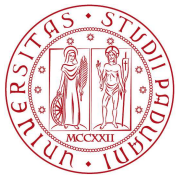


Università
degli Studi
di Padova

Sede amministrativa: Università degli Studi di Padova

Dipartimento di Matematica "Tullio Levi-Civita"

CORSO DI DOTTORATO DI RICERCA IN SCIENZE MATEMATICHE

CURRICOLO: MATEMATICA

CICLO XXX

Mathematical Modelling and Statistics of BIODIVERSITY

Coordinatore: Ch.mo Prof. Martino Bardi

Supervisore: Ch.mo Prof. Marco Favretti

Dottorando: Anna Tovo

To my family

Riassunto

La vita sulla Terra è caratterizzata da una straordinaria varietà di forme viventi in continua evoluzione per meglio adattarsi all'ambiente circostante e strettamente connesse le une alle altre. Oggigiorno, grazie all'enorme quantità di dati a disposizione, è possibile investigare a fondo su diversi sistemi viventi.
La presente tesi è il risultato di un percorso attraverso i complessi pattern della teoria ecologica. In essa trattiamo sia modelli teorici sia problematiche legate all'analisi dei dati, come anche le connessioni tra loro, tutto all'interno di un contesto matematico. Centro d'interesse sono i diversi aspetti della biodiversità di un ecosistema, termine con il quale indichiamo la varietà delle sue specie. In particolare, vogliamo investigare il modo in cui le diverse specie interagiscono le une con le altre e come, da queste connessioni, possano originarsi dei pattern macro-ecologici ricorrenti. Infatti, nonostante la loro apparente diversità e complessità, è oggi evidente che i sistemi ecologici mostrano comportamenti simili. Questo fatto suggerisce che tali sistemi evolvono secondo un meccanismo comune, insensibile ai dettagli del sistema su cui agisce. Di conseguenza, si apre la strada allo sviluppo di modelli teorici che siano abbastanza complessi da riuscire a spiegare tali fenomeni, ma che al contempo non contengano più dettagli di quelli necessari a riprodurli.
La prima parte della tesi è dedicata all'esplorazione dei fondamenti della teoria dei processi di punto, uno strumento matematico molto utile quando si va ad investigare dataset contenenti posizioni di punti nello spazio. In particolare, essendo i nostri database relativi a coordinate di alberi appartenenti a specie diverse, ci concentreremo sul cosiddetto *processo sovrapposto* e sulle sue statistiche di primo e secondo ordine. Poi studieremo un algoritmo che permette di ottenere informazioni sull'intensità di un processo di punto, capace al contempo di ridurre le fluttuazioni di campionamento e di rivelare caratteristiche importanti di un pattern spaziale, come l'anisotropia ed il clustering. Infine, esploreremo in dettaglio le nozioni di diversità e similarità e i vari indici proposti in letteratura per misurarle. In particolare, studieremo come inserire queste nozioni nel contesto dei processi di punto. L'obiettivo è quello di trovare una relazione analitica per il decadimento di similarità tra due regioni in funzione della distanza tra esse estendendo la nozione classica dell'indice di Sørensen in modo da incorporare informazioni spaziali.
Nella seconda parte della tesi, affronteremo il problema di inferire la biodiversità totale di un ecosistema avendo a disposizione solo alcuni suoi campioni. In particolare, proporremo un nuovo metodo che, sfruttando la proprietà di invarianza di scala della distribuzione binomiale negativa, permette di avere stime accurate e robuste. Testandolo sia su foreste artificiali che reali, mostreremo che il metodo è più affidabile rispetto ad altri proposti in letteratura.

Abstract

Life on Earth is characterised by an amazing variety of living forms which are in continuous evolution to better adapt to the surrounding environment and highly connected one to the other. A deep investigation of different living systems has recently been favoured by the huge quantity of data nowadays available.
The present thesis is the final result of a journey through complex patterns in theoretical ecology. We study both models and issues in data analysis as well as the connections between them within a mathematical framework. In particular, we explore the different aspects of the *biodiversity* of an ecosystem, referring with this term to the variety of its *species*. Our interest is to investigate how these species interact with each other and with the surrounding environment and how these connections can structure recurrent macro-ecological patterns. Indeed, despite their diversity and complexity, it is straightforward that ecological systems share similar behaviours. This fact suggests that such systems are driven by a common mechanism, which is insensitive to the details of the systems on which it acts. A theoretical understanding is therefore possible through the development of mathematical models rich enough to reproduce the investigated patterns, but containing only the essential ingredients able to originate them.
In the first part of the present thesis, we explore the fundamentals of spatial point process theory, a powerful mathematical tool to model data in the form of sets of spatial locations of points. In particular, since our datasets usually consist of information on trees belonging to different species, we focus on the so-called *superposed* process and its first and second-order statistics. We then study an algorithm to infer the intensity function of a point process which is capable to reduce sampling fluctuations and to capture relevant spatial characteristics of a spatial pattern, as space anisotropy and clustering. Finally, we explore in details the notions of ecological diversity and similarity and some of the most popular indexes used to measure them. In particular, we study how to insert them in the context of point processes' theory. Our aim is at finding an analytical relation for the decay of similarity between two regions of a landscape as a function of the distance between them, by extending the classic notion of *Sørensen's index* to incorporate spatial information.
In the second part of the thesis, we tackle the problem of inferring the total biodiversity of an ecosystem when only scattered samples are observed. In particular, we propose a novel upscaling method which, by exploiting the scaling invariance property of the negative binomial distribution, generates accurate and robust predictions. We test it on both computer-generated and real forests and we show that it outperforms other methods previously proposed in literature.

Introduction

"Thus, from the war of nature, from famine and death, the most exalted object which we are capable of conceiving, namely, the production of the higher animals, directly follows. There is grandeur in this view of life, with its several powers, having been originally breathed into a few forms or into one; and that, whilst this planet has gone cycling on according to the fixed law of gravity, from so simple a beginning endless forms most beautiful and most wonderful have been, and are being, evolved." (Darwin, 1859). In his masterpiece, *The Origin of Species*, the great English naturalist Charles Darwin, celebrates the beauty and the variety of life on Earth, whose wealthy complexity of forms and shapes can be appreciated across wide spatial scales: from the differences between genes at a microscopic level up to those between entire ecosystems at a macroscopic one.

Identifying and understanding the relationships between all the life on Earth are some of the greatest challenges in science. Thanks to the huge quantity of data available for several activities structured on different time and space scales (topology and dynamics of ecological and social networks, geographical positions, internal structure and growth of cities, time statistics for email and paper mail correspondence and many others), a deep investigation on the many different complex living systems populating our planet is now possible. The ability to extract relevant information from these data may be crucial for the understanding of such systems and can be used as a starting point for future mathematical models. Indeed, thanks to the analysis of these massive databases, it is becoming more and more evident that, despite their diversity and complexity, natural phenomena are characterised by the emergence of regularities that are largely independent of their biological and physiological details.

The presence of common features has been interpreted by scientists as a clue that a common mechanism exists, and thus that a theoretical understanding is possible through mathematical models rich enough to show the investigated patterns, but containing only the essential ingredients pointing to the existence of a universal mechanism for observed patterns in living systems. However, despite the sustained efforts, we are currently able to describe only very specific cases with suitably tailored models. Many issues remain open both in modelling living systems and in data analysis.

The present thesis is the final result of a journey through complex patterns in ecology. We study both models and issues in data analysis as well as the connections between them. In particular, we explore the different aspects of the *biodiversity* of an ecosystem, referring with this term to the variety of its *species*, i.e. groups of living organisms, usually plants, which can interbreed and which populate the habitat

under study. A high biodiversity is an index of sustainability since it guarantees, as underlined by the *Millennium Ecosystem Assessment* in the early 2000s, four different ecosystem services to our planet: a life support (nutrient cycle, soil formation and primary production), a better supply (food, drinking water, materials or fuel), a correct regulation (of the climate and the tides, water purification, pollination and infestation control) and, last but not least, a cultural impact, be it aesthetic, spiritual or educational. For these reasons, nowadays, it is essential to have accurate analytical methods to measure the biodiversity in all its ecological aspects: a proper control allows for immediate interventions with appropriate safeguards.
Our interest is to investigate how these species interact with each other and with the surrounding environment.
In particular, as a matter of fact, ecological systems are characterised by the recurrent emergency of patterns. Our goal is to discover, describe and analyse the elements that underlie these patterns as well as the patterns themselves from a mathematical point of view.
In this thesis we focus on three important macro-ecological patterns which emerge when studying an ecological community.
The first one is the so-called *species-abundance distribution* (SAD), which tells us how commonness and rarity are distributed among the species of an ecosystem. Its typical representation is the *Preston plot* (Preston, 1948), where the system's species are grouped into an histogram according to the following rule: the first column accounts for half of the species with only one individual (we will refer to these species as *singletons*); in the second column we insert the other half of the singletons plus half of the species with two individuals (the so-called *doubletons*); more generally, in the n^{th} column will fall half of the species with 2^{n-2} individuals, half of those with 2^{n-1} and all the other species having a population comprised between these two bounds.
As we will see, a recurrent situation happens in the investigated ecological communities: in general, half of the forest trees belong to few species (the so-called *hyper-dominants*, which are made up of a huge number of individuals). The remaining species, which are the majority, have a much smaller number of individuals. These latter species are said to be *hyper-rare* because they have only few individuals. As a consequence, the Preston plot does usually show a unimodal shape with a long tail.
A second recurrent pattern is the *species-area relationships* (SAR), which looks at how biodiversity changes with the sampled area. There is empirical evidence that such curve shows a tri-phasic behaviour in a log-log plot. In fact, it usually displays a first linear phase in correspondence to local scales (e.g. forest samples), followed by a straight one when looking at the ecosystem scale (the entire forest). At larger scales, the curve starts growing up rapidly again because of the species' turnover.
Still connected with space, the last recurrent pattern we will investigate is the *similarity decay function*, which gives us information about how similar two sampling units are, given they are distant r apart.
Below we describe in details each chapter's content.

In Chapter 1, we expose the fundamentals of point process theory, which constitutes

an important and open research mathematical field thanks to its interdisciplinary character. Indeed, spatial point processes are particularly useful when the data we wish to model are in the form of spatial patterns such as the locations of points (e.g. trees, birds, etc.) in a given region and they have been therefore widely applied to ecology and also to the context of complex system in many different scientific fields, from astronomy to seismology to bioinformatics.
In the first part of the chapter, we introduce the basic first and second order statistics of a spatial point process, which aim at describing, respectively, its characteristics around a fixed location and the spatial relations between two of its points.
When dealing with an ecological dataset as those we wish to investigate in the present thesis, data are usually compound of information inherent many different species of plants. In this case we can assume that the individuals' locations of each species are a realisation of a particular spatial point process. Thus, what we observe when looking at all the points in the study region, regardless of their species' label, is the so-called *superposed process* (Baddeley et al., 2007). In the second part of this chapter we study how to define the fundamental statistics of this latter process in terms of those of the single species' ones.
In the third part we analyse some famous examples of spatial point processes, which are the homogeneous Poisson process and the Neyman-Scott cluster processes. The first one is characterised by the property that its points are stochastically independent one from the other (Daley and Vere-Jones, 2003), so that no correlations occur. The second kind of processes, instead, aims at modelling the mechanism of plant's reproduction, where a parent tree spreads its offspring around its location (Plotkin, Potts et al., 2000; Azaele, Cornell et al., 2012), so that the first generation of trees will be found aggregated in clusters in the study region and correlations between points becomes of importance in the process's analysis. We analyse these processes in details and we see how they can be simulated on the computer.
Finally, we study how to model a species' pattern according to these types of processes through the estimation of its fundamental statistics.

In Chapter 2 we are concerned with the statistical analysis of spatial patterns describing the location of plants in tropical forests.
It has been observed that many different complex systems, among which ecological communities, share the tendency to form spatial or temporal clusters (He, Legendre et al., 1997; Condit et al., 2000; Plotkin, Potts et al., 2000; Adorisio et al., 2009). However, classifying a spatial point pattern as clustered rather than regular can be a challenging task because establishing the main features of its spatial density function strongly depends on the scale through which we look at it (Hui, McGeoch and Warren, 2006). In fact, it is intuitively clear that a very small grid includes too many inessential details to be effective and also statistics can be too poor, while a very large one could with high probability miss important characteristics of the dataset.
More generally, it is well known that the form of a data-based density function may depend on the algorithm (binning rule) used for the binning of the data (Haegeman and Etienne, 2010). Thus, in order to correctly infer the underlying structure of a dataset, the choice of the optimal number of bins must be very careful and balanced

between capturing the major features in the data and ignoring details due to fluctuations.
In our view the main flaw of many binning rules (Sturges, 1926; Yule and Kendall, 1950; Doane, 1976; Scott, 2015; Freedman and Diaconis, 1981; Stone, 1984; Haegeman and Etienne, 2010; Knuth, 2006; Tovo, 2014) is that they assume some knowledge on the data distribution. For example Sturges' rule (Sturges, 1926) assumes that the data are normally distributed. This is a key point if you have in view applications to ecological datasets. In many cases it is not reasonable to assume such knowledge and the process generating the dataset must be considered unknown. Therefore any criteria based on some prior knowledge of the true density should not be applied as it often introduces a degree of arbitrariness that may produce biased conclusions.
In this chapter we intend to use a method based on *maximum a-posteriori estimation* and *Bayes's Theorem* proposed by K. H. Knuth (Knuth, 2006) to find the optimal bin size of a two-dimensional histogram. Knuth's non-parametric method selects the optimal scale from the data without any assumption on the underlying process that generated them. We show that the Knuth method can be used to highlight relevant spatial characteristics on the underlying distribution such as space anisotropy and clusterisation. We test it against the most currently used (Epanechnikov) kernel method for two-dimensional datasets and with a non-kernel method for a one-dimensional dataset (Stone's binning rule). In both the cases it results to be more efficient in detecting *complete spatial random* (CSR) processes and in avoiding sample fluctuations. Therefore our analysis validates it as a reliable method for determining the intensity function of a spatial pattern. Additionally, it is not subject to the *virtual aggregation* phenomenon (Schiffers et al., 2008). It correctly detects homogeneity cases or the presence of a gradient in the density function and the relative difference of the rectangular bin sides can be used as a measure of the pattern's anisotropy. It also allows to infer quantitative (cluster size) information on both first and second-order statistics. Thus, it is not only a rule to choose the bin size in which to organize the data. Indeed our analysis proves that Knuth's bin size is a good indicator of how finely structured is the dataset and that it can be used as a trusted tool for the preliminary statistical analysis of a spatial dataset. We show what are the relevant information contained in the size and the shape of the optimal bin and how they are related to the spatial features of the process/dataset.
We finally test our findings to study cluster-like structures in plants' arrangement on the Barro Colorado Island (BCI) ecological dataset, which consists of the spatial coordinates of individuals belonging to 300 different species of plants located in a 50 ha rectangle of rainforest.
All these results were reported in the paper "Application of optimal data-based binning method to spatial analysis of ecological datasets" published in *Spatial Statistics* (Tovo, Formentin et al., 2016).

In Chapter 3 we see how, thanks to the theoretical tools offered by point process theory, it is possible to introduce and explore the concept of biodiversity of an ecological community within a rigorous mathematical framework. In literature, such notion is strictly connected to the scale on which data are sampled: if they are

located within a limited habitat, we talk of alpha-diversity, whereas if the dataset comprises several sampling units scattered in a larger landscape we are looking at the gamma-diversity of the community. In this latter case, one can also compare two or more sampling units (beta-diversity or species' turnover). Complementary to the notion of diversity is the concept of similarity, which instead measures how similar two samples are in terms of the species' composition.
In the first part of this chapter, we give a brief review on the most important biodiversity indexes introduced in literature to measure similarity and diversity in species' composition of ecological communities and their connections. In the second part, we focus on binary similarity indexes and we rigorously insert all these concepts within the context of point process theory and extend such notions in order to incorporate spatial information. The main sources of inspiration for our approach are the works of Shimatani (Shimatani, 2001; Shimatani and Kubota, 2004), Plotkin and al. (Plotkin, Chave et al., 2002), Morlon et al. (Morlon et al., 2008) (based on (Plotkin, Chave et al., 2002)), and also Chave et al. (Chave and Leigh, 2002).

In Chapter 4 we focus on the Sørensen's similarity index (Sørensen, 1948) and we aim at finding an analytical relation between the change in floristic composition and the distance between two plots of a tropical rainforest. Because of the many drivers of diversity acting on real landscapes on many different spatial scales, this problem is hard to reduce to a mathematical model. In chronological order, important contributions to this central problem of estimating biodiversity of forests are the seminal works of Leight et al. (Leigh et al., 1993), Nekola and White (Nekola and White, 1999) and the neutral theory approach of Hubbell (Hubbell, 1997, 2001b) (see e.g. the comprehensive book Magurran and McGill, 2011).
In this thesis we focus on a single driver of diversity, that is the tendency of plants to form clusters of individuals. The shape and extent of the cluster may vary from species to species depending on seed dispersal limiting factors, or other effects (e.g Janzen-Connell effect), which may be inter or intraspecific, but our aim is at reducing this multiplicity of biotic factors to a single statistical descriptor. Stated in more mathematical terms, our first goal is to study how the presence of spatial correlations between positions of individuals (plants) affects the change in species' composition of two small plots at a given distance.
We describe the plant arrangement by a superposition $\boldsymbol{X}$ of spatial point processes and in this framework we introduce an analytical function which represents the average spatial density of the Sørensen similarity between two infinitesimal plots distant r apart. The similarity decay function we obtain is the follows

$$\chi_{\boldsymbol{X}}(r) = \lambda_{\boldsymbol{X}}(g_{\boldsymbol{X}}(r) - 1) + \chi_{\boldsymbol{X},\infty}$$

which results to depend on the density function of the superposed process, $\lambda_{\boldsymbol{X}}$, on its pair correlation function, $g_{\boldsymbol{X}}(r)$, and on a constant, $\chi_{\boldsymbol{X},\infty}$, depending solely on the species' abundances and representing the similarity at a scale where the clustering of individuals has no effects. The pair correlation function, in turn, depends on the clustering of each species weighted by their relative abundance. Therefore, in the proposed model, the similarity decay function is dominated by the most abundant species, a feature previously recognised in other studies, but still debated.

Apart from presenting a novel analytical approach to the definition of a decay of similarity function using point processes, the main aim is to test the proposed formula against field data. Here we use the BCI and Pasoh forest databases, which register the spatial positions of respectively 222602 and 310520 plants belonging to 301 and 927 species and covering an area of 50ha each. To this end, we adopt the statistical estimator for the pair correlation function proposed in Stoyan and Stoyan, 1994 and we design a novel one for the Sørensen similarity. The former is derived from the general theory of point processes even if it does not need any hypothesis on the type of stochastic point process that we should associate to the species of the forest under study. The latter is, instead, based directly on our Sørensen's similarity formula. They therefore provide a test of the proposed formula at a very general level.
A second goal is to select the class of spatial point processes that best describe the plants' arrangement in the study area and test its effectiveness in reproducing the decay of the similarity function. The clustering of each species is described by a univariate (if we assume rotational symmetry of the two-dimensional cluster) probability density, the so-called *dispersal kernel*, which gives the probability that an individual of a cluster is located at distance r from the cluster centre. The dispersal kernel features of each species are thus the essential information for our model that have to be derived from experimental data (by the minimum contrast method in our work). We test the effectiveness of three dispersal kernels (exponential, Gaussian and Cauchy) at describing the species' clustering. More precisely, once we determine the cluster description parameters for each species, we compute the analytical form of the pair correlation function and of the similarity index for each considered dispersal kernel and compare them with the empirical curves estimated from our data.
We have reported all our results in the manuscript "The distance decay of similarity in tropical rainforests. A spatial point processes analytical formulation", which has been accepted on *Theoretical Population Biology* (Tovo and Favretti, 2017).

In Chapter 5 we tackle the problem of upscaling biodiversity from scattered samples to larger areas of tropical forests.
Recently, a semi-analytical method has been proposed to upscale species richness assuming a log-series as the species-abundance distribution (SAD) (Harte, Smith et al., 2009; Ter Steege, Pitman et al., 2013; Slik et al., 2015). This distribution is named after the great statistician Ronald A. Fisher (Fisher et al., 1943) as the limiting form of a negative binomial distribution to describe the probability of observing n individuals when sampling from a population belonging to different species, excluding zero observations. The log-series distribution is often used to describe SAD patterns in ecological communities, including tropical tree communities. The robustness of the upscaling method relies on the stability property of Fisher's α (approximately reflecting the number of observed singleton species (Fisher et al., 1943)), which ought not to depend on the forest sample size and is given by

$$\frac{N_p}{S_p} = (e^{S_p/\alpha} - 1),$$

where N_p and S_p are the total number of individuals and species, respectively, when

sampling a fraction p of the forest ($N_1 = N$ and $S_1 = S$ corresponds to the total number of individuals and species when sampling the whole forest). The method proposed in Slik et al., 2015 is composed of three main steps: 1) Fisher's α is calculated assuming that the species have a log-series distribution, and using as input the observed species S_p and number of trees N_p. 2) The total number of stems N for the whole area of interest is extrapolated (this is not a trivial task and there is no consensus on the best methods to implement it. Generally, constant average stem density is assumed (Ter Steege, Pitman et al., 2013; Slik et al., 2015)). 3) Estimate the number of species at the largest scale using the formula $S = \alpha \ln(1 + N/\alpha)$ (Fisher et al., 1943).

This method has been used to estimate the species richness of the Amazonia (Ter Steege, Pitman et al., 2013) and the global tropical tree species richness (Slik et al., 2015). In the latter case, Slik et al. noted that, when merging forests in different tropical regions, the value of Fisher's α shows an asymptotic behaviour for large areas, as if converging to its asymptote for each region (Slik et al., 2015). From this limiting value, it is then possible to infer the total species richness of the different tropical regions. Non-parametric approaches have also been proposed in the literature to infer species richness. Instead of assuming a specific functional form for the SAD and fitting data to arrive at the parameters, such methods are based on the intuitive idea that it is only the rare species that carry information on the undetected species in a sample. A successful example is the method introduced by Chao (Chao, 2005; Chao, Colwell et al., 2009; Chao and Chiu, 2016), which takes into account just the number of singletons and doubletons observed at the sample scale to infer the total species richness of the whole forest.

Based on theoretical and computational analysis as well as using data from 15 tropical forests located all over the globe, we show that the LS method suffers from important limitations. Often the SAD - especially at large scales or with increasing sampling effort (Chisholm, 2007) - displays an interior mode (Azaele, Suweis et al., 2016), which a log-series cannot capture. Indeed, the Fisher's distribution is not flexible enough (Azaele, Maritan et al., 2015) to describe different SAD patterns found in tropical forests (Chave, 2004; Magurran, 2005; Chave, Alonso et al., 2006; Volkov et al., 2007; Magurran, 2013; Matthews and Whittaker, 2014; Azaele, Suweis et al., 2016).

In this chapter, we present a more general analytical framework to extrapolate species richness from local to whole forest scales. This framework, derived from first principles on the basis of biological processes, outperforms previously proposed methods and it assumes a negative binomial distribution as the SAD of the ecosystem

$$\mathcal{P}(n|r,\xi) = \binom{n+r-1}{n} \xi^n (1-\xi)^r.$$

This distribution arises naturally as the steady-state SAD of a system which undergoes simple birth and death dynamics, with an effective birth rate accounting for the effects of immigration events and/or intraspecific interactions (Volkov, Banavar, He et al., 2005; Azaele, Suweis et al., 2016), and under the neutral hypothesis that individuals are demographically identical (Volkov et al., 2007). Moreover, it is also able to adequately fit the SADs of diverse ecosystems such as tropical forests and coral

reefs (Volkov et al., 2007; Azaele, Suweis et al., 2016). The fundamental property on which our framework is based, is that the functional form of a negative binomial does not change when sampling different fractions of areas – *form invariance* under different sampling efforts – although the parameters of the distribution do change according to a computable deterministic function. More precisely, the negative binomial at different scales has the same r parameters, but different ξ, which is a function of the scale. Thus, we obtain an analytical expression of the upscaled SAD at the whole forest scale from the data at the sample scale p^* (denoting the fraction of the surveyed forest area with respect to its total extension). Using the SAD at the local scale, a maximum likelihood method is used to estimate the parameters of the SAD, and we use our upscaling equations to predict the species richness of the entire forest, i.e., at the largest scale $p = 1$. In particular, we found that the total number of species S is related to the number of species at the sampling scale p^*, S_{p^*}, by the following relation:

$$S = S_{p^*} \frac{1 - (1 - \xi)^r}{1 - (1 - \xi_{p^*})^r},$$

where ξ_{p^*} and r are the fitted parameters of the SAD at scale p^*. As noted above, r is scale invariant and hence independent of p^*, whereas the parameter ξ at the largest scale ($p = 1$) is given by

$$\xi = \frac{\xi_{p^*}}{p^* + (1 - p^*)\xi_{p^*}} \quad .$$

The framework resembles the renormalisation group technique in critical phenomena in which the behaviour of a system at different scales is described in terms of equations for the model parameters, similarly to what has been suggested here (Stanley, 1999). By using our framework, we are able to generate accurate and robust predictions for computer-generated forests and for 15 empirical tropical forests.
Our framework is also able to give a quantitative estimate of the sampling effort needed for achieving species richness predictions with error bars below approximately 5% (this percentage was arbitrarily chosen as an illustration and our approach can be straightforwardly used for any other percentage of error). These estimates have been obtained through Monte Carlo simulations that test the self-consistency of the negative binomial method and allow us to infer these critical sampling thresholds.
The results contained in this chapter have been reported in the paper "Upscaling species richness and abundances in tropical forests", published in *Science Advances* (Tovo, Suweis et al., 2017).

