ALMA MATER STUDIORUM
UNIVERSITÀ DI BOLOGNA

DOTTORATO DI RICERCA IN

SCIENZE DELLA TERRA, DELLA VITA E DELL'AMBIENTE

Ciclo 36

Settore Concorsuale: 05/A1 - BOTANICA

Settore Scientifico Disciplinare: BIO/03 - BOTANICA AMBIENTALE E APPLICATA

SPATIOTEMPORAL ALGORITHMS FOR MEASURING ECOSYSTEM HETEROGENEITY FROM SPACE

Presentata da: Elisa Thouverai

Coordinatore Dottorato

Barbara Cavalazzi

Supervisore

Duccio Rocchini

Co-supervisore

Alessandro Chiarucci

Esame finale anno 2024

Abstract

Aim: This thesis integrates new open-source algorithms for the monitoring of ecosystem heterogeneity through remote sensing. The project is organized into three distinct parts, focusing on the measurement of spatial patterns, temporal trends, and spatiotemporal patterns.

Methods: Due to its widespread use in ecological research, the algorithms presented in this thesis were developed using the R software. **i)** Chapter 1 introduces the functions included in `rasterdiv` package for the calculation of spatial heterogeneity. In Chapter 2 is described a new function (`RaoAUC()`) for the computation of spatial heterogeneity that summarizes the information of parametric Rao index in a single metric. **ii)** Chapter 3 introduces the helical graphs, a novel visualization method for temporal trends in biodiversity drivers, plotting the mean values of a variable calculated at various points in time against the corresponding rate of change of the selected variable. **iii)** Chapter 4 presents a new method to quantify and visualize spatiotemporal heterogeneity change of an area exploiting beta diversity measures.

Study Areas: **i)** For testing the algorithms, in Chapter 1 a Global image from Copernicus Proba-V NDVI was used to test all `rasterdiv` metrics. `RaoAUC()` function was tested on two areas afflicted by fire in Chapter 2: 1) two multispectral images of an area in Kangaroo Island (AU) from Copernicus Sentinel 2 before and after the fire; 2) two hyperspectral images from an area near Santa Barbara (California, USA) before and two years after the fire. **ii)** In Chapter 3, the helical graphs were tested on two forest biomes, tropical and boreal forests of Northern Emisphere. **ii)** The spatiotemporal map from two CORINE landcover images (1990 and 2018) of Italy in Chapter 4.

Results and Discussions: **i)** The metrics tested in Chapter 1 offer insights into various facets of spatial heterogeneity, integrating available information of Earth surface properties, including aspects of functional (structural,

biophysical and biochemical), taxonomic, phylogenetic and genetic diversity. The `RaoAUC()` function tested in Chapter 2, binding different heterogeneity metrics, emerges as a valuable tool for identifying areas susceptible to environmental changes. **ii)** Chapter 3 proved that helical graphs efficiently highlight temporal trends of environmental variables and can be exploited in various applications. **iii)** The spatiotemporal maps developed in Chapter 4 are not only intuitive and easily interpretable but also provide a quantitative measure that seamlessly integrates into modeling frameworks. These maps contribute to a comprehensive understanding of temporal landscape dynamics, enhancing the ability to monitor and analyze environmental changes over time.

Conclusion: The algorithms presented in this thesis have proven their efficacy, interpretability, and versatility, contributing valuable insights into distinct aspects of ecosystem heterogeneity.

Contents

To my beloved grandfather,
Angelo Maluta

Introduction

Ecosystem Heterogeneity

The current global state of biodiversity has emerged as a central concern in social and political discourse (IPBES, 2019). In contemporary times, biodiversity is undergoing a decline at an unprecedented rate, with the extinction of species reaching exponential proportions (Pimm et al. , 2014). Human activities pose a significant threat to the integrity of Earth's ecological systems, resulting in profound alterations to natural environments and disruptions of ecosystems' equilibria (Sage and Kubien, 2007; Steffen et al., 2004). These activities contribute directly to habitat degradation, overexploitation of land and sea, the proliferation of invasive species, increased pollution, and climate change associated with global warming (Risser et al., 2000). Consequently, there is an urgent need for coordinated actions and policies directed towards the conservation and restoration of ecosystems and species (Chiarucci , 2007). In this context, monitoring key variables influencing biodiversity patterns and changes is essential for making well-informed decisions regarding conservation policies (Jetz et al. , 2019).

Recognized as a significant driver of biodiversity, ecosystem heterogeneity has been established as a key factor governing species diversity patterns across space (Stein et al. , 2014), as it plays a pivotal role in driving various ecological processes and functions. This influence extends to species diversity patterns and changes (Rocchini et al., 2018a), metapopulation dynamics (Fahrig , 2007), population connectivity (Malanson and Cramer , 1999), and gene flow (Lozier et al., 2013). Ecosystem heterogeneity enhances the availability of niche space, provides refuges, and creates opportunities for isolation and adaptation. Consequently, it contributes to the promotion of species coexistence, persistence, and diversification (Stein et al. , 2014; Tews et al., 2004). Therefore, its study is fundamental to understand spatial patterns in the distribution of species diversity.

Ecosystem heterogeneity has been conceptualized in diverse ways within the literature: (i) as the horizontal component of habitat variation (Li & Reynolds, 1995), (ii) encompassing the spatial and temporal variation of qualitative and

quantitative descriptors of the variable of interest (August, 1983), (iii) defined as the spatially structured variability of the habitat (Ettema and Wardle , 2002), or (iv) as within-habitat variability (Heaney, 2001; Hortal et al., 2009). In this thesis, ecosystem heterogeneity will serve as an umbrella concept encapsulating the degree of non-uniformity in land cover, vegetation, and physical factors such as topography, soil, topoclimate, and microclimate (Stein et al. , 2014).

Remote Sensing

Remote sensing involves acquiring data from a distance, often employing aircraft or satellites equipped with diverse sensors capable of detecting and recording information about the Earth's surface or atmosphere (NASA Earthdata , 2024). These sensors can capture a spectrum of data in the form of electromagnetic radiation, spanning visible light, infrared, or microwave signals (Zhu et al. , 2018). The satellite images obtained through these sensors are called bands (Zhu et al. , 2018). These bands can be characterized as two-dimensional arrays of pixels, with each pixel assigned a specific value representing the reflectance of a portion of the electromagnetic radiation, which is then converted into a digital number (Gomarasca , 2009).

The spectral resolution of a satellite image is determined by the number of bands captured by the sensors (Zhu et al. , 2018). This resolution can be classified into two main categories: multispectral and hyperspectral (Nalepa , 2021). Multispectral imaging, characterized by a relatively small number of bands, is particularly useful for land cover classification, providing valuable information about different surface features (Nalepa , 2021). On the other hand, hyperspectral imaging involves capturing tens to hundreds of bands, enabling a more detailed analysis of materials and their specific spectral signatures (Nalepa , 2021).

Satellite images also provide a dual perspective with both spatial and temporal resolutions (Wulder et al. , 2008). Spatial resolution defines the size of pixels detected by sensors, measured in meters, determining the level of detail captured in the imagery (Wulder et al. , 2008). On the other hand, temporal resolution specifies how frequently a satellite revisits or captures images of the same location over time, influencing the frequency of updates and monitoring intervals (Wulder et al. , 2008).

Numerous satellite missions play a crucial role in Earth observation, with some of the most notable ones being Landsat, Sentinel and MODIS (Fu et al.

, 2020). The Landsat program, inaugurated in 1972, stands as a cornerstone, delivering multispectral imagery with moderate spatial resolution (NASA , 1972). Complementing this, the Sentinel series, integral to the Copernicus program led by the European Space Agency (ESA), encompasses satellites such as Sentinel-1 and Sentinel-2, offering a comprehensive suite of radar and optical imagery suitable for all-weather monitoring (ESA , 2014). Another significant contributor is the Moderate Resolution Imaging Spectroradiometer (MODIS), situated aboard NASA's Terra and Aqua satellites, providing global coverage essential for climate studies and environmental monitoring (NASA MODIS Land Science Team , 2020).

This huge amount of data collected through remote sensing serves a myriad of applications, proving invaluable in ecosystem monitoring, agriculture, urban planning, disaster management, forestry, and geological exploration (Khorram et al. , 2012). This thanks to the estimation of several metrics from different compartments of the Earth system (land, ocean, atmosphere and cryosphere), such as precipitation patterns, global temperatures, snow cover, aerosol concentrations and vegetation indexes (Khorram et al. , 2012).

Ecosystem heterogeneity, a key factor influencing biodiversity, can be effectively quantified using remote sensing imagery (Rocchini et al., 2021). The Spectral Variation Hypothesis (SVH) serves as the theoretical foundation for this calculation, positing that the spatial variability in the remotely sensed signal, manifested as spectral heterogeneity in an image, correlates with ecosystem heterogeneity (Palmer et al., 2002). This relationship positions spectral heterogeneity as a potent proxy for species diversity (Palmer et al., 2002).

Spectral heterogeneity in a satellite image refers to the continuous variability of pixel values (Rocchini et al., , 2010). This variability in spectral reflectance is influenced by various factors, including the scale dimensions of the sensor (spatial, spectral, radiometric), viewing/illumination geometry, and biophysical properties of vegetation in the environment (such as structure, chemical constituents, and physiological processes) (Rocchini et al., , 2010). Leveraging spectral heterogeneity allows for the detection of hotspots of plant biodiversity over space (Rocchini et al., 2018b). By capturing and quantifying the intricate spectral variations, remote sensing proves to be a valuable tool in unraveling the complexities of ecosystem heterogeneity and its role in supporting biodiversity (Rocchini et al., 2021).

R software in ecological research

The widespread use of the R software for ecological analyses is owed to the collective efforts of numerous researchers who have contributed to the creation and documentation of well-established packages (Borcard et al. , 2011). Introduced in 2000, R has significantly transformed the approach and dissemination of research, fostering global collaborations and swiftly resolving common research challenges. At its core, R represents an integrated suite of software facilities for data manipulation, calculation, and graphical display (Venables et al. , 2019). It utilizes a programming language called 'S' and encompasses tools for data handling, analysis, and graphical representation (Venables et al. , 2019). Recognized as a statistical system by many users, R hosts a plethora of classical and modern statistical techniques, with some integrated into the base R environment, while many others are contributed by users and uploaded to repositories like CRAN (`https://CRAN.R-project.org`) or platforms such as Github (`https://www.github.com`). With thousands of contributed packages, written by diverse authors, the R ecosystem serves various purposes. These packages implement specialized statistical methods, provide access to data or hardware, and complement textbooks (Venables et al. , 2019).

A distinctive feature of R lies in its adherence to the software freedom philosophy, a concept founded by Richard Stallman in 1985. Stallman's philosophy, embodied in the concept of 'copyleft', opposes copyright restrictions (Rocchini & Neteler , 2012). Following the Free Software Definition, R upholds four fundamental freedoms that software must embody to be considered free (Vainio & Vaden , 2007):

- Freedom 0: The freedom to run the program for any purpose.
- Freedom 1: he freedom to study how the program works and adapt it to user needs, with access to the source code.
- Freedom 2: The freedom to redistribute copies, enabling collaboration.
- Freedom 3: The freedom to improve the program and release improvements to the public, with access to the source code.

The explicit use of free and open-source software, coupled with code availability, is essential for fostering completely open science. R, in this context, enables the development of packages and the sharing of code with other

users through repositories like CRAN (https://CRAN.R-project.org and its mirrors), Bioconductor (https://www.bioconductor.org), GitHub (https://www.github.com) and Omegahat (http://www.omegahat.net) (Venables et al. , 2019).

Outline of the thesis

The objective of the thesis is to advance open-source algorithms for the monitoring of spatiotemporal patterns of ecosystem heterogeneity using remote sensing. The work is organized into three distinct parts: the first comprises two chapters dedicated to the measurement of spatial heterogeneity. The second part delves into the representation of temporal trends, while the third part introduces a method designed for monitoring spatiotemporal heterogeneity, retaining information on both temporal trends and spatial patterns. This structured approach allows for a comprehensive exploration of ecosystem dynamics, providing valuable insights into the interplay between spatial and temporal aspects of environmental heterogeneity.

Part I

- Chapter 1: This paper introduces the R package `rasterdiv`, a comprehensive toolkit featuring various functions for the calculation of spatial heterogeneity. Each metric employed in these functions is meticulously described, accompanied by a global application showcasing the versatility and applicability of each metric. The package serves as a valuable resource for researchers and practitioners seeking robust tools for the assessment and analysis of ecosystem heterogeneity using R.

- Chapter 2: This paper presents a novel function, (`RaoAUC()`), designed for the computation of spatial heterogeneity. The newly introduced metric, AUC, serves to consolidate the information offered by the parametric Rao index, which can compute multiple metrics by altering a single parameter. The algorithm underwent testing on both theoretical and empirical examples, leveraging multi and hyperspectral images to demonstrate its efficacy and versatility. This contribution enriches the `rasterdiv` R package, offering a new refined and flexible metric for their investigations.

Part II

- Chapter 3: This paper introduces a novel visualization method for depicting temporal trends in biodiversity drivers. Inspired by the work of statistician Danny Dorling and illustrator Kirsten McClure, the helical graphs illustrate the mean values of a variable calculated at various points in time against the corresponding rate of change of the selected variable. To evaluate their effectiveness, the helical graphs were applied to a 20-year time series of the NDVI index (an indicator of vegetation productivity) in two distinct forest biomes: tropical and boreal. This innovative visualization approach provides a nuanced and insightful representation of temporal trends, contributing to the exploration and understanding of biodiversity dynamics over time.

Part III

- Chapter 4: This chapter presents a novel method for visualizing maps that depict spatiotemporal heterogeneity patterns. The algorithm applies the Jaccard index, a beta diversity metric, on landcover maps, where the land cover classes are used as proxies for species. The metric is computed between two images captured at distinct time points, providing a quantitative measure of the change in land cover composition. The chapter includes a case study that demonstrates the application and effectiveness of the proposed method in capturing and visualizing spatiotemporal heterogeneity patterns over time.

The algorithms presented in this thesis have demonstrated their utility, interpretability, and versatility. Each algorithm provides valuable insights into different facets of ecosystem heterogeneity, effectively exploiting both spatial and temporal dimensions. Furthermore, these algorithms have been made accessible to the scientific community by being published in public repositories (CRAN and GitHub), facilitating their widespread use and contribution to advancing research in the field.

Bibliography

August, P.V. (1983) The role of habitat complexity and heterogeneity in structuring tropical mammal communities. Ecology, 64, 1495-1507.

Borcard D., Gillet F., Legendre P. (2011) Numerical Ecology with R. New York, NY: Springer New York.

Chiarucci, A. (2007). To sample or not to sample? That is the question ... for the vegetation scientist. Folia Geobot., 42, 209–216.

European Space Agency (2014) Copernicus [Online] Available at: https://www.esa.int/Applications/Observing_the_Earth/Copernicus/

Ettema, C.H., Wardle, D.A. (2002) Spatial soil ecology. Trends in Ecology & Evolution, 17, 177-183.

Fahrig, L. (2007) Landscape heterogeneity and metapopulation dynamics. In Wu, J. and Hobbs, R. (Eds.). Key Topics in Landscape Ecology, pp. 78-91. Cambridge University Press, Cambridge, UK.

Fu, W., Ma, J., Chen, P. and Chen, F., (2020) Remote sensing satellites for digital earth. Manual of digital earth, pp.55-123.

Gomarasca, M.A. (2009) Digital Image Processing. In: Basics of Geomatics. Springer, Dordrecht.

Heaney, L.R. (2001) Small mammal diversity along elevational gradients in the Philippines: an assessment of patterns and hypotheses. Global Ecology and Biogeography, 10, 15-39.

Hortal, J., Triantis, K.A., Meiri, S., Th´ebault, E., Sfenthourakis, S. (2009) Island species richness increases with habitat diversity. The American Naturalist 174, E205–E217.

IPBES (2019) Summary for policymakers of the global assessment report on biodiversity and ecosystem services of the Intergovernmental Science-Policy Platform on Biodiversity and Ecosystem Services. S. Díaz, J. Settele,

E. S. Brondízio E.S., H. T. Ngo, M. Guèze, J. Agard, A. Arneth, P. Balvanera, K. A. Brauman, S. H. M. Butchart, K. M. A. Chan, L. A. Garibaldi, K. Ichii, J. Liu, S. M. Subramanian, G. F. Midgley, P. Miloslavich, Z. Molnár, D. Obura, A. Pfaff, S. Polasky, A. Purvis, J. Razzaque, B. Reyers, R. Roy Chowdhury, Y. J. Shin, I. J. Visseren-Hamakers, K. J. Willis, and C. N. Zayas (eds.). IPBES secretariat, Bonn, Germany. 56 pages.

Jetz, W., McGeoch, M.A., Guralnick, R., Ferrier, S., Beck, J., Costello, M.J., Fernandez, M., Geller, G.N., Keil, P., Merow, C. and Meyer, C., (2019) Essential biodiversity variables for mapping and monitoring species populations. Nature ecology & evolution, 3(4), pp.539-551.

Khorram, S., Koch, F.H., van der Wiele, C.F., Nelson, S.A.C., (2012) Remote Sensing. Springer.

Li, H., Reynolds, J. (1995) On definition and quantification of heterogeneity. Oikos, 73, 280-284.

Lozier, J.D., Strange, J.P., Koch, J.B. (2013) Landscape heterogeneity predicts gene flow in a widespread polymorphic bumble bee, *Bombus bifarius* (Hymenoptera: Apidae). Conservation Genetics, 14, 1099-1110.

Malanson, G.P., Cramer, B.E. (1999) Landscape heterogeneity, connectivity, and critical landscapes for conservation. Diversity and Distributions, 5, 27-39.

Nakamura, G., Goncalves, L.O., Duarte, L.d.S. (2020) Revisiting the dimensionality of biological diversity. Ecography, 43, 539-548.

Nalepa, J., (2021) Recent advances in multi-and hyperspectral image analysis. Sensors, 21(18), p.6002.

NASA (1972) Landsat. [Online] Available at: https://landsat.gsfc.nasa.gov/

NASA Earthdata. Remote Sensing. NASA Earthdata: A Global Discovery and Visualization System. https://www.earthdata.nasa.gov/learn/backgrounders/remote-sensing

NASA MODIS Land Science Team (2020) MODIS Land Products. [Online] Available at: https://modis.gsfc.nasa.gov/data/

Palmer, M.W., Earls, P., Hoagland, B.W., White, P.S., Wohlgemuth, T. (2002). Quantitative tools for perfecting species lists. Environmetrics, 13, 121137.

Pimm, S.L., Jenkins, C.N., Abell, R., Brooks, T.M., Gittleman, J.L., Joppa, L.N., Raven, P.H., Roberts, C.M. and Sexton, J.O., (2014) The biodiversity of species and their rates of extinction, distribution, and protection. science, 344(6187), p.1246752.

Risser, G.P., Clarke, J.N., Dale, V., Field, C., Lewis, M.W. Jr., Lubchenco, J., ..., Ustin, S., (2000) Global Change Ecosystems Research (Chapter: Definitions and Implications of Global Change). National Academy Press Washington DC.

Rocchini, D., Balkenhol, N., Carter, G.A., Foody, G.M., Gillespie, T.W., He, K.S., Kark, S., Levin, N., Lucas, K., Luoto, M. and Nagendra, H. (2010) Remotely sensed spectral heterogeneity as a proxy of species diversity: recent advances and open challenges. Ecological Informatics, 5(5), pp.318-329.

Rocchini D. and Neteler M. (2012) Let the four freedoms paradigm apply to ecology. Trends in Ecology & Evolution 27(6), 310-311.

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., Wegmann, M., Nagendra, H. (2018) Measuring β-diversity by remote sensing: a challenge for biodiversity monitoring. Methods in Ecology and Evolution, 9, 1787-1798.

Rocchini D., Bacaro G., Chirici G., Da Re D., Feilhauer H., Foody G.M., Galluzzi M., Garzon-Lopez C.X., Gillespie T.W., He K.S., Lenoir J., Marcantonio M., Nagendra H., Ricotta C., Rommel E., Schmidtlein S., Skidmore A.K., Van De Kerchove R., Wegmann M., Rugani B. (2018) Remotely sensed spatial heterogeneity as an exploratory tool for taxonomic and functional diversity study. Ecol Indic., 85, 983–990.

Rocchini, D., Salvatori, N., Beierkuhnlein, C., Chiarucci, A., De Boissieu, F., Förster, M., Garzon-Lopez, C.X., Gillespie, T.W., Hauffe, H.C., He, K.S. and Kleinschmit, B., (2021) From local spectral species to global spectral communities: A benchmark for ecosystem diversity estimate by remote sensing. Ecological informatics, 61, p.101195.

Sage, R.F., Kubien, D.S. (2007) The temperature response of C3 and C4 photosynthesis. Plant, Cell and Environment, 30:1086-1106. https://doi.org/10.1111/j.1365-3040.2007.01682.x.

Steffen, W., Sanderson, R.A., Tyson, P.D., Jäger, J., Matson, P.A., Moore III, B., Oldfield, F., Richardson, K., Schellnhuber, H.J., Turner, B.L., Wasson, R.J., (2004) Global Change and the Earth System. A Planet Under Pressure. Springer.

Stein, A., Gerstner, K., Kreft, H. (2014) Environmental heterogeneity as a universal driver of species richness across taxa, biomes and spatial scales. Ecology Letters, 17, 866-880.

Tews, J., Brose, U., Grimm, V., Tielbörger, K., Wichmann, M., Schwager, M. and Jeltsch, F. (2004) Animal species diversity driven by habitat heterogeneity/diversity: the importance of keystone structures. Journal of Biogeography, 31, 79–92.

Vainio N. and Vaden T.(2007) Free Software Philosophy and Open Source. In: St.Amant K. & Still B. (eds.) Handbook of Research on Open Source Software: Technological, Economic, and Social Perspectives (pp. 1-11). Hershey, PA: Information Science Reference.

Venables W.N., Smith D.M., R Core Team. (2019) An Introduction to R. http://www.r-project.org/.

Wulder, M.A., White, J.C., Coops, N.C. and Butson, C.R., (2008) Multi-temporal analysis of high spatial resolution imagery for disturbance monitoring. Remote Sensing of Environment, 112(6), pp.2729-2740.

Zhu L, Suomalainen, J., Liu, J., Hyyppä, J., Kaartinen, H., and Haggren, H. (2018) A Review: Remote Sensing Sensors. Multi-purposeful Application of Geospatial Data. InTech. Available at: http://dx.doi.org/10.5772/intechopen.71049.

Chapter 1

Measuring diversity from space: a global view of the free and open source `rasterdiv` R package under a coding perspective

Published as:
Thouverai, E., Marcantonio, M., Bacaro, G., Re, D.D., Iannacito, M., Marchetto, E., Ricotta, C., Tattoni, C., Vicario, S. and Rocchini, D. (2021). Measuring diversity from space: A global view of the free and open source rasterdiv R package under a coding perspective. Community Ecology, 22, pp.1-11.

