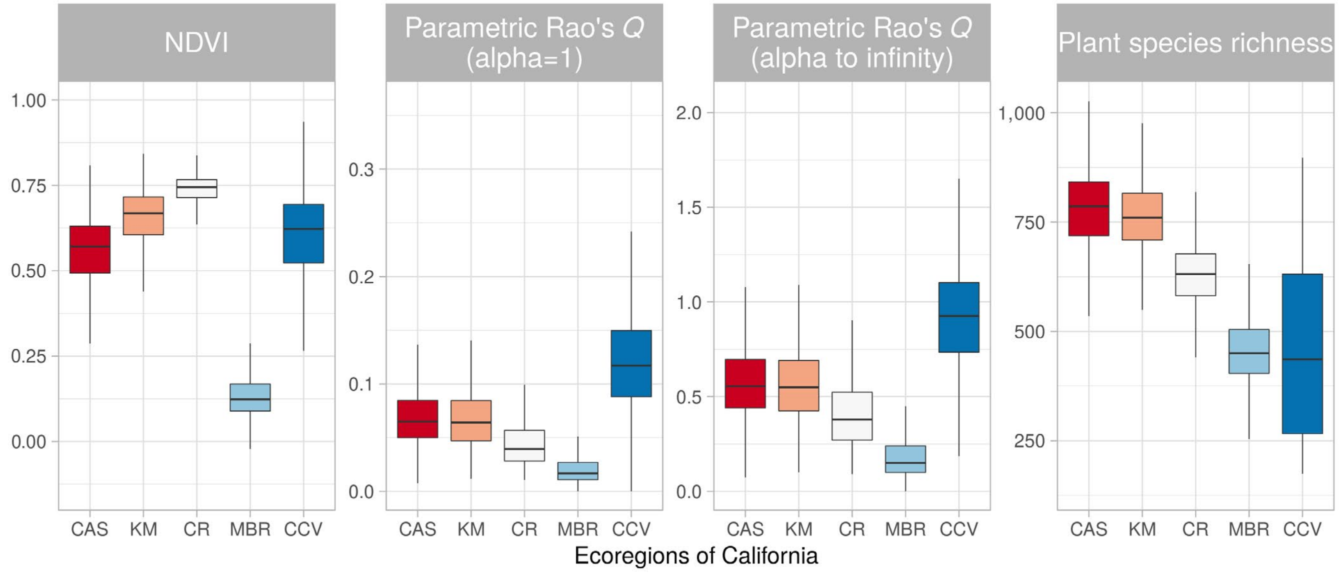

FIGURE 3 Boxplot showing the distribution of values for the normalized difference vegetation index (NDVI), the parametric Rao's *Q* (with $\alpha = 1$ and $\alpha \to \infty$) and plant species richness. The Rao's *Q* index matched native plant species diversity in those ecoregions characterized by a lower human impact, such as Cascades (CAS), Klamath Mountains (KM), Coast Range (CR), Mohave Basin and Range (MBR). On the contrary, Rao's *Q* performed poorly when compared with native plant species richness in agricultural and developed lands with high productivity (i.e., high NDVI) and heterogeneity (California Central Valley, CCV). See also Figure 2 and the main text for additional information [Colour figure can be viewed at wileyonlinelibrary.com]

similar gradients of the spatial heterogeneity of California (including BIOCLIM variables, NDVI, elevation) as well as environmental DNA (eDNA) data are found in Lin et al. (in press).

4 | DISCUSSION

In this paper, we have provided a straightforward solution to: (a) account for distances in information theory based metrics, and (b) provide a generalized formula in order to avoid point description and account for the continuum of diversity. Diversity can be represented by different dimensions (Nakamura et al., 2020). Considering one single measure to account for the whole continuum of diversity metrics might be a powerful addition to the main framework. On the contrary, fragmenting the concept of diversity when trying to capture single aspects of the whole spectrum could be counterproductive.

The proposed unifying measure succeeded in integrating abundance- and distance-based algorithms over a wide variety of diversity metrics. We demonstrated that such integration is not only theoretical, but also applicable to real spatial data, considering several dimensions of diversity at the same time. Being part of the `rasterdiv` R package, the proposed method is expected to ensure high robustness and reproducibility.