Contents

Part I

Application of Spatial Point Process Theory to Ecology

> *To such an extent does nature delight and abound in variety that among her trees there is not one plant to be found which is exactly like another; and not only among the plants, but among the boughs, the leaves and the fruits, you will not find one which is exactly similar to another.*

Leonardo Da Vinci, *Thoughts on Art and Life*

1 Point Processes

1.1 Introduction to point processes

Point processes are a particularly useful tool of probability theory when dealing with data in the form of set of spatial or temporal location of points in a one or more-dimensional space (Diggle, 2003; Møller and Waagepetersen, 2004; Baddeley et al., 2007).
Application of point process theory can be found in many diverse scientific disciplines far beyond ecology, such as bioinformatics (Cha and Zhou, 2014), genetics (Shimatani, 2002), social network (Zipkin et al., 2016), archaeology (Hodder and Orton, 1976), geography (Cliff and Ord, 1981), epidemiology (Quesada et al., 2017), seismology (Vere-Jones, 1970), astronomy (Peebles, 1974), computational neuroscience (Brown et al., 2004), economics (Lunde and Engle, 1998) and many others.
Here we are interested in application of point processes to theoretical ecology and our typical dataset will consist in the locations of trees within a surveyed region $\mathcal{W}$, which we consider a subset of $\mathbb{R}^2$. We will refer to such kind of datasets as *spatial patterns* or *realisations* of the point process with which we wish to model our data.
While a point process is a powerful tool for describing, for example, the instants along the time-line of a particular event (e.g. an earthquake, an hospital call, etc.), a *spatial point process* comes of interest when studying random patterns of points in a $d-$dimensional space, where $d \geq 2$ (Stoyan and Stoyan, 1994; Baddeley et al., 2007; Illian et al., 2008; Chiu et al., 2013), as those we are interested in. Henceforth, by point processes we mean spatial point processes.
In this chapter we will explore the basic concepts about spatial point processes and their first and second-order statistics following Baddeley et al., 2007. The former refers to properties which describe the characteristics of a point process around a single location, whereas the latter aim to describe spatial relations between pairs of points of a process.

1.1.1 The definition of a spatial point process

The notion of point process is strictly connected to the one of *random measure*, which we recall here.

Definition 1.1. Let S be a complete separable metric space and $\mathfrak{B}(S)$ the σ-algebra of its borel sets. Let $\mathcal{M}_S$ be the space of all boundedly finite measures on $\mathfrak{B}(S)$.
A **random measure** is a map $\mathcal{N}$ from a probability space $(\Omega, \mathfrak{F}, \mathbb{P})$ to the measurable space $(\mathcal{M}_S, \mathfrak{B}(\mathcal{M}_S))$.

As stated in the introduction, spatial point processes are very useful in analysing and modelling data in the form of spatial pattern such as the locations of trees or birds in a given region of $\mathbb{R}^2$ (here we set ourselves in a two-dimensional framework but all the following notions and results hold also in $\mathbb{R}^d$, with $d \geq 2$).
An immediate description of such kind of databases can be given by defining, for every bounded closed subregion $B \subset \mathbb{R}^2$, the following counting variable:

$$\mathcal{N}(B) = \text{number of points within } B.$$

The collection of all the variables $\{\mathcal{N}(B)\}_{B \subseteq \mathbb{R}^2}$ contains all the information about a point process's realisation, since the locations of its points are uniquely determined by the set of $x \in \mathbb{R}^2$ such that $\mathcal{N}(\{x\}) > 0$. One can show that $\mathcal{N}$ is a random measure defined on $S = \mathbb{R}^2$ (Stoyan and Stoyan, 1994; Daley and Vere-Jones, 2007; Chiu et al., 2013) satisfying the following properties:

- $\mathcal{N}(\varnothing) = 0$
- ***additivity***: $\mathcal{N}(B_1 \cup B_2) = \mathcal{N}(B_1) + \mathcal{N}(B_2)$ if $(B_1 \cap B_2) = \varnothing$, for $B_1,\ B_2 \in \mathbb{R}^2$
- ***continuity***: given a decreasing sequence of closed, bounded set $B_n \in \mathbb{R}^2$ having limit $\cap_n B_n = B$, we have that $\mathcal{N}(B_n) \to \mathcal{N}(B)$.

This let us formally introduce the concept of a spatial point process as follows:

Definition 1.2. A **point process** (also called a **counting random measure**) on a set $\mathbb{R}^2$ is a random measure $\mathcal{N}_{\boldsymbol{X}}$ taking natural values $\mathcal{N}_{\boldsymbol{X}}(B) \in \mathbb{N}$ for every $B \subseteq \mathbb{R}^2$.
Moreover, we say that the point process $\boldsymbol{X}$ is:

- **locally finite**, if each bounded subset $B \subseteq \mathbb{R}^2$ must contain only a finite number of points of the process:

$$\mathcal{N}_{\boldsymbol{X}}(B) < \infty$$

- **simple**, if two points of the process cannot have the same location:

$$\mathcal{N}_{\boldsymbol{X}}(\{x\}) \leq 1 \qquad \forall x \in \mathbb{R}^2.$$

From now on we will assume that $\boldsymbol{X}$ satisfies both the locally finiteness and simplicity regularity conditions for each of its realisations.
We remark that another way to define a point process is through the so-called *void* or *vacancy indicator* (see Definition 1.18)

$$v_{\boldsymbol{X}}^{B} = \mathbb{I}(\mathcal{N}_{\boldsymbol{X}}(B) = 0),$$

where $\mathbb{I}(\cdot)$ is the indicator function:

$$\mathbb{I}(\mathcal{C}) = \begin{cases} 1 & \text{if condition } \mathcal{C} \text{ holds} \\ 0 & \text{otherwise.} \end{cases}$$

Indeed, the location of the points of a process's realisation are determined by the set of all $x \in \mathbb{R}^2$ such that $v_{\boldsymbol{X}}^{\{x\}} = 1$. In analogy to the additivity property of the random measure $\mathcal{N}_{\boldsymbol{X}}$, the void indicator satisfies a **multiplicative** property:

$$\mathcal{N}_{\boldsymbol{X}}(B_1 \cup B_2) = v_{\boldsymbol{X}}^{B_1} v_{\boldsymbol{X}}^{B_2} \qquad \text{for any sets } B_1, B_2 \subseteq \mathbb{R}^2.$$

The concept of a random measure is also strictly connected to the one of *random set* (Daley and Vere-Jones, 2007; Illian et al., 2008). Indeed, in the hypothesis of simplicity, to each measure $\mathcal{N}_{\boldsymbol{X}}$ describing a point process can be uniquely associated the set of points $\mathcal{S}_{\boldsymbol{X}} = \{x_1, x_2, \dots\} \subset \mathbb{R}^2$ such that $\mathcal{N}_{\boldsymbol{X}}(\{x_i\}) > 0 \ \forall i$. Such set $\mathcal{S}_{\boldsymbol{X}}$ is called the *support* of $\mathcal{N}_{\boldsymbol{X}}$. We will denote with $\boldsymbol{X}$ the spatial point process under study referring with such symbol both to the random measure $\mathcal{N}_{\boldsymbol{X}}$ and to the corresponding random subset $\mathcal{S}_{\boldsymbol{X}}$. A realisation of a point process in the sense of a random set $\mathcal{S}_{\boldsymbol{X}}$ is called a *spatial pattern*.

1.1.2 The number distribution

The *space of outcomes* of a point process in $\mathbb{R}^2$ can now be formally described in terms of the random measure $\mathcal{N}_{\boldsymbol{X}}$ previously introduced.
Let us denote with $\mathscr{N}$ the set of all counting measures on $\mathbb{R}^2$, i.e. measures which assume a non-negative finite integer value on every compact set $B \subseteq \mathbb{R}^2$. Then a spatial point process can be seen also as a random element $\mathcal{N}_{\boldsymbol{X}}$ of $\mathscr{N}$.
Given a compact subset $B \subset \mathbb{R}^2$ and a non-negative integer n, we can then define a basic *event* as

$$E_{B,n} = \{\mathcal{N}_{\boldsymbol{X}} \in \mathscr{N} : \ \mathcal{N}_{\boldsymbol{X}}(B) = n\}. \tag{1.1}$$

Let us consider the σ-field of subsets of $\mathscr{N}$ generated by all the events of the form (1.1). We will denote it with $\mathcal{N}$. Then we can give the following

Definition 1.3. The **outcome** or **canonical space** of a point process defined in $\mathbb{R}^2$ is the pair $(\mathscr{N}, \ \mathcal{N})$.

We wish now to associate a distribution to the point process $\boldsymbol{X}$. In order to do that, let $(\Omega, \mathcal{A}, \mathbb{P})$ be a probability space. We know from Definition 1.1 that a point process $\boldsymbol{X}$, being a random counting measure, is a measurable map $\mathcal{N}_{\boldsymbol{X}} : \Omega \to \mathscr{N}$ associating to each elementary outcome $\omega \in \Omega$ an outcome $\mathcal{N}_{\boldsymbol{X}}(\omega) \in \mathscr{N}$.

We can thus introduce the so-called *number distribution* $\mathbf{P}_{\boldsymbol{X}}(\cdot)$, specifying the probability

$$\mathbf{P}_{\boldsymbol{X}}(\mathscr{A}) = \mathbb{P}(\mathcal{N}_{\boldsymbol{X}} \in \mathscr{A})$$

that the random measure $\mathcal{N}_{\boldsymbol{X}}$ of the point process $\boldsymbol{X}$ satisfies the properties described by the subset $\mathscr{A} \in \mathcal{N}$. Let us notice that the hypothesis of measurability of $\mathcal{N}_{\boldsymbol{X}}$ guarantees that the event $\{\mathcal{N}_{\boldsymbol{X}} \in \mathscr{A}\} = \{\omega \in \Omega : \mathcal{N}_{\boldsymbol{X}}(\omega) \in \mathscr{A}\}$ belongs to $\mathcal{A}$.
We have then the following

Definition 1.4. The joint probability distributions of $(\mathcal{N}_{\boldsymbol{X}}(B_1), \ldots, \mathcal{N}_{\boldsymbol{X}}(B_m))$, where $m > 0$ and $B_1, \ldots, B_m \subset \mathbb{R}^2$ are called the **finite-dimensional distributions** or **fidis** of a point process $\boldsymbol{X}$.

In particular, the fidis of a point process specify the value of $\mathbf{P}_{\boldsymbol{X}}(\mathscr{A})$, where $\mathscr{A}$ is an event in $\mathcal{N}$ of the form $\{\mathcal{N}_{\boldsymbol{X}} \in \mathscr{N} : \mathcal{N}_{\boldsymbol{X}}(B_1) = n_1, \ldots, \mathcal{N}_{\boldsymbol{X}}(B_m) = n_m\}$.
We have now the tools to compare two point processes. Indeed the following uniqueness theorem holds:

Theorem 1.5. *The number distribution $\mathbf{P}_{\boldsymbol{X}}(\cdot)$ of a point process is uniquely determined by its fidis. In other worlds, two point processes $\boldsymbol{X}$ and $\boldsymbol{Y}$ coincide if and only if they have the same fidis.*

We have noticed that a point process can be seen either as a random measure $\mathcal{N}_{\boldsymbol{X}}$ or as a random set $\mathcal{S}_{\boldsymbol{X}}$. In this latter case it is useful to introduce the so-called *vacancy probabilities*. Let us consider the event $\mathscr{A} = \{\mathcal{N}_{\boldsymbol{X}} \in \mathscr{N} : \mathcal{N}_{\boldsymbol{X}}(B) = 0 \ \forall B \subset \mathbb{R}^2\}$. This is clearly an element of the σ-field $\mathcal{N}$ since it can be obtained by the intersection of the countably many events of the form $E_{\mathcal{B}_r,0}$, where $r \in \mathbb{N}$ and $\mathcal{B}_r \in \mathbb{R}^2$ is the ball of centre the origin and radius r. Given a compact set $B \subset \mathbb{R}^2$, we call $\mathbb{P}(\mathcal{N}_{\boldsymbol{X}}(B) = 0)$ the vacancy probability of B.

Definition 1.6. Given a compact subset $B \subset \mathbb{R}^2$, the **capacity functional** of a point process $\boldsymbol{X}$ computed in B is the functional

$$T_{\boldsymbol{X}}(B) = \mathbb{P}(\mathcal{N}_{\boldsymbol{X}}(B) > 0) = 1 - \mathbb{P}(\mathcal{N}_{\boldsymbol{X}}(B) = 0).$$

We can then rewrite Theorem 1.5 in terms of the capacity functional:

Theorem 1.7. *The number distribution $\mathbf{P}_{\boldsymbol{X}}(\cdot)$ of a spatial point process is uniquely determined by its capacity functional $T_{\boldsymbol{X}}(\cdot)$. In other worlds, two point processes $\boldsymbol{X}$ and $\boldsymbol{Y}$ coincide if and only if they have the same capacity functional.*

Therefore, we have two possible ways to describe a property of a spatial point process: either in terms of its fidis or in terms of its capacity functional. An example are the important concepts of a stationary and of an isotropic point process, which we introduce through the following definitions.

Definition 1.8. A spatial point process is said to be **stationary** if the fidis of original process $\boldsymbol{X}$ is the same of the process obtained by shifting each point of $\boldsymbol{X}$ by any vector $v \in \mathbb{R}^2$. Equivalently, a spatial point process is said to be **stationary** if its capacity functional $T_{\boldsymbol{X}}(\cdot)$ is invariant under translation, i.e. if it satisfies the relation $T_{\boldsymbol{X}}(B) = T_{\boldsymbol{X}}(B + v)$ for any compact sets B and for any vector $v \in \mathbb{R}^2$.

Definition 1.9. A spatial point process is said to be **isotropic** if the fidis of original process $\boldsymbol{X}$ is the same of the process obtained by applying any rotations of $\mathbb{R}^2$ to each point of $\boldsymbol{X}$. Equivalently, a spatial point process is said to be **isotropic** if its capacity functional $T_{\boldsymbol{X}}(\cdot)$ is invariant under rotation.

In other words, if a point process $\boldsymbol{X}$ defined on $\mathbb{R}^2$ is both stationary and isotropy, then its statistical properties are the same over the whole bi-dimensional space and they do not depend on the direction considered. These properties will be of particular importance when investigating both first and second order statistics of point process (see Section 1.2).
We will now introduce some other quantities which naturally come out when exploring and analysing point processes.

1.1.3 The concept of distance in point patterns

When analysing a point process, one of the first quantity we can introduce is the distance between points. Let u be an arbitrary point of $\mathbb{R}^2$. We denote with $d(\boldsymbol{X}, u)$ the distance between u and the closest point to u belonging to the process $\boldsymbol{X}$:

$$d(\boldsymbol{X}, u) = \min_{x \in \boldsymbol{X}} dist(\boldsymbol{X}, u), \tag{1.2}$$

In eq. (1.2), x is any points of the process $\boldsymbol{X}$ seen as a random set $\mathcal{S}_{\boldsymbol{X}}$ and *dist* denotes any distance on $\mathbb{R}^2$. In what follows, unless differently specified, we will consider the Euclidean distance in $\mathbb{R}^2$. The variable $d(\boldsymbol{X}, u)$ is called the *contact distance* between the process $\boldsymbol{X}$ and a point $u \in \mathbb{R}^2$.
Let us consider the event $\mathscr{A} = \{\mathcal{N}_{\boldsymbol{X}} \in \mathscr{N} : \ \mathcal{N}_{\boldsymbol{X}}(\mathcal{B}(u, r)) > 0\} \subset \mathscr{N}$. For what we have already seen, $\mathscr{A}$ is measurable since $\mathbf{P}_{\boldsymbol{X}}(\mathscr{A}) = \mathbf{T}_{\boldsymbol{X}}(\mathcal{B}(u, r))$ is well defined. This implies that also the event $\mathscr{E} = \{\mathcal{N}_{\boldsymbol{X}} \in \mathscr{N} : \ d(\boldsymbol{X}, u) \leq r\}$ is measurable and that $\mathbf{P}_{\boldsymbol{X}}(\mathscr{A}) = \mathbf{P}_{\boldsymbol{X}}(\mathscr{E})$.
We can now introduce the following definition.

Definition 1.10. Given a stationary point process $\boldsymbol{X}$, its **contact distribution function** or **empty space function** $F_{\boldsymbol{X}}(\cdot)$ is the cumulative distribution function of the random variable $d(\boldsymbol{X}, u)$, where $u \in \mathbb{R}^2$:

$$F_{\boldsymbol{X}}(r) = \mathbb{P}(d(\boldsymbol{X}, u) \leq r).$$

Let us remark that the hypothesis of stationarity guarantees that this distribution does not depend on the fixed point u and that therefore the wording $F_{\boldsymbol{X}}(r)$ instead of $F_{\boldsymbol{X}}(r, u)$ is justified.

1.2 Moments of a point process

First and second moments are at the basis of the theory of point processes and they are the starting point of exploratory analysis of point patterns (Stoyan and Stoyan, 1994; Baddeley et al., 2007; Illian et al., 2008; Chiu et al., 2013; Diggle, 2013; Wiegand and Moloney, 2013; Cressie, 2015).

From these quantities we can derive other important properties of point processes which have been deeply studied since they can be applied for describing spatial dataset of different kind (Illian et al., 2008; Wiegand and Moloney, 2013; Cressie, 2015).

1.2.1 The intensity function

Given a spatial point process $\boldsymbol{X}$ defined on $\mathbb{R}^2$, its *intensity measure* is a first-order statistic counting the mean number of points per unit area.

Definition 1.11. Given a subset $B \subset \mathbb{R}^2$, the **intensity measure** $\nu_{\boldsymbol{X}}$ of a point process $\boldsymbol{X}$ evaluated in B gives the mean number of random points of $\boldsymbol{X}$ falling within the subset:

$$\nu_{\boldsymbol{X}}(B) = \mathbb{E}[\mathcal{N}_{\boldsymbol{X}}(B)],$$

provided that $\mathbb{E}[\mathcal{N}_{\boldsymbol{X}}(B)]$ is finite for every compact $B \subset \mathbb{R}^2$.
Moreover, we call **intensity function** of the point process $\boldsymbol{X}$ the function $\lambda_{\boldsymbol{X}}$ such that:

$$\nu_{\boldsymbol{X}}(B) = \int_B \lambda_{\boldsymbol{X}}(u)du,$$

given that it exists.

We have then the following definition

Definition 1.12. A point process $\boldsymbol{X}$ is called **homogeneous process** if its intensity function is constant $\lambda_{\boldsymbol{X}}(u) \equiv \lambda_{\boldsymbol{X}}$ for every $u \in \mathbb{R}^2$, i.e. if its intensity measure is a constant multiple of the Lebesgue measure on $\mathbb{R}^2$.

Let us remark that if $\boldsymbol{X}$ is a stationary process, than it is also homogeneous. Indeed we have the following result:

Theorem 1.13. *Let $\boldsymbol{X}$ be a stationary process defined in $\mathbb{R}^2$. Then its intensity function is a constant multiple of the Lebesgue measure on $\mathbb{R}^2$.*

Proof. Let $\boldsymbol{X}$ be a stationary process defined in $\mathbb{R}^2$. Then, by definition,we know that the fidis of $\boldsymbol{X}$ and of the shifted process $\boldsymbol{X} + v$ are the same for every vector $v \in \mathbb{R}^2$. Thus $\mathbf{P}_{\boldsymbol{X}}(\mathscr{A}) = \mathbf{P}_{\boldsymbol{X}+v}(\mathscr{A})$, for any event $\mathscr{A}$ in $\mathcal{N}$ of the form $\{\mathcal{N}_{\boldsymbol{X}} \in \mathscr{N} : \mathcal{N}_{\boldsymbol{X}}(B_1) = n_1, \ldots, \mathcal{N}_{\boldsymbol{X}}(B_m) = n_m\}$, with $m > 0$. By taking $m = 1$ and $B_1 = B + v$ this implies that $\mathbf{P}_{\boldsymbol{X}}(\{\mathcal{N}_{\boldsymbol{X}} \in \mathscr{N} : \mathcal{N}_{\boldsymbol{X}}(B + v) = n\}) = \mathbf{P}_{\boldsymbol{X}+v}(\{\mathcal{N}_{\boldsymbol{X}} \in \mathscr{N} : \mathcal{N}_{\boldsymbol{X}}(B + v) = n\}) = \mathbf{P}_{\boldsymbol{X}}(\{\mathcal{N}_{\boldsymbol{X}} \in \mathscr{N} : \mathcal{N}_{\boldsymbol{X}}(B) = n\})$. Thus

$$\nu_{\boldsymbol{X}}(B + v) = \mathbb{E}[\mathcal{N}_{\boldsymbol{X}}(B + v)] = \mathbb{E}[\mathcal{N}_{\boldsymbol{X}}(B)] = \nu_{\boldsymbol{X}}(B).$$

But the only measures satisfying such relation are the multiples of the Lebesgue one. □

A local interpretation of the intensity function is the following. Let us consider a ball $\mathcal{B}(u, r_u)$ in $\mathbb{R}^2$ centred in u and having radius r_u and infinitesimally small size

$\mathrm{d}u$. Then $\lambda_{\boldsymbol{X}}(u)\mathrm{d}u$ approximates the probability that one point of the process will fall within $\mathcal{B}(u, r_u)$:

$$\lambda_{\boldsymbol{X}}(u)\mathrm{d}u \sim \mathbb{P}(\mathcal{N}_{\boldsymbol{X}}(\mathcal{B}(u, r_u)) > 0) = T_{\boldsymbol{X}}(\mathcal{B}(u, r_u)).$$

More precisely $\mathbb{P}(\mathcal{N}_{\boldsymbol{X}}(\mathcal{B}(u, r_u)) > 0) = \lambda_{\boldsymbol{X}}(u)\mathrm{d}u + o(\mathrm{d}u)$, where $o(\mathrm{d}u)$ gives the probability that $\mathcal{B}(u, r_u)$ contain more than one point of the process (Stoyan and Stoyan, 1994).
In Chapter 2 we will discuss how to infer the intensity function of a point process, given one of its realisations.
The importance of the intensity function can be better understood looking at the following result (Stoyan and Stoyan, 1994; Baddeley et al., 2007; Daley and Vere-Jones, 2003, 2007; Chiu et al., 2013).

Theorem 1.14 (Campbell's Formula)**.** *Given a spatial point process $\boldsymbol{X}$ defined on a locally compact subset $\mathcal{W} \subseteq \mathbb{R}^2$ with intensity measure $\nu_{\boldsymbol{X}}(\cdot)$, let $f : \mathcal{W} \to \mathbb{R}$ be any real valued measurable function. Then the random variable*

$$\mathcal{S} = \sum_{x \in \boldsymbol{X}} f(x)$$

has expected value given by

$$\mathbb{E}[\mathcal{S}] = \int_{\mathcal{W}} f(u)\nu_{\boldsymbol{X}}(du) \tag{1.3}$$

Proof. By monotone approximation of measurable functions, it suffices to show that the result holds for any step function:

$$\hat{f}_m(u) = \sum_{i=1}^{m} c_i \chi_{B_i}(u),$$

where $\chi_{B_i}(\cdot)$ is the characteristic function of the compact subset $B_i \subset \mathcal{W}$, $m \in \mathbb{N}$ is greater than zero and $c_i \in \mathbb{R}$.
We have that

$$\mathcal{S} = \sum_{x \in \boldsymbol{X}} \hat{f}_m(x) = \sum_{x \in \boldsymbol{X}} \sum_{i=1}^{m} c_i \chi_{B_i}(x) = \sum_{i=1}^{m} c_i \sum_{x \in \boldsymbol{X}} \chi_{B_i}(x) = \sum_{i=1}^{m} c_i \mathcal{N}_{\boldsymbol{X}}(B_i),$$

since $\sum_{x \in \boldsymbol{X}} \chi_{B_i}(x)$ is the number of points of $\boldsymbol{X}$ falling in B_i, i.e. $\mathcal{N}_{\boldsymbol{X}}(B_i)$.
Thus, taking the mean value, we have that

$$\mathbb{E}[\mathcal{S}] = \mathbb{E}\Big[\sum_{i=1}^{m} c_i \mathcal{N}_{\boldsymbol{X}}(B_i)\Big] = \sum_{i=1}^{m} c_i \mathbb{E}[\mathcal{N}_{\boldsymbol{X}}(B_i)] = \sum_{i=1}^{m} c_i \nu_{\boldsymbol{X}}(B_i) = \int_{\mathcal{W}} \hat{f}_m(u)\nu_{\boldsymbol{X}}(\mathrm{d}u),$$

where we used, in the second equality, the linearity of the expected value. □

Let us remark the process $\boldsymbol{X}$ admits intensity function $\lambda_{\boldsymbol{X}}(\cdot)$, then eq. (1.3) becomes

$$\mathbb{E}[\mathcal{S}] = \int_{\mathcal{W}} f(u)\lambda_{\boldsymbol{X}}(u)\mathrm{d}u.$$

The intensity function of a point process is just the first step in analysing a point pattern. Other important pattern's characteristics, such us the tendency to aggregation or dispersion (Illian et al., 2008; Wiegand and Moloney, 2013; Diggle, 2013; Tovo, Formentin et al., 2016) are revealed only by second-order statistics, which take into account the correlations between pair of points due to possible interactions. Below we investigate the most common second order statistics and their use.