1.1 Abstract

The variation of species diversity over space and time has been widely recognised as a key challenge in ecology. However, measuring species diversity over large areas might be difficult for logistic reasons related to both time and cost savings for sampling, as well as accessibility of remote ecosystems. In this paper, we present a new `R` package - `rasterdiv` - to calculate diversity indices based on remotely sensed data, by discussing the theory beyond the developed algorithms. Obviously, measures of diversity from space should not be viewed as a replacement of in-situ data on biological diversity, but they are rather complementary to existing data and approaches. In practice, they integrate available information of Earth surface properties, including aspects of functional (structural, biophysical and biochemical), taxonomic, phylogenetic and genetic diversity. Making use of the `rasterdiv` package can result useful in making multiple calculations based on reproducible open source algorithms, robustly rooted in Information Theory.

Keywords: biodiversity; ecological informatics; modelling; remote sensing; satellite imagery.

1.2 Introduction

Back in 1872, Ludwig Eduard Boltzmann (Boltzmann, 1872) introduced the first measure of entropy, later called marginal entropy and restructured by Claude Elwood Shannon under a mathematical theory umbrella (Shannon, 1949). As such, it became one of the cornerstones of ecological theory and was adopted widely in ecological practice for measuring biodiversity and its change. Concerning biological entropy, the variation of species diversity over space and time has been widely recognized as a key challenge in ecology and was associated with analytic geometric models focusing either on the spatial component of species dispersal (Palmer, 2007; Gorelick, 2008) or on environmental drivers (Kreft and Jetz, 2007).

To address this issue, many spatio-statistical models have been proposed to model biological entropy using data from ecological surveys (Bachl et al., 2019). However, the statistical clarity (sensu Dushoff et al. (2019)) of such models strictly depends on a high in-situ data uncertainty, which propagates through all inferential steps (Meyer et al., 2016; Rocchini et al., 2019). Furthermore, measuring species diversity over wide areas might be difficult for

logistic reasons related both to time and sampling costs (Chiarucci et al., 2011; Hernandez-Stefanoni et al., 2012) and to theoretical and practical constraints, which are mainly related to two sources of uncertainty. The first is the uncertainty associated with the detectability and the determination of individual plants or animals species. The second is the one linked to different sampling strategies (McGlinn and Palmer, 2009) or efforts (Rocchini et al., 2019) per area, or, in the worst case, to the impossibility of getting information about the real grain (sensu Scheiner et al. (2000)) sampled (Hobohm, 2003). In the absence of such information, it becomes excessively challenging to properly address the modifiable areal unit problem (MAUP), which in this case is the sensitivity of biodiversity to scale (Jelinski and Wu, 1996). This is true, even though evidence exists for a chance to rely, in some instances, on expert knowledge to build straightforward and robust diversity maps worldwide (Hobohm et al., 2019).

Accordingly, algorithms based on remote sensing and spatial ecology might help estimating the variation of biodiversity over space and time (Skidmore et al., 2011; Schimel and Scheiner, 2019) and represent a powerful first exploratory tool to detect the spatial variability accross the landscape. The relationship between ecological processes (and functions) and the remotely sensed diversity can rely on the definition of niche proposed by Kroes (1977), and according to which a niche is the biotic structural and functional part of the ecosystem. Strictly speaking, such definition can be profitably used to measure spatial heterogeneity in ecosystems in order to convey information on their potential functions (Schneider et al., 2017).

From this point of view, the development of Free and Open Source algorithms to measure diversity from space would be beneficial to allow high robustness and reproducibility of the proposed approaches (Rocchini and Neteler, 2012). Furthermore, their intrinsic transparency, community-vetoing options, sharing and rapid availability are also valuable additions and reasons to move to open source options. Among the different open source packages, the `R` software environment is certainly one of the most widespread worldwide and different packages have been devoted to remote sensing for: i) raster data management (`raster` package, Hijmans and van Etten (2020)), ii) remote sensing data analysis (`RStoolbox` package, Leutner et al. (2019)), iii) spectral species diversity (`biodivMapR` package, Féret and Boissieu (2020)), iv) Sparse Generalized Dissimilarity Modelling based on remote sensing data (`sgdm` package, Leitao et al. (2012)), v) entropy-based local spatial association (`ELSA` package, Naimi et al. (2019)), vi) landscape metrics calculation

(landscapemetrics package, Hesselbarth et al. (2019)), to name just a few. Reader can also refer to https://cran.r-project.org/web/views/Spatial.html for the CRAN Task View on analysis of spatial data.

However, currently no package provides a flow of functions grounded on Information Theory related to abundance based measures, by further introducing distances and going back to Information Theory again by generalised entropy. In this paper we introduce a new R package which provides such a functions' throughput workflow. The aim of this manuscript is to encompass the theory beyond the algorithms developed in the rasterdiv package (currently available at: https://github.com/mattmar/rasterdiv), relying on the definition given by Gorelick (2011b):

"Theory is neither mathematical nor abstract. Theory is the creative, inductive, and synthetic discipline of forming hypothesis [...]"

1.3 Information Theory

One of the mostly used metrics for measuring remotely sensed diversity is related to the entropy measurement firstly introduced by Shannon (Shannon, 1949).

Given a sample area with N pixel values and p_i relative abundances for every $i \in \{1, \ldots, N\}$, in decreasing order, the Shannon index is calculated as:

$$H = -\sum_{i=1}^{N} p_i \ln p_i \tag{1.1}$$

Taking into account only the most abundant pixel value, the Berger-Parker (Berger and Parker , 1970) index is given by:

$$I_{BP} = p_1 \tag{1.2}$$

In remote sensing applications, the derivation of synthetic indices of any sort (i.e., diversity) is often performed by sequentially considering only small chunks of the whole image. These chunks are commonly defined as 'kernel', 'windows' or 'moving windows'. From now on, we will use this terminology to indicate the local space of analysis.

Both indices can be calculated in rasterdiv on a numerical matrix by using a moving window and applying the command Shannon and BergerParker.

```
Shannon(x, window=3)
```

```
BergerParker(x, window=3)
```

where `x` and `window` represent the input numerical matrix (raster file) and the side of the moving window on which the calculation is performed. Additional arguments common to all functions in the package are `np` which sets the number of parallel processes among which to split the calculation of the index, `cluster.type` which defines the type of protocol to spawn parallal processes (default is "`MPI`") and `debugging` that, if set as `TRUE`, will run the function in verbose mode.

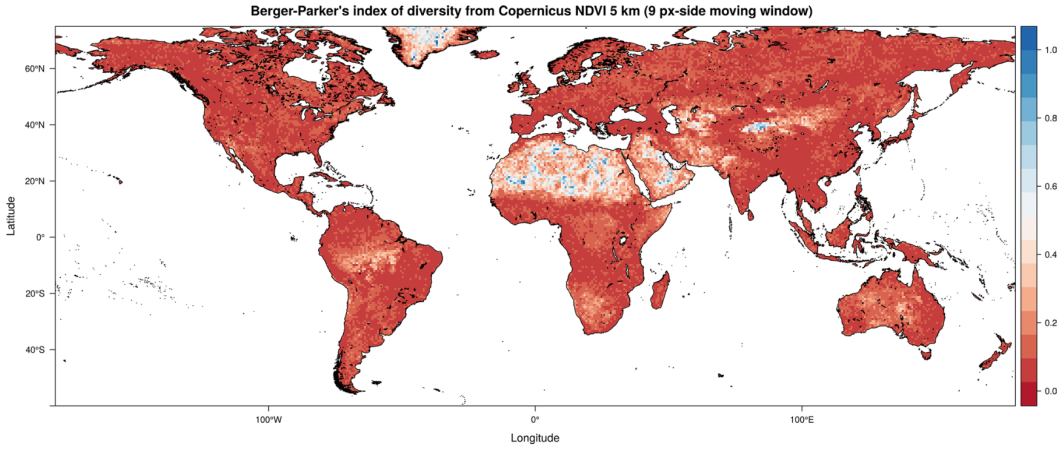

FIGURE 1.1: Berger-Parker index measuring the most abundant spectral value (Equation 1.2). All the indices in this paper are calculated starting from a Copernicus Proba-V NDVI (Normalised Difference Vegetation Index) long term average image (June 21st 1999-2017) at 5km grain, also provided into the `rasterdiv` package as a free default set which can be loaded by the function `data()`. A generally low value of the index (based on the most abundant spectral value) is found, since spectral input values are generally different from each other in a moving window. This figure has been generated by the command `bpa <- BergerParker(ndvi17_r,window=9,np=8,cluster.type="SOCK")`.

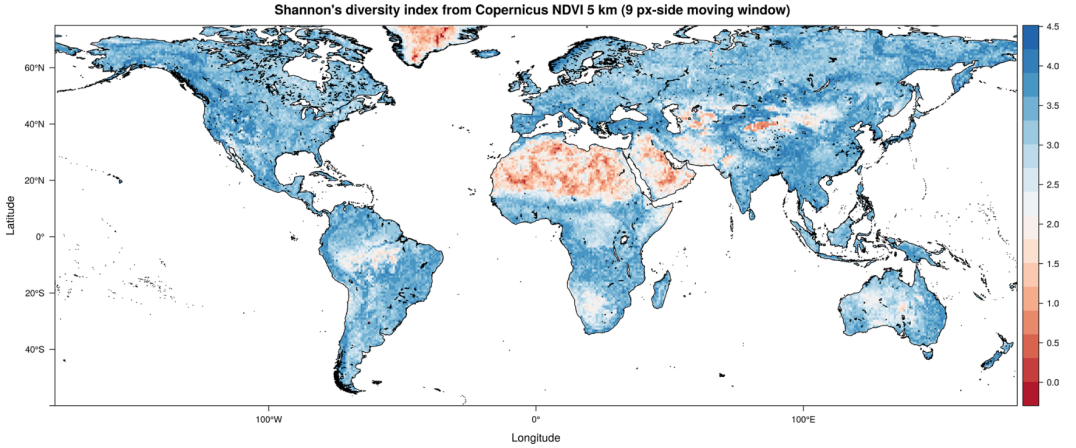

FIGURE 1.2: Shannon index calculated on a Copernicus Proba-V NDVI image at 5km. Shannon's H is generally high since it only considers relative abundance of spectral values, which are generally different from each other. This figure has been generated by the command `sha <- Shannon(ndvi17_r,window=9,np=8,cluster.type="SOCK")`.

Both indices obey to the relative abundance of values. The most simple Berger-Parker index equals the relative proportion of the most abundant class in a moving window (Figure 1.1). Hence, low values of Berger-Parker are expected for continuous satellite data, given the high variability of reflectance values. In contrast, in the Shannon index, the abundance of every single numerical category (pixel value) is taken into account. This might lead to taking into account the turnover among values, since the higher the turnover the lower the dominance of a single class (Figure 1.2). However, Shannon's H is unable to discern situations where there is a high richness (number of numerical categories) and a low evenness from those where there is a low richness but a high evenness.

To better account for evenness, the Pielou index (Pielou, 1966) can be calculated by simply standardising the Shannon index on the maximum possible Shannon index attainable given the same richness value. The latter is attained when the maximum potential evenness of pixel values/numerical categories is reached, i.e. when they are equally distributed over space.

$$E = \frac{H}{H_{max}} \tag{1.3}$$

H_{max} corresponds to the natural logarithm of the number of pixel values. Using `rasterdiv`, the Pielou index can simply be calculated as:

```
Pielou(input, window=3)
```

Figure 1.3 reports an example with a moving window of 9x9 pixels.

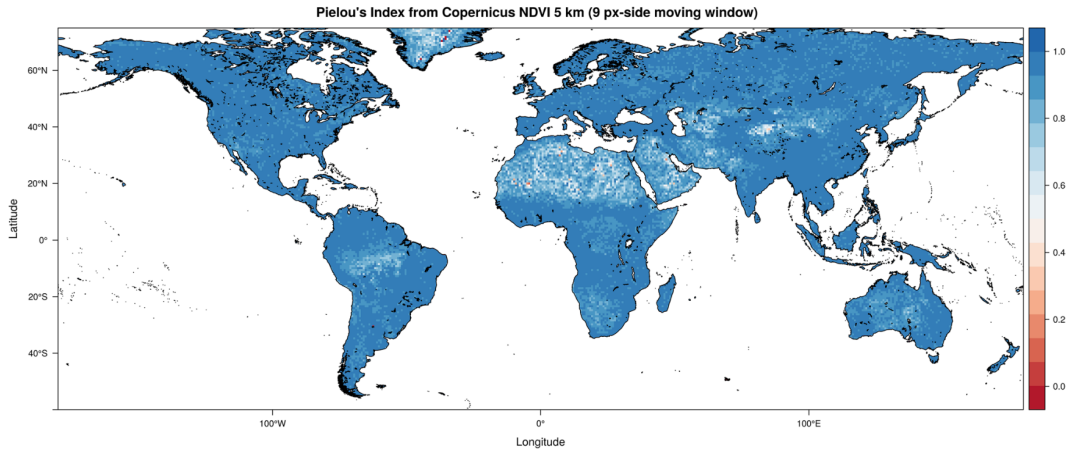

FIGURE 1.3: Pielou index calculated on a Copernicus Proba-V NDVI image at 5km. A flattening effect with respect to Shannon's H is found, due to the standardisation on the maximum possible Shannon entropy (see Equation 1.3). This figure has been generated by the command `pie <- Pielou(ndvi17_r,window=9,np=8,cluster.type="SOCK")`.

1.4 Solving the non-dimensionality of Shannon's H': the Rao's Q diversity index

Both Shannon's H and Pielou's E are dimensionless. In other words, they consider differences in the relative abundance among pixel values, but not their relative spectral distance, i.e. the distance among spectral values. For instance, let $A = (1, 2, 3, 4, 5, 6, 7, 8, 9)$ and $B = (1, 10^2, 10^3, 10^4, 10^5, 10^6, 10^7, 10^8, 10^9)$ be two theoretical arrays of values. In both cases, values are different from each other; hence, despite their relative numerical distance the Shannon index will always be maximum, i.e. $H = \log(9) = 2.197225$ reducing $E = H/H_{max} = 1$.

In remotely sensed imagery this is a crucial point since it might happen that contiguous zones could have similar (but not equal) reflectance values. For instance, the diversity of a homogeneous surface like water could be overestimated if spectral distances are not considered.

To overcome this issue, the Rao's Quadratic diversity (hereafter Rao's Q, Rao (1982)) could be applied by not only taking into account relative abundance but also the spectral distance among different pixel values.

Given the values of different pixels i and j, the Rao's Q consider their pairwaise distance d_{ij} as:

$$Q = \sum_{i=1}^{N} \sum_{j=1}^{N} d_{ij} \times p_i \times p_j \tag{1.4}$$

Hence, an array with different but spectrally close values will lead to a high Shannon's H but a low Rao's Q. On the contrary, an array with different and distant values in the spectral space will lead to both a high Shannon's H and a high Raos' Q.

Moving towards a 2D spatial extent, let M be a 2D matrix $M = \begin{pmatrix} \lambda_1 & \lambda_2 & \lambda_3 \\ \lambda_4 & \lambda_5 & \lambda_6 \\ \lambda_7 & \lambda_8 & \lambda_9 \end{pmatrix}$ formed by pixels with a certain reflectance value λ in a single band for instance. For simplicity, let us consider an 8-bit band, i.e. containing 256 possible values. As a consequence, deriving Rao's Q involves calculating a distance matrix M_d for all the pixel values:

$$M_d = \begin{pmatrix} d_{\lambda_1,\lambda_1} & d_{\lambda_1,\lambda_2} & d_{\lambda_1,\lambda_3} & \cdots & d_{\lambda_1,\lambda_n} \\ d_{\lambda_2,\lambda_1} & d_{\lambda_2,\lambda_2} & d_{\lambda_2,\lambda_3} & \cdots & d_{\lambda_2,\lambda_n} \\ d_{\lambda_3,\lambda_1} & d_{\lambda_3,\lambda_2} & d_{\lambda_3,\lambda_3} & \cdots & d_{\lambda_3,\lambda_n} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ d_{\lambda_n,\lambda_1} & d_{\lambda_n,\lambda_2} & d_{\lambda_n,\lambda_3} & \cdots & d_{\lambda_n,\lambda_n} \end{pmatrix} \tag{1.5}$$

Thus, according to Equation 2.2, Rao's Q is related to the sum of all the pixel values pairwise distances, each of which is multiplied by the relative abundance of each pair of pixels in the analysed image $d \times (1/N^2)$. In other words, Rao's Q is the expected difference in reflectance values between two pixels drawn randomly with replacement from the evaluated set of pixels. The distance matrix can be built in several dimensions (layers), thus allowing to consider more than one band at a time. As a consequence, Rao's Q can be calculated in a multidimensional (multi-layers) system.

In `rasterdiv` package Rao's Q is calculated as:

```
 spectralrao(x, distance_m="euclidean", window=3,
mode="classic", shannon=FALSE, na.tolerance=0.0)
```

The `distance_m` argument includes different types of distances, such as the Euclidean, Manhattan and Canberra distances, to make the calculation.

Obviously, it is suggested to make use of a different distance when a multispectral set is used; otherwise all the considered distances in just one band will reduce to the Euclidean one (Figure 1.4).

In fact it is automatically demonstrated that in one dimension $D_M : (x,y) \mapsto \sum |x_i - y_i| = D_E : (x,y) \mapsto \sqrt{\sum (x_i - y_i)^2}$ where D_M and D_E are the Manhattan and the Euclidean distances, respectively.

The Canberra distance is derived from the Manhattan distance by standardizing separately the absolute differences of each band with the sum of both values, and thus will also equal D_E in one dimension, such that: $D_C : (x,y) \mapsto \sum \frac{|x_i - y_i|}{|x_i| + |y_i|} = D_E : (x,y) \mapsto \sqrt{\sum (x_i - y_i)^2}$.

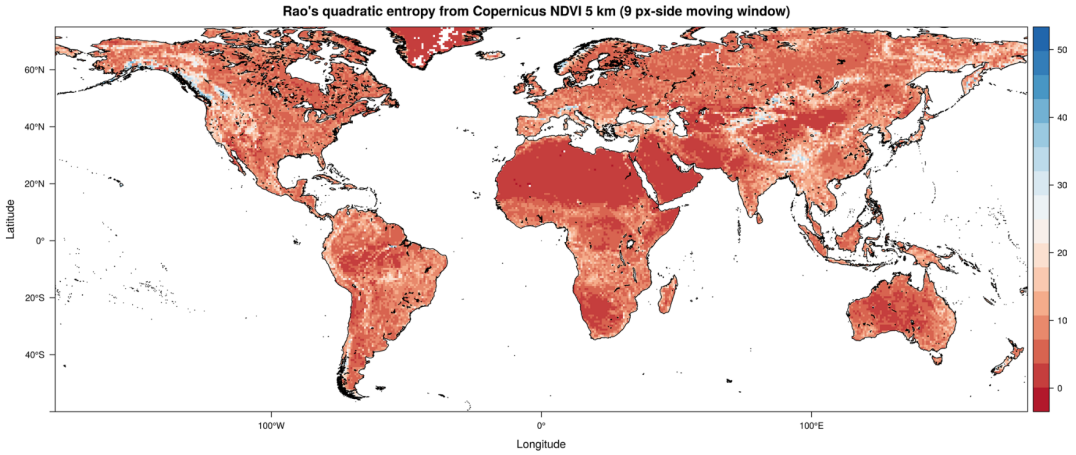

FIGURE 1.4: Rao's Q index calculated on a Copernicus Proba-V NDVI image at 5km. Differenly from the original Shannon's formula, Rao's Q also considers the distance among different values by better discriminating the queues of the diversity distribution from very low diversity (e.g. deserts and ice fields) to very high diversity (e.g. upper highly-complex mountain ranges). This figure has been generated by the command `rao <- Rao(ndvi17_r,dist_m="euclidean",window=9,shannon=FALSE,np=8, cluster.type="SOCK",na.tolerance=0.5)`.

1.5 Solving the intrinsic continuity of spectral data: Cumulative Residual Entropy

As previously stated, spectral data are continuous variables that are approximate to discrete (the so called "digital number") for practical reasons. As such, the fact that two different pixels should be counted or not in a category depends from the whim of the normalisation of the signal when Digital

Numbers (DNs) are generated. Shannon index is built strictly for a non ordered finite set of categories. For continuous variables, a derivative version of Shannon index was proposed, but soon it was clear that it had very different properties than categorical formulation (Jumarie, 1990; Michalowicz et al., 2013). Rao et al. (2004) proposed a Cumulative Residual Entropy (CRE) to build a consistent Shannon-like index for continuous variables. It is based on residual cumulative probability ($P(X >= x_i)$), that can be estimated in a robust manner from empirical mono-dimensional distributions by counting for each value the number of observations with equal or larger values and then dividing by the total. CRE is defined as follow:

$$CRE = -\int_0^{\inf} P(X \geq x) \log P(X \geq x) dx \tag{1.6}$$

and to estimate it from an empirical distribution, the following approach is advised:

$$\begin{aligned} CRE &= -\sum_{i=1}^{N} P(X \geq x_i) \log P(X \geq x_i) dx \\ dx &= (x_i - x_{i-1}) \end{aligned} \tag{1.7}$$

where X is the sorted vector of N observations. In practice, the approach is similar to the Rao's Q, given that a coefficient d, representing the disparity of the observations, is used to weight the diversity estimate based on probability. The difference resides in that the disparity in this continuous measure is absolute, while in Rao'Q it is relative between two observations.

This difference makes more complex the generalisation to a multi-layer, where this time the uni-dimensional cumulative residual probability is substituted with a multivariate one. For instance, here is an example making use of three layers / bands:

$$\begin{aligned} X = [x_0, x_1, ..., x_N], \quad Y = [y_0, y_1, ..., y_N], \quad Z = [z_0, z_1, ..., z_N] \\ CRE = -\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{k=1}^{N} P_{cr}(X,Y,Z)_{i,j,k} \log P_{cr}(X,Y,Z)_{i,j,k} dx_i dy_i dz_i \\ dx_i = (x_i - x_{i-1}) \\ P_{cr}(X,Y,Z)_{i,j,k} = P(X \geq x_i; Y \geq y_j; Z \geq z_k) \end{aligned} \tag{1.8}$$

The calculation of the cumulative residual probability $P_{cr}(X,Y,Z)$ in an efficient way is based on: i) calculating a contingency array with a certain dimension for each band, and then ii) performing a reverse cumulative sum

along each dimension as follows:

$$\begin{aligned} &\forall I, J, K \in [0, ..., N] \\ &P(X, Y, Z)_{I,J,K} = P(X = x_I, Y = y_J, Z = z_K) \\ &P_{cr}(X|Y, Z)_{I,J,K} = \sum_{i=0}^{I} P(X, Y, Z)_{(N-i),J,K} \\ &P_{cr}(X, Y|Z)_{I,J,K} = \sum_{j=0}^{J} P(X|Y, Z)_{(I,(N-j),K} \\ &P_{cr}(X, Y, Z)_{I,J,K} = \sum_{k=0}^{K} P(X, Y|Z)_{I,J,(N-k)} \end{aligned} \tag{1.9}$$

In `rasterdiv` Cumulative Residual Entropy can be calculated as:

```
CRE(x, window=3)
```

producing a map such as that achieved in Figure 1.5.