Remote sensing is obviously not a panacea for all the organismic-based diversities like taxonomic, functional and genetic diversity but it can represent an important exploratory tool to detect diversity hotspots and their changes in space and time at the ecosystem level (Rocchini, 2007). First of all, it measures heterogeneity of the environment with indirect links to the biodiversity of both plant and animal taxa, but also with potential discrepancies with species diversity, as in the presented case study of the native plant species diversity of California. This said, depending on the complexity and the resolution at which the proposed parameterized Rao's *Q* is applied, it might allow new insights into the ecological processes acting in a certain ecosystem to shape its diversity. In this paper, the examples provided were based on a single NDVI layer since (a) it is a valuable index of vegetation health and (b) it is freely available in the `rasterdiv` package to reproduce the code proposed in this paper (see Supporting Information Appendix S2 for an application at worldwide spatial scale based on the Copernicus Proba-V NDVI freely available in the package). We are aware that NDVI has very limited capacity to track diversity in some habitats like dense forests, because it is saturated at dense vegetation (Mutanga & Skidmore, 2004). From this point of view, imaging spectroscopy offers higher information content, also enabling plant functional trait retrievals (Jetz et al., 2016; Schneider et al., 2019) as well as structural traits by LiDAR data (Schneider et al., 2020). The application of the proposed algorithm to future spaceborne imaging spectroscopy is promising. In other words, the algorithm has been designed to be used with multiple layers, like a whole multispectral image or the most meaningful principal components (Peres-Neto et al., 2005), or land use classes probabilities derived from fuzzy set theory (Feoli, 2018; Rocchini & Ricotta, 2007). This is even one of the major advantages of the Rao's *Q* index as it allows consideration of both abundance and distance among pixel values, thus being applicable to any continuous raster layer, or to any matrix combination, even in a multiple spectral system.

Creating a unique 'umbrella' under which all of the potential metrics of diversity can be used is highly beneficial for, for example, monitoring the variation in time of biological systems considering two major axes: (a) the α parameter in Equation 4 providing information about the type of diversity at time t_0, (b) the temporal dimension from time t_0 to time t_n given the same α parameter. For the future, exploring this temporal dimension would allow information on ecosystem changes in different diversity types to be gathered at a glance.

Moreover, generalized entropy allows us to characterize the dimensionality of diversity (sensu Stevens & Tello, 2014) of different habitats/ecosystems. Those areas with a higher diversity dimensionality, namely a higher variability into the diversity spectrum, would need a generalized measure to be fully undertaken. On the contrary, ecosystems with a lower dimensionality would have a lower difference among diversity measures and thus a flat curve of the diversity spectrum (Nakamura et al., 2020).

From a functional point of view, when all indices of diversity are highly correlated to each other (low dimensionality), it is expected that the ecological processes underlying diversity are just a few. On the contrary, with a lower correlation among indices (higher dimensionality) there might be a higher number of axes of variation coming out from different processes shaping ecological heterogeneity in space (Stevens & Tello, 2014).

There might be the possibility that a completely random matrix produces a pattern of diversity (Type I error). On the other hand, a structured matrix could produce a very low diversity pattern (Type II error; Gotelli, 2000). In both cases, the parametric Rao's *Q* could allow researchers to determine, thanks to the use of a continuum of diversities, (a) why a diversity pattern is still produced even in the case of a random matrix, and (b) why a certain landscape shows a very low diversity at a certain point of the whole diversity spectrum. With point descriptors of diversity such inference cannot be done since the investigation is limited to a small window of the entire diversity spectrum, by basically relying on a single final number. In other words, the commonly asked question about which index best describes diversity has no certain answer (Gorelick, 2011). Hence, the use of a trend of diversities will lead to the comprehension of hidden parts of the whole diversity dimensionality.

Furthermore, it is expected that the ecological processes shaping diversity should act at defined spatial scales (Borcard & Legendre, 2002). Hence, different diversity types of the whole dimensionality spectrum are expected to show scale dependent patterns, being apparent only at certain scales and not at some others. The use of a continuum allows measurement of all of the different diversity types together in a single step. Changing the extent of the analysis by using different moving windows would then allow researchers to encompass different spatial structures at different scales.