1.2.2 Second order statistics

Let $\boldsymbol{X}$ be a spatial point process defined on $\mathbb{R}^2$. As $\mathcal{N}_{\boldsymbol{X}}(B)$, for $B \subseteq \mathbb{R}^2$, is a random variable, we can define its *variance* and *covariance* as:

$$\mathbf{var}(\mathcal{N}_{\boldsymbol{X}}(B)) = \mathbb{E}[\mathcal{N}_{\boldsymbol{X}}(B)^2] - \mathbb{E}[\mathcal{N}_{\boldsymbol{X}}(B)]^2$$
$$\mathbf{covar}(\mathcal{N}_{\boldsymbol{X}}(B_1), \mathcal{N}_{\boldsymbol{X}}(B_2)) = \mathbb{E}[\mathcal{N}_{\boldsymbol{X}}(B_1)\mathcal{N}_{\boldsymbol{X}}(B_2)] - \mathbb{E}[\mathcal{N}_{\boldsymbol{X}}(B_1)]\mathbb{E}[\mathcal{N}_{\boldsymbol{X}}(B_2)]$$

Let us notice that if $\boldsymbol{X}$ is a spatial point process defined on $\mathcal{R}^2$, then $\boldsymbol{X} \times \boldsymbol{X}$ is a point process defined on $\mathcal{R}^2 \times \mathcal{R}^2$ whose points are all ordered pairs (u_1, u_2) with $u_i \in \boldsymbol{X}$ for $i = 1, 2$. Thus the product $\mathcal{N}_{\boldsymbol{X}}(B_1) \times \mathcal{N}_{\boldsymbol{X}}(B_2)$ is a random variable counting the number of ordered pairs (u_1, u_2) with $u_1 \in B_1$ and $u_2 \in B_2$. We can then introduce the following definitions.

Definition 1.15. Let $\boldsymbol{X}$ be a spatial point process and $\mathcal{N}_{\boldsymbol{X}}$ its associated random measure. The **second moment measure** $\nu_{2,\boldsymbol{X}}(\cdot)$ of the point process $\boldsymbol{X}$ is the intensity measure of $\boldsymbol{X} \times \boldsymbol{X}$ defined as:

$$\nu_{2,\boldsymbol{X}}(B_1 \times B_2) = \nu_{\boldsymbol{X}\times\boldsymbol{X}}(B_1 \times B_2) = \mathbb{E}[\mathcal{N}_{\boldsymbol{X}}(B_1)\mathcal{N}_{\boldsymbol{X}}(B_2)].$$

Definition 1.16. Let $\boldsymbol{X}$ be a spatial point process and $\mathcal{N}_{\boldsymbol{X}}$ its associated random measure. The **second factorial moment measure** $\nu_{[2],\boldsymbol{X}}(\cdot)$ of the point process $\boldsymbol{X}$ is defined as:

$$\nu_{[2],\boldsymbol{X}}(B_1 \times B_2) = \mathbb{E}[\mathcal{N}_{\boldsymbol{X}}(B_1)\mathcal{N}_{\boldsymbol{X}}(B_2)] - \mathbb{E}[\mathcal{N}_{\boldsymbol{X}}(B_1 \cap B_2)].$$

Moreover, let us assume that there exists a function $\rho_{\boldsymbol{X}}(\cdot, \cdot)$ such that

$$\nu_{[2],\boldsymbol{X}}(C) = \int_C \rho_{\boldsymbol{X}}(u, v)\mathrm{d}u\mathrm{d}v.$$

for any compact subset $C \subset \mathbb{R}^2 \times \mathbb{R}^2$. Then $\rho_{\boldsymbol{X}}(\cdot, \cdot)$ is called **second moment density** of the point process $\boldsymbol{X}$.

As for the intensity function, we can also give an informal interpretation of the second moment density. Consider two disjoints balls $\mathcal{B}(u, r_u)$ and $\mathcal{B}(v, r_v)$ in $\mathbb{R}^2$ of infinitesimally small radii r_u and r_v and centred on two different locations u and v, respectively. Let us denote with $\mathrm{d}u$ and $\mathrm{d}v$ the sizes of the two balls. Then the quantity $\rho_{\boldsymbol{X}}(u, v)\mathrm{d}u\mathrm{d}v$ gives an approximation of the joint probability that one point of a given point process will occur within $\mathcal{B}(u, r_u)$ and a second point within $\mathcal{B}(v, r_v)$:

$$\rho_{\boldsymbol{X}}(u, v)\mathrm{d}u\mathrm{d}v \sim \mathbb{P}(\mathcal{N}_{\boldsymbol{X}}(\mathcal{B}(u, r_u)) > 0, \mathcal{N}_{\boldsymbol{X}}(\mathcal{B}(v, r_v)) > 0)$$

In the particular case of a stationary and isotropic point process $\boldsymbol{X}$ in $\mathbb{R}^2$, we know that the fidis of the process are invariant under both translations and rotations of $\mathbb{R}^2$. Thus the second moment $\rho_{\boldsymbol{X}}(u, v)$ does not depend on the exact locations of the considered two points u and v, but only on their distance $dist(u, v) = r$. In this case we will write $\rho_{\boldsymbol{X}}(r)$ in the place of $\rho_{\boldsymbol{X}}(u, v)$.
The notion of the second moment density leads to another object, fundamental when studying the spatial relation between points of a process $\boldsymbol{X}$, which is the *pair correlation function*.

Definition 1.17. Let $\boldsymbol{X}$ be a spatial point process defined in $\mathbb{R}^2$ with intensity function $\lambda_{\boldsymbol{X}}(\cdot)$ and second moment density $\rho_{\boldsymbol{X}}(\cdot,\cdot)$. The **pair correlation function** $g_{\boldsymbol{X}}(\cdot,\cdot)$ of $\boldsymbol{X}$ is the ratio

$$g_{\boldsymbol{X}}(u,v) = \frac{\rho_{\boldsymbol{X}}(u,v)}{\lambda_{\boldsymbol{X}}(u)\lambda_{\boldsymbol{X}}(v)}$$

where $u, v \in \mathbb{R}^2$.

Again, if $\boldsymbol{X}$ is a stationary and isotropic, we have that $g_{\boldsymbol{X}}(u,v) = g_{\boldsymbol{X}}(r)$ depends only on the distance r between the two points u and v. Moreover, we know that if a point process is stationary, it is also homogeneous, so that its intensity function $\lambda_{\boldsymbol{X}}(\cdot)$ takes a constant value $\lambda_{\boldsymbol{X}}(u) \equiv \lambda_{\boldsymbol{X}}$ over all $u \in \mathbb{R}^2$. Therefore, in this special case we get

$$g_{\boldsymbol{X}}(x,y) = g_{\boldsymbol{X}}(r) = \frac{\rho_{\boldsymbol{X}}(r)}{\lambda_{\boldsymbol{X}}^2}$$

We now define two other random variables associated to a point process which will become useful in what follows.

Definition 1.18. Let $\boldsymbol{X}$ be a spatial point process defined in $\mathbb{R}^2$ and $\mathcal{N}_{\boldsymbol{X}}$ its associated random measure. Let then $B \subset \mathbb{R}^2$.

- The **presence indicator** of $\boldsymbol{X}$ associated with B is a random variable defined as:

$$1_{\boldsymbol{X}}^B = \begin{cases} 1 & \text{if } \mathcal{N}_{\boldsymbol{X}}(B) > 0 \\ 0 & \text{if } \mathcal{N}_{\boldsymbol{X}}(B) = 0 \end{cases}$$

- The **vacancy indicator** of $\boldsymbol{X}$ associated with B is a random variable defined as:

$$v_{\boldsymbol{X}}^B = \begin{cases} 1 & \text{if } \mathcal{N}_{\boldsymbol{X}}(B) = 0 \\ 0 & \text{if } \mathcal{N}_{\boldsymbol{X}}(B) > 0 \end{cases}$$

Let us notice that the following relation holds for every $B \subset \mathbb{R}^2$.:

$$1_{\boldsymbol{X}}^B + v_{\boldsymbol{X}}^B \equiv 1,$$

The presence indicator of a point process $\boldsymbol{X}$ is strictly connected to its capacity functional and its pair correlation function. Indeed we have the following relation

$$\mathbb{E}[1_{\boldsymbol{X}}^B] = 1 \cdot \mathbb{P}(1_{\boldsymbol{X}}^B = 1) + 0 \cdot \mathbb{P}(1_{\boldsymbol{X}}^B = 0) = \mathbb{P}(1_{\boldsymbol{X}}^B = 1) = T_{\boldsymbol{X}}(B). \tag{1.4}$$

Thus, the expectation of the presence indicator associated with the region $B \subset \mathbb{R}^2$ gives the probability that at least one point of the process falls within it.
Let us now consider two disjoint regions A and B of $\mathbb{R}^2$. We have that

$$\begin{aligned}\mathbb{E}[1_{\boldsymbol{X}}^A 1_{\boldsymbol{X}}^B] &= 1 \cdot \mathbb{P}(1_{\boldsymbol{X}}^A = 1, 1_{\boldsymbol{X}}^B = 1) + 0 \cdot \mathbb{P}(1_{\boldsymbol{X}}^A = 1, 1_{\boldsymbol{X}}^B = 0) + \\ &\quad + 0 \cdot \mathbb{P}(1_{\boldsymbol{X}}^A = 0, 1_{\boldsymbol{X}}^B = 1) + 0 \cdot \mathbb{P}(1_{\boldsymbol{X}}^A = 0, 1_{\boldsymbol{X}}^B = 0) \\ &= \mathbb{P}(1_{\boldsymbol{X}}^A = 1, 1_{\boldsymbol{X}}^B = 1) = \mathbb{P}(\mathcal{N}_{\boldsymbol{X}}(A) > 0, \mathcal{N}_{\boldsymbol{X}}(B) > 0). \end{aligned} \tag{1.5}$$

Let $A = \mathcal{B}(u, r_u)$ and $B = \mathcal{B}(v, r_v)$ be two balls of infinitesimally small sizes $\mathrm{d}u$ and $\mathrm{d}v$ centred in u and v and with radii r_u and r_v, respectively. Eq. (1.5) yields

$$\mathbb{E}[1_{\boldsymbol{X}}^A 1_{\boldsymbol{X}}^B] \sim \rho_{\boldsymbol{X}}(A,B)\mathrm{d}u\mathrm{d}v.$$

1.3 The Palm distribution

All the statistics we have considered up to now are sometimes called *location-related* (Wiegand and Moloney, 2013), meaning that they aim to describe the properties of a point process around an arbitrary location or between arbitrary locations of the space where the process is defined. In contrast, we call *point-related* those summary statistics which describe the properties of the point process around one *typical point* $x \in \boldsymbol{X}$.
In this case it is useful to introduce the concept of the *Palm distribution*, whose formal definition is the following:

Definition 1.19. Let $\boldsymbol{X}$ be a spatial point process defined in $\mathbb{R}^2$ and let x be an arbitrary point of $\boldsymbol{X}$. The **Palm distribution** $\mathbf{P}^x_{\boldsymbol{X}}(\cdot)$ of $\boldsymbol{X}$ at the location x is defined by

$$\mathbf{P}^x_{\boldsymbol{X}}(\mathscr{A}) = \mathbb{P}(\mathcal{N}_{\boldsymbol{X}} \in \mathscr{A} | x \in \boldsymbol{X}) \qquad \forall \mathscr{A} \in \mathcal{N}. \tag{1.6}$$

We write $\mathbb{P}(\mathcal{N}_{\boldsymbol{X}} \in \mathscr{A} | x \in \boldsymbol{X}) = \mathbb{P}^x(\mathcal{N}_{\boldsymbol{X}} \in \mathscr{A})$ and refer to $\mathbb{P}^x(\cdot)$ as the **Palm probability measure**, which is the conditional probability that the random measure $\mathcal{N}_{\boldsymbol{X}}$ satisfies the properties specified by $\mathscr{A}$ given that x is a point of the process $\boldsymbol{X}$, i.e. $\mathcal{N}_{\boldsymbol{X}}(\{x\}) > 0$.

By using sophisticated theoretical tools as the Radon-Nikodym theorem and the Campbell measure (see e.g. Stoyan and Stoyan, 1994; Baddeley et al., 2007; Chiu et al., 2013), it can be rigorously shown that $\mathbf{P}^x_{\boldsymbol{X}}(\mathscr{A})$ in eq. (1.6) is well defined.
The distribution $\mathbf{P}^x_{\boldsymbol{X}}(\cdot)$ takes its name after the Swedish electrical engineer and statistician Conrad Palm, who introduced it in his Ph.D thesis about telephone traffic problems (Palm, 1943).
From now on we will assume that $\boldsymbol{X}$ is a stationary point process with constant intensity function $\lambda_{\boldsymbol{X}}(u) = \lambda_{\boldsymbol{X}}$ for every $u \in \mathbb{R}^2$. In this case we know that if we translate each point of $\boldsymbol{X}$ by any vector v we get an equivalent process in the distribution sense. Therefore, in this case, we will set the typical point $x = 0 \in \mathbb{R}^2$ without loss of generality.

1.3.1 Point-related statistics

The notion of the Palm distribution let us define, for stationary processes, other second order statistics both local and not, which have been introduced in the theory of point processes and have found large applications in point patterns' analysis (Dale, 2000; Wiegand and Moloney, 2013; Tovo, Formentin et al., 2016). In Section 1.1.3 we have introduced the concept of the contact distribution of a stationary point process, which is the location-related statistics describing the distance of the process from an arbitrary point $u \in \mathbb{R}^2$. Let us now introduce the corresponding point-related statistic.

Definition 1.20. Let $\boldsymbol{X}$ be a stationary point process defined in $\mathbb{R}^2$, with Palm probability measure $\mathbb{P}^0(\cdot)$ and constant intensity function $\lambda_{\boldsymbol{X}}(u) \equiv \lambda_{\boldsymbol{X}}$, $u \in \mathbb{R}^2$. The **nearest-neighbour distance distribution function** $G_{\boldsymbol{X}}(\cdot)$ is defined as:

$$G_{\boldsymbol{X}}(r) = \mathbb{P}^0(\mathcal{N}_{\boldsymbol{X}}(\mathcal{B}_r) > 1). \tag{1.7}$$

In other words, $G_{\boldsymbol{X}}(r)$ is the cumulative distribution function of the distances between an arbitrary point of the process and its nearest neighbour point.
Combining the location-related contact distribution function $F_{\boldsymbol{X}}(r)$ and the point-related nearest-neighbour distance distribution function $G_{\boldsymbol{X}}(r)$ we obtain another second order statistics called the J-function, defined as

$$J_{\boldsymbol{X}}(r) = \frac{1 - G_{\boldsymbol{X}}(r)}{1 - F_{\boldsymbol{X}}(r)},$$

which has found applications in the field of point processes (Baddeley et al., 2007; Illian et al., 2008).
The Palm distribution let us define another fundamental distribution function widely used in point patterns' analysis, which is the *reduced second moment function*.

Definition 1.21. Let $\boldsymbol{X}$ be a stationary point process defined in $\mathbb{R}^2$, with Palm probability measure $\mathbb{P}^0(\cdot)$ and constant intensity function $\lambda_{\boldsymbol{X}}(u) \equiv \lambda_{\boldsymbol{X}}$, $u \in \mathbb{R}^2$. The **reduced second moment function** or **K-function** of the point process $\boldsymbol{X}$, $K_{\boldsymbol{X}}(\cdot)$, is defined as:

$$K_{\boldsymbol{X}}(r) = \frac{\mathbb{E}[\mathcal{N}_{\boldsymbol{X}}(\mathcal{B}_r)|0 \in \boldsymbol{X}] - 1}{\lambda_{\boldsymbol{X}}}, \tag{1.8}$$

where $\mathbb{E}[\mathcal{N}_{\boldsymbol{X}}(\mathcal{B}_r)|0 \in \boldsymbol{X}]$ is the conditional expectation of $\mathcal{N}_{\boldsymbol{X}}(\mathcal{B}_r)$, given that 0 is a point of the point process $\boldsymbol{X}$.

Intuitively speaking, the K-function evaluated at r and multiplied by the intensity function gives the mean extra number of points within a ball centred at a process point's location and of radius r.
The K-function was firstly introduced by Ripley (Ripley, 1976, 1977, 2005) – indeed, sometimes it is referred as Ripley's K-function – and it is strictly connected to the so-called *reduced second moment measure* of $\boldsymbol{X}$ (Baddeley et al., 2007; Cressie, 2015; Ornstein and Zernike, 1914). It can be proved (Ripley, 1976, 2005; Baddeley et al., 2007) that the K-function of a stationary point process $\boldsymbol{X}$ defined on $\mathbb{R}^2$ can also be written in terms of the pair correlation function $g_{\boldsymbol{X}}(\cdot,\cdot)$ as follows:

$$K_{\boldsymbol{X}}(r) = \int_{\mathcal{B}_r} g_{\boldsymbol{X}}(0,u)\mathrm{d}u. \tag{1.9}$$

If $\boldsymbol{X}$ is also isotropic, from eq. (1.9) we deduce the following additional relation

$$g_{\boldsymbol{X}}(r) = \frac{1}{2\pi r}\frac{\mathrm{d}}{\mathrm{d}r}K_{\boldsymbol{X}}(r).$$

Other important second-order statistics introduced in literature are the Besag's L-function (Besag, 1977; Illian et al., 2008; Wiegand and Moloney, 2013) and Schiffers's K_2 index (Schiffers et al., 2008; Tovo, Formentin et al., 2016). They will be studied in more details in Chapter 2. Here we only give their formal definitions.

Definition 1.22. Let $\boldsymbol{X}$ be a stationary point process defined in $\mathbb{R}^2$ and let $K_{\boldsymbol{X}}(\cdot)$ be its reduced second moment function. The **L-function** of the point process $\boldsymbol{X}$, $L_{\boldsymbol{X}}(\cdot)$, is defined as:

$$L_{\boldsymbol{X}}(r) = \sqrt{\frac{K(r)}{\pi}} - r. \tag{1.10}$$

Definition 1.23. Let $\boldsymbol{X}$ be a stationary and isotropic point process defined in $\mathbb{R}^2$ and let $g_{\boldsymbol{X}}(\cdot)$ be its pair correlation function. Schiffers's $\boldsymbol{K_2}$**-index** of the point process $\boldsymbol{X}$, $K_{2,\boldsymbol{X}}(\cdot)$, is defined as:

$$K_{2,\boldsymbol{X}}(r) = \frac{\mathrm{d}}{\mathrm{d}r} g(r), \tag{1.11}$$

given that such derivative exists.

1.4 Superposition of point processes

Given a spatial point process $\boldsymbol{X}$, there are different operations which can be used to transform it into another process $\boldsymbol{Y}$ which may better model a spatial pattern under study. Such operations include mapping, thinning, clustering and superposition (Baddeley et al., 2007).
The first one consists in applying a fixed transformation (e.g. a translation or a rotation) to each point of $\boldsymbol{X}$. The second one is obtained by removing some points of the process according to a precise rule. In the third case, each point of $\boldsymbol{X}$ is replaced by a random set of points (cluster) representing a possibly different point process (see Section 1.5.2 for an example).
Here we are interested in investigating the last operation, that is the superposition of spatial point processes.
Indeed, up to now we have worked with only one point process. Nevertheless, ecological datasets usually consist of information (coordinates, diameter, status, etc.) about trees which belong to many different species of plants. To each tree residing within the surveyed region $\mathcal{W}$ of the forest, which can be thought as a closed subset of $\mathbb{R}^2$, we thus have an additional label s indicating its species. In this case we can assume that the individuals' locations within $\mathcal{W}$ of each species s are a realisation of a particular point process, $\boldsymbol{X}_s$. We now wish to investigate the properties of the so-called *superposed* process, which can be obtained by looking at all the points in $\mathcal{W}$, regardless of their label s (see figure 1.1).
Let us give its formal definition. In what follows we only consider the superposition of two spatial point processes. Nevertheless, all the results can be extended to a superposition of an arbitrary number of point processes.

Definition 1.24. Let $\boldsymbol{X}$ and $\boldsymbol{Y}$ be spatial point processes defined in $\mathbb{R}^2$ such that $x \neq y$, for all $x \in \boldsymbol{X}$ and for all $y \in \boldsymbol{Y}$. Let $\mathcal{N}_{\boldsymbol{X}}$ and $\mathcal{N}_{\boldsymbol{Y}}$ be the random measures associated to the two processes. The **superposed process** $\boldsymbol{X} \cup \boldsymbol{Y}$ of $\boldsymbol{X}$ and $\boldsymbol{Y}$ is the spatial point process consisting of all the points in the union of the two random sets $\mathcal{S}_{\boldsymbol{X}}$ and $\mathcal{S}_{\boldsymbol{Y}}$ and with random measure $\mathcal{N}_{\boldsymbol{X}\cup\boldsymbol{Y}}$ defined by

$$\mathcal{N}_{\boldsymbol{X}\cup\boldsymbol{Y}}(B) = \mathcal{N}_{\boldsymbol{X}}(B) + \mathcal{N}_{\boldsymbol{Y}}(B),$$

for all $B \subset \mathbb{R}^2$.

In the particular case of $\mathcal{N}_{\boldsymbol{X}}$ and $\mathcal{N}_{\boldsymbol{Y}}$ independent, it is useful to describe the superposed process through its capacity functional $T_{\boldsymbol{X}\cup\boldsymbol{Y}}(\cdot)$. Indeed, under the independence assumption, we have that, given $B \subset \mathbb{R}^2$:

$$T_{\boldsymbol{X}\cup\boldsymbol{Y}}(B) = 1 - [1 - T_{\boldsymbol{X}}(B)][1 - T_{\boldsymbol{Y}}(B)]. \tag{1.12}$$

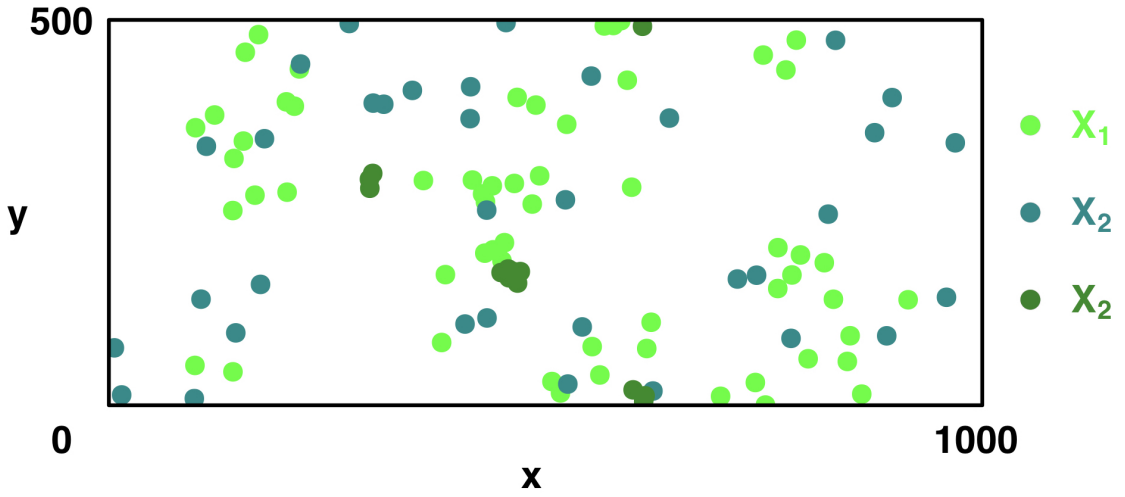


Figure 1.1: Example of a superposed process. Given the realisations of the single processes $\boldsymbol{X}_s$, $s \in \{1,2,3\}$ (different colours in the graphics) within the surveyed window (here a 1000×500 rectangle), the corresponding realisation of the superposed process $\boldsymbol{X} = \boldsymbol{X}_1 \cup \boldsymbol{X}_2 \cup \boldsymbol{X}_3$ consists of all the points belonging to $\boldsymbol{X}_s$, $s \in \{1,2,3\}$, regardless of their process' label.