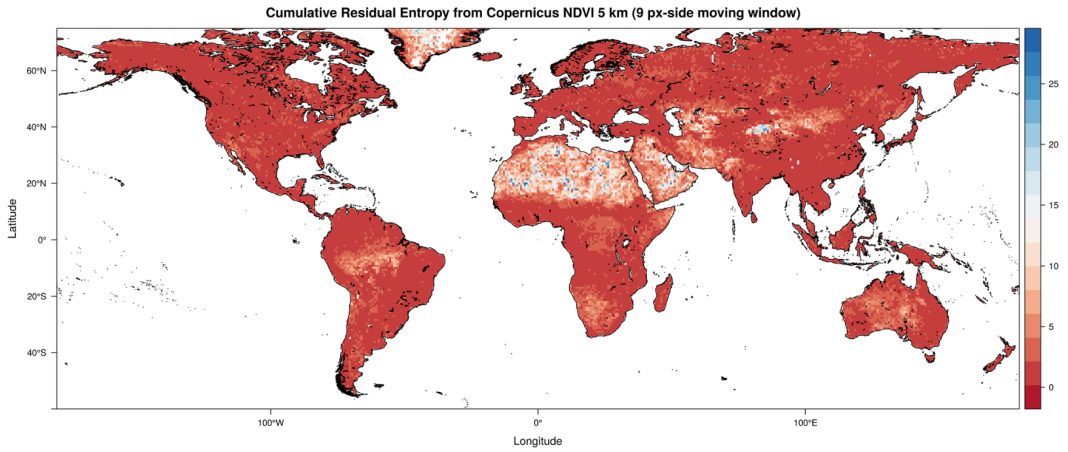

FIGURE 1.5: Cumulative Residual Entropy calculated on a Copernicus Proba-V NDVI image at 5km. This figure has been generated by the command `cre <- CRE(ndvi17_r,window=9,np=8,cluster.type="SOCK",na.tolerance=0.5)`.

1.6 Solving point descriptors of diversity: the Rényi and Hill generalised entropy

The metrics described above represent point descriptors of diversity, each of which is able to represent only a part of the whole diversity spectrum that can be attained. There is actually no single measure that could be adopted

to represent all the different aspects of diversity with an intrinsic fallacy in considering a 'true' diversity (Gorelick, 2011a).

Rényi (1970) firstly proposed a measure which is able to represent several diversity metrics in just one formula, by only changing one parameter (α in the original version of his manuscript). Given a sample area with N pixel values and p_i relative abundances for every $i \in \{1, \ldots, N\}$, the Rényi index is:

$$H_\alpha = \frac{1}{1-\alpha} \times \ln \sum_{i=1}^{N} p_i^\alpha \tag{1.10}$$

Changing the parameter α will lead to different indices starting from the same formula (Hill, 1973). As an example, when α=0, $H_0 = \ln(N)$ where N=richness, namely the maximum possible Shannon index (H_{max}). In practice, with $\alpha = 0$, all the spectral values equally contribute to the index, without making use of their relative abundance. For $\alpha \to 1$, the Rényi will equal Shannon H, according to the l'Hospital's rule, while for α=2 the Rényi index will equal the $\ln(1/D)$ where D is the Simpson's dominance (Simpson, 1949). The theoretical curve relating the Rényi index and α is a negative exponential, i.e. it decays until flattening for higher values of α, where the weight of the most abundant spectral values is higher with small differences among the attained diversity maps (Ricotta et al. (2003a)).

In `rasterdiv` the Rényi index is calculated as:

```
Renyi(x, window=3, mode ="single", alpha=1, base=exp(1))
```

where `x` and `window` are the input set and the window size of analysis, as in previous functions (Figure 1.6). The `mode` can be i) `"single"` to compute the Rényi index for just one alpha value, ii) `"iterative"` to compute the Rényi index for all the integer values of alpha in a given interval, or `"sequential"` to compute the Rényi index for all the alpha values in a given vector. `alpha` indicates the α vale in Equation 1.10. Its default value is 1. In `"single"` mode, alpha has to be a numerical value greater than 0; in `"iterative"` mode, alpha has to be a length 2 vector and in `"sequential"` mode, alpha has to be a vector of length at least 2. `base` is a numerical value, which let the user choose the base of the logarithm in Rényi index formula. Its default value is exp(1).

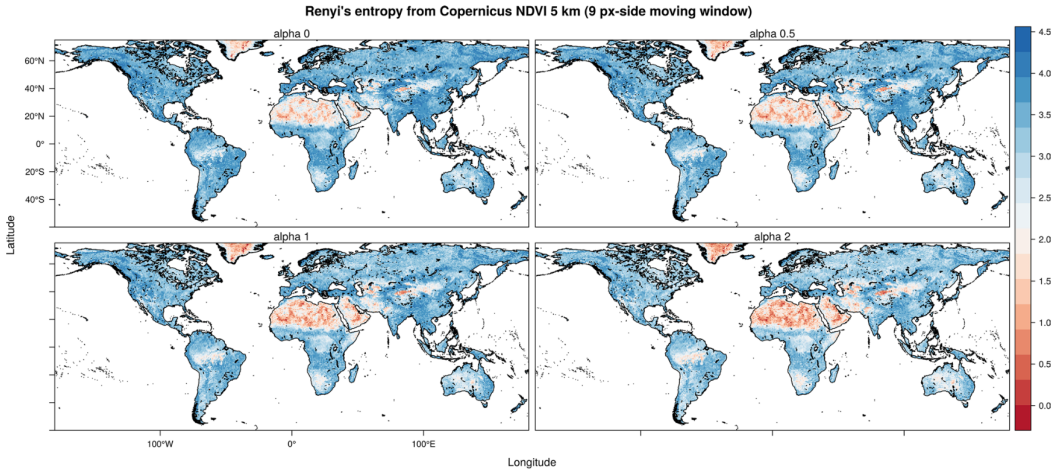

FIGURE 1.6: Rényi index calculated on a Copernicus Proba-V NDVI image at 5km, considering different α values, from 0 to 2. With $\alpha \rightarrow 1$, the diversity map is equal to the Shannon's map of Figure 1.2. Increasing α will create a flattening of the index with a lower ability to discern differences among different maps (Ricotta et al., 2003a). This figure has been generated by the command `ren <- Renyi(ndvi17_r,window=9,np=8,cluster.type="SOCK",alpha=c(0,0.5,1,2))`.

Hill (1973) was the first ecologist applying the generalised entropy concept initially developed by Rényi (1970). In particular, since no particular formula would have a preeminent advantage over the others (Hill, 1973), the Hill's generalised entropy N_α was based on the effective number of species of H_α, namely the number of species that would lead to H_α if they were equally abundant. In our case, the "species concept" is translated to the "spectral values" concept. Hence, N_α is the effective number of spectral values that would give H_α as an output. N_α can thus be calculated as:

$$N_\alpha = \left(\sum_{i=1}^{N} p_i^\alpha \right)^{\frac{1}{1-\alpha}} \tag{1.11}$$

As for the Rényi generalised entropy, changing α will let the index transform in many other widely used indices, which are point descriptions of diversity, i.e. peculiar cases of the Hill's generalised theory. Hence, for $\alpha = 0$, $N_0 = N$, where N is the total number of spectral values in the window of analysis; for $\alpha = 1$, $N_1 = \exp H$; for $\alpha = 2$, $N_2 = 1/S$, where S is the Simpson's index, and for $\alpha = \infty$, $N_\infty = \frac{1}{I_{BP}}$, where I_{BP} is the Berger-Parker index (Figure 1.7). We refer to Ricotta et al. (2003a) and Ricotta et al. (2003b) for a concise review on the theoretical properties of the Rényi and the Hill's generalised entropy,

respectively. In rasterdiv, the Hill's generalised entropy can be calculated as:

```
Hill(x, window=3, mode ="single", alpha=1, base=exp(1))
```

The Hill's generalised entropy can be computed for i) one α value in single mode, or ii) all the integer values of α in a given interval (iterative mode). We refer to Chao et al. (2016) for a complete overview of the Hill's numbers application in ecology.

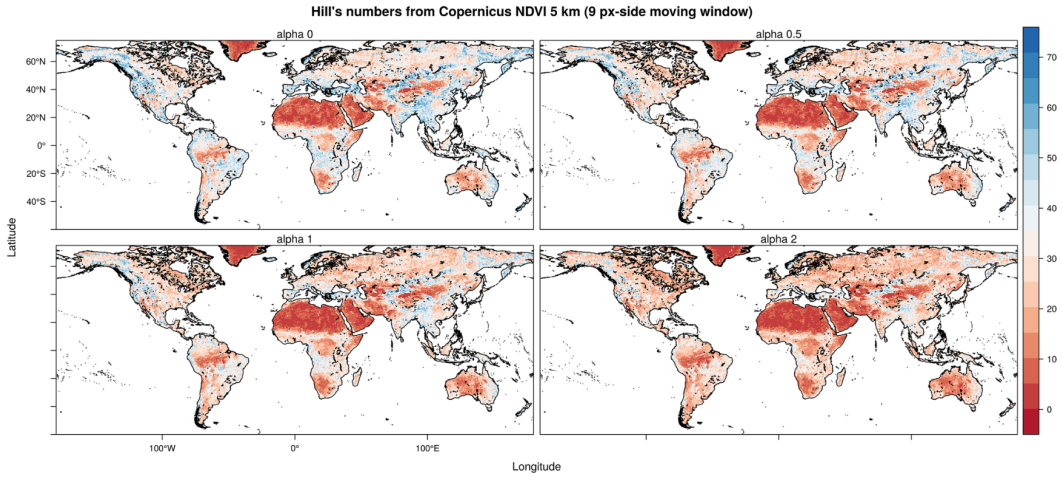

FIGURE 1.7: Another generalised entropy measure of diversity: the Hill index, for which the same reasoning of the Rényi index holds true. The maps are derived from a Copernicus Proba-V NDVI image at 5km. This figure has been generated by the command `hil <- Hill(ndvi17_r,window=9,np=8,cluster.type="SOCK",alpha=c(0,0.5,1,2))`.

1.7 Discussion

In this paper, we provided a full description of the main functionalities of the new R package rasterdiv. The rasterdiv package provides an unprecedented suite of functions to calculate different indicies for estimating diversity from space and to perform a first exploration of potential biodiversity hotspots worldwide at a glance. Of course, measures of diversity from space should not be viewed as a replacement of in-situ data on biological diversity, but they are rather complementary to existing data and approaches. In practice, they integrate available information of Earth surface properties, including aspects of functional (structural, biophysical and biochemical), taxonomic, phylogenetic and genetic diversity (Laliberté et al., 2019).

Obviously, in most of the Information Theory based metrics, only one layer can be used, considering those indices related to relative abundance, apart from the Rao's Q and the Cumulative Residual Entropy (CRE). In the Rao's Q index, multidimensional systems can be used to calculate spectral distance (see also Nakamura et al. (2020) on the dimensionality of diversity), while in the CRE it is possible to calculate a multidimensional cumulative distribution to be used in the estimates (Drissi et al., 2008). In general, remotely sensed data are actually the approximation of more complex systems, which depends on the original radiometric and spectral resolution. In ecological terms, such original spectral space formed by many bands is analogous to the Hutchinson's hypervolume, in which a geometrical order is given to those variables shaping species' niches (Hutchinson, 1959; Blonder, 2018). In this case, the spectral space is expected to be related to both species niches and their relative diversity. The use of such spaces is an efficient approach to figure out the diversity of an area and potentially guide field sampling and monitoring schemes (Rocchini et al., 2008, 2018).

Concerning the data being used, spectral diversity measures computed from satellite images represent a valid alternative to class-based land cover maps for investigating landscapes heterogeneity (Rocchini et al., 2014). For instance, a highly fragmented landscape characterised by a mosaic of crops and seminatural forests suffers from oversimplification when investigated through land cover classes (Amici et al., 2018), while it should present higher spectral diversity values compared to more homogeneous landscapes within the same study area (Rocchini, 2007). Several studies have already acknowledged the importance of computing continuous spectral diversity measures from spectral bands in order to better understand and discriminate the various landscape components (Karlson et al., 2015; Godinho et al., 2018; Ribeiro et al., 2019; Doxa et al., 2020). This said, caution is warranted when making use of continuous data, by seriously considering the radiometry of pixel values. As an example, relying on continuous NDVI (Normalised Difference Vegetation Index) values, ranging from -1 to 1 with float (decimal) precision data, will lead to a high neighbouring diversity which could actually be the effect of data binning rather than of a biological underlying pattern. In general, an 8-bit image with a range of integer values/classes from 0 to 255 would be preferable. In this paper, we made use of an 8-bit NDVI layer rescaled from Copernicus data. However, a multispectral system reduced to

one single layer through the first component of a Principal Component Analysis, or similar multidimensionality reduction techniques, would also be useful (Féret and Boissieu, 2020). In fact, NDVI assumes a biomass-grounded reflectance model, while the direct use of the original spectral data (digital numbers) does not generally require any assumptions about the biology of objects being sensed.

As remotely sensed estimates of diversity are currently based on relatively long time series, they might allow a more general forecasting framework of future shifts in rates of diversity change. This is particularly important when aiming at finding potential indicators of diversity change in time (Schmeller et al., 2018). On this point, it has been widely demonstrated that remotely sensed diversity might be in line with most of the spatially-constrained Essential Biodiversity Variables proposed by Skidmore et al. (2015).

The `rasterdiv` package might also be particularly useful when aiming at calculating diversity directly from climate data, derived from remote sensing (Metz et al., 2014). This could allow analysing diversity based on the main drivers of biological diversity in the field, rather than on the patterns resulting from pure spectral response. This is true when considering both wide climatic variations at global scale and microclimate variations at the scale of individuals (Zellweger et al., 2019). Due to unprecedented rates of climatic changes, the adaptation of species to climate change is a benchmark in ecology. Hence, estimating diversity from climate gridded data could improve our understanding of the variability of species ranges at different spatial and temporal scales (Senner et al., 2018).

1.8 Conclusion

Measuring diversity from above and delivering rapid and robust knowledge about diversity over wide regions could be of crucial importance for guiding management practices. From this point of view, the spatial variation of the spectral signal has an intrinsic cumbersome relation with the spatial autocorrelation (sensu Laliberté (2008)) of pixel values over space (and time, e.g. Rocchini et al. (2019)), which renders the proposed `rasterdiv` package a powerful tool to monitor the variation of ecosystems properties over space and time, and thus their change (Rocchini et al., 2018).

As previously stated, no single measure provides a full description of all the different aspects of diversity. That is why, the `rasterdiv` package can result useful in making multiple calculations based on reproducible open

source algorithms, robustly rooted on Information Theory from which the different indices are extracted.

Bibliography

Amici, V., Filibeck, G., Rocchini, D., Geri, F., Landi, S., Giorgini, D., Scoppola, A., Chiarucci, A. (2018). Are CORINE land cover classes reliable proxies of plant species assemblages? A test in Mediterranean forest landscapes. Plant Biosystems, 152: 994-1001.

Bachl, F.E., Lindgren, F., Borchers, D.L., Illian, J.B. (2019). inlabru: an R package for Bayesian spatial modelling from ecological survey data. Methods in Ecology and Evolution, 10: 760-766.

Berger, W.H., Parker, F.L. (1970). Diversity of planktonic foraminifera in deep-sea sediments". Science, 168: 1345-1347.

Blonder, B. (2018), Hypervolume concepts in niche- and trait-based ecology. Ecography, 41: 1441-1455.

Boltzmann, L.E. (1872). Weitere studien über das waärmegleichgewicht unter gasmolekälen. S. K. Akad. Wiss. Wein 66: 275-370.

Chao, A., Chiu, C.-H., Jost, L. (2106). Phylogenetic diversity measures and their decomposition: a framework based on Hill numbers. In: Pellens, R., Grandcolas, P. (2016). Biodiversity Conservation and Phylogenetic Systematics - Preserving our evolutionary heritage in an extinction crisis. Springer, Basel, Switzerland.

Chiarucci, A., Bacaro, G., Scheiner, S.M. (2011). Old and new challenges in using species diversity for assessing biodiversity. Philosophical transactions of the Royal Society of London. Series B, 366: 2426-2437.

Doxa, A., Prastacos, P. (2020). Using Rao's quadratic entropy to define environmental heterogeneity priority areas in the European Mediterranean biome. Biological Conservation, 241: 108366.

Drissi, N., Chonavel, T., Boucher, J.M. (2008). Generalized cumulative residual entropy for distributions with unrestricted supports. Research Letters in Signal Processing, 2008: 1-5.

Dushoff, J., Kain, M.P., Bolker, B.M. (2019). I can see clearly now: Reinterpreting statistical significance. Methods in Ecology and Evolution, 10: 756-759.

Féret, J.-B., de Boissieu, F. (2020). `biodivMapR`: An R package for α-and β-diversity mapping using remotely sensed images. Methods in Ecology and Evolution, 11: 64-70.

Godinho, S., Guiomar, N., Gil, A. (2018). Estimating tree canopy cover percentage in a mediterranean silvopastoral systems using Sentinel-2A imagery and the stochastic gradient boosting algorithm. International Journal of Remote Sensing, 39: 4640-4662.

Gorelick, R. (2008).Species richness and the analytic geometry of latitudinal and altitudinal gradients. Acta Biotheoretica, 56: 197-203.

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

Gorelick, R. (2011b). What is theory? Ideas in Ecology and Evolution, 4: 1-10.

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.

Hesselbarth, M.H.K., Sciaini, M., With, K.A., Wiegand, K., Nowosad, J. (2019). *landscapemetrics*: an open-source R tool to calculate landscape metrics. Ecography, 42: 1648-1657.

Hill, M.O. (1973). Diversity and evenness: a unifying notation and its consequences. Ecology 54, 427-431.

Hijmans, R.J., van Etten, J. (2020). `raster`: Geographic analysis and modeling with raster data. R package version 3.0-12.
http://CRAN.R-project.org/package=raster

Hobohm, C. (2003). Characterization and ranking of biodiversity hotspots: centres of species richness and endemism, Biodiversity and Conservation, 12: 279-287.

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., Palmer, M.W., Fragomeni Simon, M., Stolz, C., Chiarucci, A. (2019). Global endemics-area relationships of vascular plants. Perspectives in Ecology and Conservation, 17, 41-49.

Hutchinson, G. 1959. Homage to Santa Rosalia or why are there so many kinds of animals? American Naturalist, 93: 145-159.

Jelinski, D.E., Wu, J. (1996). The modifiable areal unit problem and implications for landscape ecology. Landscape Ecology, 11: 129-140.

Jumarie, G. (1990). Relative Information. Springer, Berlin, Germany.

Karlson, M., Ostwald, M., Reese, H., Sanou, J., Tankoano, B., Mattsson, E. (2015). Mapping tree canopy cover and aboveground biomass in Sudano-Sahelian woodlands using Landsat 8 and Random Forest. Remote Sensing, 7: 10017-10041.

Kreft, H., Jetz, W. (2007). Global patterns and determinants of vascular plant diversity. Proceedings of the National Academy of Sciences, 104: 5925-5930.

Kroes, H.W. (1977). The niche structure of ecosystems. Journal of Theoretical Biology, 65: 317-326.

Laliberté, E., Schweiger, A.K., Legendre, P: (2019). Partitioning plant spectral diversity into alpha and beta components. Ecology Letters, 23: 370-380.

Laliberté, E. (2008). Analyzing or explaining beta diversity? Comment. Ecology, 89, 3232-3237.

Leitao, P.J., Schwieder, M., Senf, C. (2017). `sgdm`: An R package for performing Sparse Generalized Dissimilarity Modelling with tools for gdm. ISPRS International Journal of Geo-Information, 6: 23.

Leutner, B., Horning, N., Schwalb-Willmann, J., Hijmans, R.J. (209). `RStoolbox`: Tools for remote sensing data analysis. R package version 0.2.6. http://CRAN.R-project.org/package=RStoolbox

McGlinn, D.J., Palmer, M.W. (2009). Modeling the sampling effect in the species-time-area relationship. Ecology, 90: 836-846.

Meyer, C., Weigelt, P., Kreft, H. (2016). Multidimensional biases, gaps and uncertainties in global plant occurrence information. Ecology Letters, 19: 992-1006.

Metz, M., Rocchini, D., Neteler, M. (2014). Surface temperatures at the continental scale: tracking changes with remote sensing at unprecedented detail. Remote Sensing, 6: 3822-3840.

Michalowicz, J.V., Nichols, J.M., Bucholtz, F. (2013). Handbook of Differential Entropy. Chapman and Hall/CRC, London. UK.

Nakamura, G., Goncalves, L.O., Duarte, L.d.S. (2020). Revisiting the dimensionality of biological diversity. Ecography, in press.

Naimi, B., Hamm, N.A.S., Groen, T.A., Skidmore, A.K., Toxopeus, A.G., Alibakhshi, S. (2019). ELSA: Entropy-based local indicator of spatial association. Spatial Statistics, 29: 66-88.

Palmer, M.W. (2007). Species-area curves and the geometry of nature. in: Storch, D., Marquet, P., Brown, J. (Eds.) Scaling Biodiversity, Cambridge University Press, Cambridge, UK.

Pielou, E.C. (1966). The measurement of diversity in different types of biological collections. Journal of Theoretical Biology, 13: 131-144.

Rao, C.R. (1982). Diversity and dissimilarity coefficients: a unified approach. Theoretical Population Biology, 21: 24-43.

Rao, M., Chen, Y., Vemuri, B.C., Wang, F. (2004). Cumulative Residual Entropy: a new measure of information. IEEE Transactions in Information Theory, 50: 1220-1228.

Rényi, A., 1970. Probability Theory. North Holland Publishing Company, Amsterdam.

Ribeiro, I, Proenca, V., Serra, P., Palma, J., Domingo-Marimon, C., Pons, X., Domingos, T. (2019). Remotely sensed indicators and open-access biodiversity data to assess bird diversity patterns in Mediterranean rural landscapes- Scientific Reports, 9: 1-13.

Ricotta, C., Corona; P., Marchetti, M., Chirici, G., Innamorati, S. (2003a). LaDy: software for assessing local landscape diversity profiles of raster land cover maps using geographic windows. Environmental Modelling & Software, 18: 373-378.