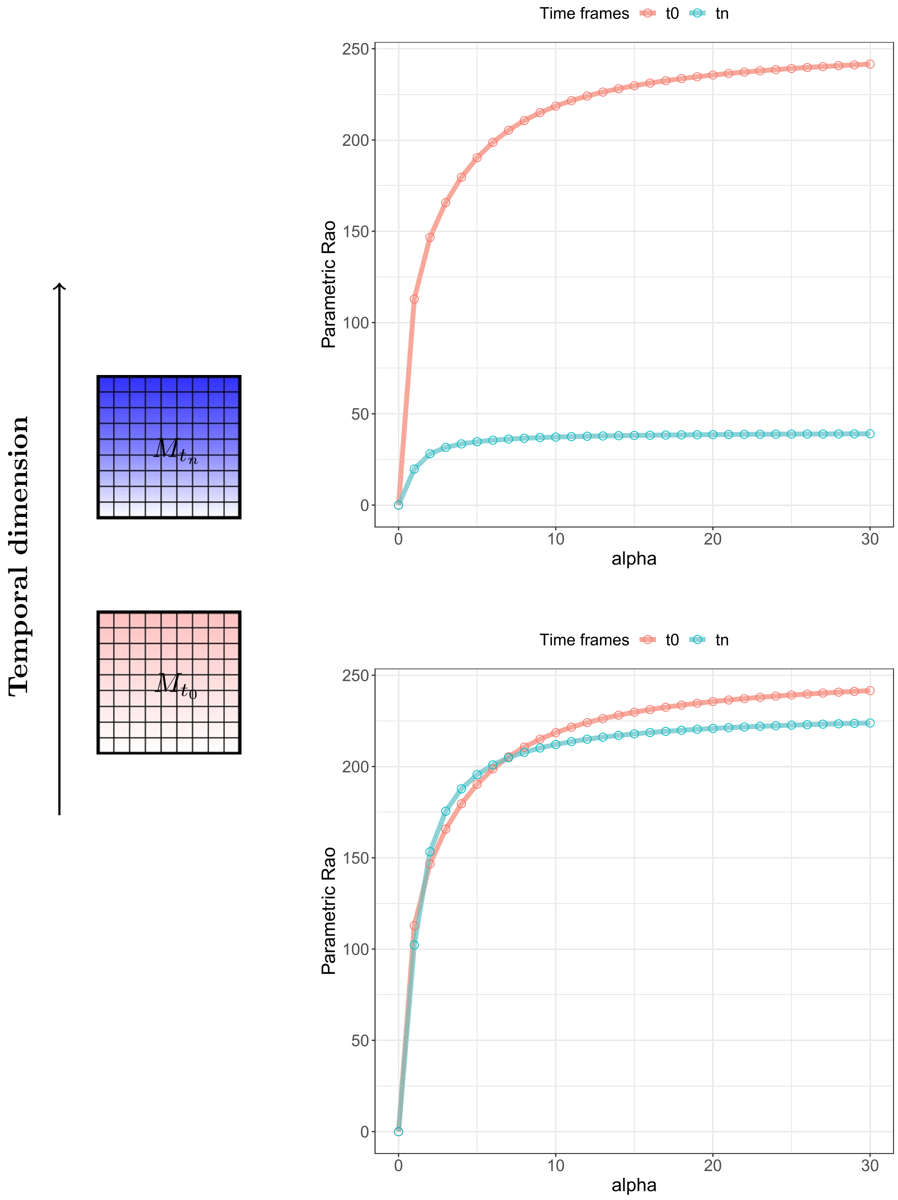

FIGURE 4 A theoretical example of the power of using generalized entropy for monitoring purposes. Given a landscape at times t_0 (pink) and t_n (blue), calculating generalized entropy will allow the formation of a graph showing the continuum of Rao's Q values observed over a range of values for α. The same landscape in different times might show an abrupt change (e.g., a catastrophic event) with an apparent diversity decrease (top). In this case, point descriptors (e.g., single α values) of diversity may be sufficient to describe this pattern. When the change in diversity is subtle (bottom), using a point descriptor might fail to detect it but it becomes manifest in the continuum of diversities based on generalized entropy. The complete code for reproducing this theoretical example is available in Supporting Information Appendix S3 [Colour figure can be viewed at wileyonlinelibrary.com]

While geographical gradients of diversity over space are complex to catch in their very nature, biodiversity measurement has mainly relied in the past on few formulas which represented a hegemony (Stevens et al., 2013). In this paper, we have demonstrated that diversity is actually multifaceted and should necessarily be approached through a generalized approach. Furthermore, the proposed generalized parametric Rao's Q index might be profitably used to plot multitemporal trends (see e.g., Dornelas et al., 2014) of diversity metrics and discover previously imperceptible differences when making use of single metrics (Figure 4).