From now on we will assume that $\boldsymbol{X}$ and $\boldsymbol{Y}$ are independent.
Let $\nu_{\boldsymbol{X}}(\cdot)$ and $\nu_{\boldsymbol{Y}}(\cdot)$ be their intensity measures. Then, by linearity of the expectation, we can find the intensity measure of the superposed process $\boldsymbol{X} \cup \boldsymbol{Y}$:

$$\begin{aligned}\nu_{\boldsymbol{X}\cup\boldsymbol{Y}}(B) &= \mathbb{E}[\mathcal{N}_{\boldsymbol{X}\cup\boldsymbol{Y}}(B)] = \mathbb{E}[\mathcal{N}_{\boldsymbol{X}}(B) + \mathcal{N}_{\boldsymbol{Y}}(B)] \\ &= \mathbb{E}[\mathcal{N}_{\boldsymbol{X}}(B)] + \mathbb{E}[\mathcal{N}_{\boldsymbol{Y}}(B)] = \nu_{\boldsymbol{X}}(B) + \nu_{\boldsymbol{Y}}(B),\end{aligned}$$

for every $B \subset \mathbb{R}^2$.
Moreover, if both the point processes admit intensity functions $\lambda_{\boldsymbol{X}}(\cdot)$ and $\lambda_{\boldsymbol{Y}}(\cdot)$, respectively, then also their superposition admits it:

$$\nu_{\boldsymbol{X}\cup\boldsymbol{Y}}(B) = \nu_{\boldsymbol{X}}(B) + \nu_{\boldsymbol{Y}}(B) = \int_B \lambda_{\boldsymbol{X}}(u)\mathrm{d}u + \int_B \lambda_{\boldsymbol{Y}}(u)\mathrm{d}u = \int_B \lambda_{\boldsymbol{X}\cup\boldsymbol{Y}}(u)\mathrm{d}u,$$

where we have set

$$\lambda_{\boldsymbol{X}\cup\boldsymbol{Y}} = \lambda_{\boldsymbol{X}} + \lambda_{\boldsymbol{Y}}. \tag{1.13}$$

We call $\lambda_{\boldsymbol{X}\cup\boldsymbol{Y}}(\cdot)$ the intensity function of the process $\boldsymbol{X} \cup \boldsymbol{Y}$.
Let us remark that, by eq. (1.13), if the original processes $\boldsymbol{X}$ and $\boldsymbol{Y}$ are homogeneous then so it is their superposition. Indeed, we have that

$$\nu_{\boldsymbol{X}\cup\boldsymbol{Y}}(B) = \int_B [\lambda_{\boldsymbol{X}}(u) + \lambda_{\boldsymbol{Y}}(u)]\mathrm{d}u = (\lambda_{\boldsymbol{X}} + \lambda_{\boldsymbol{Y}}) \cdot \mu(B) = \lambda_{\boldsymbol{X}\cup\boldsymbol{Y}} \cdot \mu(B),$$

where we have denoted with $\mu(B)$ the Lebesgue measure of the subset $B \subset \mathbb{R}^2$.
We wish now to define the second-order measures of the superposed process $\boldsymbol{X} \cup \boldsymbol{Y}$. In order to do that, let us firstly compute the quantity $\mathbb{E}[\mathcal{N}_{\boldsymbol{X}\cup\boldsymbol{Y}}(B_1)\mathcal{N}_{\boldsymbol{X}\cup\boldsymbol{Y}}(B_2)]$, with

$B_1,\ B_2 \subset \mathbb{R}^2$.

$$\begin{aligned}\mathbb{E}[\mathcal{N}_{\boldsymbol{X}\cup\boldsymbol{Y}}(B_1)\mathcal{N}_{\boldsymbol{X}\cup\boldsymbol{Y}}(B_2)] &= \mathbb{E}[(\mathcal{N}_{\boldsymbol{X}}(B_1)+\mathcal{N}_{\boldsymbol{Y}}(B_1))(\mathcal{N}_{\boldsymbol{X}}(B_2)+\mathcal{N}_{\boldsymbol{Y}}(B_2))]\\ &= \mathbb{E}[\mathcal{N}_{\boldsymbol{X}}(B_1)\mathcal{N}_{\boldsymbol{X}}(B_2)] + \mathbb{E}[\mathcal{N}_{\boldsymbol{X}}(B_1)\mathcal{N}_{\boldsymbol{Y}}(B_2)]+\\ &\quad + \mathbb{E}[\mathcal{N}_{\boldsymbol{X}}(B_2)\mathcal{N}_{\boldsymbol{Y}}(B_1)] + \mathbb{E}[\mathcal{N}_{\boldsymbol{Y}}(B_1)\mathcal{N}_{\boldsymbol{Y}}(B_2)]\\ &= \mathbb{E}[\mathcal{N}_{\boldsymbol{X}}(B_1)\mathcal{N}_{\boldsymbol{X}}(B_2)] + \mathbb{E}[\mathcal{N}_{\boldsymbol{X}}(B_1)]\mathbb{E}[\mathcal{N}_{\boldsymbol{Y}}(B_2)]+\\ &\quad + \mathbb{E}[\mathcal{N}_{\boldsymbol{X}}(B_2)\mathbb{E}[\mathcal{N}_{\boldsymbol{Y}}(B_1)] + \mathbb{E}[\mathcal{N}_{\boldsymbol{Y}}(B_1)]\mathcal{N}_{\boldsymbol{Y}}(B_2)].\end{aligned}$$

From the result above we can get the second moment measure $\nu_{2,\boldsymbol{X}\cup\boldsymbol{Y}}(\cdot)$ of the process $\boldsymbol{X}\cup\boldsymbol{Y}$:

$$\begin{aligned}\nu_{2,\boldsymbol{X}\cup\boldsymbol{Y}}(B_1\times B_2) = \nu_{2,\boldsymbol{X}}(B_1\times B_2) + \nu_{2,\boldsymbol{X}}(B_1\times B_2)+\\ + \nu_{\boldsymbol{X}}(B_1)\nu_{\boldsymbol{Y}}(B_2) + \nu_{\boldsymbol{X}}(B_2)\nu_{\boldsymbol{Y}}(B_1),\end{aligned} \tag{1.14}$$

for $B_1,\ B_2 \subset \mathbb{R}^2$.
By subtracting the quantity

$$\mathbb{E}[\mathcal{N}_{\boldsymbol{X}\cup\boldsymbol{Y}}(B_1\cap B_2)] = \mathbb{E}[\mathcal{N}_{\boldsymbol{X}}(B_1\cap B_2)] + \mathbb{E}[\mathcal{N}_{\boldsymbol{Y}}(B_1\cap B_2)]$$

to eq. (1.14), we obtain the second factorial moment measure of $\boldsymbol{X}\cup\boldsymbol{Y}$:

$$\begin{aligned}\nu_{[2],\boldsymbol{X}\cup\boldsymbol{Y}}(B_1\times B_2) = \nu_{[2],\boldsymbol{X}}(B_1\times B_2) + \nu_{[2],\boldsymbol{X}}(B_1\times B_2)+\\ + \nu_{\boldsymbol{X}}(B_1)\nu_{\boldsymbol{Y}}(B_2) + \nu_{\boldsymbol{X}}(B_2)\nu_{\boldsymbol{Y}}(B_1).\end{aligned} \tag{1.15}$$

If we now assume that $\boldsymbol{X}$ and $\boldsymbol{Y}$ admit second moment densities $\rho_{\boldsymbol{X}}(\cdot,\cdot)$ and $\rho_{\boldsymbol{Y}}(\cdot,\cdot)$ we can rewrite eq. (1.15) as follows:

$$\begin{aligned}\nu_{[2],\boldsymbol{X}\cup\boldsymbol{Y}}(B_1\times B_2) &= \int_{B_1}\int_{B_2}\rho_{\boldsymbol{X}}(u,v)\mathrm{d}u\mathrm{d}v + \int_{B_1}\int_{B_2}\rho_{\boldsymbol{Y}}(u,v)\mathrm{d}u\mathrm{d}v+\\ &\quad + \int_{B_1}\lambda_{\boldsymbol{X}}(u)\mathrm{d}u\int_{B_2}\lambda_{\boldsymbol{Y}}(v)\mathrm{d}v + \int_{B_2}\lambda_{\boldsymbol{X}}(v)\mathrm{d}v\int_{B_1}\lambda_{\boldsymbol{Y}}(u)\mathrm{d}u\\ &= \int_{B_1}\int_{B_2}\rho_{\boldsymbol{X}}(u,v)\mathrm{d}u\mathrm{d}v + \int_{B_1}\int_{B_2}\rho_{\boldsymbol{Y}}(u,v)\mathrm{d}u\mathrm{d}v+\\ &\quad + \int_{B_1}\int_{B_2}\lambda_{\boldsymbol{X}}(u)\lambda_{\boldsymbol{Y}}(v)\mathrm{d}u\mathrm{d}v + \int_{B_2}\int_{B_1}\lambda_{\boldsymbol{X}}(v)\lambda_{\boldsymbol{Y}}(u)\mathrm{d}u\mathrm{d}v\\ &= \int_{B_1}\int_{B_2}[\rho_{\boldsymbol{X}}(u,v)+\rho_{\boldsymbol{Y}}(u,v)+\lambda_{\boldsymbol{X}}(u)\lambda_{\boldsymbol{Y}}(v)+\lambda_{\boldsymbol{X}}(v)\lambda_{\boldsymbol{Y}}(u)]\mathrm{d}u\mathrm{d}v,\end{aligned} \tag{1.16}$$

where in the second equality we resort to Fubini's theorem.
Eq. (1.16) implies that also the process $\boldsymbol{X}\cup\boldsymbol{Y}$ admits second moment density. Indeed we can define it as follows:

$$\rho_{\boldsymbol{X}\cup\boldsymbol{Y}}(u,v) = \rho_{\boldsymbol{X}}(u,v)+\rho_{\boldsymbol{Y}}(u,v)+\lambda_{\boldsymbol{X}}(u)\lambda_{\boldsymbol{Y}}(v)+\lambda_{\boldsymbol{X}}(v)\lambda_{\boldsymbol{Y}}(u), \tag{1.17}$$

for any $u, v \in \mathbb{R}^2$.
We can thus define also the pair correlation function of $\boldsymbol{X}\cup\boldsymbol{Y}$ as the ratio

$$g_{\boldsymbol{X}\cup\boldsymbol{Y}}(u,v) = \frac{\rho_{\boldsymbol{X}\cup\boldsymbol{Y}}(u,v)}{\lambda_{\boldsymbol{X}\cup\boldsymbol{Y}}(u)\lambda_{\boldsymbol{X}\cup\boldsymbol{Y}}(v)}. \tag{1.18}$$

For stationary and isotropic processes $\boldsymbol{X}$ and $\boldsymbol{Y}$ with constant densities $\lambda_{\boldsymbol{X}}$ and $\lambda_{\boldsymbol{Y}}$, respectively, eqs. (1.17) and (1.18) take the form

$$\rho_{\boldsymbol{X}\cup\boldsymbol{Y}}(r) = \rho_{\boldsymbol{X}}(r) + \rho_{\boldsymbol{Y}}(r) + 2\lambda_{\boldsymbol{X}}\lambda_{\boldsymbol{Y}}$$

and

$$g_{\boldsymbol{X}\cup\boldsymbol{Y}}(r) = \frac{\rho_{\boldsymbol{X}\cup\boldsymbol{Y}}(r)}{(\lambda_{\boldsymbol{X}}+\lambda_{\boldsymbol{Y}})^2} = \frac{\lambda_{\boldsymbol{X}}^2 g_{\boldsymbol{X}}(r) + \lambda_{\boldsymbol{Y}}^2 g_{\boldsymbol{Y}}(r) + 2\lambda_{\boldsymbol{X}}\lambda_{\boldsymbol{Y}}}{(\lambda_{\boldsymbol{X}}+\lambda_{\boldsymbol{Y}})^2}.$$

Moreover, let $1_{\boldsymbol{X}}^B$ and $v_{\boldsymbol{X}}^B$ be the presence indicator and the vacancy indicator of the point process $\boldsymbol{X}$ associated with $B \subset \mathbb{R}^2$ and let $1_{\boldsymbol{Y}}^B$ and $v_{\boldsymbol{Y}}^B$ be the corresponding random variables of $\boldsymbol{Y}$. Then, the presence indicator and the vacancy indicator of $\boldsymbol{X}\cup\boldsymbol{Y}$ are respectively given by

$$1_{\boldsymbol{X}\cup\boldsymbol{Y}}^B = \min(1, 1_{\boldsymbol{X}}^B + 1_{\boldsymbol{Y}}^B) \tag{1.19}$$

and

$$v_{\boldsymbol{X}\cup\boldsymbol{Y}}^B = v_{\boldsymbol{X}}^B \cdot v_{\boldsymbol{Y}}^B. \tag{1.20}$$

Let us now assume that both $\boldsymbol{X}$ and $\boldsymbol{Y}$ are also stationary. We wish compute the contact distribution function of $\boldsymbol{X}\cup\boldsymbol{Y}$.
We know that for a point process $\boldsymbol{X}$, its contact distribution function is defined as $F_{\boldsymbol{X}}(r) = \mathbb{P}(d(\boldsymbol{X},0) \leq r)$, i.e. the probability that there is at least one point of the process fall within the ball $\mathcal{B}_r$ centred in the origin and having radius r. Such probability is equivalent to $\mathbb{P}(\mathcal{N}_{\boldsymbol{X}}(\mathcal{B}_r) \geq 1) = T_{\boldsymbol{X}}(\mathcal{B}_r) = 1 - \mathbb{P}(\mathcal{N}_{\boldsymbol{X}}(\mathcal{B}_r) = 0)$. This last equality is particularly useful, since we have already computed the capacity functional of the superposed process. Indeed, from eq. (1.12), we get

$$\begin{aligned} F_{\boldsymbol{X}\cup\boldsymbol{Y}}(r) &= 1 - [1 - T_{\boldsymbol{X}}(\mathcal{B}_r)][1 - T_{\boldsymbol{Y}}(\mathcal{B}_r)] \\ &= 1 - [1 - F_{\boldsymbol{X}}(r)][1 - F_{\boldsymbol{Y}}(r)]. \end{aligned}$$

It remains to define the point-related statistics of $\boldsymbol{X}\cup\boldsymbol{Y}$. From Section 1.3.1, we know that we must firstly introduce the Palm distribution $\boldsymbol{P}_{\boldsymbol{X}\cup\boldsymbol{Y}}^x$ which specifies the probability that $\boldsymbol{X}\cup\boldsymbol{Y}$ satisfies some properties given that the point $x \in \mathbb{R}^2$ belongs to $\boldsymbol{X}\cup\boldsymbol{Y}$. Since we are working under the assumption that no points of $\boldsymbol{X}$ and $\boldsymbol{Y}$ are coincident, either $x \in \boldsymbol{X}$ or $x \in \boldsymbol{Y}$. Therefore, we can write the probability

$\boldsymbol{P}^x_{\boldsymbol{X}\cup\boldsymbol{Y}}$ conditioning on whether the first or the second case occurs:

$$\begin{aligned}
\boldsymbol{P}^x_{\boldsymbol{X}\cup\boldsymbol{Y}}(\mathcal{A}) &= \mathbb{P}(\mathcal{N}_{\boldsymbol{X}\cup\boldsymbol{Y}} \in \mathcal{A} | x \in \boldsymbol{X}\cup\boldsymbol{Y}) \\
&= \mathbb{P}(\mathcal{N}_{\boldsymbol{X}\cup\boldsymbol{Y}} \in \mathcal{A} | x \in \boldsymbol{X}, x \in \boldsymbol{X}\cup\boldsymbol{Y})\mathbb{P}(x \in \boldsymbol{X} | x \in \boldsymbol{X}\cup\boldsymbol{Y}) + \\
&\quad + \mathbb{P}(\mathcal{N}_{\boldsymbol{X}\cup\boldsymbol{Y}} \in \mathcal{A} | x \in \boldsymbol{Y}, x \in \boldsymbol{X}\cup\boldsymbol{Y})\mathbb{P}(x \in \boldsymbol{Y} | x \in \boldsymbol{X}\cup\boldsymbol{Y}) \\
&= \mathbb{P}(\mathcal{N}_{\boldsymbol{X}\cup\boldsymbol{Y}} \in \mathcal{A} | x \in \boldsymbol{X})\frac{\mathbb{P}(x \in \boldsymbol{X}, x \in \boldsymbol{X}\cup\boldsymbol{Y})}{\mathbb{P}(x \in \boldsymbol{X}\cup\boldsymbol{Y})} + \\
&\quad + \mathbb{P}(\mathcal{N}_{\boldsymbol{X}\cup\boldsymbol{Y}} \in \mathcal{A} | x \in \boldsymbol{Y})\frac{\mathbb{P}(x \in \boldsymbol{Y}, x \in \boldsymbol{X}\cup\boldsymbol{Y})}{\mathbb{P}(x \in \boldsymbol{X}\cup\boldsymbol{Y})} \\
&= \mathbb{P}(\mathcal{N}_{\boldsymbol{X}\cup\boldsymbol{Y}} \in \mathcal{A} | x \in \boldsymbol{X})\frac{\mathbb{P}(x \in \boldsymbol{X})}{\mathbb{P}(x \in \boldsymbol{X}\cup\boldsymbol{Y})} + \\
&\quad + \mathbb{P}(\mathcal{N}_{\boldsymbol{X}\cup\boldsymbol{Y}} \in \mathcal{A} | x \in \boldsymbol{Y})\frac{\mathbb{P}(x \in \boldsymbol{Y})}{\mathbb{P}(x \in \boldsymbol{X}\cup\boldsymbol{Y})} \\
&= \mathbb{P}(\mathcal{N}_{\boldsymbol{X}\cup\boldsymbol{Y}} \in \mathcal{A} | x \in \boldsymbol{X})\frac{\lambda_{\boldsymbol{X}}(x)}{\lambda_{\boldsymbol{X}\cup\boldsymbol{Y}}(x)} + \\
&\quad + \mathbb{P}(\mathcal{N}_{\boldsymbol{X}\cup\boldsymbol{Y}} \in \mathcal{A} | x \in \boldsymbol{Y})\frac{\lambda_{\boldsymbol{Y}}(x)}{\lambda_{\boldsymbol{X}\cup\boldsymbol{Y}}(x)}.
\end{aligned} \tag{1.21}$$

By using eq. (1.21) we can obtain all the point-related statistic of the superposed process $\boldsymbol{X}\cup\boldsymbol{Y}$, under the assumption of $\boldsymbol{X}$ and $\boldsymbol{Y}$ stationary and isotropic.
Let us start with the nearest neighbour distance distribution function $G_{\boldsymbol{X}\cup\boldsymbol{Y}}(r) = \mathbb{P}^0(\mathcal{N}_{\boldsymbol{X}\cup\boldsymbol{Y}}(\mathcal{B}_r) > 1) = 1 - \mathbb{P}^0(\mathcal{N}_{\boldsymbol{X}\cup\boldsymbol{Y}}(\mathcal{B}_r) = 1)$. We firstly need to compute the conditional probability $\mathbb{P}^0(\mathcal{N}_{\boldsymbol{X}\cup\boldsymbol{Y}}(\mathcal{B}_r) = 1)$:

$$\begin{aligned}
\mathbb{P}^0(\mathcal{N}_{\boldsymbol{X}\cup\boldsymbol{Y}}(\mathcal{B}_r) = 1) &= \mathbb{P}(\mathcal{N}_{\boldsymbol{X}\cup\boldsymbol{Y}}(\mathcal{B}_r) = 1 | 0 \in \boldsymbol{X})\frac{\lambda_{\boldsymbol{X}}(0)}{\lambda_{\boldsymbol{X}\cup\boldsymbol{Y}}(0)} + \\
&\quad + \mathbb{P}(\mathcal{N}_{\boldsymbol{X}\cup\boldsymbol{Y}}(\mathcal{B}_r) = 1 | 0 \in \boldsymbol{Y})\frac{\lambda_{\boldsymbol{Y}}(0)}{\lambda_{\boldsymbol{X}\cup\boldsymbol{Y}}(0)} \\
&= \mathbb{P}(\mathcal{N}_{\boldsymbol{X}}(\mathcal{B}_r) = 1 | 0 \in \boldsymbol{X})\mathbb{P}(\mathcal{N}_{\boldsymbol{Y}}(\mathcal{B}_R) = 0)\frac{\lambda_{\boldsymbol{X}}}{\lambda_{\boldsymbol{X}\cup\boldsymbol{Y}}} + \\
&\quad + \mathbb{P}(\mathcal{N}_{\boldsymbol{X}}(\mathcal{B}_r) = 0)\mathbb{P}(\mathcal{N}_{\boldsymbol{Y}}(\mathcal{B}_r) = 1 | 0 \in \boldsymbol{Y})\frac{\lambda_{\boldsymbol{Y}}}{\lambda_{\boldsymbol{X}\cup\boldsymbol{Y}}} \\
&= \mathbb{P}^0(\mathcal{N}_{\boldsymbol{X}}(\mathcal{B}_r) = 1)\mathbb{P}(\mathcal{N}_{\boldsymbol{Y}}(\mathcal{B}_r) = 0)\frac{\lambda_{\boldsymbol{X}}}{\lambda_{\boldsymbol{X}\cup\boldsymbol{Y}}} + \\
&\quad + \mathbb{P}^0(\mathcal{N}_{\boldsymbol{Y}}(\mathcal{B}_r) = 1)\mathbb{P}(\mathcal{N}_{\boldsymbol{X}}(\mathcal{B}_r) = 0)\frac{\lambda_{\boldsymbol{Y}}}{\lambda_{\boldsymbol{X}\cup\boldsymbol{Y}}} \\
&= [1 - G_{\boldsymbol{X}}(r)][1 - F_{\boldsymbol{Y}}(r)]\frac{\lambda_{\boldsymbol{X}}}{\lambda_{\boldsymbol{X}\cup\boldsymbol{Y}}} + \\
&\quad + [1 - G_{\boldsymbol{Y}}(r)][1 - F_{\boldsymbol{X}}(r)]\frac{\lambda_{\boldsymbol{X}}}{\lambda_{\boldsymbol{X}\cup\boldsymbol{Y}}}.
\end{aligned} \tag{1.22}$$

Thus, $G_{\boldsymbol{X}\cup\boldsymbol{Y}}(\cdot)$ is the complementary probability of 1.22.

For the J-function is is sufficient to compute the ratio

$$\begin{aligned} J_{\boldsymbol{X}\cup\boldsymbol{Y}}(r) &= \frac{1-G_{\boldsymbol{X}\cup\boldsymbol{Y}}(r)}{1-F_{\boldsymbol{X}\cup\boldsymbol{Y}}(r)} \\ &= \frac{\lambda_{\boldsymbol{X}}}{\lambda_{\boldsymbol{X}\cup\boldsymbol{Y}}} J_{\boldsymbol{X}}(r) + \frac{\lambda_{\boldsymbol{X}}}{\lambda_{\boldsymbol{X}\cup\boldsymbol{Y}}} J_{\boldsymbol{Y}}(r). \end{aligned}$$

Let us now compute the Ripley K-function by using its relation with the pair correlation function, eq. (1.9):

$$\begin{aligned} K_{\boldsymbol{X}\cup\boldsymbol{Y}}(r) &= \int_{\mathcal{B}_r} g_{\boldsymbol{X}\cup\boldsymbol{Y}}(0,u)\mathrm{d}u \\ &= \int_{\mathcal{B}_r} \frac{\lambda_{\boldsymbol{X}}^2 g_{\boldsymbol{X}}(0,u) + \lambda_{\boldsymbol{Y}}^2 g_{\boldsymbol{Y}}(0,u) + 2\lambda_{\boldsymbol{X}}\lambda_{\boldsymbol{Y}}}{(\lambda_{\boldsymbol{X}}+\lambda_{\boldsymbol{Y}})^2}\mathrm{d}u \\ &= \frac{1}{(\lambda_{\boldsymbol{X}}+\lambda_{\boldsymbol{Y}})^2}\left[\lambda_{\boldsymbol{X}}^2\int_{\mathcal{B}_r} g_{\boldsymbol{X}}(0,u)\mathrm{d}u + \lambda_{\boldsymbol{Y}}^2\int_{\mathcal{B}_r} g_{\boldsymbol{Y}}(0,u)\mathrm{d}u + 2\lambda_{\boldsymbol{X}}\lambda_{\boldsymbol{Y}}\int_{\mathcal{B}_r}\mathrm{d}u\right] \\ &= \frac{1}{(\lambda_{\boldsymbol{X}}+\lambda_{\boldsymbol{Y}})^2}\left[\lambda_{\boldsymbol{X}}^2 K_{\boldsymbol{X}}(0,u) + \lambda_{\boldsymbol{Y}}^2 K_{\boldsymbol{Y}}(0,u) + 2\lambda_{\boldsymbol{X}}\lambda_{\boldsymbol{Y}}\mu(\mathcal{B}_r)\right], \end{aligned}$$

where we denoted with $\mu(\mathcal{B}_r) = \int_{\mathcal{B}_r}\mathrm{d}u$ the Lebesgue measure of the 2-dimensional ball $\mathcal{B}_r$ centred in the origin and having radius r.

1.5 Examples of spatial point processes

1.5.1 Homogeneous Poisson process

The homogeneous Poisson process is sometimes called the *zero* or *completely random process* after its property that each point is stochastically independent to all the others (Daley and Vere-Jones, 2003). Therefore, in ecological theory, it represents the special case where there are no intraspecific spatial interactions between individuals. In particular, we say that a point pattern satisfies the *complete spatial randomness hypothesis* (CSR) if it is well described by a homogeneous Poisson process.
Let us thus formally introduce it.