Ricotta, C., Avena, G. (2003b). On the relationship between Pielou's evenness and landscape dominance within the context of Hill's diversity profiles. Ecological Indicators, 2: 361-365.

Rocchini, D., Dadalt, L., Delucchi, L., Neteler, M., Palmer, M.W. (2014) Disentangling the role of remotely sensed spectral heterogeneity as a proxy for North American plant species richness. Community Ecology, 15: 37-43.

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., Wegmann, M., Nagendra, H. (2018). Measuring β-diversity by remote sensing: a challenge for biodiversity monitoring. Methods in Ecology and Evolution, 9: 1787-1798.

Rocchini, D., Marcantonio, M., Arhonditsis, G., Lo Cacciato, A., Hauffe, H.C., He, K.S. (2019). Cartogramming uncertainty in species distribution models: A Bayesian approach. Ecological Complexity, 38: 146-155.

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.

Rocchini, D., Neteler, M. (2012). Let the four freedoms paradigm apply to ecology. Trends in Ecology & Evolution, 27: 310-311.

Rocchini, D., Ricotta, C. (2007). Are landscapes as crisp as we may think? Ecological Modelling, 204: 535-539.

Rocchini, D., Wohlgemuth, T., Ghisleni, S., Chiarucci, A. (2008). Spectral rarefaction: linking ecological variability and plant species diversity. Community Ecology, 9: 169-176.

Schmeller, D., Weatherdon, L., Loyau, A., Bondeau, A., Brotons, L., Brummitt, N., Geijzendorffer, I., Haase, P., Kuemmerlen, M., Martin, C., Mihoub, J.-B., Rocchini, D., Saarenmaa, H., Stoll, S., Regan, E. (2018). A suite of essential biodiversity variables for detecting critical biodiversity change. Biological Reviews, 93: 55-71.

Scheiner, S.M., Cox, S.B., Willig, M., Mittelbach, G.G., Osenberg, C., Kaspari, m. (2000). Species richness, species–area curves and Simpson's paradox. Evolutionary Ecology Research, 2: 791-802.

Schneider, F.D., Morsdorf, F., Schmid, B., Petchey, O.L., Hueni, A., Schimel, D.S., Schaepman, M.E. (2017). Mapping functional diversity from remotely

sensed morphological and physiological forest traits. Nature Communications, 8: 1441.

Schimel, D., Schneider, F.D. (2019). Flux towers in the sky: global ecology from space. New Phytologist, 224, 570-584.

Senner, N.R., Stager, M., Cheviron, Z.A. (2018). Spatial and temporal heterogeneity in climate change limits species' dispersal capabilities and adaptive potential. Ecography, 41: 1428-1440.

Shannon, C.E. (1948). A mathematical theory of communication. Bell System Technical Journal, 27: 379-423, 623-656.

Skidmore, A.K., Pettorelli, N., Coops, N.C:, Geller, G.N., Hansen, M., Lucas, R., Mucher, C.A., O'Connor, B., Paganini, M., Pereira, H.M., Schaepman, M.E., Turner, W., Wang, T., Wegmann, M. (2015). Agree on biodiversity metrics to track from space. Nature, 523: 403-405.

Skidmore, A.K., Franklin, J., Dawsonc, T.P., Pilesjo, P. (2011). Geospatial tools address emerging issues in spatial ecology: a review and commentary on the Special Issue. International Journal of Geographical Information Science, 25: 337-365.

Simpson, E.H. (1949). Measurement of diversity. Nature. 163: 688.

Zellweger, F., De Frenne, P., Lenoir, J., Rocchini, D., Coomes, D. (2019). Advances in microclimate ecology arising from remote sensing. Trends in Ecology & Evolution, 34: 327-341.

Chapter 2

Integrals of life: tracking ecosystem spatial heterogeneity from space through the area under the curve of the parametric Rao's Q index

Published as:
Thouverai, E., Marcantonio, M., Lenoir, J., Galfré, M., Marchetto, E., Bacaro, G., Gatti, R.C., Da Re, D., Di Musciano, M., Furrer, R. and Malavasi, M., Moudrý, V., Nowosad, J., Pedrotti, F., Pelorosso, R., Pezzi, G., Šímová, P., Ricotta, C., Silvestri, S., Tordoni, E., Torresani, M., Vacchiano, G., Zannini, P., Rocchini D. (2023). Integrals of life: Tracking ecosystem spatial heterogeneity from space through the area under the curve of the parametric Rao's Q index. Ecological Complexity, 52, p.101029.

2.1 Abstract

Spatio-ecological heterogeneity is strongly linked to many ecological processes and functions such as plant species diversity patterns and change, metapopulation dynamics, and gene flow. Remote sensing is particularly useful for measuring spatial heterogeneity of ecosystems over wide regions with repeated measurements in space and time. Besides, developing free and open source algorithms for ecological modelling from space is vital to allow to prove workflows of analysis reproducible. From this point of view, NASA developed programs like the Surface Biology and Geology (SBG) to support the development of algorithms for exploiting spaceborne remotely sensed data to provide a relatively fast but accurate estimate of ecological properties in vast areas over time. Most of the indices to measure heterogeneity from space are point descriptors : they catch only part of the whole heterogeneity spectrum. Under the SBG umbrella, in this paper we provide a new R function part of the `rasterdiv` R package which allows to calculate spatio-ecological heterogeneity and its variation over time by considering all its possible facets. The new function was tested on two different case studies, on multi- and hyperspectral images, proving to be an effective tool to measure heterogeneity and detect its changes over time.

2.2 Introduction

The concept of spatiotemporal heterogeneity is crucial in ecological modelling to link spatial patterns to the generating processes and to the functional networking among organisms (Borcard et al., 1992). In ecological research, the search for new methods underlying spatiotemporal patterns in ecosystem heterogeneity has been a recurring theme (Rocchini and Ricotta, 2007; Atluri et al., 2018). Spatio-ecological heterogeneity, in this paper considered as the degree of non-uniformity in vegetation, land cover, and physical factors (soil, topography, microclimate and topoclimate; (Stein et al. , 2014), has been proven to be strongly linked to many ecological processes and functions such as plant species diversity patterns and change (Rocchini et al., 2018), metapopulation dynamics (Fahrig , 2007), and gene flow (Lozier et al., 2013). Indeed, an increase of spatial heterogeneity means an increase in the availability of ecological niches, provision of refuges at relatively short distances and opportunities for spatial isolation and local adaptation (Stein et al. , 2014). As a consequence, species coexistence, persistence and diversification

are generally in strict relation with the degree of environmental heterogeneity available within the landscape (Stein et al. , 2014; Tews et al., 2004). The development of new methods for measuring spatio-ecological heterogeneity is also fundamental to make estimations of its change in time in order to improve conservation planning (Skidmore et al., 2021).

In this context, NASA developed programs like the Global Ecosystem Dynamics Investigation (GEDI, `gedi.umd.edu/`) or the Surface Biology and Geology (SBG) mission (`science.nasa.gov/earth-science/decadal-sbg`) exploiting spaceborne remotely sensed data to provide a relatively fast but accurate estimate of spatio-ecological heterogeneity in vast areas over time. In fact, spectral heterogeneity of an optical image - associated with the reflectance values of the pixels - can be a proxy of the spatio-ecological heterogeneity (Rocchini, 2007). Hence, the variation of spatio-ecological heterogeneity in space and time (e.g., phenological cycles) can be effectively inferred using remote sensing (Schneider et al., 2017).

Therefore, the measure of ecosystem heterogeneity over time from satellite through Free and Open Source Software and algorithms allows robust, reproducible and standardized estimates of ecosystem patterns and processes (Rocchini and Neteler, 2012). Also, its use brings many advantages: availability, transparency and shareability. In this context, the R platform is one of the most used statistical and computational environment in ecology, partially thanks to the continuous development of relevant packages. In particular, the `rasterdiv` package (Marcantonio et al. , 2021; Rocchini et al., 2021; Thouverai et al., 2021) allows to calculate a plethora of different indices to measure spatio-ecological heterogeneity from space.

Most of the algorithms have been related to Information Theory relying on abundance-based metrics, starting from Shannon's index (Shannon, 1949) (see section 2.3). However, some information about the spectral distance among pixel reflectance values might be lost if not considered in the calculation (Rocchini et al., 2017). Currently, the candidate for solving the problem is Rao's Quadratic Entropy index (hereafter Rao's Q) (Rao, 1982): this index, besides the relative abundance of pixel values in a given moving window or polygonal area, incorporates also their spectral distances (section 2.3). Both Shannon and Rao's Q indices are point descriptors of heterogeneity, namely they can only show part of the whole heterogeneity spectrum. Recently Rocchini et al. (2021) proposed an implementation of the Rao's Q index by parameterizing the original formula, and allowing the whole continuum of heterogeneity to be measured thanks to Rao's Q continuous profiles (see section

2.3).

This paper aims to show how to make proper use of the Rao's continuum heterogeneity variation profile by proposing a new R function – integrated into the `rasterdiv` R package (Marcantonio et al. , 2021) which calculates AUC, the area under the curve formed by applying the parametric Rao's Q index (see section 2.3). Two case studies on multi- and hyperspectral satellite images are also provided in order to verify if the new metric proposed could be an effective tool for the study of spatio-ecological heterogeneity.

2.3 The algorithm

2.3.1 The theory

Algorithms that aim to measure environmental heterogeneity through remote sensing data can rely on the moving window technique, which divides remotely sensed imagery into user-defined squares (windows) to derive measures of heterogeneity. Examples are included in the `rasterdiv` R package (Rocchini et al., 2021). One of the most used metrics included in the package is the Shannon entropy index H (Shannon, 1949):

$$H = -\sum_{i=1}^{N} p_i \ln p_i \tag{2.1}$$

where the relative abundance of every pixel reflectance value calculated as the ratio between the actual value of the pixel $i \in \{1, ..., N\}$ and the sum of the pixel values of the moving window (p_i) in an image of N pixels is considered. It is usally calculated of one layer images, such as a vegetation index or the first axis of a PCA. However, Shannon's H does not consider the spectral distances among pixel reflectance values, overestimating the heterogeneity of homogeneous surfaces (Rocchini et al., 2017). For instance, when using Shannon's H, spectral values differing by a few decimals will be treated the same as spectral values differing by several order of magnitudes. To overcome this issue, Rao's Q index (Rao, 1982) can be used to include the pixel's spectral distances in the calculation:

$$Q = \sum_{i=1}^{N} \sum_{j=1}^{N} d_{ij} \times p_i \times p_j \tag{2.2}$$

where d_{ij} is the spectral distance between pixel i and pixel j and p_i and p_j are the relative abundances of the pixels i and j in an assemblage of N pixels.

The spectral distance between pixels d_{ij} can be calculated over any number of layers and using any metric for the calculation of pairwise distances. For example, in the `rasterdiv` package, the `Rao` function permits the calculation of Rao's Q chosing from "euclidean", "manhattan", "canberra", "minkowski" and "mahalanobis" as the type of distance calculated (Marcantonio et al. , 2021). Both Shannon's H and Rao's Q are point descriptors of heterogeneity, showing only one part of its potential spectrum. Therefore, the use of generalized entropies, where one single formula represents a parameterized version of an index, provides a continuum of heterogeneity metrics reflecting all the characteristics of the heterogeneity spectrum. Rocchini et al. (2021) presented a parametric version of Rao's Q allowing the characterisation of the dimensionality of heterogeneity in different ecosystems:

$$Q_\alpha = \left(\sum_{i,j=1}^{N} \omega_{ij} d_{ij}^{\alpha} \right)^{\frac{1}{\alpha}} \tag{2.3}$$

where d_{ij} is the spectral distance between pixel i and pixel j and ω_{ij} is the combined probability $(1/N^2)$ of extracting pixels i and j in this order in an image of N pixels. In other words, parametric Rao's Q is a generalized mean that measures the expected distance between two randomly chosen pixels regulated by the parameter α. The α parameter provides a continuum of potential diversity indices by regulating the weight of d_{ij} with the highest values obtaining different types of means as it is increasing ($[\alpha \to 0] \Rightarrow$ geometric, $[\alpha = 1] \Rightarrow$ arithmetic, $[\alpha = 2] \Rightarrow$ quadratic, $[\alpha = 3] \Rightarrow$ cubic, and so on till $[\alpha \to \infty] \Rightarrow max_d$).

In this paper, we propose to calculate the area under the curve (AUC) constructed by applying the index parametric Rao's Q over a sequence of α values. We want to verify if AUC can be used to quantify the width of the diversity spectrum calculated with parametric Q for each pixel, resulting in an image that can be exploited to monitor the change in the heterogeneity spectrum over time for a selected area.

2.3.2 The R function

The function `rasterdiv::RaoAUC()` exploits the function `rasterdiv::paRao()` to define the values of the parametric Rao's Q using a vector of alphas decided by the user. Accordingly, the values of parametric Rao's Q are calculated building a moving window around every pixel of the remote sensing

image for every alpha selected. Then, the integral of the curve formed by the values of the parametric Rao's Q index obtained for every pixel is calculated.

2.4 Examples

In this section, we present one theoretical examples and two case studies for the new R function proposed (`RaoAUC()`). Specifically, AUC was calculated for one layer, multi- and hyperspectral satellite images of areas afflicted by a sudden event that changed the spatio-ecological heterogeneity of the area. We choose two images per case study of two different moments in time and calculated the difference between the two, highlighting the increase in heterogeneity.

2.4.1 A theoretical example

In this section, we will show how to use the function `accRao()` from the `rasterdiv` package to calculate the accumulation function (integral) of Rao values obtained using a range of alpha-values. We used a raster for the global average NDVI rescaled at 8-bit available from `rasterdiv`. This raster was first cropped on the islands of Sardinia and Corsica. In order to simulate the effects of an ecological perturbation, for example widespread drought, we created a new raster with perturbed NDVI values for these two islands. Pixels with NDVI higher than 150 were decreased using values from a normal distribution centered on 50 with a standard deviation of 5. Then, we applied `accRao()` both on the original and simulated raster by using alphas ranging from 1 to 10:

```
RaoAUC.before ← accRao(
  alphas = 1:10, #range of alphas

  x = ndvi.before, #raster layer

  dist_m = "euclidean", #method for the
                        #calculation of the
                        #spectral distance

  window = 3, #dimension of the moving window

  method = "classic", #specifies if the function
                      #is applied on a single
                      #layer or on a
                      #multidimensional system

  rasterAUC = TRUE, #specifies if the output
                    #will be a raster layer or
                    #a matrix

  na.tolerance = 0.4, #proportion of NA values
                      #tolerated

  np = 1 #number of cores which will be spawned
)

RaoAUC.after ← accRao(alphas=1:10,
  x = ndvi.after,
  dist_m = "euclidean",
  window = 3,
  method = "classic",
  rasterAUC=TRUE,
  na.tolerance=0.4, np=1)
```

Afterwards, the difference between the two rasters, before and after the simulated perturbation, was calculated (Figure 2.1). Also, the average parametric Rao of the images in Figure 2.1 was calculated for every α value, and the resulting curves are showed in Figure 2.2.

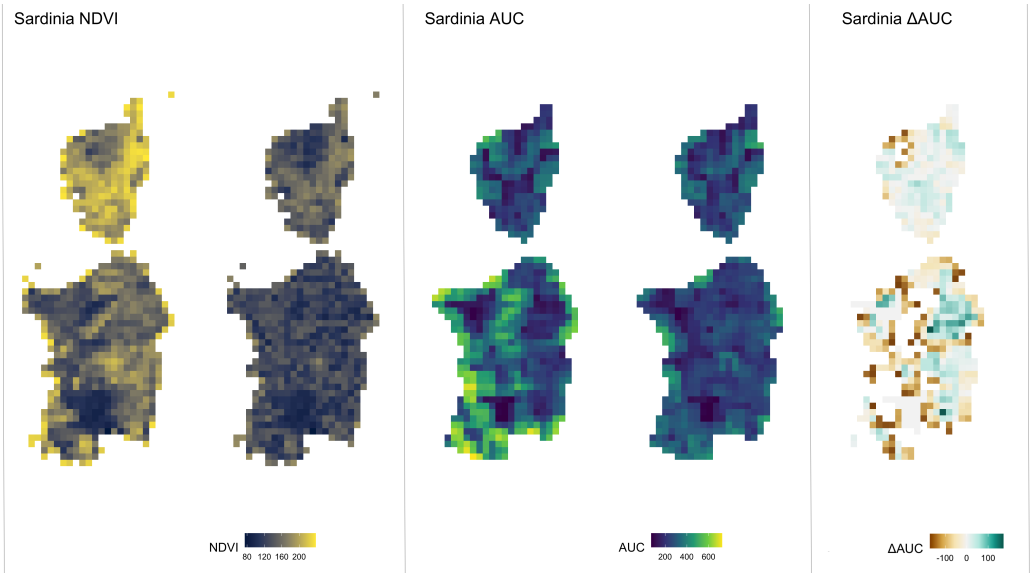

FIGURE 2.1: From left to right: the NDVI images of Sardinia and Corsica before and after the simulated perturbation, the correspondent AUC images and their difference after - before the simulated perturbation.

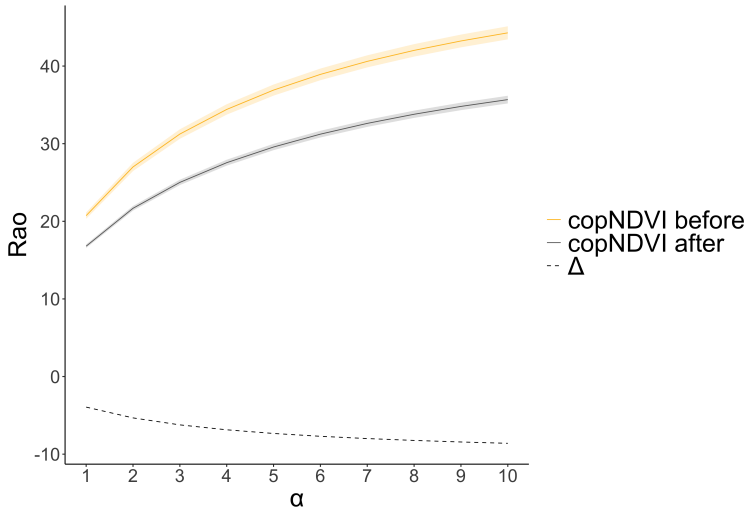

FIGURE 2.2: Three curves representing respectively: the mean values of parametric Rao's Q (i) before (yellow) and (ii) after (grey) the simulated ecological perturbation (drought) of Figure 2.1, their correspondent confidence intervals and (iii) their difference (after - before, dashed line) over increasing alphas.

`accRao()` function derives the value of parametric Rao for each pixel using a moving window algorithm. To illustrate how this methodology works, we applied `paRao()` on a single group of neighbor pixels, which represents a moving window, from the two NDVI rasters and with alphas ranging from 1 to 10 as follows:

```
#Selection of the 3x3 window
ndvi.pix.b ← ndvi.before[41:43,
  21:23,drop = FALSE]
ndvi.pix.a ← ndvi.after[41:43,   21:23,drop=FALSE]

#Set the alpha interval
alphas ← 1:10

#Set the number of pixels in the selected window
N ← 3^2

#Function to calculate paRao over the set alphas
RaoFx ← function(alpha,N,D) {
  ( sum((1/(N^4)) * D^alpha )*2)^(1/alpha)
}

#Calculation of paRao before
rao.b ← sapply(alphas, function(a) {
  RaoFx(alpha = a, N = N,
  D = as.vector(ndvi.pix.b))})

#Calculation of paRao after
rao.a ← sapply(alphas, function(a) {
  RaoFx(alpha = a, N = N,
  D = as.vector(ndvi.pix.a))})
```

From the values obtained (a parametric value for each alpha), the area under the curve was calculated integrating the results (Figure 2.3):

```
#Calculation of AUC before
RaoAUC.bf ← approxfun(x = alphas, y = rao.b)
RaoAUC.b ← integrate(RaoAUC.bf, lower = 1,
  upper = 10, subdivisions = 500)

#Calculation of AUC after
RaoAUC.af ← approxfun(x = alphas, y = rao.a)
RaoAUC.a ← integrate(RaoAUC.af, lower = 1,
  upper = 10, subdivisions = 500)
```

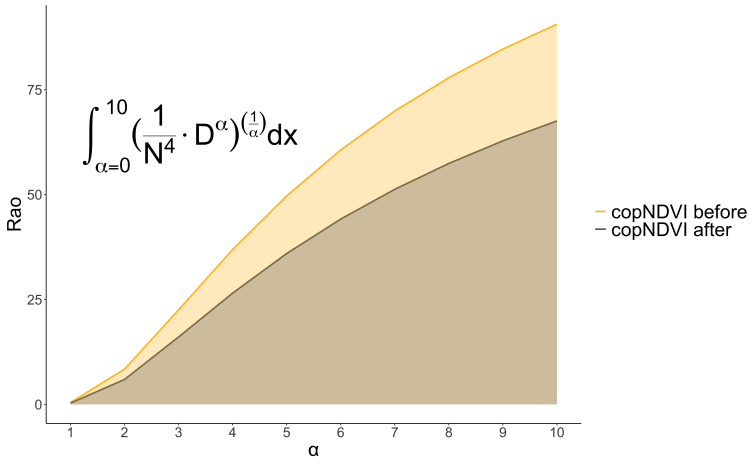

FIGURE 2.3: Curves representing the values of parametric Rao's Q for one pixel before (yellow) and after (grey) the simulated ecological perturbation (drought) of Figure 2.1 over increasing alphas. The area under the curve (AUC) is highlighted.

2.4.2 Empirical examples

In this section, the `accRao()` function is tested on two real-world case studies by comparing remotely sensed images before and after a perturbation event. AUC is calculated on multi- and hyperspectral images, exploiting the information that every band holds to estimate the spatio-ecological heterogeneity.

Example 1: Fire spread in the Kangaroo island (Australia)

This section focuses on the major fire-affected area of Kangaroo Island in January 2020, in particular on Flinders Chase NP and the associated Ravine Des Casoars Wilderness Protection Area. Two cloudless images from Copernicus Sentinel-2 (https://scihub.copernicus.eu/) with a spatial resolution

of 10m before (January 2019) and after (January 2021) were compared (Figure 2.4).