5 | CONCLUSION

In order to unfold the dimensionality of diversity, methods to directly account for several aspects of diversity at a time are needed. From this point of view, generalized entropy undoubtedly represents a powerful approach for mapping the diversity continuum.

Metrics grounded in information theory ensure to make use of relative abundance of pixel values given the same richness in the moving window of analysis. However, distance metrics also allow the relative dispersion in the spectral space of the cloud of pixels in a certain area to be accounted for (Laliberté et al., 2020). The proposed parameterization of the Rao's *Q* explicitly considers the dispersion of pixel values in a spectral space (and their relative abundance) by allowing the whole dimensionality of diversity to be captured.

ACKNOWLEDGMENTS

DR and DK were partially supported by the H2020 project SHOWCASE (Grant agreement No862480). DR was also supported by the H2020 COST Action CA17134 'Optical synergies for spatiotemporal sensing of scalable ecophysiological traits (SENSECO)'. RF was supported by the Swiss National Science Foundation (SNSF 175529). SV was partially supported by the H2020 ESHAPE project (grant agreement 820852). The research carried out at the Jet Propulsion Laboratory, California Institute of Technology, was under a contract with the National Aeronautics and Space Administration (80NM0018D0004). Government sponsorship is acknowledged.

AUTHOR CONTRIBUTIONS

DR and MMar equally contributed to this manuscript in the analysis and writing phase. DR, MMar, DDR, MI, ML, EM, AHT, ET contributed to the analysis of the data provided in this paper. All authors contributed to the scientific writing of this manuscript.

DATA AVAILABILITY STATEMENT

The code and the data used in this paper are based on completely free and open source software, and they are available at the CRAN repository of the R package `rasterdiv`: https://CRAN.R-project.org/package=rasterdiv

ORCID

Duccio Rocchini https://orcid.org/0000-0003-0087-0594
Matteo Marcantonio https://orcid.org/0000-0003-3896-2355
Daniele Da Re https://orcid.org/0000-0002-3398-9295
Giovanni Bacaro https://orcid.org/0000-0003-0946-4496
Enrico Feoli https://orcid.org/0000-0002-5898-9777
Giles M. Foody https://orcid.org/0000-0001-6464-3054
Reinhard Furrer https://orcid.org/0000-0002-6319-2332
Ryan J. Harrigan https://orcid.org/0000-0002-4504-7158
David Kleijn https://orcid.org/0000-0003-2500-7164
Jonathan Lenoir https://orcid.org/0000-0003-0638-9582
Meixi Lin https://orcid.org/0000-0001-5233-0675
Marco Malavasi https://orcid.org/0000-0002-9639-1784
Rachel S. Meyer https://orcid.org/0000-0002-4907-5797
Vítězslav Moudry https://orcid.org/0000-0002-3194-451X
Fabian D. Schneider https://orcid.org/0000-0003-1791-2041
Petra Šímová https://orcid.org/0000-0003-2480-1171
Andrew H. Thornhill https://orcid.org/0000-0002-0325-5725
Saverio Vicario https://orcid.org/0000-0003-1140-0483
Robert K. Wayne https://orcid.org/0000-0003-3537-2245
Carlo Ricotta https://orcid.org/0000-0003-0818-3959

REFERENCES

Aerts R., Ewald M., Nicolas M., Piat J., Skowronek S., Lenoir J., Hattab T., Garzón-López C. X., Feilhauer H., Schmidtlein S., Rocchini D., Decocq G., Somers B., Van De Kerchove R., Denef K., Honnay O. (2017). Invasion by the Alien Tree Prunus serotina Alters Ecosystem Functions in a Temperate Deciduous Forest. *Frontiers in Plant Science*, *8*, 1–11.