Definition 1.25 (homogeneous Poisson process)**.** A spatial point process $\boldsymbol{X}$ defined in $\mathbb{R}^2$ is called a homogeneous Poisson process of intensity $\lambda_{\boldsymbol{X}}$ if the associated counting measure $\mathcal{N}_{\boldsymbol{X}}$ satisfies the following properties:

- for every compact subset $B \subset \mathbb{R}^2$ having Lebesgue measure $\mu(B)$, $\mathcal{N}_{\boldsymbol{X}}(B)$ has a Poisson distribution with mean $e^{-\lambda_{\boldsymbol{X}}\mu(B)}$:

$$\mathbb{P}(\mathcal{N}_{\boldsymbol{X}}(B) = n) = e^{-\lambda_{\boldsymbol{X}}\mu(B)}\frac{(\lambda_{\boldsymbol{X}}\mu(B))^n}{n!};$$

- given two disjoint compact subsets B_1 and B_2 of $\mathbb{R}^2$, $\mathcal{N}_{\boldsymbol{X}}(B_1)$ and $\mathcal{N}_{\boldsymbol{X}}(B_2)$ are independent.

The Poisson process is therefore a first example of a motion-invariant process.
Let us give the expressions of the fundamental statistics of the process.[1]
Since the random measure $\mathcal{N}_{\boldsymbol{X}}$ is Poisson distributed, the probability that the number of points of the process falling within a ball $\mathcal{B}_r$ of radius r centred in the origin is given by

$$\mathbb{P}(\mathcal{N}_{\boldsymbol{X}}(\mathcal{B}_r) = n) = \frac{e^{-\lambda_{\boldsymbol{X}}\pi r^2}(\lambda_{\boldsymbol{X}}\pi r^2)^n}{n!},$$

from which we get the expression of the contact distribution function

$$F_{\boldsymbol{X}}(r) = \mathbb{P}(\mathcal{N}_{\boldsymbol{X}}(\mathcal{B}_r) > 0) = 1 - \mathbb{P}(\mathcal{N}_{\boldsymbol{X}}(\mathcal{B}_r) = 0) = 1 - e^{-\lambda_{\boldsymbol{X}}\pi r^2}.$$

Let us now consider the basic second-order statistics. Because of the absence of spatial correlations, we have that the second moment density factorises into

$$\rho_{\boldsymbol{X}}(u, v) = \lambda_{\boldsymbol{X}}(u)\lambda_{\boldsymbol{X}}(v) = \lambda_{\boldsymbol{X}}^2.$$

It follows that the pair correlation function has constant value:

$$g_{\boldsymbol{X}}(x, y) \equiv 1. \tag{1.23}$$

On the left panel of figure 1.2 we show a realisation of a homogeneous Poisson process in a rectangular window (a 1000×500 area as for BCI and Pasoh forests' samples) and the corresponding empirical pair correlation function, which is, as expected, in good agreement with the theoretical prediction.
Anyway, in applications, the complete spatial randomness hypothesis mostly comes to fail due both to environmental variables and to seed dispersal mechanisms, which may lead to inhomogeneous pattern as well as clustered or dispersed structures.
On the right panel of figure 1.2 we insert the distribution within our Barro Colorado Island sample of the *Acalypha diversifolia* species. It is immediately visible that its pair correlation function is not constant, at least for distances smaller than a certain threshold, r_c. This is a common characteristic when looking at real species which shows the inadequacy of the Poisson model. Anyway, in exploratory analysis, the first step to analyse a database is to compare it with the zero-process.

1.5.2 Neyman-Scott processes

A much more interesting example are the Neyman-Scott processes. They have been firstly introduced in 1958 by Neyman and Scott (Neyman and Scott, 1958) to describe the locations of galaxies in space and have then found large applications in ecological theory due to its ability to model the clumping mechanism of plants' species in which daughter seeds are spread around a parent tree's location (Plotkin, Potts et al., 2000; Azaele, Cornell et al., 2012; Tovo, Formentin et al., 2016). A Neyman-Scott process is the result of three steps (see figure 1.3):

[1] As before, here we are working in $\mathbb{R}^2$ since we are interested in applications to ecological databases. The definition of a homogeneous Poisson process as well as all the presented results can be easily generalised to $\mathbb{R}^d$, with $d \geq 2$.

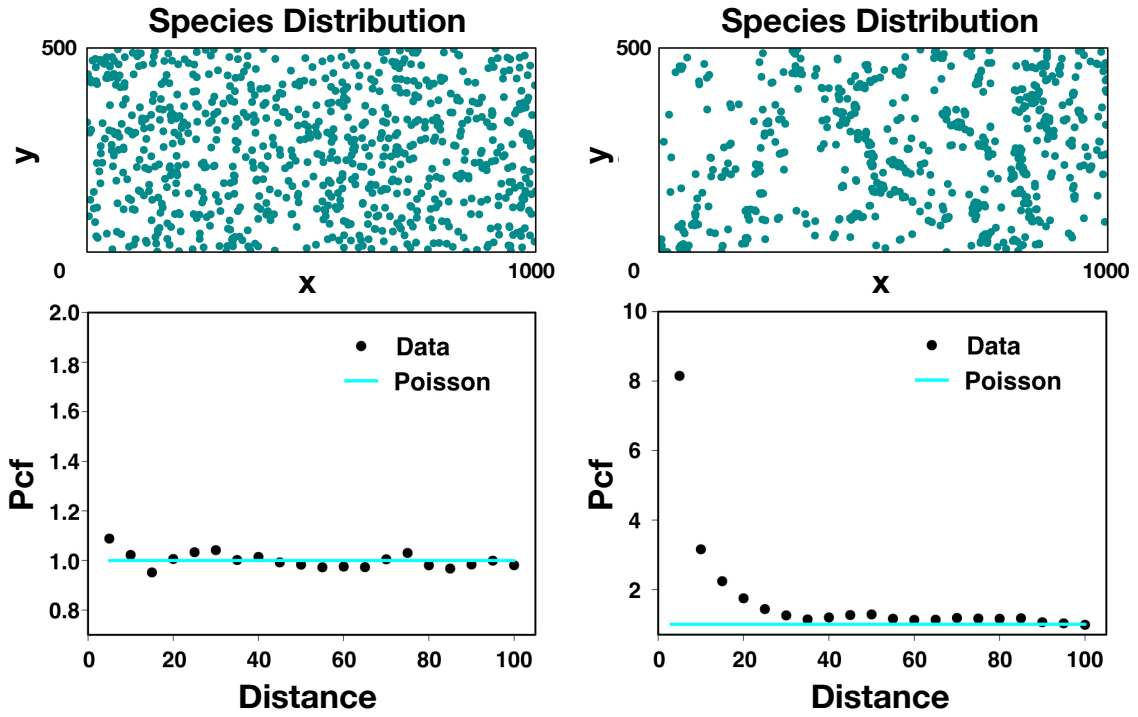


Figure 1.2: Pair correlation function for an artificial Poisson generated species (on the left) and for the *Acalypha diversifolia* in the BCI 50ha surveyed plot. While for the first one the pair correlation function shows an empirical behaviour well approximated by the theoretical constant straight line $y = 1$, in the second case it is characterised, up to a distance of around 40 m, a monotonically decreasing behaviour.

- Parent points are distributed according to a homogeneous Poisson process with intensity $\rho_{\boldsymbol{X}}$.
- To each parent a random number of offspring is assigned, drawn from a Poisson distribution of intensity $\mu_{\boldsymbol{X}}$.
- The offspring are identically and independently scattered around their parents with a fixed spatial probability density given by a radial function $d_{\gamma_{\boldsymbol{X}}}(\cdot)$, the so-called *dispersal kernel*, depending on some parameters $\gamma_{\boldsymbol{X}}$.

The resulting process is formed only by the offspring's locations.
Let us notice that with this model we are assuming that the formation of clusters is due to an isotropic local propagation of offspring from the parent, which is clearly an oversimplification of the complex natural mechanisms which actually determines their rise. Several generalisations of Neyman-Scott processes have been proposed in literature. For example, for the parent's generating process, one can consider, instead of taking $\lambda_{\boldsymbol{X}}$ as a constant function for distributing the parents, one may define it as a function of the space (in this case we speak about *heterogeneous Poisson processes* or *Poisson cluster processes* , see Wiegand and Moloney, 2013) or even you may threat them as a realisation of another completely different stochastic process (in this case we speak about *Cox-processes* or doubly stochastic Poisson processes, see Cox and Isham, 1980; Diggle, 2003). Another possible generalisation

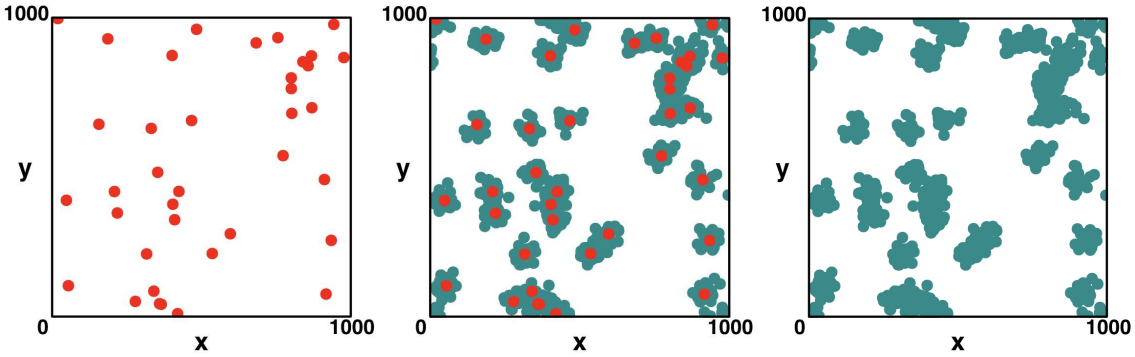


Figure 1.3: The three-steps procedure generating a Neyman-Scott process on a 1000×1000 window. Left panel: parent points (red dots) are randomly distributed according to a homogeneous Poisson process. Middle panel: a random number of points is radially distributed around each parent. Right panel: Parent points are removed from the plot and the locations of their offspring solely are considered.

is to consider, when distributing the offspring, different types of clusters, which may be characterised by different sizes (see, e.g. Stoyan and Stoyan, 1996; Tanaka et al., 2008). Finally, instead of considering only the first generation of offspring, one can extend the model by considering a second generation, where each daughter tree gives birth to its own cluster according to the same point process but with possibly different parameters (Wiegand, Gunatilleke et al., 2007; Stoyan and Stoyan, 1996; Diggle, 2003; Watson et al., 2007). This latter model can then be further extended by considering the superposition of the first n^{th} generations obtained by this reproductive mechanism (Shimatani, 2002, 2010). Here, for our purposes, we will limit ourselves to the case of homogeneous Poisson cluster processes.
Henceforth we will denote with $\boldsymbol{X}_p$ the parents' process and with $\boldsymbol{X}_c$ a representative cluster's process with parent's location in the origin. A general cluster $\boldsymbol{X}_c^x$ of parent $x \in \boldsymbol{X}_p$ can thus be obtained by translating the representative cluster's centre to its parent's location x:

$$\boldsymbol{X}_c^x = \boldsymbol{X}_c + x.$$

The final process is then given by the superposition of all the clusters:

$$\boldsymbol{X} = \bigcup_{x \in \boldsymbol{X}_p} \boldsymbol{X}_c^x.$$

We call this approach *homogeneous independent clustering*.
Independently of the daughters' distribution, the intensity function of the superposed process $\boldsymbol{X}$ is given by the product

$$\lambda_{\boldsymbol{X}} = \rho_{\boldsymbol{X}} \cdot \mu_{\boldsymbol{X}}. \tag{1.24}$$

In order to compute the second order characteristics of $\boldsymbol{X}$, we firstly need some more notations.
Given the radially symmetric density function $d_{\gamma_{\boldsymbol{X}}}(\cdot)$ describing the locations of the offspring around each parent's coordinates, we can compute the probability density function of the vector difference between the positions of two offspring from the same

parent (Chiu et al., 2013). Indeed, let $u \in \mathbb{R}^2$ be the the position of an offspring belonging to the representative cluster ($u \in \boldsymbol{X}_c$). Then the probability density that another point will be found in a position $v \in \mathbb{R}^2$ such that the vector difference is $u - v = y$ is given by the product

$$d_{\gamma_{\boldsymbol{X}}}(u)d_{\gamma_{\boldsymbol{X}}}(v) = d_{\gamma_{\boldsymbol{X}}}(u)d_{\gamma_{\boldsymbol{X}}}(u-y). \tag{1.25}$$

Therefore, the density probability of v can be obtained as the convolution of the bi-variate dispersal kernel $d_{\gamma_{\boldsymbol{X}}}$, $d_{\gamma_{\boldsymbol{X}}} \star d_{\gamma_{\boldsymbol{X}}}$, i.e. by integrating the quantity (1.25) over the position of the first offspring u:

$$h_{\gamma_{\boldsymbol{X}}}(y) = \int_{\mathbb{R}^2} d_{\gamma_{\boldsymbol{X}}}(u)d_{\gamma_{\boldsymbol{X}}}(u-y)\mathrm{d}u,$$

Let us remark that, since $d_{\gamma_{\boldsymbol{X}}}(\cdot)$ is a radial function, so is $h_{\gamma_{\boldsymbol{X}}}(\cdot)$.
Since we are studying isotropic processes, it is easier to work with polar coordinates. Thus the random location of an offspring $u = (u_1, u_2)$ will henceforth be described by the pair (r, θ), where $r \in (0, \infty)$ and $\theta \in [0, \pi]$ are the random angle and modulus, respectively, of the ray connecting the representative parent's coordinates (the origin) with (u_1, u_2) in the Cartesian system. We will then indicate $d^{\mathrm{pol}}_{\gamma_{\boldsymbol{X}}}(\cdot)$ and $h^{\mathrm{pol}}_{\gamma_{\boldsymbol{X}}}(\cdot)$ the density functions in polar coordinates of the distance between a daughter and its parent and between two daughters of the representative cluster, respectively[2]. Finally, let $H^{\mathrm{pol}}_{\gamma_{\boldsymbol{X}}}(\cdot)$ be the distribution function of the random distance between offspring, having density $h^{\mathrm{pol}}_{\gamma_{\boldsymbol{X}}}(r)$.
Then, the K-function and the pair correlation function of $\boldsymbol{X}$ are respectively given by (Illian et al., 2008; Chiu et al., 2013)

$$K_{\boldsymbol{X}}(r) = \pi r^2 + \frac{\mu_{\boldsymbol{X}}}{\lambda_{\boldsymbol{X}}}\mathbb{E}[n(n-1)]H^{\mathrm{pol}}_{\gamma_{\boldsymbol{X}}}(r),$$

$$g_{\boldsymbol{X}}(r) = 1 + \frac{1}{\lambda_{\boldsymbol{X}}\mu_{\boldsymbol{X}}}\mathbb{E}[n(n-1)]\frac{h^{\mathrm{pol}}_{\gamma_{\boldsymbol{X}}}(r)}{2\pi r}. \tag{1.26}$$

Since the number of points within each cluster is distributed according to a Poisson density function of intensity $\mu_{\boldsymbol{X}}$, we have that

$$\mathbb{P}(\mathcal{N}_{\boldsymbol{X}_c} = n) = e^{-\mu_{\boldsymbol{X}}}\frac{\mu^n_{\boldsymbol{X}}}{n!},$$

from which

$$\mathbb{E}[n(n-1)] = \sum_{n=2}^{\infty}\mathbb{P}(\mathcal{N}_{\boldsymbol{X}_c} = n)n(n-1) = e^{-\mu_{\boldsymbol{X}}}\sum_{n=2}^{\infty}\frac{\mu^n_{\boldsymbol{X}}}{(n-2)!} = e^{-\mu_{\boldsymbol{X}}}e^{\mu_{\boldsymbol{X}}}\mu^2_{\boldsymbol{X}} = \mu^2_{\boldsymbol{X}}.$$

Therefore we get the following expressions for the functions $K_{\boldsymbol{X}}(\cdot)$ and $g_{\boldsymbol{X}}(\cdot)$:

$$K_{\boldsymbol{X}}(r) = \pi r^2 + \frac{1}{\rho_{\boldsymbol{X}}}\mu^2_{\boldsymbol{X}}H^{\mathrm{pol}}_{\boldsymbol{X}}(r), \tag{1.27}$$

[2]Let us remark that, if $f(x)$is a radially symmetric function, then the following relation holds: $f(r) = \frac{1}{2\pi r}f^{\mathrm{pol}}(r)$, where $r = |x|$.

$$g_{\boldsymbol{X}}(r) = 1 + \frac{1}{\rho_{\boldsymbol{X}}\mu_{\boldsymbol{X}}^2}\mu_{\boldsymbol{X}}^2\frac{h_{\boldsymbol{X}}^{\text{pol}}(r)}{2\pi r} = 1 + \frac{1}{\rho_{\boldsymbol{X}}}\frac{h_{\boldsymbol{X}}^{\text{pol}}(r)}{2\pi r}. \tag{1.28}$$

It only remains to compute the distribution $H_{\gamma_{\boldsymbol{X}}}^{\text{pol}}(\cdot)$ and its density $h_{\gamma_{\boldsymbol{X}}}^{\text{pol}}(\cdot)$ as functions of $d_{\gamma_{\boldsymbol{X}}}^{\text{pol}}$ (see Stoyan and Stoyan, 1994). By definition, $H_{\gamma_{\boldsymbol{X}}}^{\text{pol}}(\cdot)$ is the probability that the distance between two individuals of the representative cluster $\boldsymbol{X}_c$ is less or equal than r. In order to compute it, we condition on the position of the two offspring u and v, having polar coordinates (r_u, θ_u) and (r_v, θ_v), respectively:

$$H_{\gamma_{\boldsymbol{X}}}^{\text{pol}}(r) = \int_0^\infty \int_0^\infty H_{\gamma_{\boldsymbol{X}}}^{\text{pol}}(r|r_u, r_v) d_{\gamma_{\boldsymbol{X}}}^{\text{pol}}(r_u) d_{\gamma_{\boldsymbol{X}}}^{\text{pol}}(r_v) \mathrm{d}r_u \mathrm{d}r_v.$$

By symmetry, we can rewrite the above integral by considering the case $r_u \leq r_v$

$$H_{\gamma_{\boldsymbol{X}}}^{\text{pol}}(r) = 2\int_0^\infty \int_0^\infty H_{\gamma_{\boldsymbol{X}}}^{\text{pol}}(r|r_u \leq r_v) d_{\gamma_{\boldsymbol{X}}}^{\text{pol}}(r_u) d_{\gamma_{\boldsymbol{X}}}^{\text{pol}}(r_v) \mathrm{d}r_u \mathrm{d}r_v. \tag{1.29}$$

Let now ϕ be the random angle between the two rays connecting the origin and the daughter points u and v and let ρ be the modulus of the vector distance $u - v$. Then, the following relation must hold

$$r_v - r_u \leq \rho \leq r_u + r_v,$$

so that $H_{\gamma_{\boldsymbol{X}}}^{\text{pol}}(r|r_u \leq r_v) = 0$ if $r \leq r_v - r_u$ and $H_{\gamma_{\boldsymbol{X}}}^{\text{pol}}(r|r_u \leq r_v) = 1$ if $r \geq r_u + r_v$. Let us then consider the case where $r_v - r_u \leq r \leq r_u + r_v$. Moreover, by symmetry, let us restrict to the case $\phi \in [0, \pi]$. By Carnot Theorem, we have that

$$\rho^2 = r_u^2 + r_v^2 - 2r_u r_v \cos\phi,$$

Therefore

$$\phi = \arccos\left(\frac{r_u^2 + r_v^2 - \rho^2}{2r_u r_v}\right).$$

Since ρ is an increasing function with ϕ, it follows that, $\rho \leq r$ if and only if $\phi \leq \arccos\left(\frac{r_u^2 + r_v^2 - r^2}{2r_u r_v}\right)$. Thus, given ϕ and $r_u \leq r_v$, we have that

$$H_{\gamma_{\boldsymbol{X}}}(r|\phi, r_u \leq r_v) = \mathbb{I}\left(\phi \leq \arccos\left(\frac{r_u^2 + r_v^2 - r^2}{2r_u r_v}\right)\right),$$

where $\mathbb{I}(\mathcal{C})$ is the indicator function, taking value 1 if condition $\mathcal{C}$ holds, and 0 otherwise.
Then, conditioning on ϕ, the probability $H_{\gamma_{\boldsymbol{X}}}^{\text{pol}}(r|r_u \leq r_v)$ can be computed as

$$\begin{aligned}
H_{\gamma_{\boldsymbol{X}}}^{\text{pol}}(r|r_u \leq r_v) &= \int_0^\pi H_{\gamma_{\boldsymbol{X}}}^{\text{pol}}(r|\phi, r_u \leq r_v)\mathbb{P}(\phi)\mathrm{d}\phi \\
&= \frac{1}{\pi}\int_0^\pi H_{\gamma_{\boldsymbol{X}}}^{\text{pol}}(r|\phi, r_u \leq r_v)\mathrm{d}\phi \\
&= \frac{1}{\pi}\int_0^\pi \mathbb{I}\left(\phi \leq \arccos\left(\frac{r_u^2 + r_v^2 - \rho^2}{2r_u r_v}\right)\right)\mathrm{d}\phi \\
&= \frac{1}{\pi}\int_0^{\arccos\left(\frac{r_u^2 + r_v^2 - \rho^2}{2r_u r_v}\right)} \mathrm{d}\phi \\
&= \frac{1}{\pi}\arccos\left(\frac{r_u^2 + r_v^2 - \rho^2}{2r_u r_v}\right),
\end{aligned}$$

where the second equality follows by the fact that the random angle ϕ is uniformly distributed within $[0, \pi]$: $\mathbb{P}(\phi) = 1/\pi$.
Let us now split the integral in eq. (1.29):

$$\begin{aligned} H^{\text{pol}}_{\gamma_{\boldsymbol{X}}}(r) &= 2\int_0^{\infty}\int_0^{\infty} H^{\text{pol}}_{\gamma_{\boldsymbol{X}}}(r|r_u \le r_v) d^{\text{pol}}_{\gamma_{\boldsymbol{X}}}(r_u) d^{\text{pol}}_{\gamma_{\boldsymbol{X}}}(r_v) \mathrm{d}r_u \mathrm{d}r_v \\ &= 2\int_0^{r/2}\int_{r_u}^{\infty} H^{\text{pol}}_{\gamma_{\boldsymbol{X}}}(r|r_u, r_v) d^{\text{pol}}_{\gamma_{\boldsymbol{X}}}(r_u) d^{\text{pol}}_{\gamma_{\boldsymbol{X}}}(r_v) \mathrm{d}r_u \mathrm{d}r_v \\ &\quad + 2\int_{r/2}^{\infty}\int_{r_v}^{\infty} H^{\text{pol}}_{\gamma_{\boldsymbol{X}}}(r|r_u, r_v) d^{\text{pol}}_{\gamma_{\boldsymbol{X}}}(r_u) d^{\text{pol}}_{\gamma_{\boldsymbol{X}}}(r_v) \mathrm{d}r_u \mathrm{d}r_v. \end{aligned}$$

Let us compute separately the two integrals.
As for the first one, we can split the interval $[r_u, \infty]$ into the union $[r_u, r - r_u] \cup [r - r_u, r + r_u] \cup [r + r_u, \infty]$, where r is bigger than $r_v - r_u$, between $r_v - r_u$ and $r_u + r_v$ and lower than $r_v - r_u$, respectively. Consequently $H^{\text{pol}}_{\gamma_{\boldsymbol{X}}}(r|r_u, r_v)$ takes value equal to 1, $\arccos\left(\frac{r_u^2 + r_v^2 - \rho^2}{2r_u r_v}\right)$ and 0.
As for the second integral, we split $[r_u, \infty] = [r_u, r_u + r] \cup [r_u + r, \infty]$, where $r \in [r_v - r_u, r_u + r_v]$ and $H^{\text{pol}}_{\gamma_{\boldsymbol{X}}}(r|r_u, r_v) = \arccos\left(\frac{r_u^2 + r_v^2 - \rho^2}{2r_u r_v}\right)$ within the first interval and $r \in [0, r_v - r_u]$ and $H^{\text{pol}}_{\gamma_{\boldsymbol{X}}}(r|r_u, r_v) = 0$ within the second one.
In conclusion:

$$\begin{aligned} H^{\text{pol}}_{\gamma_{\boldsymbol{X}}}(r) &= 2\int_0^{r/2}\int_{r-r_u}^{r+r_u} \frac{1}{\pi}\arccos\left(\frac{r_u^2 + r_v^2 - \rho^2}{2r_u r_v}\right) d^{\text{pol}}_{\gamma_{\boldsymbol{X}}}(r_u) d^{\text{pol}}_{\gamma_{\boldsymbol{X}}}(r_v) \mathrm{d}r_u \mathrm{d}r_v \\ &\quad + 2\int_0^{r/2}\int_{r_u}^{r-r_u} \frac{1}{\pi}\arccos\left(\frac{r_u^2 + r_v^2 - \rho^2}{2r_u r_v}\right) d^{\text{pol}}_{\gamma_{\boldsymbol{X}}}(r_u) d^{\text{pol}}_{\gamma_{\boldsymbol{X}}}(r_v) \mathrm{d}r_u \mathrm{d}r_v \\ &\quad + 2\int_{r/2}^{\infty}\int_{r_u}^{r-r_u} d^{\text{pol}}_{\gamma_{\boldsymbol{X}}}(r_u) d^{\text{pol}}_{\gamma_{\boldsymbol{X}}}(r_v) \mathrm{d}r_u \mathrm{d}r_v. \end{aligned} \tag{1.30}$$

It is not always possible to get an exact analytical formula for $H^{\text{pol}}_{\gamma_{\boldsymbol{X}}}(r)$ and its derivative $h^{\text{pol}}_{\gamma_{\boldsymbol{X}}}(r)$ through which we could compute both the K and the pair correlation function of $\boldsymbol{X}$. In such cases a numerical approach is needed. Nevertheless, there are famous examples of Neyman-Scott processes where analytical computations are instead possible (Illian et al., 2008; Baddeley et al., 2007; Chiu et al., 2013; Wiegand and Moloney, 2013; Cressie, 2015). In what follows we will explore some of them, together with their fundamental statistics.