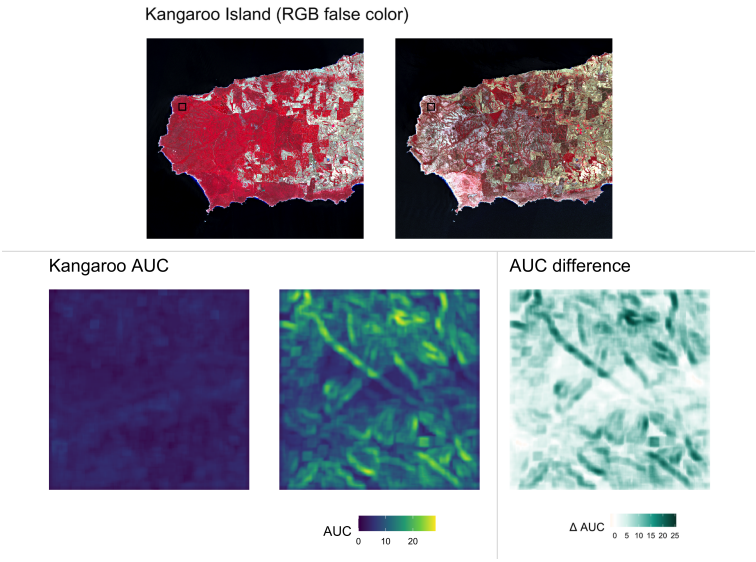

FIGURE 2.4: On top left, the Kangaroo Island before and after the fires (the area used for the analysis is highlighted) and the selected area before and after the fire in RGB false color (NIR, red, green); on the right the correspondent AUC images and their difference after - before the fire.

The accRao() function was applied on the 2 multispectral images (Red, Green, Blue and NIR bands) using a moving window of 9×9 pixels and the parameter alpha was set to a range of 1 to 5:

```
#accRao() function
accRao(alphas = 1:5, x = kanga_multi,
       dist_m = "euclidean", window = 9,
       method = "multidimension", rasterAUC = TRUE,
       na.tolerance = 0.9, np = 1)
```

Subsequently, the difference between the obtained AUC images was calculated, with positive values meaning an increase in spatio-ecological heterogeneity (Figure 2.4). In this case, the AUC of Rao's Q profiles succeeded to highlight areas where the perturbation (fire) event caused an increase of spatial heterogeneity of vegetation which was more homogeneous (continuous woodland cover) before the perturbation.

Example 2: Post fire in Santa Barbara, California

For the last empirical examples two hyperspectral images of a postfire scene in Santa Barbara (California) were downloaded from AVIRIS https://aviris.jpl.nasa.gov/ platform. The first image is from June 2009, the second from June 2011 in order to visualize the recovery of the vegetation after the fire event (see Figure 2.5).

FIGURE 2.5: Post fire in Santa Barbara 2009 (left) and 2011 (right). The area within the square is the studied area.

The `accRao()` function was applied over all the 224 bands of the two images using a moving window of 9×9 pixels and setting the α parameter to a range of one to 5:

```
#accRao() function
accRao(alphas = 1:5, x = santabarbara_hyper,
       dist_m = "euclidean", window = 9,
       method = "multidimension", rasterAUC = TRUE,
       na.tolerance = 0.9, np = 1)
```

Subsequently, the difference between the obtained AUC images was calculated as in the previous examples (Figure 2.6). The difference between the

obtained AUC highlights subtle changes of spatio-ecolocical heterogeneity in the studied area between 2009 and 2011.

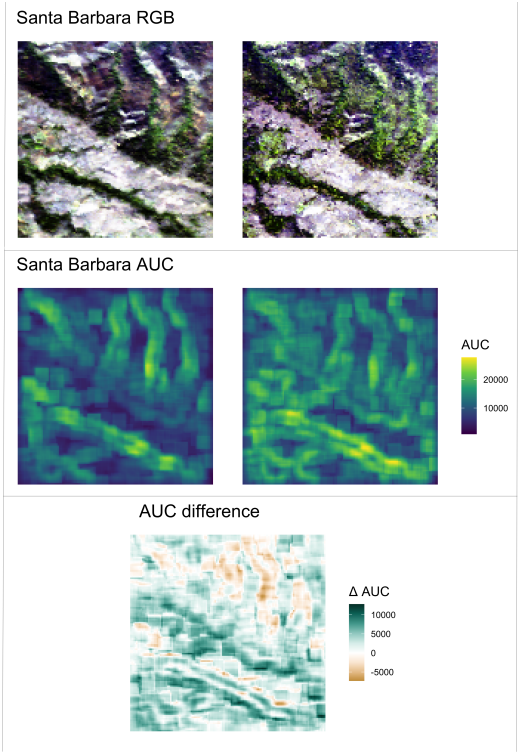

FIGURE 2.6: From the top: RGB images of the study area (Santa Barbara, CA) in 2009 and 2011, the correspondent AUC images and their difference 2011 - 2009.

2.5 Discussion

The study of landscape structure has been steadily growing in recent years (e.g., Lichstein et al., 2002; Saravia, 2015) with the development of several methodologies and approaches, which have been tested ecosystems and supported in the scientific literature (see Bar-Massada and Wood, 2014). In particular, the use and availability of remote sensing data have made it possible to assess specific heterogeneity patterns over various ecosystems, with increasing performance in terms of spectral/spatial/temporal characteristics, opening up new possibilities for exploring complex ecological processes.

Using our algorithm, environmental heterogeneity is estimated by the range of spectral values associated to the spatial variability within a given habitat. Hence, environmental heterogeneity can be evaluated contiguously, from regional to continental extents, according to the remote sensing data used and the spatial extent of the analysis. Among the heterogeneity metrics, parametric Rao's Q adds a layer of information to classical estimates of

heterogeneity from remotely sensed multispectral data. This index considers pairwise pixel spectral distance to separate areas with high richness but low evenness from those with low richness but high evenness (Rocchini et al., 2017).

In addition, the parametric Rao's Q can be calculated in a multivariate system such as a multi-temporal system, i.e. long time series, in order to improve the assessments and prediction of changes in spatio-ecological heterogeneity over space and time (Rugani and Rocchini, 2016). Also, by considering multiple bands, it has a higher capability to discern subtle diversity changes over the landscape (Torresani et al., 2019).

In this paper, all the potential facets of heterogeneity were investigated by parameterizing the Rao's Q metric and calculating the area under the curve of continuous entropy profiles. This would be particularly useful when dealing with multitemporal sets, with increases or decreases of heterogeneity provoked by different ecological processes like drought (subsection 2.4.1, see also Jiao et al., 2020) and fire (subsection 2.4.2, see also Chuvieco and Kasischke, 2007; subsection 2.4.2).

The application of AUC on Rao's Q in before / after ecological perturbation scenarios can help pointing out areas with the highest difference in spectral heterogeneity, by considering the whole heterogeneity continuum. For example, subsection 2.4.2 of two postfire scenes shows the sensibility of the algorithm in highlight even subtle landscape changes using multiple bands for the analysis. Heterogeneity of ecosystems is multifaceted in its very nature. As stressed by (Gorelick, 2011) there is no "true heterogeneity" measurement since important holistic aspects of ecosystems are inevitably lost once making use of single metrics. From this point of view, the proposed generalized entropy, based on a parameterization of the Rao's Q entropy (and its area under the curve) can help catching the multidimensionality of ecosystem heterogeneity components (Nakamura et al., 2020), avoiding the intrinsic fallacy of a single best index of true heterogeneity (Gorelick, 2011).

Moreover, the Rao's Q original formula directly takes into account the distance among values (pixel reflectances once applied to remote sensing imagery). This leads to the possibility of accounting for the turnover among reflectances, also known as beta-diversity in ecology (Rocchini et al., 2018). Since little consensus has been reached as to general measures of heterogeneity / beta-diversity measurement in literature (Koleff et al., 2003), the aforementioned use of a generalized metric like the parametric Rao's Q helps detecting gradients in reflectance beta-diversity change (turnover) over space,

otherwise hidden when relying on point descriptors of heterogeneity, i.e. single metrics like the commonly used Shannon and Simpson indices in remote sensing applications (Nagendra, 2002). In other words, while a wide range of approaches has been used to catch the variation of ecosystem properties, finding ways to generalize heterogeneity measurement could represent a consistent approach to describe heterogeneity patterns change in space and time (Haralick & Kelly , 1969).

The use of a continuum of diversities as in the parametric Rao's Q leads to the understanding of hidden parts of the whole diversity of dimensionalities (Nakamura et al., 2020). Increasing alpha in Equation 2.3 will increase the weight of higher distances among different values until reaching the maximum distance value possible (Rocchini et al., 2021). For this reason, spatio-ecological heterogeneity values of the parametric Rao's Q increase with each alpha progressively added to the calculation constructing a curve for every moving window built around each pixel (Rocchini et al., 2021). Consequently, applying an integral, it is possible to calculate the area under every pixel's window area curve obtaining a new spatio-ecological heterogeneity metric, AUC. Hence, the `accRao()` function can highlight the differences before and after an ecological perturbation both in the theoretical and in the empirical examples (Figures 2.1, 2.4 and 2.6) showing the change in the whole heterogeneity continuum and being able to detect both: (i) spatially wide heterogeneity change patterns, as in the Kangaroo Island's fires example (see subsection 2.4.2), as well as (ii) spatially localized differences in space and time, as in the post fire in Santa Barbara example (see subsection 2.4.2).

The three examples proposed in section 2.4, show the application of AUC on one layer (subsection 2.4.1), multispectral (subsection 2.4.2) and hyperspectral 2.4.2 satellite images. However, for the hyperspectral images it is difficult to address a cause for the heterogeneity change: because of the high number of bands exploited for the analysis we can't know which ones weight more in the measure of the index. Analysis like the Principal Component Analysis (PCA) or correlation matrices can help to highlight the bands which give more contribution in the calculation of the spatio-ecological heterogeneity.

Also, in the empirical case studies only a range of alpha between 1 and 5 was tested because of the high computational complexity of the function `accRao()` as it is now. We are actually working to speed up the algorithm, so it would be interesting in a future study to test different ranges of alpha. In this context, it would also be helpful the study of the influence of the number

of bands and their resolution on the measure of AUC, as highlighted by the Santa Barbara subsection (see subsection 2.4.2).

In conclusion, the integration over an alpha range is more convenient than having to choose a single alpha level as the most representative level of diversity. This task is often complicated as there is no direct interpretation for the meaning of indexes calculated with different alphas. Here, we propose the way forward to re-conciliate the advantage of having a single metrics without the need of choosing a single alpha value.

2.6 Conclusion

In this paper, we provided a practical demonstration of the effectiveness of a method that can supply meauseres of generalized entropy at different spatial scales and in different contexts. Generalized means represent an effective tool to develop a unifying notation for a large family of parametric diversity and dissimilarity functions (Ricotta et al., 2021). Indeed, binding different heterogeneity metrics in order to analyze ecosystem changes proved to be a reliable approach to enhance the output information. Although remote sensing data have long held the promise of transforming environmental monitoring efforts, publicly accessible tools leveraging these data to achieve actionable in-sights have been lacking. We suggest that Rao's AUC can be useful to identify areas more vulnerable to environmental changes , and to develop and implement appropriate habitat management plans and environmental policies.

2.7 Acknowledgements

This study has received funding from the project SHOWCASE (SHOW - CASingsynergies between agriculture, biodiversity and ecosystems ser- vices tohelp farmers capitalising on native biodiversity) within the EuropeanUnion's Horizon 2020 Researcher and Innovation Programme under grantagreement No. 862480. DR was also supported by the Horizon Europe project Earthbridge.

PZ was supported by LifeWatch Italy through the project LifeWatch- PLUS (CIR01_00028).

RF is supported by the Swiss National Science Foundation, Switzer- land SNSF175529.

ETo is supported by the Estonian Research Council, grant code MOBJD1030.

MM is supported by the FRS-FNRS and by the "Action de Recherche concertée" grant number 17/22-086.

DR, MM, MG and ET contributed to the development of the R function and to the analysis. DR, PZ and ET contributed to the implementation of the figures. All the authors contributed to the writing of the manuscript.

The rasterdiv package that contains the new function proposed can be downloaded at https://github.com/mattmar/rasterdiv or directly from the CRAN (https://CRAN.R-project.org/package=rasterdiv). The hyperspectral images of Santa Barbara of June 2009 and 2011 can be respectively retrieved from https://popo.jpl.nasa.gov/avcl/y09_data/f090826t01p00r08.tar.gz and https://popo.jpl.nasa.gov/avcl/y11_data/f110719t01p00r19.tar.gz. The images of Kangaroo Island can be retrieved from https://scihub.copernicus.eu/dhus/#/home.

Bibliography

Atluri, G., Karpatne, A., & Kumar, V. (2018). Spatio-Temporal Data Mining: A Survey of Problems and Methods. *ACM Computing Surveys*, 51, 4-45.

Bar-Massada, A., & Wood, E.M. (2014). The richness-heterogeneity relationship differs between heterogeneity measures within and among habitats. *Ecography*, 37, 1-8.

Borcard, D., Legendre, P., & Drapeau, P. (1992). Partialling out the spatial component of ecological variation. *Ecology*, 73, 1045-1055.

Chuvieco, E., & Kasischke, E.S. (2007). Remote sensing information for fire management and fire effects assessment. *Journal of Geophysical Research: Biogeosciences*, 112, G01S90.

Fahrig, L. (2007). Landscape heterogeneity and metapopulation dynamics. In Wu, J., & Hobbs, R. (Eds.), *Key Topics in Landscape Ecology*(pp. 78-91). Cambridge, UK: Cambridge University Press.

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

Haralick, R.M., & Kelly, G.L. (1969). Pattern recognition with measurement space and spatial clustering for multiple images. *Proceedings of the IEEE*, 57(4), 654–665.

Koleff, P., Gaston, K.J., & Lennon, J.J. (2003). Measuring beta diversity for presence–absence data. *Journal of Animal Ecology*, 72, 367-382.

Jiao, T., Williams, C.A., Rogan, J., De Kauwe, M.G., & Medlyn, B.E. (2020). Drought impacts on Australian vegetation during the millennium drought measured with multisource spaceborne remote sensing. *Journal of Geophysical Research: Biogeosciences*, 125, e2019JG005145.

Lichstein, J.W., Simons, T.R., Shriner, S.A., & Franzreb, K.E. (2002). Spatial autocorrelation and autoregressive models in ecology. *Ecological Monographs*, 72, 445-463.

Lozier, J.D., Strange, J.P., & Koch, J.B. (2013). Landscape heterogeneity predicts gene flow in a widespread polymorphic bumble bee, *Bombus bifarius* (Hymenoptera: Apidae). *Conservation Genetics*, 14, 1099-1110.

Marcantonio, M., Iannacito, M., Thouverai, E., Da Re, D., Tattoni, C., Bacaro, G., Vicario, S., Ricotta, C. & Rocchini, D. (2021). rasterdiv: Diversity Indices for Numerical Matrices. R package version 0.2-3. https://CRAN.R-project.org/package=rasterdiv

Nagendra, H. (2002). Opposite trends in response for the Shannon and Simpson indices of landscape diversity. *Applied Geography*, 22, 175-186.

Nakamura, G., Gon calves, L.O., & Duarte, L.d.S. (2020). Revisiting the dimensionality of biological diversity. *Ecography*, 43, 539-548.

Negron-Juarez, R., Baker, D.B., Zeng, H., Henkel, T.K., & Chambers, J.Q. (2010). Assessing hurricane-induced tree mortality in U.S. Gulf Coast forest ecosystems. *Journal of Geophysical Research: Biogeosciences*, 115, G04030.

Rao, C.R. (1982). Diversity and dissimilarity coefficients: A unified approach. *Theoretical Population Biology*, 21, 24–43.

Ricotta, C., Szeidl, L., & Pavoine, S. (2021). Towards a unifying framework for diversity and dissimilarity coefficients. *Ecological Indicators*, 129, 107971.

Rocchini, D. (2007). Effects of spatial and spectral resolution in estimating ecosystem alpha-diversity by satellite imagery. *Remote Sensing of Environment*, 111, 423-434.

Rocchini, D., Luque, S., Pettorelli, N., Bastin, L., Doktor, D., Faedi, N., et al. (2018). Measuring beta-diversity by remote sensing: a challenge for biodiversity monitoring. *Methods in Ecology and Evolution*, 9, 1787-1798.

Rocchini, D., Marcantonio, M., Da Re, D., Bacaro, G., Feoli, E., Foody, G. M., et al. (2021). From zero to infinity: minimum to maximum diversity of the planet by spatio-parametric Rao's quadratic entropy. Global Ecology and Biogeography. *Global Ecology and Biogeography*, 30, 1153– 1162.

Rocchini, D., Marcantonio, M., & Ricotta, C. (2017). Measuring Rao's Q diversity index from remote sensing: An open source solution, *Ecological Indicators*, 72, 234-238.

Rocchini, D. & Neteler, M. (2012). Let the four freedoms paradigm apply to ecology. *Trends in Ecology & Evolution*, 27, 310-311.

Rocchini, D., & Ricotta, C. (2007). Are landscapes as crisp as we may think? *Ecological Modelling*, 204, 535-539.

Rocchini, D., Thouverai, E., Marcantonio, M., Iannacito, M., Da Re, D., Torresani, M., et al. (2021). rasterdiv - an Information Theory tailored R package for measuring ecosystem heterogeneity from space: to the origin and back. *Methods in Ecology and Evolution*, 12, 1093- 1102.

Rugani, B., & Rocchini, D. (2016). Positioning of remotely sensed spectral heterogeneity in the framework of life cycle impact assessment on biodiversity. *Ecological Indicators*, 61, 923-927.

Saravia, L.A. (2015). A new method to analyse species abundances in space using generalized dimensions. *Methods in Ecology and Evolution*, 6, 1298-1310.

Schneider, F.D., Morsdorf, F., Schmid, B., Petchey, O.L., Hueni, A., Schimel, D.S., & Schaepman, M.E. (2017). Mapping functional diversity from remotely sensed morphological and physiological forest traits. *Nature Communications*, 1441.

Shannon, C.E. (1948). A Mathematical Theory of Communication. *Bell System Technical Journal*, 27: 379-423.

Simpson, E. H. (1949). Measurement of diversity. *Nature*, 163, 688.

Skidmore, A.K., Coops, N.C., Neinavaz, E., Ali1, A., Schaepman, M.E., Paganini, M., Kissling, W.D., Vihervaara, P., Darvishzadeh, R., Feilhauer, H., Fernandez, M., Fernandez, N., Gorelick, N., Geizendorffer, I., Heiden, U., Heurich, M., Hobern, D., Holzwarth, S., Muller-Karger, F.E., Van De Kerchove, R., Lausch, A., Leitao, P.J., Lock, M., Mucher, C.A., O'Connor, B., Rocchini, D., Turner, W., Vis, J.-K., Wang, T., Wegmann, M., Wingate, V. (2021). Priority list of biodiversity metrics to observe from space. *Nature Ecology & Evolution*, 5, 896-906.

Stein, A., Gerstner, K., & Kreft, H. (2014). Environmental heterogeneity as a universal driver of species richness across taxa, biomes and spatial scales. *Ecology Letters*, 17, 866-880.

Tews, J., Brose, U., Grimm, V., Tielbörger, K., Wichmann, M., Schwager, M., & Jeltsch, F. (2004). Animal species diversity driven by habitat heterogeneity/diversity: the importance of keystone structures. *Journal of Biogeography*, 31, 79–92.

Thouverai, E., Marcantonio, M., Bacaro, G., Da Re, D., Iannacito, M., Marchetto, E., Ricotta, C., et al. (2021). Measuring diversity from space: a global view of the free and open source rasterdiv R package under a coding perspective. *Community Ecology*, 22, 1-11.

Torresani, M., Rocchini, D., Zebisch, M., Sonnenschein, R., Marcantonio, M., Ricotta, C., & Tonon, G. (2019). Estimating tree species diversity from space in an alpine conifer forest: the Rao's Q diversity index meets the Spectral Variation Hypothesis. *Ecological Informatics*, 52, 26-34.

Chapter 3

Helical graphs to visualize the NDVI temporal variation of forest vegetation in an open source space

Published as:
Thouverai, E., Marcantonio, M., Cosma, E., Bottegoni, F., Gatti, R.C., Conti, L., Di Musciano, M., Malavasi, M., Moudrý, V., Šímová, P. and Testolin, R., Zannini, P. (2023). Helical graphs to visualize the NDVI temporal variation of forest vegetation in an open source space. Ecological Informatics, 74, p.101956.

3.1 Abstract

Global change caused by human activity (e.g., land fragmentation, deforestation, pollution, anthropization of natural landscapes) has several effects on the biomes of the Earth, leading to alterations in the functioning of ecological systems. In this context, remote sensing represents an important tool to assess ecosystem changes, as it allows to collect a huge amount of data at different temporal and spatial resolutions concerning various compartments of the Earth system (land, ocean, atmosphere, and cryosphere). This information can be used to estimate precipitation patterns, global temperatures, snow cover and aerosol concentrations. The aim of this work is to exploit this wide availability of data to display the ecosystem changes using a new visualization method: the helical graphs. The helical graphs represent the change of a variable over time, reporting on the y-axis its moving averages and on the x-axis its rates of change. These new charts were tested on the NDVI index retrieved from Google Earth Engine (https://earthengine.google.com/) to visualize trends on selected biomes of the Earth (tropical and boreal forests). The results show that the helical graphs are a useful tool to highlight trends that might not be easily detected in a time series. In conclusion, the helical graphs can have a lot of application in ecology, especially exploiting the wide amount of data available thanks to the remote sensing.

3.2 Introduction

Nowadays, human activities are menacing the integrity of Earth's ecological systems, leading to heavy alterations of natural environments and to the disruption of ecosystems' equilibria (Sage and Kubien, 2007; Steffen et al., 2004). This phenomenon, known as "Global Change", includes land fragmentation, deforestation, pollution, anthropization of natural landscapes and alterations in the functioning of ecological systems (Risser et al., 2000). Among the latter, world's forests play a significant role in biodiversity maintenance and climate regulation, so that monitoring their change is important for biodiversity conservation, management of ecosystem services, and climate protection (Anderson-Teixeira et al., 2021). The impact of global change on forests is altering their function across different latitudes, so it is important to study its effects under different climatic conditions (Hansen et al., 2013). The consequences of global change are spreading faster and at a larger scale, becoming

more difficult to track them using solely ground-based monitoring (Rocchini et al., 2021).

In this context, remote sensing represents an important tool to assess ecosystem changes, as it allows to collect a huge amount of data at different temporal and spatial resolutions (Campbell and Wynne , 2011). This technology allows the estimation of several metrics from different compartments of the Earth system (land, ocean, atmosphere and cryosphere), such as precipitation patterns, global temperatures, snow cover and aerosol concentrations (Khorram et al. , 2012). Therefore, satellite images can be employed in a wide variety of studies, for instance to measure ecosystem (Rocchini et al., 2021) and species (Chapangu et al., 2020) diversity, study species (Alessi et al., 2021) and biomass (Santoro et al., 2021) distribution, evaluate variations in temperature, precipitation, wind direction, UV-B radiations and pollutants (Carey et al., 2001), assess changes in land cover (Boyd et al., 2002), estimate carbon stocks (Chapangu et al., 2020), assess endangered species' habitat alteration (Cazzolla Gatti and Velichevskaya, 2020), and monitor spatial status of coastal areas threatened by sea level rise and human activities (Ahmed et al., 2009).