Baldwin, B. G., Thornhill, A. H., Freyman, W. A., Ackerly, D. D., Kling, M. M., Morueta-Holme, N., & Mishler, B. D. (2017). Species richness and endemism in the native flora of California. *American Journal of Botany*, *104*, 1–15. https://doi.org/10.3732/ajb.1600326

Borcard, D., & Legendre, P. (2002). All-scale spatial analysis of ecological data by means of principal coordinates of neighbour matrices. *Ecological Modelling*, *153*, 51–68. https://doi.org/10.1016/S0304-3800(01)00501-4

Dornelas, M., Gotelli, N. J., McGill, B., Shimadzu, H., Moyes, F., Sievers, C., & Magurran, A. E. (2014). Assemblage time series reveal biodiversity change but not systematic loss. *Science*, *344*, 296–299. https://doi.org/10.1126/science.1248484

Dungan, J. L., Perry, J. N., Dale, M. R. T., Legendre, P., Citron-Pousty, S., Fortin, M.-J., Jakomulska, A., Miriti, M., & Rosenberg, M. S. (2002). A balanced view of scale in spatial statistical analysis. *Ecography*, *25*, 626–640. https://doi.org/10.1034/j.1600-0587.2002.250510.x

Feoli, E. (2018). Classification of plant communities and fuzzy diversity of vegetation systems. *Community Ecology*, *19*, 186–198. https://doi.org/10.1556/168.2018.19.2.11

Gillespie, T. W. (2005). Predicting woody-plant species richness in tropical dry forests: A case study from South Florida, USA. *Ecological Applications*, *15*, 27–37. https://doi.org/10.1890/03-5304

Goodall, D. W. (1970). Statistical ecology. *Annual Review of Ecology and Systematics*, *1*, 99–124.

Gorelick, R. (2011). Do we have a consistent terminology for species diversity? The fallacy of true diversity. *Oecologia*, *167*, 885–888.

Gotelli, N. J. (2000). Null model analysis of species co-occurrence patterns. *Ecology*, *81*(9), 2606–2621

Griffith, G. E., Omernik, J. M., Smith, D. W., Cook, T. D., Tallyn, E., Moseley, K., & Johnson, C. B. (2016). *Ecoregions of California*. USGS. http://pubs.er.usgs.gov/publication/ofr20161021

Guiasu, R. C., & Guiasu, S. (2011). The weighted quadratic index of biodiversity for pairs of species: A generalization of Rao's index. *Natural Science*, *3*, 795–801. https://doi.org/10.4236/ns.2011.39104

Hardy, G., Littlewood, J. E., & Polya, G. (1952). *Inequalities*. Cambridge University Press.

Hernandez-Stefanoni, J. L., Gallardo-Cruz, J. A., Meave, J. A., Rocchini, D., Bello-Pineda, J., & Lopez-Martinez, J. O. (2012). Modeling alpha- and beta-diversity in a tropical forest from remotely sensed and spatial data. *International Journal of Applied Earth Observation and Geoinformation*, *19*, 359–368.

Hill, M. O. (1973). Diversity and evenness: A unifying notation and its consequences. *Ecology*, *54*, 427–431. https://doi.org/10.2307/1934352

Hobohm, C., Janisova, M., Steinbauer, M., Landi, S., Field, R., Vanderplank, S., Beierkuhnlein, C., Grytnes, J.-A., Vetaas, R. O., Fidelis, A., de Nascimento, L., Clark, V. P., Fernandez-Palacios, J. M., Franklin, S., Guarino, R., Huang, J., Krestov, P., Ma, K., Onipchenko, V., … Chiarucci, A. (2019). Global endemics-area relationships of vascular plants. *Perspectives in Ecology and Conservation*, *17*, 41–49. https://doi.org/10.1016/j.pecon.2019.04.002

Hutchinson, G. (1959). Homage to Santa Rosalia or why are there so many kinds of animals? *The American Naturalist*, *93*, 145–159.

Jetz, W., Cavender-Bares, J., Pavlick, R., Schimel, D., Davis, F. W., Asner, G. P., Guralnick, R., Kattge, J., Latimer, A. M., Moorcroft, P., Schaepman, M. E., Schildhauer, M. P., Schneider, F. D., Schrodt, F., Stahl, U., & Ustin, S. L. (2016). Monitoring plant functional diversity from space. *Nature Plants*, *2*, 16024.

Jost, L. (2006). Entropy and diversity. *Oikos*, *113*, 363–375.