Convolution of 2-dimensional radial functions: from $d_{\gamma_{\boldsymbol{X}}}(\cdot)$ to $h_{\gamma_{\boldsymbol{X}}}(\cdot)$

Firstly, let us report the theory necessary for computing the convolution function $h_{\gamma_{\boldsymbol{X}}}(\cdot)$ of the dispersal kernel $d_{\gamma_{\boldsymbol{X}}}(\cdot)$ (see also Bracewell, 1986 and Birkinshaw, 1994). Let $f(u)$ be a function on the plane, $u \in \mathbb{R}^2$, and let us denote with $F(f)(k) = \tilde{f}(k)$ its Fourier transform, $k \in \mathbb{R}^2$:

$$\tilde{f}(k) = \frac{1}{2\pi}\int_{\mathbb{R}^2} f(u) e^{-ik\cdot u} \mathrm{d}u, \tag{1.31}$$

where $k \cdot u$ is the scalar product in $\mathbb{R}^2$.
Let us show that, if $f(u)$ is a radial function, that is $f(u) = f(r_u)$, with $r_u = |u|$,

than also $\tilde{f}(k)$ is radial, i.e. $\tilde{f}(k) = \tilde{f}(r_k)$, $r_k = |k|$[3]. Indeed, let us write both u and k in polar coordinates as (r_u, θ_u) and (r_k, θ_k), respectively and let us call $\theta = \theta_u - \theta_k$. Then, changing the integral in eq. (1.31) into polar coordinates we have that

$$\begin{aligned}
\tilde{f}(k) &= \frac{1}{2\pi}\int_0^\infty \int_0^{2\pi} r_u f(r_u) e^{-ir_k r_u(\cos(\theta_k)\cos(\theta_u)+\sin(\theta_k)\sin(\theta_u))} \mathrm{d}\theta_u \mathrm{d}r_u \\
&= \frac{1}{2\pi}\int_0^\infty \int_0^{2\pi} r_u f(r_u) e^{-ir_k r_u \cos(\theta_k-\theta_u)} \mathrm{d}\theta_u \mathrm{d}r_u \\
&= \frac{1}{2\pi}\int_0^\infty \int_0^{2\pi} r_u f(r_u) e^{-ir_k r_u \cos(\theta)} \mathrm{d}\theta \mathrm{d}r_u \\
&= \frac{1}{2\pi}\int_0^\infty r_u f(r_u) \int_0^{2\pi} e^{-ir_k r_u \cos(\theta)} \mathrm{d}\theta \mathrm{d}r_u \\
&= \frac{1}{2\pi}\int_0^\infty r_u f(r_u) \int_0^{2\pi} \cos(r_k r_u \cos(\theta)) \mathrm{d}\theta \mathrm{d}r_u. \qquad (1.32)
\end{aligned}$$

Let us now recall the integral representation of the Bessel function of the first kind of 0^{th} order:

$$J_0(z) = \frac{1}{\pi}\int_0^\pi \cos(z\cos(\theta))\mathrm{d}\theta.$$

Then we can further carry the computations in eq. (1.32)

$$\begin{aligned}
\tilde{f}(k) &= \frac{1}{\pi}\int_0^\infty r_u f(r_u)\Big(\int_0^\pi \cos(r_k r_u \cos(\theta))\mathrm{d}\theta\Big)\mathrm{d}r_u \\
&= \int_0^\infty r_u f(r_u) J_0(r_k r_u)\mathrm{d}r_u \\
&= \tilde{f}(r_k). \qquad (1.33)
\end{aligned}$$

It is also possible to write the inverse relation between $\tilde{f}(k)$ and $f(x) = F^{-1}(\tilde{f})(x)$ in terms of the Bessel function. Indeed, the following equation holds

$$f(r_u) = \int_0^\infty r_k \tilde{f}(r_k) J_0(r_k r_u)\mathrm{d}r_k,$$

which implies that it, if the Fourier transform of a function is radial, so is the function itself.
Of course, we are interested in the convolution of functions in order to compute $h_{\gamma_{\boldsymbol{X}}}(\cdot)$ from $d_{\gamma_{\boldsymbol{X}}}(\cdot)$. It is thus useful to recall that, by the Convolution Theorem, we have that following implication

$$\text{if} \quad f(u) = f_1(u) \star f_2(u) \quad \text{then} \quad \tilde{f}(k) = 2\pi \tilde{f}_1(k)\tilde{f}_2(k). \qquad (1.34)$$

Now, since $d_{\gamma_{\boldsymbol{X}}}(u) = d_{\gamma_{\boldsymbol{X}}}(r_u)$, is radial, by relation (1.34) above, we have that

$$h_{\gamma_{\boldsymbol{X}}}(r_u) = F^{-1}(\tilde{h}_{\gamma_{\boldsymbol{X}}})(r_u) = F^{-1}(F(d_{\gamma_{\boldsymbol{X}}} \star d_{\gamma_{\boldsymbol{X}}}))(r_u) = F^{-1}(2\pi(\tilde{d}_{\gamma_{\boldsymbol{X}}})^2)(r_u).$$

By eq. (1.33), we finally have that

$$h_{\gamma_{\boldsymbol{X}}}(r_u) = 2\pi \int_0^{+\infty} r_k (\tilde{d}_{\gamma_{\boldsymbol{X}}}(r_k))^2 J_0(r_k r_u) dr_k,$$

where

$$\tilde{d}_{\gamma_{\boldsymbol{X}}}(r_k) = \int_0^{+\infty} r_u h_{\gamma_{\boldsymbol{X}}}(r_u) J_0(r_k r_u) dr_u.$$

[3]Note that the Fourier transform of a radial function $f(u) = f(r_u)$, $r_u = |u|$ only exists if $\int_0^{+\infty} \sqrt{r_u}|f(r_u)|dr_u < +\infty$.

Matérn cluster process

In a Matérn cluster process $\boldsymbol{X}$, the daughters are randomly located within a ball centred in their parent's location and with fixed radius $R_{\boldsymbol{X}}$, which is an additional parameter of the model (Illian et al., 2008; Chiu et al., 2013). Let us briefly introduce the main statistics of the process.
We know from eq. (1.24) that the intensity function is given by the product of the parent's Poisson process density $\rho_{\boldsymbol{X}}$ and the mean number of points per cluster $\mu_{\boldsymbol{X}}$. This result can also be found as a nice application of Campbell's formula (see Theorem 1.14). Indeed, since the process $\boldsymbol{X}$ is the result of a superposition of point processes (clusters), we know that, conditionally on the parent's location $x \in \boldsymbol{X}_p$, the intensity function of $\boldsymbol{X}$ is given by:

$$\lambda_{\boldsymbol{X}|\boldsymbol{X}_p}(u) = \sum_{x \in \boldsymbol{X}_p} \lambda_{\boldsymbol{X}_c^x}(u),$$

where $\lambda_{\boldsymbol{X}_c^x}(\cdot)$ is the intensity function of the cluster whose parent is located at $x \in \mathbb{R}^2$. In the case under study, the offspring of a parent at location $x \in \boldsymbol{X}_p$ is randomly distributed within the ball $\mathcal{B}(x, R_{\boldsymbol{X}})$, centred in x and having radius $R_{\boldsymbol{X}}$, according to a Poisson process of intensity $\mu_{\boldsymbol{X}}$. Thus, we have that

$$\lambda_{\boldsymbol{X}_c^x}(u) = \frac{\mu_{\boldsymbol{X}}}{\pi R_{\boldsymbol{X}}^2} \cdot \chi_{\mathcal{B}(x,R_{\boldsymbol{X}})}(u),$$

where $\chi_{\mathcal{B}(x,R_{\boldsymbol{X}})}(\cdot)$ is the characteristic function of $\mathcal{B}(x, R_{\boldsymbol{X}})$.
In order to find the unconditional density $\lambda_{\boldsymbol{X}}(\cdot)$ we have to average over all parents' locations:

$$\lambda_{\boldsymbol{X}}(u) = \mathbb{E}[\lambda_{\boldsymbol{X}|\boldsymbol{X}_p}(u)] = \mathbb{E}\left[\sum_{x \in \boldsymbol{X}_p} \lambda_{\boldsymbol{X}_c^x}(u)\right] = \mathbb{E}\left[\sum_{x \in \boldsymbol{X}_p} \frac{\mu_{\boldsymbol{X}}}{\pi R_{\boldsymbol{X}}^2} \cdot \chi_{\mathcal{B}(x,R_{\boldsymbol{X}})}(u)\right].$$

By Campbell's formula we have that

$$\begin{aligned}
\mathbb{E}\left[\sum_{x \in \boldsymbol{X}_p} \frac{\mu}{\pi R_{\boldsymbol{X}}^2} \cdot \chi_{\mathcal{B}(x,R_{\boldsymbol{X}})}(u)\right] &= \int_{\mathbb{R}^2} \frac{\mu_{\boldsymbol{X}}}{\pi R_{\boldsymbol{X}}^2} \cdot \chi_{\mathcal{B}(v,R_{\boldsymbol{X}})}(u) \lambda_{\boldsymbol{X}_p}(v) \mathrm{d}v \\
&= \frac{\mu_{\boldsymbol{X}} \rho_{\boldsymbol{X}}}{\pi R_{\boldsymbol{X}}^2} \int_{\mathbb{R}^2} \chi_{\mathcal{B}(v,R_{\boldsymbol{X}})}(u) \mathrm{d}v \\
&= \frac{\mu_{\boldsymbol{X}} \rho_{\boldsymbol{X}}}{\pi R_{\boldsymbol{X}}^2} \int_{\mathbb{R}^2} \chi_{\mathcal{B}(u,R_{\boldsymbol{X}})}(v) \mathrm{d}v \\
&= \frac{\mu_{\boldsymbol{X}} \rho_{\boldsymbol{X}}}{\pi R_{\boldsymbol{X}}^2} \cdot \pi R_{\boldsymbol{X}}^2 = \mu_{\boldsymbol{X}} \rho_{\boldsymbol{X}}.
\end{aligned}$$

The probability density function of the distances from the cluster centre, $d_{\gamma_{\boldsymbol{X}}}(\cdot)$ is given by

$$d_{\gamma_{\boldsymbol{X}}}(r) = \begin{cases} \dfrac{1}{\pi R_{\boldsymbol{X}}^2} & \text{if } 0 \leq r \leq R_{\boldsymbol{X}} \\ 0 & \text{otherwise,} \end{cases}$$

from which we can compute the distribution function $H_{\gamma_{\boldsymbol{X}}}^{\mathrm{pol}}(r)$ and its density $h_{\gamma_{\boldsymbol{X}}}^{\mathrm{pol}}(r)$ (see Section 1.5.2):

$$h_{\gamma_{\boldsymbol{X}}}^{\mathrm{pol}}(r) = \frac{4r}{\pi R_{\boldsymbol{X}}^2} \left(\arccos\left(\frac{r}{2R_{\boldsymbol{X}}}\right) - \frac{r}{2R_{\boldsymbol{X}}} \sqrt{1 - \frac{r^2}{4R_{\boldsymbol{X}}^2}} \right). \tag{1.35}$$

Here the $\gamma_{\boldsymbol{X}}$ parameter is the offspring's dispersal radius $R_{\boldsymbol{X}}$.
Finally, from eq. (1.28), we can obtain the pair correlation function

$$\begin{aligned} g_{\boldsymbol{X}}(r) &= 1 + \frac{1}{\rho_{\boldsymbol{X}}} \frac{h_{\boldsymbol{X}}^{\text{pol}}(r)}{2\pi r} \\ &= 1 + \frac{1}{\rho_{\boldsymbol{X}} 2\pi r} \frac{4r}{\pi R_{\boldsymbol{X}}^2} \left(\arccos\left(\frac{r}{2R_{\boldsymbol{X}}}\right) - \frac{r}{2R_{\boldsymbol{X}}} \sqrt{1 - \frac{r^2}{4R_{\boldsymbol{X}}^2}} \right) \\ &= 1 + \frac{2}{\rho_{\boldsymbol{X}} \pi^2 R^2} \left(\arccos\left(\frac{r}{2R_{\boldsymbol{X}}}\right) - \frac{r}{2R_{\boldsymbol{X}}} \sqrt{1 - \frac{r^2}{4R_{\boldsymbol{X}}^2}} \right). \end{aligned}$$

Modified Thomas process

In the modified Thomas process (Thomas, 1949; Plotkin, Potts et al., 2000; Morlon et al., 2008; Azaele, Cornell et al., 2012; Tovo, Formentin et al., 2016; Tovo, Suweis et al., 2017), the daughters of the representative parent are located around its location (the origin) according to a radially symmetric Gaussian distribution $d_{\gamma_{\boldsymbol{X}}}(\cdot)$:

$$d_{\gamma_{\boldsymbol{X}}}(r) = \frac{1}{2\pi\sigma_{\boldsymbol{X}}^2} e^{-\frac{r^2}{2\sigma_{\boldsymbol{X}}^2}}. \tag{1.36}$$

In this case $\gamma_{\boldsymbol{X}}$ equals the standard deviation of the bi-variate Gaussian, $\sigma_{\boldsymbol{X}}$.
Again, the intensity of the superposed process formed by all the clusters is given by

$$\lambda_{\boldsymbol{X}}(u) = \rho_{\boldsymbol{X}} \mu_{\boldsymbol{X}}.$$

The density function of $H_{\gamma_{\boldsymbol{X}}}^{\text{pol}}(r)$ and the pair correlation function are given, respectively, by

$$h_{\gamma_{\boldsymbol{X}}}^{\text{pol}}(r) = \frac{r}{2\sigma_{\boldsymbol{X}}^2} e^{-\frac{r^2}{4\sigma_{\boldsymbol{X}}^2}},$$

and

$$\begin{aligned} g_{\boldsymbol{X}}(r) &= 1 + \frac{1}{\rho_{\boldsymbol{X}}} \frac{h_{\boldsymbol{X}}^{\text{pol}}(r)}{2\pi r} \\ &= 1 + \frac{1}{\rho_{\boldsymbol{X}} 2\pi r} \frac{r}{2\sigma_{\boldsymbol{X}}^2} e^{-\frac{r^2}{4\sigma_{\boldsymbol{X}}^2}} \\ &= 1 + \frac{1}{\rho_{\boldsymbol{X}} 4\pi\sigma_{\boldsymbol{X}}^2} e^{-\frac{r^2}{4\sigma_{\boldsymbol{X}}^2}}. \end{aligned}$$

The average of r with respect to $d_{\gamma_{\boldsymbol{X}}}(\cdot)$ gives the mean cluster radius, $r_{\boldsymbol{X}}$, that is the average distance of a daughter seed from its parent.

$$\begin{aligned} r_{\boldsymbol{X}} &= \int_{-\infty}^{+\infty} |x| d_{\gamma_{\boldsymbol{X}}}(x) \mathrm{d}x \\ &= \int_0^{2\pi} \int_0^{+\infty} r^2 d_{\gamma_{\boldsymbol{X}}}(r) \mathrm{d}r \\ &= \frac{1}{2\pi\sigma_{\boldsymbol{X}}^2} \int_0^{2\pi} \int_0^{+\infty} r^2 e^{-\frac{r^2}{2\sigma_{\boldsymbol{X}}^2}} \mathrm{d}r \\ &= \frac{1}{2\pi\sigma_{\boldsymbol{X}}^2} \cdot 2\pi\sqrt{\frac{\pi}{2}}\sigma_{\boldsymbol{X}}^3 \\ &= \sigma_{\boldsymbol{X}}\sqrt{\frac{\pi}{2}}. \end{aligned}$$

Exponential process

Let us now assume that the daughters of the representative cluster are exponentially distributed around their parent, located at the origin:

$$d_{\gamma_{\boldsymbol{X}}}(r) = \frac{\beta_{\boldsymbol{X}}^2}{2\pi} e^{-\beta_{\boldsymbol{X}} r}.$$

Here the cluster parameter $\gamma_{\boldsymbol{X}}$ equals $\beta_{\boldsymbol{X}}$.
By performing the computations seen in Section 1.5.2, we have that convolution of $d_{\gamma_{\boldsymbol{X}}}(\cdot)$ is given by

$$h^{\mathrm{pol}}_{\gamma_{\boldsymbol{X}}}(r) = \frac{\beta_{\boldsymbol{X}}^4 r^3}{8} K_2(\beta_{\boldsymbol{X}} r),$$

where $K_2(\cdot)$ is the modified Bessel function of the second kind and order 2 (Abramowitz and Stegun, 1964). Note that, for $z \to 0$, it holds that $K_n(z) \sim \frac{\Gamma(n)}{2}(\frac{2}{z})^n$. Hence $h^{\mathrm{pol}}_{\gamma_{\boldsymbol{X}}}(0)$ is a finite quantity.
As for all the cluster processes, the intensity is given by the product $\lambda_{\boldsymbol{X}} = \rho_{\boldsymbol{X}}\mu_{\boldsymbol{X}}$.
In this case the pair correlation function is given by

$$\begin{aligned} g_{\boldsymbol{X}}(r) &= 1 + \frac{1}{\rho_{\boldsymbol{X}}} \frac{h^{\mathrm{pol}}_{\boldsymbol{X}}(r)}{2\pi r} \\ &= 1 + \frac{1}{\rho_{\boldsymbol{X}} 2\pi r} \frac{\beta_{\boldsymbol{X}}^4 r^3}{8} K_2(\beta_{\boldsymbol{X}} r) \\ &= 1 + \frac{\beta^4 r^2}{\rho_{\boldsymbol{X}} 16\pi} K_2(\beta_{\boldsymbol{X}} r), \end{aligned}$$

Let us compute the average clumping radius as for the modified Thomas process:

$$\begin{aligned} r_{\boldsymbol{X}} &= \int_{-\infty}^{+\infty} |x| d_{\gamma_{\boldsymbol{X}}}(x) \mathrm{d}x \\ &= \int_0^{2\pi} \int_0^{+\infty} r^2 d_{\gamma_{\boldsymbol{X}}}(r) \mathrm{d}r \\ &= \frac{\beta_{\boldsymbol{X}}^2}{2\pi} \int_0^{2\pi} \int_0^{+\infty} r^2 e^{-\beta_{\boldsymbol{X}} r} \mathrm{d}r \\ &= \frac{\beta_{\boldsymbol{X}}^2}{2\pi} 2\pi \frac{2}{\beta_{\boldsymbol{X}}^3} \\ &= \frac{2}{\beta_{\boldsymbol{X}}^2}. \end{aligned}$$

Gaussian mixture process

For the last Neyman-Scott process we will study in this thesis, we consider a dispersal kernel obtained as a continuous mixture of Gaussians having the variance parameter distributed as the inverse Gamma (see Clarke and Lidgard, 2000; Chave and Leigh, 2002). The resulting radial probability density is given by

$$d_{\gamma_{\boldsymbol{X}}}(r) = \frac{p}{\pi b^2 \left(1 + \frac{r^2}{b_{\boldsymbol{X}}^2}\right)^{p_{\boldsymbol{X}}+1}},$$

where $b_{\boldsymbol{X}}^2$ and $p_{\boldsymbol{X}}$ constitute the set of cluster parameters $\gamma_{\boldsymbol{X}}$.
The corresponding convolution is

$$h_{\gamma_{\boldsymbol{X}}}^{\mathrm{pol}}(r) = \frac{\sqrt{\pi} p_{\boldsymbol{X}}^2 \Gamma(2p_{\boldsymbol{X}}+1) \tilde{F}_1\left(p_{\boldsymbol{X}}+1, 2p_{\boldsymbol{X}}+1; p_{\boldsymbol{X}}+\frac{3}{2}; -\frac{r^2}{4b_{\boldsymbol{X}}^2}\right) r}{2^{2p_{\boldsymbol{X}}} b_{\boldsymbol{X}}^2 \, \Gamma(p_{\boldsymbol{X}}+1)},$$

where $\tilde{F}_1(\cdot)$ is the hypergeometric $2F1$ regularised function.
Note that the average cluster radius in this case is

$$r_{\boldsymbol{X}} = \frac{b_{\boldsymbol{X}} \sqrt{\pi} \Gamma\left(p_{\boldsymbol{X}} - \frac{1}{2}\right)}{2\Gamma(p_{\boldsymbol{X}})},$$

showing that $b_{\boldsymbol{X}}$ is a scale parameter and that the average radius is *infinite* for $p_{\boldsymbol{X}} \leq 1/2$.
Here we focus on the Cauchy cluster processes, which can be obtained by fixing the $p_{\boldsymbol{X}}$ parameter equal to 2. In this special case, $d_{\gamma_{\boldsymbol{X}}}(\cdot)$ is a 2-dimensional radial probability density of the form

$$d_{\gamma_{\boldsymbol{X}}}(r) = \frac{1}{2\pi b_{\boldsymbol{X}}^2 (1 + \frac{r^2}{b_{\boldsymbol{X}}^2})^{\frac{3}{2}}},$$

depending only on $b_{\boldsymbol{X}}$. Its convolution yields

$$h_{\gamma_{\boldsymbol{X}}}^{\mathrm{pol}}(r) = \frac{2b^2 r}{(4 + \frac{r^2}{b^2})^{\frac{3}{2}}},$$

from which we can get the analytic form of pair correlation function

$$\begin{aligned} g_{\boldsymbol{X}}(r) &= 1 + \frac{1}{\rho_{\boldsymbol{X}}} \frac{h^{\text{pol}}_{\boldsymbol{X}}(r)}{2\pi r} \\ &= 1 + \frac{1}{\rho_{\boldsymbol{X}} 2\pi r} \frac{2b^2 r}{(4 + \frac{r^2}{b^2})^{\frac{3}{2}}} \\ &= 1 + \frac{1}{\rho_{\boldsymbol{X}} \pi} \frac{b^2}{(4 + \frac{r^2}{b^2})^{\frac{3}{2}}}. \end{aligned}$$

We have already noticed that in this case the average clump radius is infinite. Indeed the radial density $d_{\gamma_{\boldsymbol{X}}}(\cdot)$ has a fat tail (a power-law), hence this type of cluster is well suited for describing species with long range intraspecific correlations.

Simulation of a Neyman-Scott process

Let us see how to simulate a Neyman-Scott process $\boldsymbol{X}$ having n points within a window $\mathcal{W}$. Let us assume that the intensity function of the Poisson process generating the parents' point equals $\rho_{\boldsymbol{X}}$ and that the clusters' dispersal kernel is a fixed radial function $d_{\gamma_{\boldsymbol{X}}}$. Then we adopt the following procedure, consisting of four steps (Plotkin, Potts et al., 2000):

- We simulate the parents' Poisson cluster process by placing

  $$\lfloor \rho_{\boldsymbol{X}} \times |\mathcal{W}| + 1/2 \rfloor$$

  points randomly distributed within the plot, with $|\mathcal{W}|$ the area of the study region.

- We randomly assign each of the n individuals of the process to one of the previously generated parents.

- For each parent, we locate the associated stems according to the two-dimensional kernel function $d_{\gamma_{\boldsymbol{X}}}$ around the location of the parent. In the case that the offspring falls out of the plot, we impose toroidal boundary conditions.

- We remove the parents from the plot, so that only the daughter seeds remains within $\mathcal{W}$.

1.6 From a point pattern to a point process

In this section we study how to model a spatial point pattern, e.g. the locations of trees within a forest, through point processes' theory.
The first step is to infer the basic first and second-order statistics of the process of which the pattern is considered a single realisation. Once these functions have been estimated from empirical data, the second step is to find, for each prefixed model, the corresponding parameters which best describe the pattern under study. In the following sections, we will explore in details these two key points.