In all these applications, global change is monitored by observing and reporting differences in the value of some important systemic variables (i.e., precipitation, diversity indices, fragmentation, etc.) between two or more points in time. To visualize these trends, different graphical methods have been proposed. One example is the use of different colors for different rates of change between two satellite images taken at two points in time (Ahmad, 2012), anomaly maps (Das and Srinivasan, 2009), and time-series plots (where time is reported on the x-axis and the target variable on the y-axis) (Du Toit et al., 1986). However, these visualization methods sacrifice some of the information for the sake of simplicity, while in some applications it may be critical to report additional details to understand ecological changes (Cooksey, 2020). This information should still be condensed in a single graph to remain easy to interpret. A solution is represented by the "Helical Graphs" (H-graphs) proposed by the statisticians Danny Dorling and the illustrator Kirsten McClure, which merge information on the direction of a metric together with the direction of the change between time points (Dorling, 2020).

In this paper, we focus on the vegetation state of tropical and boreal forests, given their ecological importance and their different climatic conditions. Also, as for the tropical forests, we report a local study in Brazil comparing different land use types. the Normalised Difference Vegetation Index (NDVI) is

employed to monitor the ecosystems over time. NDVI represents the productivity of the area over which is calculated, serving as a proxy for the photosynthetic portion of the ground biomass, and it is often used to detect changes in ecosystems over large areas (Huang et al., 2020).

The aim of this paper is to explore the use of H-graphs to visualize the change in vegetation biomass on two world's major forests biomes, in order to evaluate the opportunity of applying this alternative display for ecological temporal trends.

3.3 Study Areas

3.3.1 Tropical forests

Tropical forests are located between 23.5orth and 23.5outh of the equator (Gillman et al., 2015). They are characterised by dense tree cover, typically with a high diversity of trees, lianas and epiphytes (Murphy and Bowman, 2012). The canopy isgenerally dominated by evergreen broadleaves (Lewis, 2006), although moist- and drought-adapted deciduous trees are also present (Ravindranath and Sukumar, 1998).The climate of tropical forests is characterized by wet and dry seasons, which can vary greatly in their length and their intensity (Windsor, 1990). However, tropical forests are known for being the wettest of all biomes, with more than 1.5 m of rainfall annually (Lewis, 2006). Temperatures span between 20° and 30°C, but there are also areas where temperatures can be less than 20°C (Taylor et al., 2017).

This biome, even though covering only ca. 10% of the Earth's land surface, is of global importance as it stores and processes large quantities of carbon via photosynthesis and respiration (Malhi and Grace, 2000). For this reason, small changes within the tropical forest biome can potentially lead to major impacts at a global scale, on both the rate and magnitude of climate change and the conservation of biodiversity (Lewis, 2006). Nowadays, tropical forests are under pressure by human activities and their consequences, which include climate change. At the same time, they form the most diverse biome on Earth, housing over half of the planet's biodiversity (Lewis et al., 2015), including the majority of tree species (Cazzolla Gatti et al., 2022). Tropical rainforest ecozones account for 32% of total forest loss (Hansen et al., 2013), mainly due to agricultural production (Gibbs et al., 2010).

3.3.2 Boreal forests

Theh boreal forests biome is located around the northern circumpolar belt and dominated by coniferous vegetation (Apps et al., 1993). The climate is governed by strong seasonal variation (Thiffault, 2019) and the mean annual precipitation is less than 900 mm/yr (Binkley and Fisher, 2012). Despite this reduced rainfall regime, low temperatures and high cloud cover lead to low evaporative stress (Binkley and Fisher, 2012). Snow cover persists at least five to eight months (Shugart et al., 1992). Boreal forests are particularly endangered by climate change. As a matter of fact, it is warming faster than other biomes, approximately twice as quick as the global average (IPCC, 2007). The main cause of forest loss and decreased vegetation biomass in boreal forests are fires which, however, are often a consequence of logging activities (Cazzolla Gatti et al., 2021) and can be followed by higher photosynthetic activity due to the recovery of herbaceous and deciduous vegetation (Fiore et al., 2020).

3.4 Materials and Methods

3.4.1 Helical graphs theory

Helical Graphs can help visualizing the change of a variable over time. The values of the studied variable at different points in time are reported as moving averages on the y-axis, while its rate of change between measurements is reported on the x-axis. Generally speaking, the x-axis represents the rate of change of the moving average; hence the direction of the curve represents a decrease or an increase of the mean. In particular, a curve that tends to the right indicates a general increase of the mean between two points in time, while a curve that tends to the left shows a general decrease. For instance, given N values of a variable v_i to monitor over the times t_i, for every $i \in 0, 1, 2, \ldots, N$, m_i is the value represented on the y-axis of the H-graph, calculated as the moving average of v_i, v_{i-1} and v_{i+1}. Similarly, the rate of change rc_i associated to m_i, is calculated as the difference between v_{i+1} and v_{i-1}. Accordingly, the moving averages and the rates of change are calculated considering the values included between a certain distance in time before and exactly the same distance in time after the selected point in order to smooth the timelines for a better visualization of the data (Dorling, 2020).

3.4.2 Helical graphs implementation

As previously stated, two forest biomes with significantly different climatic conditions and vegetation dynamics were selected for the analysis from Dinerstein et al. (2017) classification: boreal forests and tropical forests. The tropical forests biome was retained as a macro-group including "Tropical and Subtropical Moist Broadleaf Forests" and "Tropical and Subtropical Dry Broadleaf Forests".

[b]0.9

FIGURE 3.1: Boreal forests biome

[b]0.8

FIGURE 3.2: Tropical forests biome

We also compared the tropical forest with other land use types at local scale using a land use map of Brazil for the year 2000 from MAPBIOMAS (https://mapbiomas.org/en/project) and selected three land cover types to compare (forest, forest plantation and crops) to provide a more local focus.

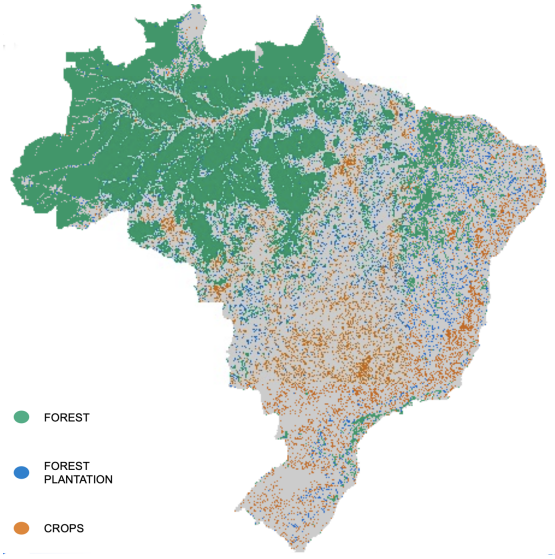

FIGURE 3.3: Land use map of Brazil downloaded from https://mapbiomas.org/en/project representing in green the tropical forest, in blue the forest plantations and in orange the crops.

We chose the NDVI index to represents the temporal trend of vegetation productivity over time for the selected areas. The monthly average NDVI for the tropical and boreal forests areas selected from the classification map of biomes and for the land cover types areas selected from the Brazil land use map was calculated from year 2000 to year 2021 using Google Earth Engine (Gorelick et al., 2017) through the R package "rgee" (R Core Team, 2021; Aybar et al., 2020). The data were retrieved from global MODIS NDVI products MOD13Q1 at a resolution of 250 m that provides for two to four images per month. Only the values of the northern hemisphere were selected taking seasonal fluctuations into consideration and excluding pixels with bad quality (ex cloudy areas or areas covered by snow).

NDVI ranges from -1 to 1 and is calculated from multispectral images as:

$$NDVI = \frac{NIR - Red}{NIR + Red} \tag{3.1}$$

where NIR is the near-infrared radiation (773 — 895 nm) and Red is the red band (610 — 690 nm).

For the construction of the helical graphs, the moving averages and the rates of change were calculated using R (R Core Team, 2021). The data were smoothed using LOESS (locally estimated scatter-plot smoothing) local regression using the function “loess” of “stats” package. The parameter span of the function indicating the degree of smoothing α was set to 0.4 and 0.1

respectively for the moving averages and rates of change of the monthly values for the biomes NDVI. As for the local study the span was set to 0.5. The degree of the smoothing is set to obtain the best possible visualization of the data in order to better fit seasonal variation depending on the case study. The complete code is available in Appendix 1.

3.5 Results and Discussion

In this study, we visualized the change in time of two of the world's major forest biomes, tropical and boreal forests. Specifically, we used NDVI (an index calculated from the NIR and Red bands of multispectral satellite images) as a proxy of vegetation productivity and biomass (Myneni et al., 1995). This index is also related to canopy properties such as leaf area index (LAI) and fractional vegetation cover (i.e. the fraction of ground covered by green vegetation) (Carlson and Arthur, 2000).

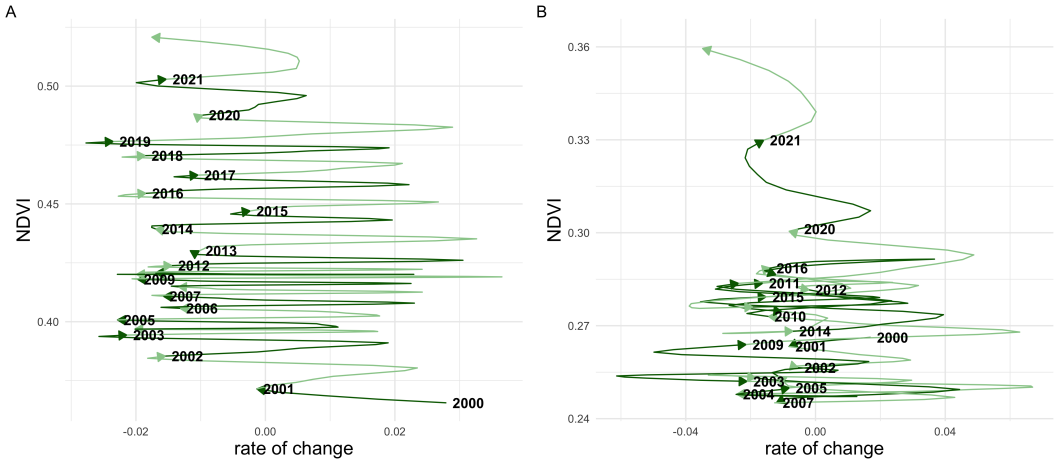

FIGURE 3.4: Helical graphs representing the variation from year 2000 to year 2021 of the tropical (A) and the boreal (B) forests. On the y-axis, the NDVI values at different time points, on the x-axis its change rate. Every arrow represents a year: even years are displayed as dark green, odd years as light green.

Figure 3.4 A, shows the helical graph of the variation of the NDVI index from year 2000 to year 2021 for the tropical forests. The rate of change oscillates between ˜0.02 and 0.03 and the NDVI goes from a value of 0.36 in the year 2000 to a value of 0.52 in 2021. The graph shows an increasing trend for the NDVI of tropical forests, which could be attributed to Global Change. In fact, rising temperatures and CO_2 concentration in the atmosphere may

lead to greaterprimary productivity and, consequently, to an overall increase in NDVI values (Norby et al., 1999; Paruelo et al., 2004). Despite this biome being increasingly threatened by deforestation, the NDVI kept increasing instead of decreasing, presumably because forested areas are often replaced by crops (Benhin, 2006). In a recent study conducted by Irteza et al. (2021), no significant NDVI differences were observed between exotic plantations and forests, so loss of natural forest is not necessarily related to the index.

The helical graph of the variation of the NDVI index from 2000 to 2021 for the boreal forests is reported in Figure 3.4 B. The rate of change oscillates between ˇ0.04 and 0.04 and the NDVI goes from a value of 0.24 in 2000 to a value of 0.36 in 2021. The graph shows an increasing trend interrupted by two decreases between years 2000 and 2005 and between years 2010 and 2013. These oscillations are probably related to El Nino–Southern Oscillation (ENSO) events, correlated with massive droughts and increased tree mortality, as well as larger and more frequent forest fires (Chunming et al., 2020). Droughts can induce the spread of massive die-back infections, whose extent and effects can be compared to those of fires (Michaelian et al., 2011). Even insect outbreaks (such as the spruce beetle which is decimating entire forests in Alaska) are increasing due to global change (Soja et al., 2007). Also, between 2011 and 2013 boreal forests experimented a huge tree cover loss mainly due to wildfires in Canada and Russia (the trend can be spotted from the helical graph in Fig. **??**) (Sizer et al., 2015). Nevertheless, the increase of NDVI values - despite the oscillations - can be caused by the recovery of herbaceous and deciduous vegetation. In addition, with boreal forest species being adapted to lower temperatures, the increase in global temperatures is menacing the native species of this biome (Frelich et al., 2021). Therefore, the observed increase in NDVI values could also be determined by the encroachment of species that are adapted to higher temperatures that replace the native species in the deforested areas (Kuuluvainen and Gauthier, 2018).

Concerning the test of the helical graphs at local scale, higher NDVI values for the forest, followed by forest plantation and crops were found (Figure 3.3).

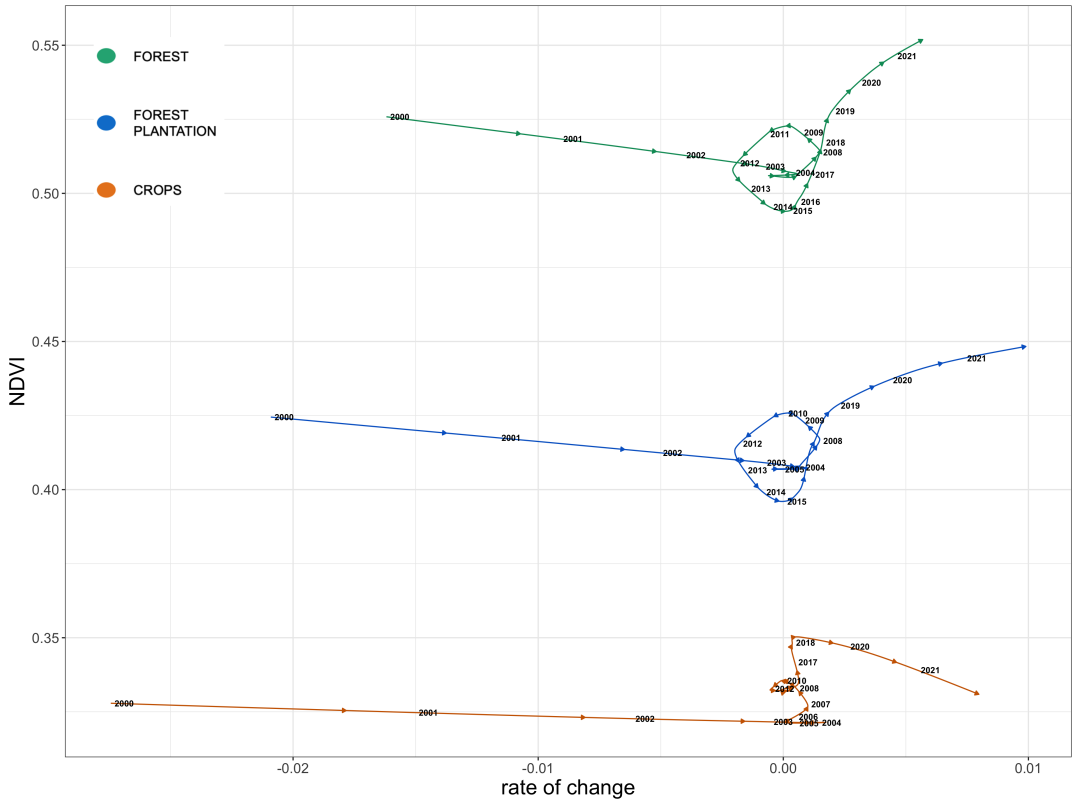

FIGURE 3.5: Helical graph of the NDVI variation from year 2000 to year 2021 of Brazil's forest, forest plantations and crops. On the y-axis, the NDVI values at different time points, on the x-axis its rate of change.

Figure 3.5 shows the highest NDVI values for forests, followed by forest plantations and crops. Moreover, we can observe almost the same curve for forests and forest plantations: an increase of NDVI beginning in 2003 until 2010, then a drop up to 2015, followed by another increase. This trend could be associated to the drought-related increase tree mortality linked to ENSO events in 2010 and at the end of 2015 (Celso et al., 2019). As for the crops, the observed trend is probably due to the crop rotation and to their conversion in pastures.

Considering the trends shown by the helical graphs, at large scale the use of NDVI to detect the changes at biome scale might not be the best choice, as there is no discrimination of landscape/land use types. Moreover, human intervention (e.g. disturbance) or local climatic effects might be effectively considered only at a regional scale. Strictly speaking, at a local scale, it might be easier to understand NDVI trends and make assumptions on landscape change, thanks to the selection of specific land use types.

This said, the helical graphs proposed in this work help visualize global change effects at biome scale by using a relatively simple but informative

variable. However, NDVI is only one of the numerous indexes that can be retrieved from satellite images to study vegetation dynamics or other phoenomena: H-graphs can be applied to different metrics and exploited for various purposes. Also, the possibility of visualizing the rates of change in the same graph as the absolute metric values has the advantage of easing the understanding of the magnitude of the change relative to a given index. Although satellite imageries and line charts are certainly easy and fast to be comprehended, in a multitemporal trend it is crucial to detect how a variable is changing, assessing both the entity and the velocity of the change. For this reason, the H-graphs should be considered particularly suited for this kind of analysis.

At the same time, this visualization technique emphasizes subtle changes, highlighting errors in the data and making them more prominent. However, to reduce the statistical noise it is sufficient to smooth the data points over a larger time period (Dorling, 2020). The implementation of smoothed data has the advantage of showing trends when they would not so easily be identifiable by simply plotting absolute values (Cleveland and Devlin, 1988).

3.6 Conclusions

This study illustrates how multitemporal remotely sensed data can be displayed in an innovative way, using the H-graphs initially proposed by Dorling (2020).

The reported trends are certainly in line with what emerges from the scientific literature: NDVI data displayed as H-graphs have given new insights on general vegetation responses to global change, in particular those related to biomass and photosynthetic activity. Both the analysed biomes showed an increasing trend in terms of their NDVI values, although there was some oscillation for Boreal Forests values. The analysis could be further improved by considering the variables influencing vegetation growth in the analytical framework, such as precipitation (Ding et al., 2007). Moreover, the availability of longer time series could better depict long-term trends.

H-graphs proved to be a practical tool to draw generalised trends relative to smoothed data, diminishing the influence of outliers. They can be exploited in many applications, such as showing trends of ecological global data in the educational literature or making early assumptions about time series trends, before deciding to expand the analytical framework or deepen an

analysis. Hence, we propose these graphs as a valid approach for visualizing trends in multitemporal ecological studies as well as other research fields.

Bibliography

Ahmad, F., 2012. Detection of change in vegetation cover using multi-spectral and multi-temporal information for district Sargodha, Pakistan. Sociedade & Natureza, 24(3). https://doi.org/10.1590/S1982-45132012000300014.

Ahmed, M.H., El Leithy, B.M., Thompson, J.R., Flower, R.J., Ramdani, M., Ayache, F. and Hassan, S.M., 2009. Application of remote sensing to site characterisation and environmental change analysis of North African coastal lagoons. Hydrobiologia, 622:147–171. https://doi.org/10.1007/s10750-008-9682-8.

Alessi, N., Wellstein, C., Rocchini, D., Midolo, G., Oeggl, K., Zerbe, S., 2021. Surface Tradeoffs and Elevational Shifts at the Largest Italian Glacier: A Thirty-Years Time Series of Remotely-Sensed Images. Remote Sensing, 13(1):134. https://doi.org/10.3390/rs13010134.

Anderson-Teixeira, K.J., Herrmann, V., Banbury Morgan, R., Bond-Lamberty, B., Cook-Patton, S.C., Ferson A.E., Muller-Landau, H.C., Wang, M.M.H., 2021. Carbon cycling in mature and regrowth forests globally. Environmental Research Letters, 16(5). https://doi.org/10.1088/1748-9326/abed01

Aybar, C., Wu, Q., Bautista, L., Yali, R., Barja, A., 2020. rgee: An R package for interacting with Google Earth Engine Journal of Open Source Software URL https://github.com/r-spatial/rgee/.

Apps, M.J., Kurz, W.A., Luxmoore, R.J., Nilsson, L.O., Sedjo, R.A., Schmidt, R., Simpson, L.G. and Vinson, T. S., 1993 Boreal forests and tundra. Water, Air and Soil Pollution, 70:39–53. https://doi.org/10.1007/BF01104987.

Benhin, J.K.A., 2006. Agriculture and Deforestation in the Tropics: A Critical Theoretical and Empirical Review. AMBIO: A Journal of the Human Environment, 35(1):9-16. https://doi.org/10.1579/0044-7447-35.1.9.

Binkley, D. and Fisher, R., 2012. Ecology and Management of Forest Soils. John Wiley & Sons Ltd. https://doi.org/10.1002/9781119455745.

Boyd, D.S., Foody, G.M., Ripple, W.J., 2002. Evaluation of approaches for forest cover estimation in the Pacific Northwest, USA, using remote sensing. Applied Geography, 22(4):375-392. https://doi.org/10.1016/S0143-6228(02)00048-6.

Campbell, J.B. and Wynne, R.H., 2011. Introduction to Remote Sensing, Fifth Edition. The Guildford Press.

Carey, C., Heyer, W.R., Wilkinson, J., Alford, R.A., Arntzen, J.W., Halliday, T., Hungerford, L., Lips, K.R., Middleton, E.M., Orchard, S.A., Rand, A.S., 2001. Amphibian Declines and Environmental Change: Use of Remote-Sensing Data to Identify Environmental Correlates. Conservation Biology, 15:903-913. https://doi.org/10.1046/j.1523-1739.2001.015004903.x.

Carlson, T. N. and Arthur, S. T., 2000. Impact of land use - land cover changes due to urbanization on surface microclimate and hydrology: A satellite perspective: Global and Planetary Change. Global and Planetary Change, 25(1):49-65. https://doi.org/10.1016/S0921-8181(00)00021-7.

Cazzolla Gatti, R., Reich, P. B., Gamarra, J. G., Crowther, T., Hui, C., Morera, A., ..., Liang, J., 2022. The number of tree species on Earth. Proceedings of the National Academy of Sciences, 119(6).

Cazzolla Gatti, R., Velichevskaya, A., Dudko, A., Fabbio, L., Notarnicola, C., 2021. The smokescreen of Russian protected areas. Science of The Total Environment, 785:147372.

Cazzolla Gatti, R., Velichevskaya, A., 2020. Certified "sustainable" palm oil took the place of endangered Bornean and Sumatran large mammals habitat and tropical forests in the last 30 years. Science of The Total Environment, 742:140712.