Laliberté, E., Schweiger, A. K., & Legendre, P. (2020). Partitioning plant spectral diversity into alpha and beta components. *Ecology Letters*, *23*, 370–380. https://doi.org/10.1111/ele.13429

Leinster, T., & Cobbold, C. A. (2012). Measuring diversity: The importance of species similarity. *Ecology*, *93*, 477–489. https://doi.org/10.1890/10-2402.1

Lin, M., Simons, A. L., Harrigan, R. J., Curd, E. E., Schneider, F. D., Ruiz-Ramos, D.V., Gold, Z., Osborne, M. G., Shirazi, S., Schweizer, T. M., Moore, T. N., Fox, E. A., Turba, R., Garcia-Vedrenne, A.E., Helman, S. K., Rutledge, K., Mejia, M. P., Ramos, M. N. M., Wetzer, R., … Meyer, R. S. (in press) Landscape Analyses Using eDNA Metabarcoding and Earth Observation Predict Community Biodiversity in California. *Ecological Applications*.

MacArthur, R. H., Recher, H., & Cody, M. (1966). On the relation between habitat selection and species diversity. *The American Naturalist*, *100*, 319–327. https://doi.org/10.1086/282425

Marcantonio, M., Iannacito, M., Thouverai, E., Da Re, D., Tattoni, C., Bacaro, G., Vicario, S., & Rocchini, D. (2020). *rasterdiv: Diversity indices for numerical matrices*. R package version 0.2-0. https://CRAN.R-project.org/package=rasterdiv

Mutanga, O., & Skidmore, A. K. (2004). Narrow band vegetation indices overcome the saturation problem in biomass estimation. *International Journal of Remote Sensing*, *25*, 3999–4014. https://doi.org/10.1080/01431160310001654923

Nakamura, G., Gonçalves, L. O., & Duarte, L. D. S. (2020). Revisiting the dimensionality of biological diversity. *Ecography*, *43*, 539–548. https://doi.org/10.1111/ecog.04574

Peres-Neto, P. R., Jackson, D. A., & Somers, K. M. (2005). How many principal components? Stopping rules for determining the number of non-trivial axes revisited. *Computational Statistics & Data Analysis*, *49*, 974–997. https://doi.org/10.1016/j.csda.2004.06.015

Rao, C. R. (1982). Diversity and dissimilarity coefficients: A unified approach. *Theoretical Population Biology*, *21*, 24–43. https://doi.org/10.1016/0040-5809(82)90004-1

R Core Team (2021). *R: A language and environment for statistical computing*, Vienna, Austria: R Foundation for Statistical Computing.

Rényi, A. (1970). *Probability theory*. North Holland Publishing Company.

Rocchini D. (2007). Distance decay in spectral space in analysing ecosystem β-diversity. *International Journal of Remote Sensing*, *28*, 2635–2644.

Rocchini, D., Andreini Butini, S., & Chiarucci, A. (2005). Maximizing plant species inventory efficiency by means of remotely sensed spectral distances. *Global Ecology and Biogeography*, *14*, 431–437. https://doi.org/10.1111/j.1466-822x.2005.00169.x

Rocchini, D., Delucchi, L., Bacaro, G., Cavallini, P., Feilhauer, H., Foody, G. M., He, K. S., Nagendra, H., Porta, C., Ricotta, C., Schmidtlein, S., Spano, L. D., Wegmann, M., & Neteler, M. (2013). Calculating landscape diversity with information-theory based indices: A GRASS GIS solution. *Ecological Informatics*, *17*, 82–93. https://doi.org/10.1016/j.ecoinf.2012.04.002

Rocchini, D., Garzon-Lopez, C. X., Marcantonio, M., Amici, V., Bacaro, G., Bastin, L., Brummitt, N., Chiarucci, A., Foody, G. M., Hauffe, H. C., He, K. S., Ricotta, C., Rizzoli, A., & Rosá, R. (2017). Anticipating species distributions: Handling sampling effort bias under a Bayesian framework. *Science of the Total Environment*, *584–585*, 282–290. https://doi.org/10.1016/j.scitotenv.2016.12.038

Rocchini, D., Hortal, J., Lengyel, S., Lobo, J. M., Jiménez-Valverde, A., Ricotta, C., Bacaro, G., & Chiarucci, A. (2011). Accounting for uncertainty when mapping species distributions: The need for maps of ignorance. *Progress in Physical Geography*, *35*, 211–226. https://doi.org/10.1177/0309133311399491