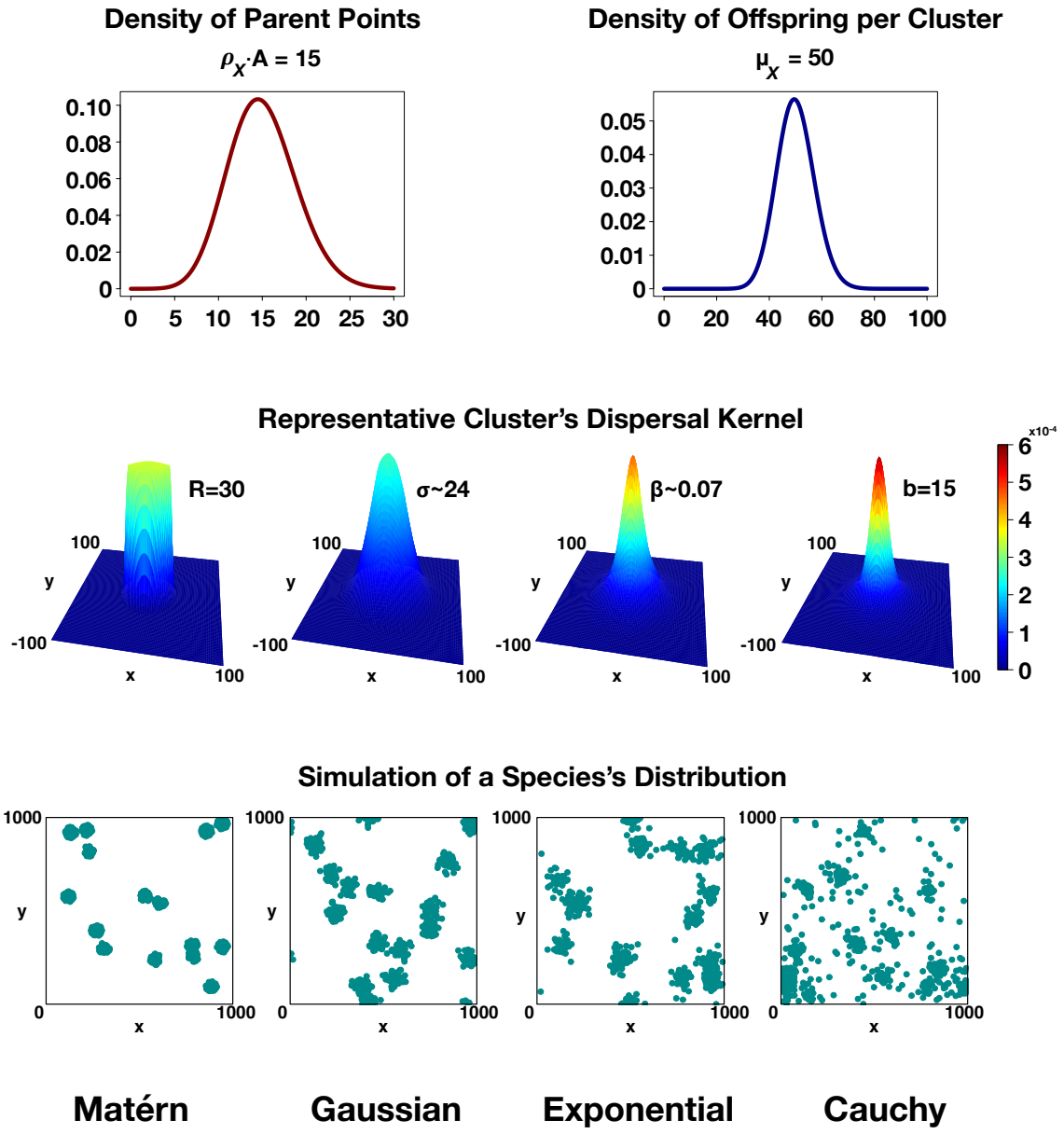


Figure 1.4: Examples of different Neyman-Scott processes: Matérn, Gaussian (or modified Thomas), exponential and Cauchy. In all cases we set the density of cluster equal to $\rho_{\boldsymbol{X}} = 15/A$, with $A = |\mathcal{W}| = 1000 \times 1000$ (in the first graphic we show the probability density function of $\rho_{\boldsymbol{X}}$ multiplied by the area A) and the average number of points per cluster equal to $\mu_{\boldsymbol{X}} = 50$. Both $\rho_{\boldsymbol{X}}$ and $\mu_{\boldsymbol{X}}$ are Poisson distributed (top panels). In the middle panels we insert the dispersal kernel of the representative cluster $\boldsymbol{X}_c$ centred in the origin. For the modified Thomas and the exponential processes, we set the clustering parameters so to have an average clumping radius equal to the chosen Matérn radius $R_{\boldsymbol{X}} = 30$. As for the Cauchy process, since the average radius is not well defined, we arbitrarily set it equal to $R_{\boldsymbol{X}}/2$. In the bottom panels we simulate a species' distribution within the 1000×1000 area for each process (see Section 1.5.2).

1.6.1 Estimators of a point process's statistics

Let us assume that we have information on the exact locations of N trees within a rectangular window $\mathcal{W}$ of sides l_x and l_y. We can consider this set of points as a realisation of a particular point process $\boldsymbol{X}$. In this section we study how to estimate the fundamental statistics of $\boldsymbol{X}$, henceforth assumed to be stationary and isotropic. Let us start with its density, $\lambda_{\boldsymbol{X}}$. An unbiased estimator for it is given by

$$\lambda_{\boldsymbol{X}} = \frac{N}{|W|}, \tag{1.37}$$

where we denoted with $|\mathcal{W}|$ the area of the surveyed region $\mathcal{W}$.
Of course, if the stationary or isotropy hypotheses do not hold for the spatial pattern under consideration, for example because of the presence of strong inhomogeneities in the environment, the above estimator (1.37) becomes inadequate. In Chapter 2 we will study in details another way to have an approximation of $\lambda_{\boldsymbol{X}}(\cdot)$ which can be applied in such cases.
As for the second-order statistics, we introduce here the estimators we will use in the rest of the thesis for the Ripley K function, $\hat{K}_{\boldsymbol{X}}(\cdot)$, and the pair correlation function, $\hat{g}_{\boldsymbol{X}}(\cdot)$.
Let us recall that the K function of a spatial point process multiplied by its intensity $\lambda_{\boldsymbol{X}}$ gives the number of extra events within a distance r from an arbitrary event.
In our case, the arbitrary event is the location of any individual belonging to the species's pattern under study. If this latter is randomly distributed (CSR or RPM), then, since the mean number of stems within a circle of radius r is $\lambda_{\boldsymbol{X}}\pi r^2$, we have that the K function equals $K_{\boldsymbol{X}}(r) = \pi r^2$, so that it grows quadratically with the distance. However, in general, this behaviour is not observed.
Let us call $x_1, ..., x_N$ the locations of the N individuals of the spatial pattern under analysis and let $\hat{\lambda}_{\boldsymbol{X}} = \frac{N}{|\mathcal{W}|}$ be our estimation of the intensity $\lambda_{\boldsymbol{X}}$. Let us then denote with $w^{(Ripley)}(x_i, x_j)$ the proportion of the circumference of the circle centred in x_i and passing through x_j which lies within $\mathcal{W}$. Lastly, let us indicate with $||x_i, x_j||$ the euclidean distance between the individuals located at x_i and x_j, respectively.
The canonical edge-corrected estimator of Ripley's K-function is then given by (Plotkin, Potts et al., 2000)

$$\hat{K}_{\boldsymbol{X}}(r) = \frac{1}{\hat{\lambda}} \sum_{i=1}^{N} \sum_{\substack{j=1 \\ j \neq i}}^{N} \frac{1}{w^{(Ripley)}(x_i, x_j)} \frac{II(||x_i - x_j|| \leq r)}{N}, \tag{1.38}$$

where $II(\cdot)$ is the indicator function.
Defining with r_{max} the larger distance at which we measure a cluster within the region $\mathcal{W}$ of our rainforest, we will evaluate the edge-corrected estimator only between 0 to r_{max}.
Finally, to estimate the pair correlation function $g_{\boldsymbol{X}}(\cdot)$, we will adopt the following estimator (Stoyan and Stoyan, 1994; Shimatani, 2001):

$$\hat{g}_{\boldsymbol{X}}(r) = \frac{1}{\hat{\lambda}_{\boldsymbol{X}} N} \sum_{\substack{j,k=1 \\ k \neq j}}^{N} \frac{w_{\boldsymbol{X}}^{(Shim)}(\parallel x_j - x_k \parallel - r)}{2\pi r B(r)}, \tag{1.39}$$

where $\|\cdot\|$ denotes again the Euclidean distance on the plane, $B(r) = 1 - r(2l_x + 2l_y - r)/l_x l_y \pi$ is the edge corrector function depending on the sides l_x and l_y of $\mathcal{W}$ (see Ohser, 1983) and $w^{(Shim)}(\cdot)$ is the Epanenchnikov kernel, defined as

$$w_{\boldsymbol{X}}^{(Shim)}(y) = \begin{cases} \frac{3}{4\delta_{\boldsymbol{X}}}\left(1 - \frac{y^2}{\delta_{\boldsymbol{X}}^2}\right) & \text{for } |y| < \delta_{\boldsymbol{X}} \\ 0 & \text{for } |y| \geq \delta_{\boldsymbol{X}}, \end{cases}$$

with $\delta_{\boldsymbol{X}} = 0.2/\sqrt{\hat{\lambda}_{\boldsymbol{X}}}$.

1.6.2 Estimating the model parameters of a Neyman-Scott process

Up to now we have made no assumptions on the process $\boldsymbol{X}$ generating the pattern under study, if not that it is isotropic and stationary.
Now we focus on the second important step in modelling a spatial pattern through point processes theory, that is to find the parameters that best describes the empirical statistics estimated as described in the previous section. In particular, we limit to the case of the Neyman-Scott processes studied in Section 1.5.2.
Let us thus assume that our spatial pattern is a realisation of a cluster process $\boldsymbol{X}$ having parameters $(\rho_{\boldsymbol{X}}, \mu_{\boldsymbol{X}}, d_{\gamma_{\boldsymbol{X}}})$.
In order to extract, from our database, the model parameters which best reproduce our pattern, we resort to the *method of minimum contrast* (Diggle and Gratton, 1984; Diggle, 2013; Plotkin, Potts et al., 2000) applied to a second order statistics $s_{\boldsymbol{X}}(\cdot)$ (either the K function $K_{\boldsymbol{X}}(\cdot)$ or the pair correlation function $g_{\boldsymbol{X}}(\cdot)$). The first step is to estimate $\hat{s}_{\boldsymbol{X}}(r)$ at some fixed distances $r = r_0, \ldots, r_{\max}$ from our data, using either eq. (1.38) or eq. (1.39) depending on the chosen statistics.
The best model parameters will then be those such that the theoretical function $s_{\boldsymbol{X}}(\cdot)$ best fits the empirical values $\hat{s}_{\boldsymbol{X}}(0), ..., \hat{s}_{\boldsymbol{X}}(r_{\max})$.
The method of minimum contrast relies on the minimisation of the following integral

$$\int_0^{r_{\max}} \left(\hat{s}_{\boldsymbol{X}}(r)^{\frac{1}{4}} - s_{\boldsymbol{X}}(r)^{\frac{1}{4}}\right)^2 \mathrm{d}r.$$

Actually, since the analytical formula of both the K function and the pair correlation function do only depend on $\rho_{\boldsymbol{X}}$ and $d_{\gamma_{\boldsymbol{X}}}$, with this procedure we still lack of an estimation of the $\mu_{\boldsymbol{X}}$ parameter. However, let us remember that the theoretical total intensity of a Neyman-Scott process $\boldsymbol{X}$ is given by $\lambda_{\boldsymbol{X}} = \rho_{\boldsymbol{X}}\mu_{\boldsymbol{X}}$, from which we get $\mu_{\boldsymbol{X}} = \lambda_{\boldsymbol{X}}/\rho_{\boldsymbol{X}}$. Using the unbiased estimator $\hat{\lambda}_{\boldsymbol{X}} = n(\mathcal{W})/|\mathcal{W}|$ and the estimated value $\hat{\rho}_{\boldsymbol{X}}$ via the minimum contrast method, one can then obtain the missing parameter as

$$\hat{\mu}_{\boldsymbol{X}} = \frac{\hat{\lambda}_{\boldsymbol{X}}}{\hat{\rho}_{\boldsymbol{X}}}.$$

Let us consider the *Acalypha diversifolia* species in BCI. In figure 1.2, we have already noticed that its empirical pair correlation function estimated via eq. (1.39) is not constant, thus it does not satisfy the CSR hypothesis. At contrast, let us model this species through the modified Thomas, the exponential and the Cauchy

cluster processes. In figure 1.5 we plot the pair correlation curves (coloured lines) obtained by inserting, into each model's analytical formula for $g_{\boldsymbol{X}}(\cdot)$, the parameters $(\hat{\rho}_{\boldsymbol{X}}, d_{\hat{\gamma}_{\boldsymbol{X}}})$ fitted through the minimum contrast method. The agreement between the predicted curves and the empirical one (black dots) is very good.

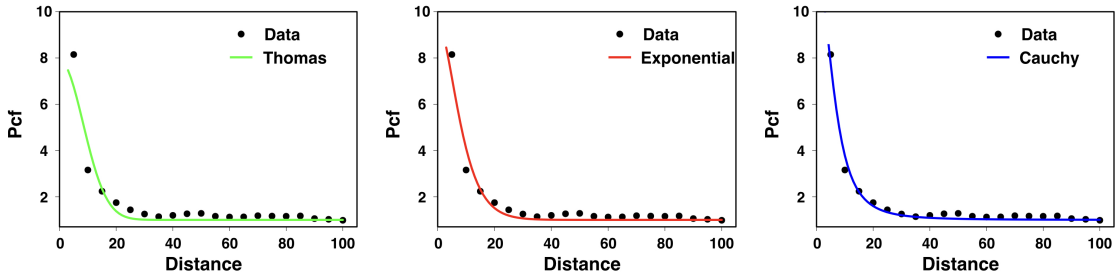


Figure 1.5: Reproduction of the pair correlation function of the *Acalypha diversifolia* species in the 50ha sample plot of BCI forest through three Neyman-Scott processes: the modified Thomas (left panel), the exponential (middle panel) and the Cauchy one (right panel). Black dots represent the empirical values obtained at different distances $r = r_0, \ldots, r_{\max}$ for the pair correlation function via eq. (1.39). Coloured lines are instead the predicted values that we get when the pair of parameters $(\hat{\rho}_{\boldsymbol{X}}, d_{\hat{\gamma}_{\boldsymbol{X}}})$ fitted by the minimum contrast method are inserted into the analytical formulas of $g_{\boldsymbol{X}}(r)$ according to the three models. For all these latter, we find a good agreement between the predictions and the empirical data. Here we have set $r_{\max} = 100$ meters.

2

Inferring the Intensity Function of a Point Process

Investigation of highly structured data sets to unveil statistical regularities is of major importance in complex system research. The first step is to choose the scale at which to observe the process. As a rule of thumb, the most informative scale is the one that includes the important features while disregards noisy details in the data. In the investigation of spatial patterns or temporal series, the optimal scale corresponds to the choice of the optimal bin size of the histogram in which to visualize the data. This is a relevant issue in spatial ecology. There exist diverse rules for data binning, many of which are heuristic. In this chapter we study an algorithm proposed by Knuth to decide the optimal bin size of an histogram (Knuth, 2006; Tovo, Formentin et al., 2016). We test it through numerical simulations on various spatial point processes which are of interest in ecology and we compare it with other popular methods proposed in literature. We show that Knuth's optimal bin size reduces noisy fluctuations and is capable to capture relevant spatial characteristics on the underlying distribution: space anisotropy and clusterisation. Moreover, when modelling data through point processes, it gives a reliable approximation of their intensity function. We then apply these findings to analyse cluster-like structures in plant arrangement of the Barro Colorado island (BCI) rainforest.

2.1 K. H. Knuth's method

There exist diverse rules to determine the optimal number of bins of a histogram. Some of the most known (Sturges, 1926; Scott, 1979; Freedman and Diaconis, 1981; Stone, 1984; Scott, 2015) rely on the minimisation of the L^2 norm between the histogram and the true underlying density on which they assume some prior knowledge. Assuming such prior knowledge is not reasonable for ecological datasets, since the process generating the data must be considered unknown. Hence, any criteria deter-

mining the optimal bin size based on prior knowledge on the true density should not be applied. Moreover, these methods work well for unimodal densities while they are known to be suboptimal for multimodal ones. Finally, some of them cannot be applied in case of particular density functions. For example, both the rule of Stone and the method of Freedman and Diaconis are not valid for uniform or piece-wise constant density functions (Freedman and Diaconis, 1981; Stone, 1984) whereas Sturges's rule (Sturges, 1926) is not suitable when the data exhibit skewness or any other non-normality. To overcome these difficulties, Knuth (Knuth, 2006) proposed a method based on *maximum a-posteriori estimation* and *Bayes's Theorem* where no prior information about the density from which the data are sampled is assumed. In what follows we briefly present the Knuth's method, referring to (Knuth, 2006) for further details.

2.1.1 From the dataset to the histogram

Let us suppose to be working with a set of N 2-dimensional data $\underline{d} = (d_1, ..., d_N)$ sampled from an unknown probability density function that we wish to estimate. In our spatial ecology applications each 2-dimensional datum will represent the location of a tree.
Let us then divide our data span V into $M_x \times M_y$ bins, where M_x is the number of bins along the x-axis and M_y the number along the y-one. We set $\underline{M} = (M_x, M_y)$ and $M = M_x \cdot M_y$. Denoting by a_x and a_y the width of each bin along the $x-$ and $y-$axis respectively, we have that all bins have equal area $a = a_x \cdot a_y$. We assume no measurement uncertainty about the data hence V is fixed. The pair $\underline{M}$ will determine the bin's area and the correspondent histogram column's height, which we call h_k, where $k = 1, ..., M$ is the bin label. After the normalisation of the volume of the histogram, h_k represents the constant value of the probability density function over the region of the bin. The volume of each histogram column is the probability mass of each bin $\pi_k = h_k a$, i.e. the probability of finding a datum in the range dictated by the k^{th} bin.
Therefore the piece-wise constant density function of the histogram with M bins is:

$$h(x, y) = \sum_{k=1}^{M} h_k \Pi_k(x, y),$$

where $\Pi_k(x, y)$ is the boxcar function, defined as

$$\Pi_k(x, y) = \begin{cases} 1 \text{ if } (x, y) \text{ falls within the } k^{th} \text{ bin} \\ 0 \text{ otherwise} \end{cases}$$

Notice that due to the normalisation, only $M - 1$ probabilities masses are independent. Indeed, the last one is determined by the relation $\pi_M = 1 - \sum_{k=1}^{M-1} \pi_k$.

2.1.2 The likelihood function

In statistical parlance, whereas probability allows to predict unknown outcomes based on known parameters, likelihood permits to solve the inverse problem, i.e.

to estimate unknown parameters based on known outcomes. More precisely, the likelihood function is defined as the probability density inferred by a set of sampled data that, when multiplied by $\mathrm{d}x$, gives the probability that a datum takes value in the infinitesimal range $[x, x + \mathrm{d}x]$.
When we arrange the data into a histogram, the probability that a datum falls within the k^{th} bin is given by the probability mass of that bin. In this case, the likelihood function is then given by the associated probability density h_k (the bin height).
As in Knuth, 2006, let us denote with $\underline{\pi} = (\pi_1, ..., \pi_{M-1})$, $0 \leq \pi_k \leq 1$, the independent probabilities masses and with n_k the number of 2-dimensional data points contained in the k^{th} bin. Assuming that the sampled data are independent, their joint likelihood reduces to the product of N factors

$$p(\underline{d}|\underline{\pi}, \underline{M}) = \left(\frac{M}{V}\right)^N \pi_1^{n_1} \pi_2^{n_2} ... \pi_{M-1}^{n_{M-1}} \left(1 - \sum_{k=1}^{M-1} \pi_k\right)^{n_M} \tag{2.1}$$

2.1.3 Prior and posterior probability

The prior probability of the number of bins represents our knowledge of it a priori. As we do not know anything but the total range V of the data, it is reasonable to set

$$P(\underline{M}) = \begin{cases} C^{-1} & \text{if } \ 1 \leq M \leq C \\ 0 & \text{otherwise} \end{cases} \tag{2.2}$$

where C is the maximum number of bins we wish to consider.
Similarly, since we do not have any clue about the underlying density function, we take as the prior probability of the bin masses $\pi_1, ..., \pi_{M-1}$ the uniform probability on the simplex defined by the corners of an $(M-1)$-dimensional hypercube with unit side lengths due to the normalisation condition:

$$p(\underline{\pi}|\underline{M}) = \frac{\Gamma\left(\frac{M}{2}\right)}{\Gamma\left(\frac{1}{2}\right)^M} \left[\pi_1 \pi_2 ... \pi_{M-1} \left(1 - \sum_{k=1}^{M-1} \pi_k\right)\right]^{-1/2} \tag{2.3}$$

where $\Gamma(\cdot)$ is the Gamma function. Eq. (2.3) is called Jeffreys's non-informative prior and it expresses complete ignorance about the form of the histogram (Box and Tiao, 2011).
To infer the posterior probability density for the number of bins M we use Bayes's Theorem, which states that, given two events A and B and provided that $\mathbb{P}(B) \neq 0$, the conditional probability $\mathbb{P}(A|B)$ is given by the ratio

$$\mathbb{P}(A|B) = \frac{\mathbb{P}(B|A) \cdot \mathbb{P}(A)}{\mathbb{P}(B)} \tag{2.4}$$

In particular, if the event B is fixed, it states that the posterior $\mathbb{P}(A|B)$ is proportional to the product of the prior $\mathbb{P}(A)$ and the likelihood $\mathbb{P}(B|A)$, $\mathbb{P}(A|B) \propto \mathbb{P}(A)\mathbb{P}(B|A)$. Applied to our case Bayes's Theorem leads to

$$p(\underline{\pi}, \underline{M}|\underline{d}) \propto p(\underline{\pi}|\underline{M}) p(\underline{M}) p(\underline{d}|\underline{\pi}, \underline{M}). \tag{2.5}$$

Inserting eqs. (2.1) to (2.3) into eq. (2.5) above, we get the joint posterior probability for the bin parameters $\underline{\pi}$ and the pair $\underline{M}$:

$$p(\underline{\pi}, \underline{M}|\underline{d}) \propto \left(\frac{M}{V}\right)^N \Gamma\left(\frac{M}{2}\right)\Gamma\left(\frac{1}{2}\right)^{-M} \cdot \\ \cdot \pi_1^{n_1-\frac{1}{2}}\pi_2^{n_2-\frac{1}{2}}...\pi_{M-1}^{n_{M-1}-\frac{1}{2}}\left(1-\sum_{k=1}^{M-1}\pi_k\right)^{n_M-\frac{1}{2}}. \tag{2.6}$$

Notice that we disregarded $p(\underline{M})$ since it is constant.
By integrating eq. (2.6) over all admissible values of $\pi_1, ..., \pi_{M-1}$ we get the posterior probability for the number of bins:

$$p(\underline{M}|\underline{d}) \propto \left(\frac{M}{V}\right)^N \frac{\Gamma(\frac{M}{2})}{\Gamma(\frac{1}{2})^M}\frac{\prod_{k=1}^M \Gamma(n_k+\frac{1}{2})}{\Gamma(N+\frac{M}{2})}. \tag{2.7}$$

Let us notice that the proportionality constant due to Bayes's Theorem, which assures the normalisation condition is satisfied by the density function we found, depends on the actual data we are working with. From now on, in particular in numerical computations, we will consider the un-normalised posterior, which we will call *relative posterior probability*.
According to the method of *maximum a-posteriori estimation*, the optimal choice for $\underline{M}$, the one that provides the best agreement of the model with the observed data, is the one which maximises the logarithm of the posterior probability. Thus, the optimal number of bins is

$$\hat{\underline{M}} = \arg\max_{\underline{M}} \log p(\underline{M}|\underline{d}).$$

A great advantage of this mathematical formalism, is that, once we have computed the optimal binning grid $\underline{M}$, from eqs. (2.6) and (2.7) (see Knuth, 2006 for details) we can analytically compute the mean and the variance of the bin probabilities which are given, respectively, by

$$\mu_k = \left(\frac{M}{V}\right)\cdot\left(\frac{n_k+1/2}{N+M/2}\right)$$

and

$$\sigma_k^2 = \left(\frac{M}{V}\right)^2\cdot\left(\frac{(n_k+1/2)(N-n_k+(M-1)/2)}{(N+M/2+1)(N+M/2)^2}\right)$$

and which allow to construct the optimal histogram with the proper error bars.