Celso H. L., S.J., Liana O., A., Alindomar L., S., Catherine T., A., Ricardo, D., Mikhaela A. J. S., P., Thales V., P., Rennan A., P., Luiz E. O. C., 2019. Fire Responses to the 2010 and 2015/2016 Amazonian Droughts. Frontiers in Earth Science, 7:2296-6463.

Chapungu, L., Nhamo, L., Cazzolla Gatti, R., Chitakira, M., 2020. Quantifying changes in plant species diversity in a savanna ecosystem through observed and remotely sensed data. Sustainability, 12(6):2345.

Chunming, S., Ying, L., Cong, G., Qiuhua, W., Lifu, S., 2020. Drought-Modulated Boreal Forest Fire Occurrence and Linkage with La Nina

Events in Altai Mountains, Northwest China. Atmosphere 11, 9:956. https://doi.org/10.3390/atmos11090956

Cleveland, W.S. and Devlin, S.J., 1988. Locally Weighted Regression: An Approach to Regression Analysis by Local Fitting. Journal of the American Statistical Association, 83(403):596-610. https://doi.org/10.1080/01621459.1988.10478639.

Cooksey R.W. (2020) Descriptive Statistics for Summarising Data. In: Illustrating Statistical Procedures: Finding Meaning in Quantitative Data. Springer, Singapore.

Das, M., Srinivasan, P. (2009), Anomaly detection and spatio-temporal analysis of global climate system. Proceedings of the Third International Workshop on Knowledge Discovery from Sensor Data, Association for Computing Machinery. https://doi.org/10.1145/1601966.1601989.

Dinerstein, E., Olson, D., Joshi, A., Vynne, C., Burgess, N. D., Wikramanayake, E., ..., Saleem, M., 2017. An Ecoregion-Based Approach to Protecting Half the Terrestrial Realm. BioScience, 67(6):534–545. https://doi.org/10.1093/biosci/bix014.

Ding, M., Zhang, Y., Liu, L., Zhang, W., Wang, Z., Bai, W., 2007. The relationship between NDVI and precipitation on the Tibetan Plateau. Journal of Geographical Sciences, 17:259–268. https://doi.org/10.1007/s11442-007-0259-7.

Dorling, D., 2020. Slowdown: The End of the Great Acceleration and Why It's Good for the Planet, the Economy, and Our Lives. Yale University Press.

Du Toit, S.H.C., Steyn, A.G.W., Stumpf, R.H., 1986. Graphical Exploratory Data Analysis. Springer-Verlag New York.

Fiore, N.M., Goulden, M.L., Czimczik, C.I., Pedron, S.A., Tayo, M.A., 2020. Do recent NDVI trends demonstrate boreal forest decline in Alaska?. Environmental Research Letters, 15:095007. https://doi.org/10.1088/1748-9326/ab9c4c.

Gibbs, H.K., Ruesch, A.S., Achard, F., Clayton, M.K., Holmgren, P., Ramankutty, N., Foley, J.A., 2010. Tropical forests were the primary sources of new agricultural land in the 1980s and 1990s, Proceedings of the National Academy of Sciences, 107(38):16732-16737. https://doi.org/10.1073/pnas.0910275107.

Gillman, L.N., Wright, S.D., Cusens, J., McBride, P.D., Malhi, Y., Whittaker, R.J., 2015. Latitude and productivity. Global Ecology and Biogeography, 24:107-117. https://doi.org/10.1111/geb.12245.

Gorelick, N., Hancher, M., Dixon, M., Ilyushchenko, S., Thau, D., Moore, R., 2017. Google Earth Engine: Planetary-scale geospatial analysis for everyone.

Hansen, M.C., Potapov, P.V., Moore, R., Hancher, M., Turubanova, S.A., Tyukavina, A., ..., Townshend, J.R.G. (2013), High-Resolution Global Maps of 21st-Century Forest Cover Change. Science, 342(6160):850–853. https://doi.org/10.1126/science.1244693.

Huang, S., Tang, L., Hupy, J.P., Wang, Y., Shao, G., 2020. A commentary review on the use of normalized difference vegetation index (NDVI) in the era of popular remote sensing. Journal of Forestry Research, https://doi:10.1007/s11676-020-01155-1.

IPCC, 2007. Climate Change 2007: The Physical Science Basis, Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change. Cambridge Univ Press, Cambridge, UK. https://www.ipcc.ch/.

Irteza, S.M., Nichol, J.E., Shi, W., Abbas, S., 2021. NDVI and Fluorescence Indicators of Seasonal and Structural Changes in a Tropical Forest Succession. Earth Systems and Environment, 5:127–133. https://doi.org/10.1007/s41748-020-00175-5.

Khorram, S., Koch, F.H., van der Wiele, C.F., Nelson, S.A.C., 2012. Remote Sensing. Springer.

Kuuluvainen, T., Gauthier, S., 2018. Young and old forest in the boreal: critical stages of ecosystem dynamics and management under global change. For. Ecosyst., 5:26. https://doi.org/10.1186/s40663-018-0142-2

Frelich, L.E., Montgomery, R.A., Reich, P.B., 2021. Seven Ways a Warming Climate Can Kill the Southern Boreal Forest. Forests 12:560. https://doi.org/10.3390/ f12050560

Lewis, S. L., 2006. Tropical forests and the changing earth system. Philosophical Transactions of the Royal Society B: Biological Sciences, 361(1465):195–210. https://doi.org/10.1098/rstb.2005.1711.

Lewis, S.L., Edwards, D.P., Galbraith D., 2015. Increasing human dominance of tropical forests. Science, 349(6250):827-832. https://doi.org/10.1126/science.aaa9932.

Malhi, Y., Grace, J., 2000. Tropical forests and atmosphericcarbon dioxide. Trends in Ecology and Evolution, 15:332–337. https://doi.org/10.1016/S0169-5347(00)01906-6.

Michaelian, M., Hogg, E.H., Hall, R.J. and Arsenault, E., 2011. Massive mortality of aspen following severe drought along the southern edge of the Canadian boreal forest. Global Change Biology, 17:2084-2094. https://doi.org/10.1111/j.1365-2486.2010.02357.x.

Murphy, B.P., Bowman, D.M., 2012. What controls the distribution of tropical forest and savanna?. Ecology Letters, 15:748-758. https://doi.org/10.1111/j.1461-0248.2012.01771.x.

Myneni, R. B., Hall, F. G., Sellers, P. J., Marshak, A. L., 1995. The interpretation of spectral vegetation indexes. IEEE Transactions on Geoscience and Remote Sensing, 33(2):481-486. https://doi.org/10.1109/TGRS.1995.8746029.

Norby, R.J., Wullschleger, S.D., Gunderson, C.A., Johnson, D.W., Ceulemans, R., 1999. Tree responses to rising CO2 in field experiments: implications for the future forest. Plant, Cell & Environment, 22:683-714. https://doi.org/10.1046/j.1365-3040.1999.00391.x.

Paruelo, J.M., Garbulsky, M.F., Guerschman, J.P. and Jobbágy, E.G., 2004. Two decades of Normalized Difference Vegetation Index changes in South America: identifying the imprint of global change. International Journal of Remote Sensing, 25(14):2793–2806. https://doi.org/10.1080/01431160310001619526.

R Core Team, 2021. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

Ravindranath, N.H., Sukumar, R., 1998. Climate Change and Tropical Forests in India. In: Markham A. (eds) Potential Impacts of Climate Change on Tropical Forest Ecosystems. Springer, Dordrecht.

Risser, G.P., Clarke, J.N., Dale, V., Field, C., Lewis, M.W. Jr., Lubchenco, J., ..., Ustin, S., 2000. Global Change Ecosystems Research (Chapter: Definitions

and Implications of Global Change). National Academy Press Washington DC.

Rocchini, D., Salvatori, N., Beierkuhnlein, C., Chiarucci, A., de Boissieu, F., Förster, M., ..., Féret, J.B., 2021. From local spectral species to global spectral communities: A benchmark for ecosystem diversity estimate by remote sensing. Ecological Informatics, 61:101195. https://doi.org/10.1016/j.ecoinf.2020.101195.

Sage, R.F., Kubien, D.S. 2007. The temperature response of C3 and C4 photosynthesis. Plant, Cell and Environment, 30:1086-1106. https://doi.org/10.1111/j.1365-3040.2007.01682.x.

Santoro, M., Cartus, O., Carvalhais, N., Rozendaal, D., Avitabile, V., Araza, A., ..., Willcock, S., 2021. The global forest above-ground biomass pool for 2010 estimated from high-resolution satellite observations. Earth System Science Data, 13(8):3927-3950.

Sizer, N., Petersen, R., Anderson, J., Hansen, M., 2015. Tree Cover Loss Spikes in Russia and Canada, Remains High Globally. https://www.wri.org/insights/tree-cover-loss-spikes-russia-and-canada-remains-high-globally (accessed 20/05/2022).

Shugart, H.H., Leemans and R., Bonan, G.B., 1992. A Systems Analysis of the Global Boreal Forest, Cambridge University Press.

Soja, A.J., Tchebakova, N.M., French, N.H.F., Flannigan, M.D., Shugart, H.H., Stocks, B.J., Sukhinin, A.I., Parfenova, E.I., Chapin, F.S., Stackhouse, P.W., 2007. Climate-induced boreal forest change: Predictions versus current observations. Global and Planetary Change, 56(3–4):274-296. https://doi.org/10.1016/j.gloplacha.2006.07.028.

Steffen, W., Sanderson, R.A., Tyson, P.D., Jäger, J., Matson, P.A., Moore III, B., Oldfield, F., Richardson, K., Schellnhuber, H.J., Turner, B.L., Wasson, R.J., 2004. Global Change and the Earth System. A Planet Under Pressure. Springer.

Taylor, P.G., Cleveland, C.C., Wieder, W.R., Sullivan, B.W., Doughty, C.E., Dobrowski, S.Z., Townsend, A.R., 2017. Temperature and rainfall interact to control carbon cycling in tropical forests. Ecology Letters, 20:779-788. https://doi.org/10.1111/ele.12765.

Thiffault, E., 2019. Chapter 5 - Boreal forests and soils, Editor(s): Busse, M., Giardina, C.P., Morris, D.M., Page-Dumroese, D.S. Developments in Soil Science, Elsevier, 36:59-82. https://doi.org/10.1016/B978-0-444-63998-1.00005-7.

Windsor, D.M., 1990. Climate and Moisture Variability in a Tropical Forest : long-term records from Barro Colorado Island, Panamá. Smithsonian Institution contributions to the earth sciences, 29: 1-145. https://doi.org/10.5479/si.00810274.29.1.

Chapter 4

Mapping diversity: how do heterogeneity patterns change in space and time?

To be published as:
Thouverai, E., Bazzichetto M., Sperandii M.G., Apruzzese M., Merelli P., and Rocchini D. Mapping diversity: how do heterogeneity patterns change in space and time?

4.1 Abstract

Human activities have several effects on the biomes of Earth, such as land fragmentation, deforestation, pollution, anthropization of natural landscapes, and alterations in the functioning of ecological systems. Remote sensing is an important tool for assessing ecosystem changes because it allows the collection of long time series of data that can be used to assess land cover and vegetation state of a chosen area. In this study, we employed Jaccard index, a beta diversity metric to produce a map that shows the variation of spatial heterogeneity over time. The moving window technique was used. Each chunk was considered as a plot and the pixels as species, and the selected metric was calculated for every image of the time series. This approach enabled us to calculate a value of beta diversity for every pixel, providing information about the variation of spatial heterogeneity over time. We chose Italy as the study area to test this workflow, specifically retrieving two landcover maps for years 1990 and 2018 from Corine Land Cover (https://land.copernicus.eu/pan-european/corine-land-cover). The resulting maps are intuitive, easy to interpret, and provide information about both the spatial pattern and the change in time of land cover.

4.2 Introduction

Landscapes serve as a manifestation of the intricate interplay between the natural environment and human activities (Antrop , 1998). These diverse terrains are delineated by distinct land cover classes, including vegetation, inland water, bare soil, and human infrastructure (Gómez et al. , 2016). Recognized over the past 15 years, changes in land cover and land-use practices have emerged as significant global environmental transformations in their own right (Turner , 2002). Moreover, these changes are intricately linked with various environmental concerns, such as climate change, carbon cycle dynamics, biodiversity loss, agricultural sustainability, and the availability of safe drinking water (Lepers et al. , 2005). The measure of non-uniformity in land cover is called heterogeneity (Stein et al. , 2014).

Heterogeneity plays a crucial role as a driver for biodiversity (Ettema and Wardle , 2002). It exerts influence over various ecological processes and functions, impacting patterns and changes in species diversity (Rocchini et al., 2018), metapopulation dynamics (Fahrig , 2007), population connectivity (Malanson and Cramer , 1999), and gene flow (Lozier et al., 2013). This

variability increases the availability of niche space, provides refuges, and offers opportunities for isolation and adaptation. Consequently, it enhances species coexistence, persistence, and diversification (Stein et al. , 2014; Tews et al., 2004).

Remote sensing emerges as a valuable tool for deciphering heterogeneity patterns across both space and time, as it allows to collect a huge amount of data at different temporal and spatial resolutions (Campbell and Wynne , 2011). This technology allows the estimation of several metrics from different compartments of the Earth system, encompassing tasks such as the generation of land cover maps (Khorram et al. , 2012). Metrics derived from remote sensing products provide a means to quantitatively discern and track temporal changes in heterogeneity and its correlation with biodiversity dynamics (Tuanmu and Jetz , 2015). In this context, the R platform stands out as one of the most widely employed statistical and computational environments in ecology, a reputation bolstered by the ongoing development of pertinent packages. Notably, the R package `rasterdiv`, as introduced by Marcantonio et al. (2021); Rocchini et al. (2021); Thouverai et al. (2021), facilitates the calculation of a wide array of indices for producing maps representing spatial heterogeneity using satellite imagery, but lacking the temporal component.

Considering the pivotal role of heterogeneity as a driver of biodiversity and the escalating rate of landscape change, the significance of developing methods to monitor its transformation, both in space and in time, is becoming increasingly imperative (Bradshaw and Fortin , 2014). In this context, beta diversity metrics, traditionally employed to discern species turnover between two communities (González-Megías and Sánchez-Piñero , 2011), can be leveraged to calculate shifts in heterogeneity.

The aim of this work is to introduce a novel method for generating maps that illustrate spatio-temporal heterogeneity. This approach aims to facilitate the monitoring of environmental heterogeneity, encompassing both spatial and temporal dimensions.

4.3 Materials and Methods

4.3.1 Data

Two landcover maps, corresponding to the years 1990 and 2018, were acquired using Google Earth Engine (GEE) to generate a map illustrating changes in heterogeneity. The selected landcover datasets were sourced from the

CORINE (Coordination of Information on the Environment) database. Italy was chosen as the primary study area (Fig. 4.1), owing to its distinctive and diverse landscape.

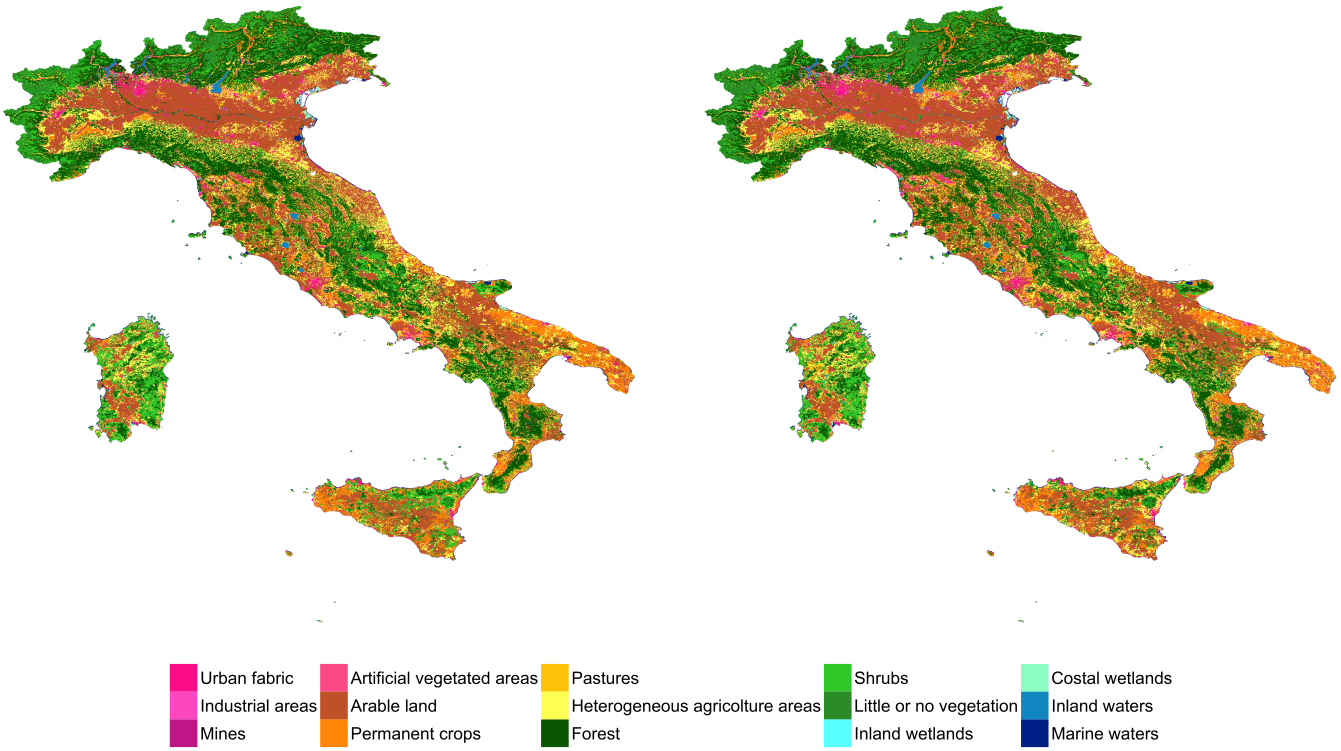

FIGURE 4.1: CORINE landcover second level map of Italy of years 1990 (left) and 2018 (right).

Additionally, a regional examination was conducted on Sicily (Fig. 4.2) to assess the algorithm's performance at a finer scale.

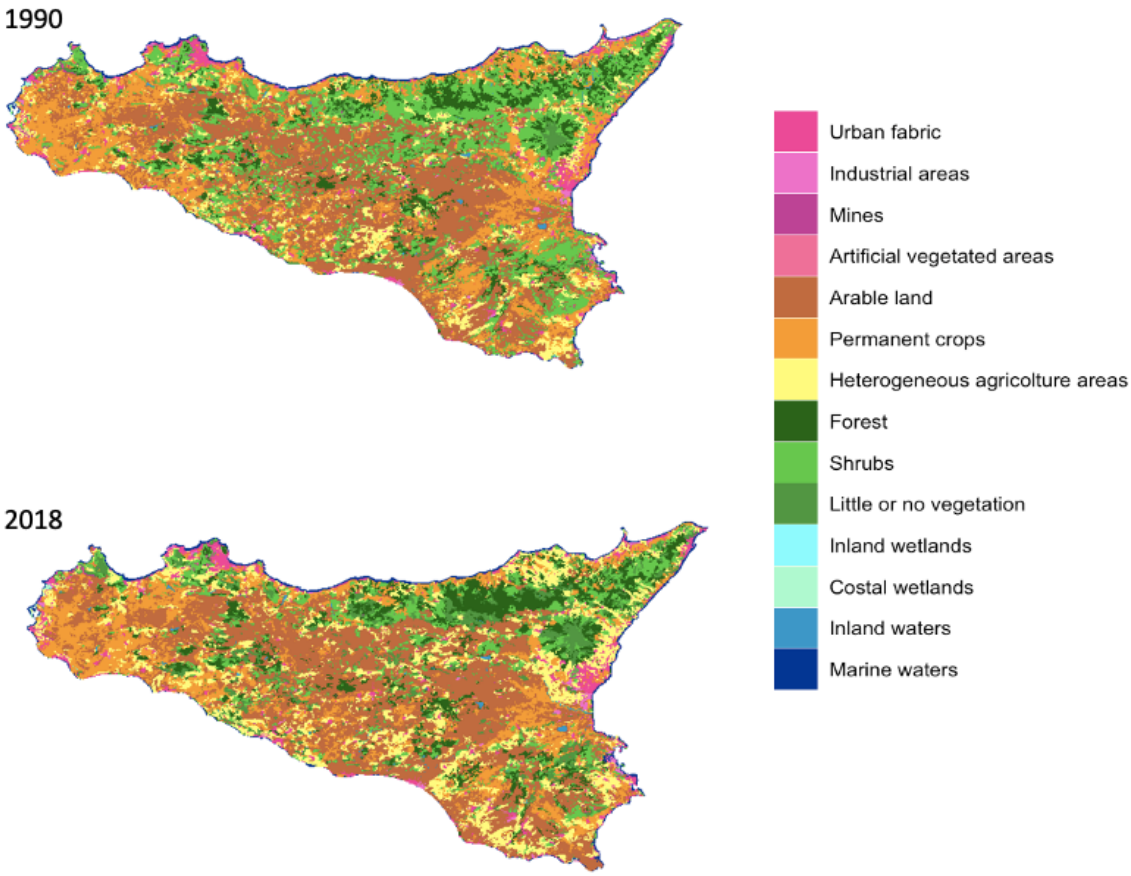

FIGURE 4.2: CORINE landcover second level map of Sicily of years 1990 and 2018.

4.3.2 The Algorithm

The temporal variation in environmental heterogeneity was assessed using the Jaccard index (Jaccard, 1901), a beta diversity metric that quantifies the dissimilarity between two sites:

$$Jaccard = \frac{b + c}{a + b + c} \tag{4.1}$$

Here, a represents the number of common species, while b and c denote the unique species found in the two compared sites.

The classes represented in the two landcover images were utilized as proxies for species. Employing the moving window technique, a specific area around each pixel was taken into account. The Jaccard index was then computed for each window, quantifying the difference in landcover composition between the two images.

A moving window of size 3x3 was used for the analysis in this paper. The choice of the moving window size is a critical factor that can significantly influence the results of the analysis. To assess its impact, a focused test was conducted on the province of Siracusa in Sicily, comparing the outcomes generated by four different moving window sizes.