Rocchini, D., Luque, S., Pettorelli, N., Bastin, L., Doktor, D., Faedi, N., Feilhauer, H., Féret, J.-B., Foody, G. M., Gavish, Y., Godinho, S., Kunin, W. E., Lausch, A., Leitão, P. J., Marcantonio, M., Neteler, M., Ricotta, C., Schmidtlein, S., Vihervaara, P., … Nagendra, H. (2018). Measuring *β*-diversity by remote sensing: A challenge for biodiversity monitoring. *Methods in Ecology and Evolution*, *9*, 1787–1798. https://doi.org/10.1111/2041-210X.12941

Rocchini, D., Marcantonio, M., Da Re, D., Chirici, G., Galluzzi, M., Lenoir, J., Ricotta, C., Torresani, M., & Ziv, G. (2019). Time-lapsing biodiversity: An open source method for measuring diversity changes by remote sensing. *Remote Sensing of Environment*, *231*, 111192. https://doi.org/10.1016/j.rse.2019.05.011

Rocchini, D., & Ricotta, C. (2007). Are landscapes as crisp as we may think? *Ecological Modelling*, *204*, 535–539. https://doi.org/10.1016/j.ecolmodel.2006.12.028

Schimel, D., Schneider, F. D., Carbon, J. P. L., & Participants, E. (2019). Flux towers in the sky: Global ecology from space. *New Phytologist*, *224*, 570–584. https://doi.org/10.1111/nph.15934

Schneider, F. D., Ferraz, A., Hancock, S., Duncanson, L. I., Dubayah, R. O., Pavlick, R. P., & Schimel, D. S. (2020). Towards mapping the diversity of canopy structure from space with GEDI. *Environmental Research Letters*, *15*, 115006. https://doi.org/10.1088/1748-9326/ab9e99

Schneider, F. D., Ferraz, A., & Schimel, D. (2019). Watching Earth's interconnected systems at work. *Eos*, *100*.

Shannon, C. E. (1948). A mathematical theory of communication. *Bell System Technical Journal*, *27*, 379–423, 623–656. https://doi.org/10.1002/j.1538-7305.1948.tb00917.x

Stevens, R. D., & Tello, J. S. (2014). On the measurement of dimensionality of biodiversity. *Global Ecology and Biogeography*, *23*, 1115–1125. https://doi.org/10.1111/geb.12192

Stevens, R. D., Tello, J. S., & Gavilanez, M. M. (2013). Stronger tests of mechanisms underlying geographic gradients of biodiversity: Insights from the dimensionality of biodiversity. *PLoS ONE*, *8*, e56853. https://doi.org/10.1371/journal.pone.0056853

Thornhill, A. H., Baldwin, B. G., Freyman, W. A., Nosratinia, S., Kling, M. M., Morueta-Holme, N., Madsen, T. P., Ackerly, D. D., & Mishler, B. D. (2017). Spatial phylogenetics of the native California flora. *BMC Biology*, *15*, 96. https://doi.org/10.1186/s12915-017-0435-x

Whittaker, R. H. (1965). Dominance and diversity in land plant communities. *Science*, *147*, 250–260.

Zellweger, F., De Frenne, P., Lenoir, J., Rocchini, D., & Coomes, D. (2019). Advances in microclimate ecology arising from remote sensing. *Trends in Ecology and Evolution*, *34*, 327–341. https://doi.org/10.1016/j.tree.2018.12.012

BIOSKETCHES

Duccio Rocchini is Full Professor at Alma Mater Studiorum University of Bologna (Italy). His main research interests include biodiversity analysis at multiple spatial scales, remote sensing and spatio-ecological modelling by open source algorithms.

Matteo Marcantonio is a post-doc at the University of California Davis. His main research interests include plant and animal diversity, remote sensing and species distribution modelling.

All authors are broadly interested in ecological research. Their research covers topics in ecology, remote sensing, plant taxonomy.

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section.

How to cite this article: Rocchini D, Marcantonio M, Da Re D, et al. From zero to infinity: Minimum to maximum diversity of the planet by spatio-parametric Rao's quadratic entropy. *Global Ecol Biogeogr*. 2021;30:1153–1162. https://doi.org/10.1111/geb.13270