4.4 Results

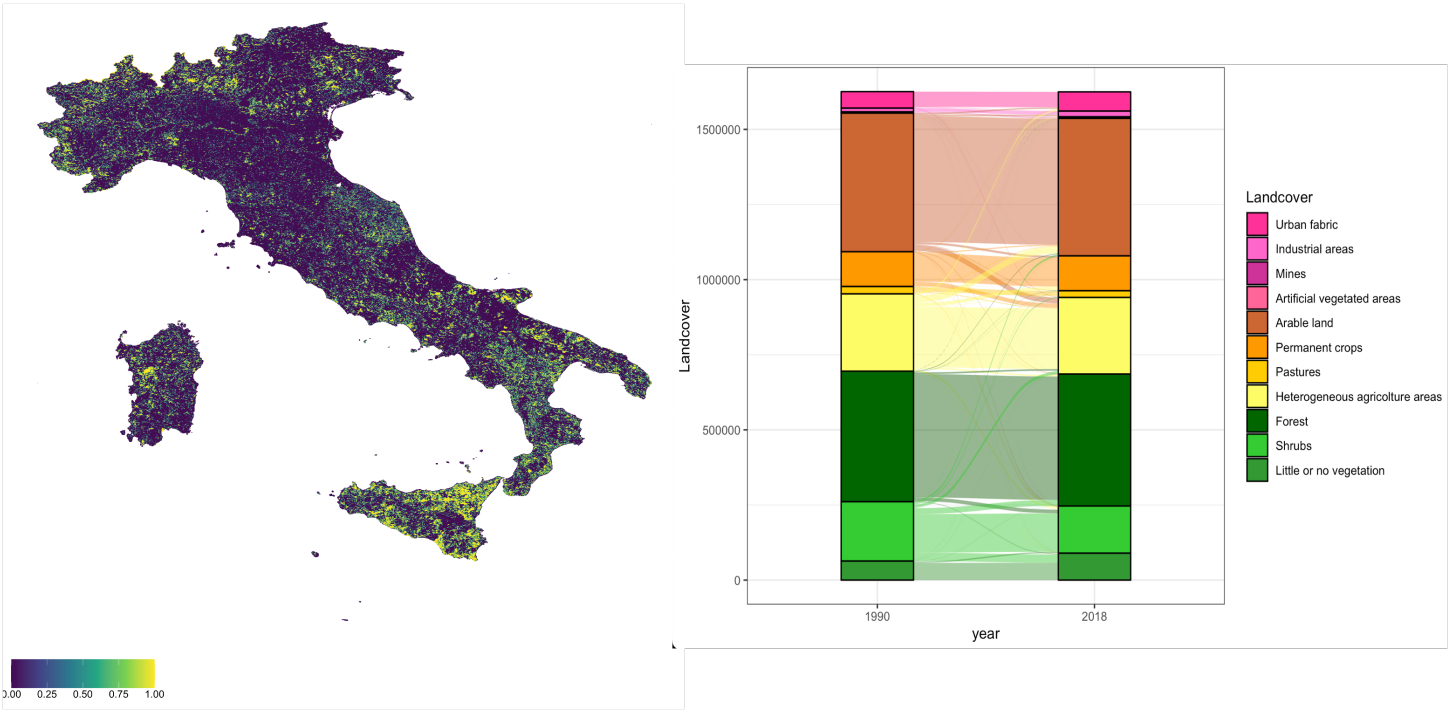

FIGURE 4.3: On the left, temporal change map of Italy calculated using the Jaccard index. On the right, the correspondent alluvial plot representing the change in class of the pixels in Sicily from 1990 to 2018.

Figure 4.3 provides a comprehensive visualization of the landcover degree of change in Italy from 1990 to 2018, employing the Jaccard Index. The map on the left illustrates the spatial distribution of change, with areas exhibiting higher degrees of change highlighted in light green/yellow. Notably, South Italy appears to undergo more pronounced changes. The alluvial plot on the right complements the map by detailing which specific land cover classes have experienced alterations. This combined representation offers a nuanced understanding of the temporal dynamics, emphasizing the regions and land cover classes contributing significantly to the observed changes in the landscape over the specified time period.

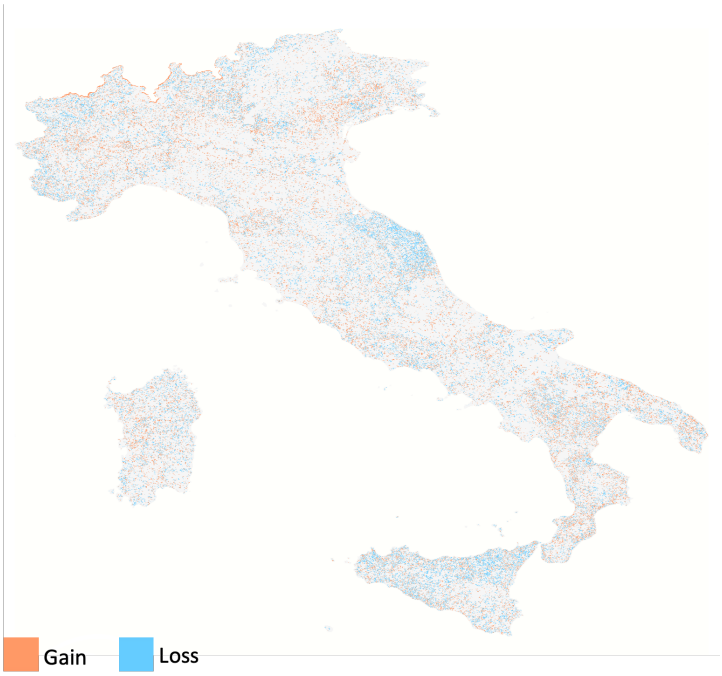

FIGURE 4.4: Map of Italy representing the Jaccard index gain and loss of landcover classes between years 1990 and 2018.

In Figure 4.4, the representation highlights regions where there is either a gain or loss of land cover classes, signifying an increase or decrease in heterogeneity, respectively. This visualization provides valuable insights into the dynamics of land cover changes, emphasizing the areas where the landscape has become more diverse (gain) or more uniform (loss) over the specified period.

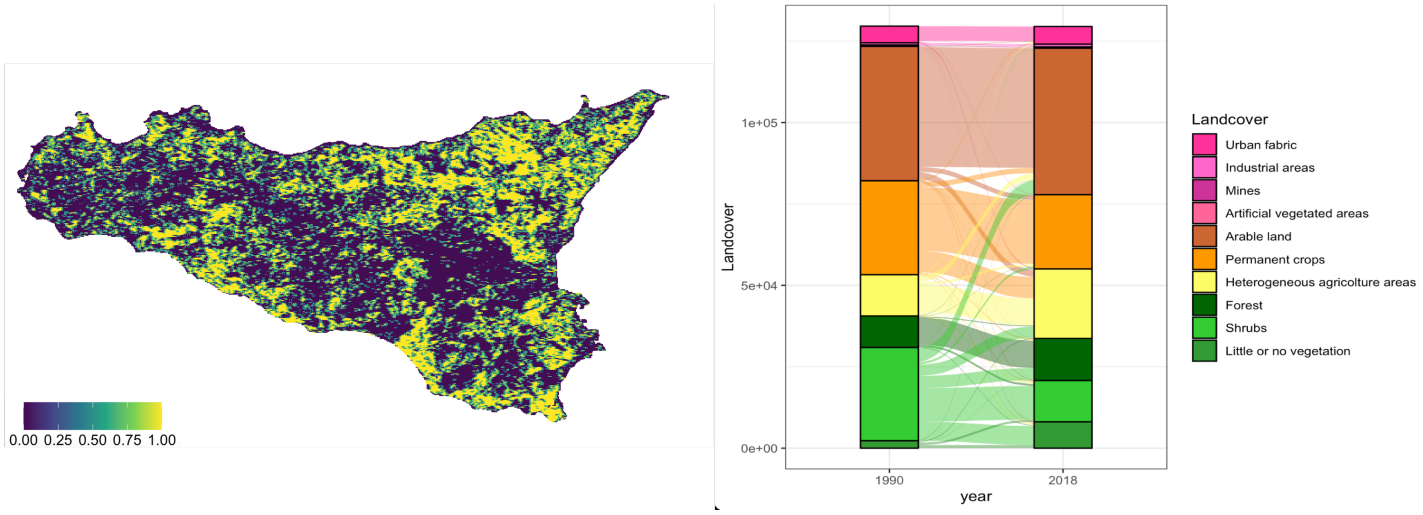

FIGURE 4.5: On the left, temporal change map calculated using the Jaccard index of Sicily. On the right, the correspondent alluvial plot representing the change in class of the pixels in Sicily from 1990 to 2018.

In Figure 4.5, the focus on Sicily is presented, emphasizing the hotspots of land cover changes within the region. The left side illustrates the Jaccard index, highlighting areas with significant alterations in land cover over time. On the right, the corresponding alluvial plot provides insights into the specific classes that underwent changes, offering a detailed breakdown of the transitions in land cover types.

4.5 Discussion

In this paper, we introduced a novel approach that harnesses beta diversity measures to generate maps depicting temporal heterogeneity based on land cover maps. The effectiveness of the method was assessed using CORINE land cover maps of Italy for the years 1990 and 2018.

The temporal beta map of Italy (Fig. 4.3) reveals a notable concentration of areas experiencing landscape changes, particularly in the North-West Alps, South Italy, and Sicily. The accompanying alluvial plot aligns with the findings of Malandra et al. (2018), indicating an increase in forest cover and a corresponding decrease in grasslands. The plot also depicts a discernible transition between different agricultural practices, likely a primary driver for the detected land cover changes. Furthermore, Figure 4.4 illustrates a general decrease in land cover classes, suggesting a reduction in overall heterogeneity across Italy.

In the Sicily case study (Fig. 4.5), the region exhibiting the most significant changes, a notable reduction in grassland areas is evident, giving way to an increase in forested areas, crops, and regions characterized by little or no vegetation. Additionally, a discernible shift is observed from arable lands and permanent crops to areas exhibiting a more heterogeneous agricultural landscape.

The novel method, utilizing beta diversity measures to depict temporal heterogeneity, has demonstrated notable efficiency in capturing areas undergoing substantial changes. The application of this method effectively highlights regions characterized by pronounced temporal shifts, providing a clear visual representation of dynamic landscape alterations over time. Moreover, the computed values of the beta diversity index offer a quantitative measure that can be seamlessly integrated into modeling frameworks. This dual capability of effectively highlighting areas with significant change and providing a quantifiable index enhances the method's utility in comprehensively analyzing and understanding temporal landscape dynamics.

In this study, temporal maps for Italy and Sicily were computed using a 3x3 moving window. The selection of an appropriate window size is a crucial aspect in remote sensing metric computation, primarily influenced by the image scale and spectral bands used in the calculation (Ozkan and Demirel , 2021). Figure 4.6 illustrates the computation of the Jaccard index for the city of Siracusa in 1990 and 2018, employing four different moving window sizes.

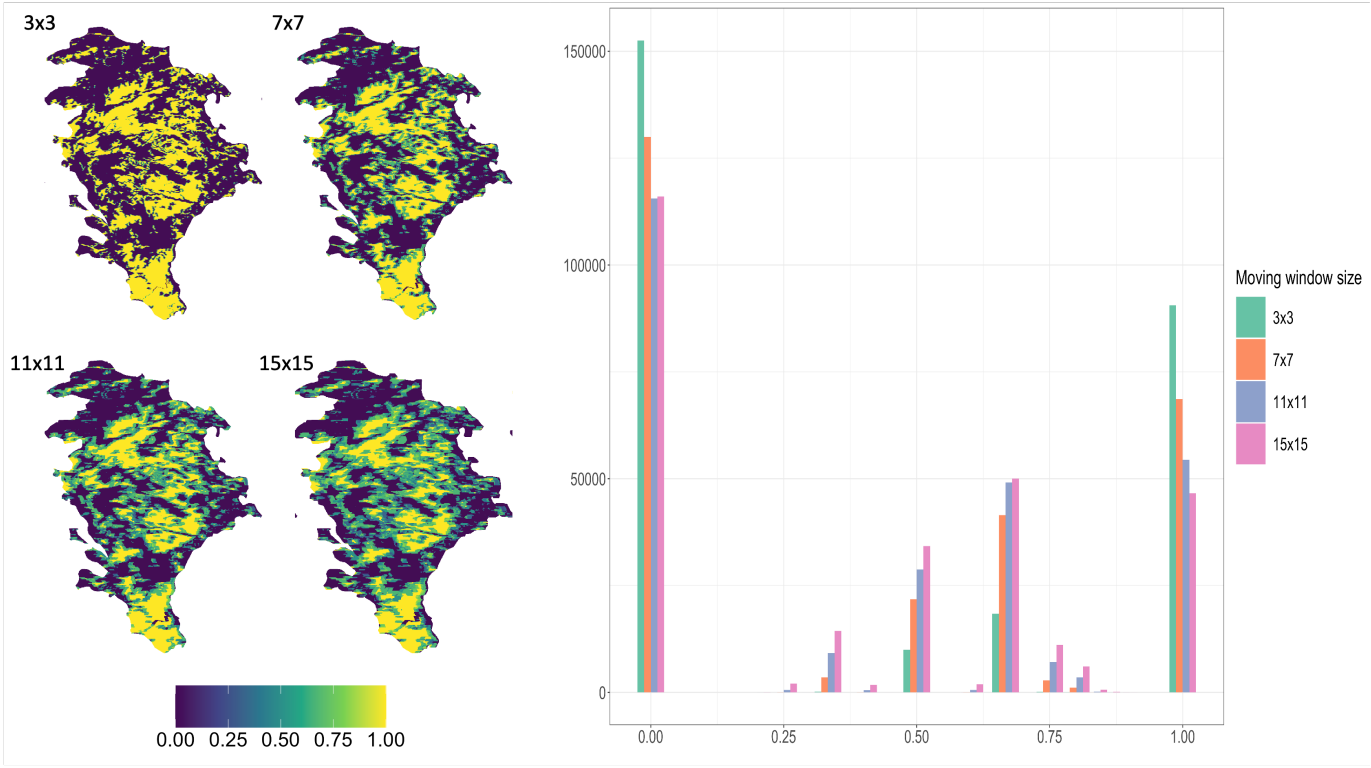

FIGURE 4.6: On the left, four temporal change maps calculated using the Jaccard index of Siracusa province (years 1990 and 2018). The temporal maps were calculated with 3x3, 7x7, 11x11 and 15x15 moving window sizes. On the right, a histogram with the pixel values frequencies of the maps.

The histogram depicting the frequency of pixel values reveals a reduction in extreme values of beta diversity (0 and 1) and an increase in intermediate values as the moving window size expands. This trend is also observable in the temporal maps, where green areas grow larger with an increase in the moving window size. The 3x3 moving window encompasses a smaller area, with fewer pixels, elevating the likelihood of either none or all of them being different in the two years. This accounts for the presence of more extreme values. However, a larger moving window includes more pixels, increasing the probability of discovering new land cover classes that underwent changes. Consequently, a zero Jaccard index in smaller windows may transform into

an intermediate value. Similarly, with a Jaccard index of one, enlarging the moving window can result in the encounter of more homogeneous areas, reducing the number of distinct land cover classes between the two years, resulting in the transition from one to an intermediate value.

At this juncture, it can be inferred that the expansion of the moving window introduces more disturbance, illustrated by the thicker green border surrounding the yellow areas in the temporal maps of the Siracusa province. Nevertheless, it's important to consider that the study area has experienced substantial changes between the years 1990 and 2018, evident in the presence of several yellow areas. In the context of a larger study area or an area undergoing more modest changes, a larger moving window may prove advantageous in detecting subtle changes that might go unnoticed with a smaller window. Hence, the selection of the moving window size should be made judiciously, taking into account the characteristics of the study area and the properties of the analyzed images. It is advisable to experiment with and test different solutions to determine the most suitable moving window size for the specific context.

In conclusion, the dual ability of our approach to effectively highlight significant changes and provide a quantifiable index offers a valuable contribution to the comprehensive understanding and analysis of temporal landscape dynamics. This study underscores the need for a thoughtful consideration of moving window size, particularly in areas with distinct temporal dynamics, to enhance the reliability and applicability of the proposed method. Additionally, it emphasizes the necessity for further development to incorporate more than two images in the analysis. This improvement would prevent the loss of valuable information regarding the temporal trend, enabling a more nuanced and accurate representation of temporal landscape dynamics.

Bibliography

Antrop, M. (1998) Landscape change: Plan or chaos?. Landscape and Urban Planning, 41(3-4), 155–161. doi:10.1016/s0169-2046(98)00068-1

Bradshaw, G.A., Fortin, M.J. (2000). Landscape Heterogeneity Effects on Scaling and Monitoring Large Areas Using Remote Sensing Data. Annals of GIS, 6(1), 61–68. doi:10.1080/10824000009480534

Campbell, J.B. and Wynne, R.H., 2011. Introduction to Remote Sensing, Fifth Edition. The Guildford Press.

Ettema, C.H., Wardle, D.A. (2002). Spatial soil ecology. Trends in Ecology & Evolution, 17, 177-183.

Fahrig, L. (2007). Landscape heterogeneity and metapopulation dynamics. In Wu, J., & Hobbs, R. (Eds.), *Key Topics in Landscape Ecology*(pp. 78-91). Cambridge, UK: Cambridge University Press.

Gómez, C., White, J.C., Wulder, M.A. (2016) Optical remotely sensed time series data for land cover classification: A review. ISPRS Journal of Photogrammetry and Remote Sensing, 116, 55–72. doi:10.1016/j.isprsjprs.2016.03.008

González-Megías, A., Gómez, J. M., Sánchez-Piñero, F. (2011). Spatio-temporal change in the relationship between habitat heterogeneity and species diversity. Acta Oecologica, 37(3), 179-186.

Jaccard P. (1901), Étude comparative de la distribution florale dans une portion des Alpes et des Jura. Bulletin de la Société vaudoise des sciences naturelles 37, 547–579.

Khorram, S., Koch, F.H., van der Wiele, C.F., Nelson, S.A.C., 2012. Remote Sensing. Springer.

Lepers, E., Lambin, E. F., Janetos, A. C., Defries, R., Achard, F., Ramankutty, N., Scholes, R.J. (2005). A Synthesis of Information on Rapid Land-cover Change for the Period 1981–2000, BioScience, 55(2), 115–124. doi:10.1641/0006-3568(2005)055[0115:ASOIOR]2.0.CO;2

Lozier, J.D., Strange, J.P., Koch, J.B. (2013). Landscape heterogeneity predicts gene flow in a widespread polymorphic bumble bee, *Bombus bifarius* (Hymenoptera: Apidae). Conservation Genetics, 14, 1099-1110.

Malandra, F., Vitali, A., Urbinati, C. and Garbarino, M., (2018) 70 years of land use/land cover changes in the Apennines (Italy): a meta-analysis. Forests, 9(9), p.551.

Malanson, G.P., Cramer, B.E. (1999). Landscape heterogeneity, connectivity, and critical landscapes for conservation. Diversity and Distributions, 5, 27-39.

Marcantonio, M., Iannacito, M., Thouverai, E., Da Re, D., Tattoni, C., Bacaro, G., Vicario, S., Ricotta, C. & Rocchini, D. (2021). rasterdiv: Diversity Indices for Numerical Matrices. R package version 0.2-3. https://CRAN.R-project.org/package=rasterdiv

Ozkan, U. Y., and Demirel, T. (2021). The influence of window size on remote sensing-based prediction of forest structural variables. Ecological Processes, 10, 1-11.

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., Wegmann, M., Nagendra, H. (2018). Measuring β-diversity by remote sensing: a challenge for biodiversity monitoring. Methods in Ecology and Evolution, 9, 1787-1798.

Rocchini, D., Thouverai, E., Marcantonio, M., Iannacito, M., Da Re, D., Torresani, M., et al. (2021). rasterdiv - an Information Theory tailored R package for measuring ecosystem heterogeneity from space: to the origin and back. *Methods in Ecology and Evolution*, 12, 1093- 1102.

Stein, A., Gerstner, K., and Kreft, H.: Environmental heterogeneity as a universal driver of species richness across taxa, biomes and spatial scales, Ecology Letters, 17(7), 866 880, 2014.

Tews, J., Brose, U., Grimm, V., Tielbörger, K., Wichmann, M., Schwager, M. and Jeltsch, F. (2004). Animal species diversity driven by habitat heterogeneity/diversity: the importance of keystone structures. Journal of Biogeography, 31, 79–92.

Thouverai, E., Marcantonio, M., Bacaro, G., Da Re, D., Iannacito, M., Marchetto, E., Ricotta, C., et al. (2021). Measuring diversity from space: a global view of the free and open source rasterdiv R package under a coding perspective. *Community Ecology*, 22, 1-11.

Tuanmu, M.N., and Jetz, W.: A global, remote sensing-based characterization of terrestrial habitat heterogeneity for biodiversity and ecosystem modelling. Global Ecology and Biogeography, 24(11), 1329-1339, 2015.

Turner BL II . 2002. Toward integrated land-change science: Advances in 1.5 decades of sustained international research on land-use and land-cover change. Pages 21–26 in Steffen W Jäger J Carson DJ Bradhsaw C eds. Challenges of a Changing Earth. Berlin: Springer.

Ringraziamenti

Questi tre anni sono stati un'avventura grazie alla quale ho imparato un sacco di cose nuove e sono cresciuta e maturata sia come ricercatrice sia come persona, superando tutte le sfide incontrate. Ho anche avuto l'opportunità di viaggiare tra congressi e periodo all'estero, e di conoscere tante persone meravigliose.

Per primo ci tengo a ringraziare il mio supervisore, il Prof. Duccio Rocchini per aver creduto in me e avermi dato l'occasione di intraprendere questo percorso e di mettermi alla prova. Ringrazio anche tutte le persone che hanno fatto parte del BIOME Lab in questi tre anni, in particolare i giovani biomers per aver condiviso con me questo percorso e avermi sempre fatta sentire in famiglia.

Tra i biomers, un ringraziamento speciale va alla mia amica Francesca Bottegoni, per la sua capacità di mettermi sempre di buon umore, le sue parole di incoraggiamento e soprattutto per i discorsi fatti al momento giusto. Un altro ringraziamento speciale va a Michele Di Musciano, Piero Zannini, Nicola Alessi e Marco Cervellini, per avermi fatto da mentori durante questo percorso trovando sempre il tempo di ascoltarmi ed aiutarmi, e per i preziosi consigli che mi hanno dato da Zapap di fronte ad una birra.

Un grazie speciale va anche al gruppo di ricerca che mi ha accolta a Praga, in particolare al Prof. Vítězslav Moudrý e alla Michela Perrone amica, collega e coinquilina ed al suo piccolo cane.

A seguire grazie alle mie coinquiline di casa Carducci a Bologna, per avermi fatta sempre sentire a casa. In particolare ringrazio la mia coinqui forever Elisa Arigliano, che mi è stata vicina e mi ha sopportata non solo durante questo dottorato, ma durante tutto il mio percorso universitario.

Ringrazio anche tutte le mie amiche sparse per l'Italia e per l'Europa che non mi hanno mai lasciata sola nei momenti difficili, sempre a portata di telefono per lunghe chiacchierate.

Grazie poi alla mia famiglia, in particolare ai miei pilastri, ovvero i miei genitori che non hanno mai smesso di credere in me.

Gli ultimi due ringraziamenti spettano a mia nonna per avermi sempre viziata e a mio nonno a cui è dedicata la tesi, per avermi fatto amare la natura

e senza il quale non avrei intrapreso quest' avventura.