2.2 Estimation of a point process's intensity function

In this section we see how Knuth's method permits to test complete spatial randomness (CSR) hypothesis[1], to reduce sampling fluctuations and to give information

[1]We recall that in ecology, the term complete spatial randomness hypothesis refers to the assumption that the pattern under study is a realisation of a homogeneous Poisson process.

regarding important characteristics of the underlying probability density function of a process such as its possible inhomogeneity or anisotropy.
As a first step, we test it on datasets generated from known processes. The point distribution is thus known a priori.
Given a spatial point process $\boldsymbol{X}$ defined on $\mathbb{R}^2$, from Chapter 1 we know that its *intensity function* $\lambda_{\boldsymbol{X}}(\cdot)$ is the first-order statistic measuring the mean number of points per unit area. Denoting with $\mathcal{N}_{\boldsymbol{X}}(\mathcal{W})$ the number of the process' points falling within a region $\mathcal{W} \in \mathbb{R}^2$, we have that

$$\mathbb{E}[\mathcal{N}_{\boldsymbol{X}}(\mathcal{W})] = \int_{\mathcal{W}} \lambda_{\boldsymbol{X}}(x)\mathrm{d}x.$$

We are interested in estimating, given a realisation of the process, its intensity $\lambda_{\boldsymbol{X}}$. Recall that if the process is assumed to be homogeneous, then its intensity function is constant within the observation window $\mathcal{W}$ and it can be therefore approximated by $\hat{\lambda}_{\boldsymbol{X}} = N/|\mathcal{W}|$, where N is the total number of points in the considered window $\mathcal{W}$ and $|\mathcal{W}|$ is this latter area.
Using Knuth's method, we know that we can approximate the intensity function of the possibly non-homogeneous processes by piece-wise constant functions. In fact, by arranging data into Knuth's optimal histogram, the density function at a point x is given by the height of column over the bin containing x. Here we also compare Knuth's answer with kernel methods which are widely used in literature (Illian et al., 2008; Schiffers et al., 2008; Wiegand and Moloney, 2013). In appendix A we also prove its superiority with respect to Stone's non-kernel method.
The idea of kernel methods is to estimate $\lambda_{\boldsymbol{X}}(x)$, $x \in \mathbb{R}^2$ by looking at the number of points falling within the small disk $\mathcal{B}(x, R)$ of radius R centred in x and by dividing it by the area of the disk:

$$\hat{\lambda}_{\boldsymbol{X}}(x) = \frac{1}{\pi R^2} \sum_{x_i \in \boldsymbol{X}} k_R(|x - x_i|). \tag{2.8}$$

In eq. (2.8), x_i is a point of the process $\boldsymbol{X}$ and $k_R(|x - x_i|)$ is the so-called *kernel function*. The R parameter on which $k_R(\cdot)$ strongly depends is called the *bandwidth* of the kernel and it must be carefully chosen. Indeed, a too small R value lead to highlight noisy fluctuations due to sampling stochasticity which are not representative of the process's intensity function. In contrast, a too high value of it will miss important characteristic details of $\lambda_{\boldsymbol{X}}$ which may be relevant for the process under study. Unfortunately, a general recipe to find an optimal bandwidth which gives the right smoothing of a rugged intensity function does not exist (Illian et al., 2008; Diggle, 2013). Some authors suggest the rough estimate $R \sim 1/\sqrt{\lambda_{\boldsymbol{X}}}$ and a subsequent finer tuning of R using visual inspection (Wiegand and Moloney, 2013). On the contrary, Knuth's non-parametric method selects the optimal scale from the data without any assumption on the underlying process indexspatial point process that generated the data. Notice that also simpler methods for assessing the CSR hypothesis like the quadrats count can be considered as parametric methods since their effectiveness depends on the choice of the size of the quadrats.
A first intuitive choice for the kernel function is the so-called *box kernel* (Diggle, 2003), whose estimate of the intensity function at a point x simply consists on the

count of process's points falling within $\mathcal{B}(x, R)$ divided by the area of the disk:

$$k_R^{box}(|x - x_i|) = \begin{cases} 1 & |x - x_i| \leq R \\ 0 & \text{otherwise} \end{cases}$$

As underlined in Wiegand and Moloney, 2013, the box kernel usually leads to rough estimates, particularly in regions where the density of points is low.
We therefore choose to compare Knuth's algorithm with a different kernel method which lead to smoother density estimates: the *Epanechnikov kernel* (Stoyan and Stoyan, 1994). It is defined as:

$$k_R^E(|x - x_i|) = \begin{cases} 2\left(1 - \frac{|x-x_i|^2}{R^2}\right) & \text{if } |x - x_i| \leq R \\ 0 & \text{otherwise} \end{cases}$$

This function weights the points $x_i \in \boldsymbol{X}$ within the disk $\mathcal{B}(x, R)$ according to their distance from the evaluation point x: the smaller $|x - x_i|$, the bigger the weight.
Below we compare the kernel and the Knuth's methods on datasets generated from both homogeneous and inhomogeneous processes to test their reliability at 1) reproducing the intensity of the process and 2) capturing the eventual presence of a gradient in the density functions.

2.2.1 Tests on the detection of CSR and gradients in the density function

We start by applying Knuth's method and the Epanechnikov kernel estimation to a generated CSR pattern within a window $\mathcal{W}$ of area 500×500 units (top panels of figure 2.1).
On the left panel we show the pattern generated according to a homogeneous Poisson process with $\lambda_{\boldsymbol{X}} = 1/500$, while in the middle and in the right panels we can see the intensity function estimated by Epanechnikov kernel and by Knuth's method, respectively. This latter arranges data in a unique bin, so that the intensity function it returns is constant within the plot. Therefore, it perfectly detects the underlying homogeneous structure of the process. By contrast, the kernel method results to be sensitive to sampling fluctuations not representative of the intensity function.
Bottom panels of figure 2.1 show the comparison between the two methods on an inhomogeneous Poisson point process where the true intensity function $\lambda_{\boldsymbol{X}}$ increases with the y coordinates. Although from the Epanechnikov kernel method it is evident that the number of points falling within the upper region of the window is bigger with respect to the lower region, it results still affected by sampling stochasticity. By contrast, Knuth's method arranges the data into a 1×4 grid, perfectly detecting the homogeneity of the process along the $x-$axis and the density gradient along the $y-$axis.

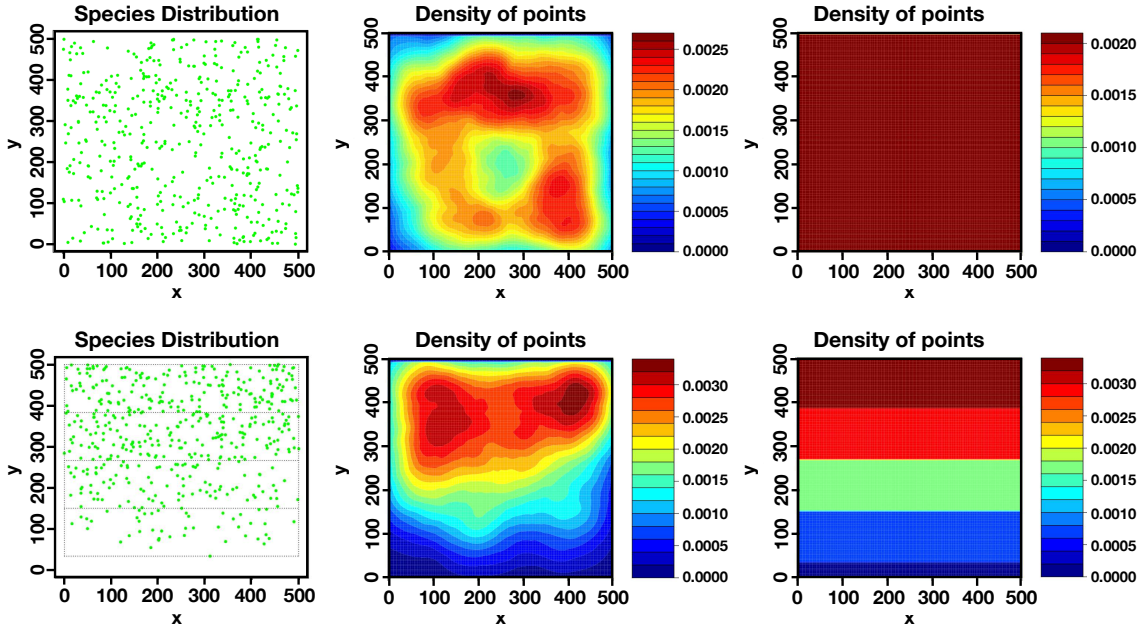


Figure 2.1: On the top: estimation of the intensity function of a Poisson point process of intensity $\lambda_{\boldsymbol{X}} = 1/500$. Epanechnikov kernel with bandwidth $R = 4.5/\sqrt{\hat{\lambda}_{\boldsymbol{X}}}$ (see Wiegand and Moloney, 2013) results to be more sensitive to sampling fluctuations, while Knuth's method arrange data in a unique bin, perfectly detecting the homogeneity of the density function. On the bottom: estimation of the intensity function of an inhomogeneous Poisson point process of intensity $\lambda_{\boldsymbol{X}} = 8 \cdot 10^{-6} y$. Epanechnikov kernel with same bandwidth as above results again sensitive to sampling fluctuations, while Knuth's method detects the homogeneity of the density function on the x-axis and the gradient along the y-axis.

2.3 Test on departure from CSR: clusterisation, dispersion and inhomogeneity

We now test Knuth's method in detecting the major characteristics of a point process' underlying structure: inhomogeneity, anisotropy, clusterisation and dispersion.

2.3.1 Knuth's method description of cluster features

As a preliminary analysis we investigate how the Knuth's method reproduces the features of three different types of clusters: square, circular and Gaussian. We consider plots of area $|\mathcal{W}| = 1000 \times 500$ units as is the BCI. For each type of cluster, we generate 100 datasets setting the number of individuals equal to $N = 1000$ and arranging them in a *unique* cluster positioned in the centre of the window. The characteristic size of the cluster (respectively the square of the cluster side l, of the cluster radius $R_{\boldsymbol{X}}$ and of the average clumping radius $\sigma_{\boldsymbol{X}}\sqrt{\pi/2}$) is made varying from 1 to 100. For each dataset, we compute Knuth's optimal binning area a and the correlation between this latter (or a/π) and the characteristic size of the cluster.

Results are displayed in figure 2.2.
In all cases, the determination coefficient shows a strong correlation between the cluster size and the optimal bin area ($R^2 > 0.9$). In the first case (square clusters) the slope equals 1, meaning that the square cluster is arranged in a unique bin and that the density of points is correctly detected as constant. In the other two cases, the slope is far from 1, meaning that the cluster is described using a higher number of bins. In particular, in the second case, Knuth's method aims at reproducing the circular boundary of the cluster while in the last case, the optimal bin serves at reproducing both the circular boundary of the cluster and the Gaussian shape of the density.

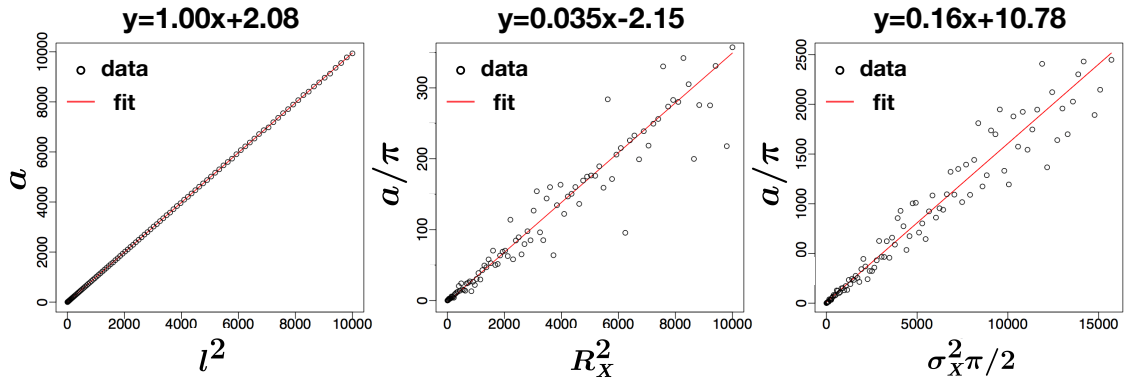


Figure 2.2: Analysis with R software of the linear relation between Knuth's optimal bin area a and the characteristic size (l^2, R_X^2 and $\sigma_X^2\pi/2$) of three type of clusters: square with uniform density (left panel), circular with uniform density as for Matérn process (middle panel) and circular with Gaussian density as for mTp process (right panel). Each dataset consists of 1000 points and for each case the characteristic size varies from 1 to 100. In the first case we plot a against the square of the cluster side l^2, getting a determination coefficient of $R^2 = 1.00$. In the second case we plot a/π against the squared radius of the cluster R_X^2 and we get $R^2 = 0.90$ and in the last case, plotting a/π against the square of the mean distance of a point from the cluster centre (average clumping radius) $\sigma_X^2 \cdot \pi/2$, we get $R^2 = 0.94$. The correlation between the two variables therefore results very strong in all cases.

In figure 2.3 we plot an example of three of the generated datasets with characteristic size equal to 50.
Here we also compare Knuth's with Epanechnikov kernel method in estimating the intensity function of these different cluster-structures.
In the first case Knuth's algorithm collects the points in a unique cluster, as we can see from the optimal grid we insert in the plot of the data distribution (left top panel). Thus the resulting estimation of the intensity perfectly coincides with the underlying one, given by the constant value 0.4 in the 50×50 cluster area and zero outside.
In the second and in the third case Knuth's method arranges data in a histogram with a higher number of bins in order to capture the circular shape of the clusters. In all cases, we remark that the bin sides are practically equal, meaning Knuth detects the isotropy of the cluster structures.

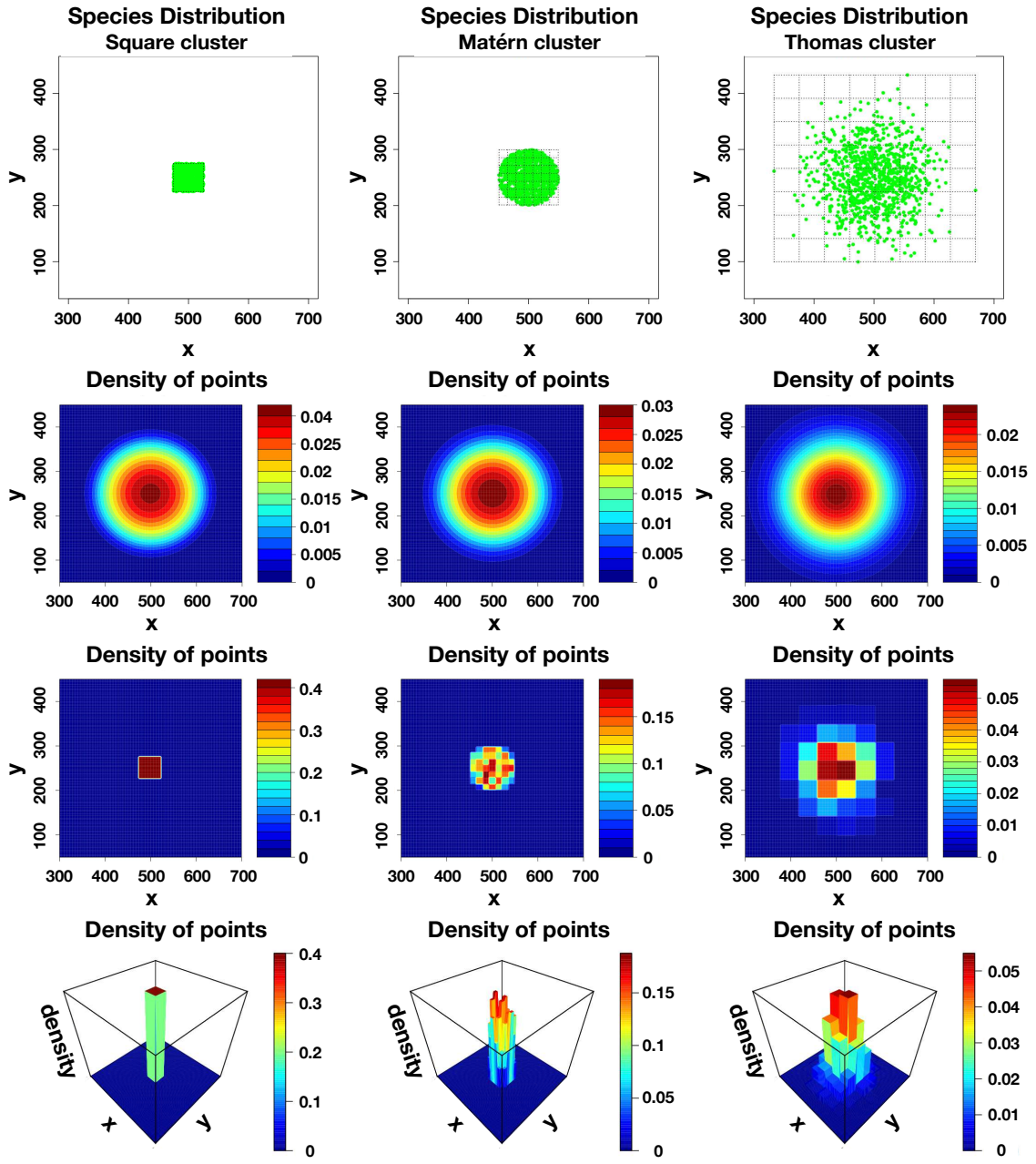


Figure 2.3: Kernel estimation versus Knuth's method on three different type of clusters: square with uniform density (left panels), circular with uniform density as for Matérn process (middle panels) and circular with Gaussian density as for mTp process (right panels). From top to bottom: distribution of points, kernel estimation of the intensity, Knuth estimation and Knuth histogram. In the first case Knuth's method collects the points in a unique cluster correctly detecting the homogeneity of the square cluster structure. In the second and in the third case it arranges data with a higher number of bins to capture the circular boundary of the clusters. On the contrary, from kernel density plots there seems to be no actual difference from the three clusters apart from their size.

By contrast, looking at the density plots obtained by the Epanechnikov kernel method, we can see that the three clusters are not quite distinguishable one from the other if not slightly for their sizes.

2.3.2 Test on the detection of anisotropies

We test Knuth's method's capability of detecting another relevant characteristic of a process: the anisotropy. By construction, Knuth algorithm is sensitive to the inversion of the orthogonal axes: if the two-dimensional pattern is tilted by a multiple of 90°, also the resulting optimal grid is rotated of the same angle.
Here we apply Knuth's method on a generated dataset consisting of 1000 individuals aggregated in a unique anisotropic Gaussian cluster. The only difference with the modified Thomas process (mTp in the following) is that the offspring is now distributed around its parent according to a bivariate Gaussian with standard deviation along the $x-$axis twice than along the $y-$axis.
In figure 2.4 we represent the generated datasets rotated of 0°, 90°, 45° and 135°. The data are also arranged in the optimal grid returned by Knuth's method. This latter well captures the anisotropic structure of the clusters: it gives 47×30 units in the first case, 30×47 units in the second, 41×39 units in the third and 39×41 units in the last one.
Notice that it results sensitive to the inversion of variance's direction between 0°-90° and 45°-135° cases, which leads to the inversion of the optimal bin sizes.
Moreover it responds well to the anisotropic structure, since in the 0° and 90° cases Knuth's method sees the greater variance along x with respect to the one along $y-$axis, in the 45° and 135° cases, it captures the isotropy along the principal axes.
These results confirm that Knuth's method is very efficient at detecting the anisotropy of the cluster structure. However, an obvious limitation of the method is that it returns a unique bin size for the whole plot. Therefore, when dealing with real patterns, where anisotropic clusters may be structured at different scales or oriented in many directions, the optimal bin size is the result of a compromise between these different sub-structures of the dataset. Notice that while different bin sizes denote that Knuth's method classifies the pattern as anisotropic, the contrary does not hold.
However, previous remarks suggest that the difference in bin width along two orthogonal axes is an index of how anisotropic is the spatial pattern. We define the anisotropy index of a pattern as $I_{an} = |a_y - a_x|/\max(a_x, a_y) \in [0, 1]$, where a_x is the bin width along the $x-$axis and a_y the bin width along the $y-$axis in order to be invariant with respect to inversion of the axes. To test the usefulness of the anisotropy index as a tool for detecting the anisotropy of a spatial pattern, we compute I_{an} for a set of patterns generated from the mTp process with 3000 points and parameters $\rho_{\boldsymbol{X}} \cdot |\mathcal{W}| = 5$ and $\sigma_{\boldsymbol{X}} \in 0, ..., 100$. The obtained frequency histogram (see figure 2.5) is highly picked around zero, meaning that Knuth's method detects the isotropic structure of the data distribution along the principal axes. In Section 2.4 below we compute the anisotropy index for the BCI dataset containing the location of the trees of around 300 species showing that Knuth's method can be efficiently used to test hypotheses on the underlying process from which real data are sampled.

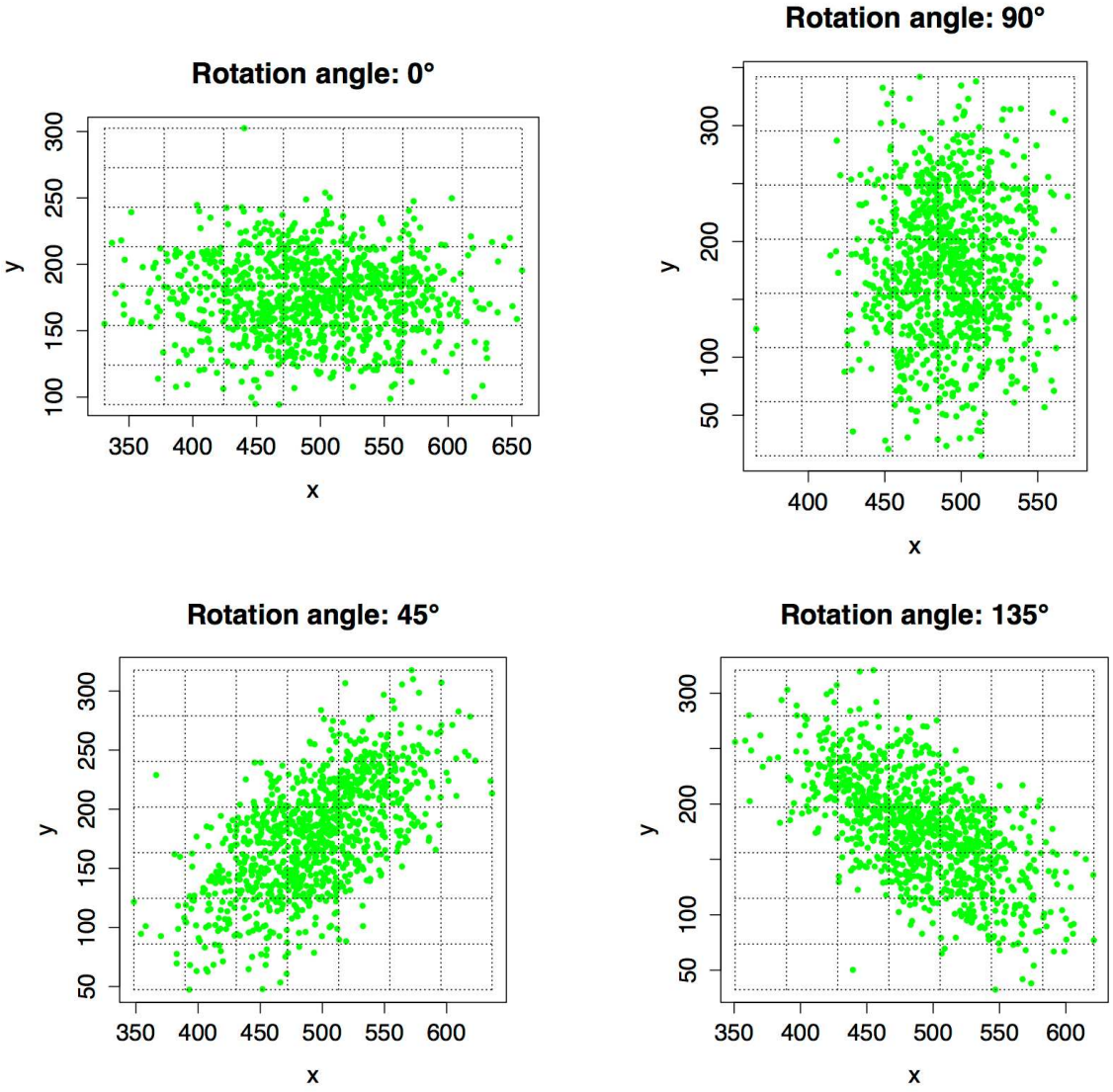


Figure 2.4: Knuth's method's answer to anisotropic Gaussian clusters: plot of a dataset consisting of 1000 individuals clumped in one cluster with $\sigma_x = 60$ units and $\sigma_y = 30$ units rotated of: 0°, 90°, 45° and 135° (top to bottom, left to right), arranged in the optimal grid returned by Knuth's method. This latter well captures the anisotropic structure of clusters: firstly, it results sensitive to the inversion of variance's direction between 0°-90° and 45°-135° cases, which leads to the inversion of the optimal bin sizes; secondly, it responds well to the anisotropic structure, since in the 0° and 90° cases Knuth's method sees the greater variance along x with respect to the one along $y-$axis, while in the 45° and 135° cases, it captures the isotropy along the principal axes.

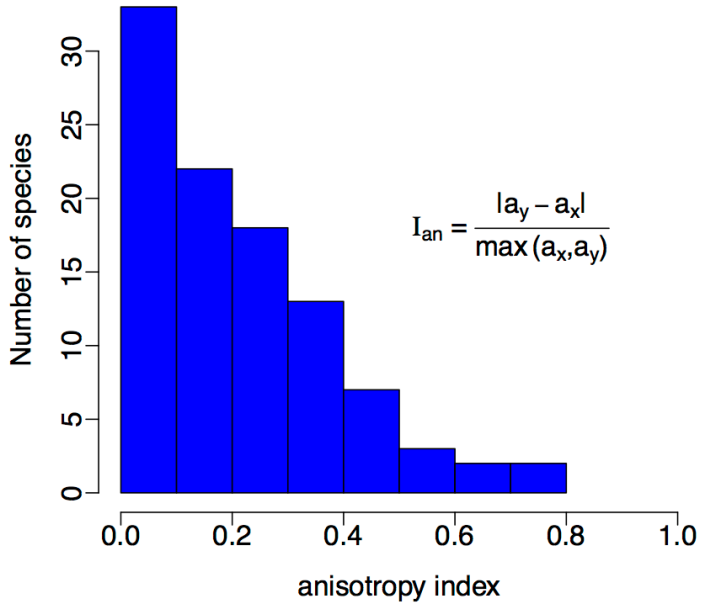


Figure 2.5: Knuth anisotropy index for the database consisting of 100 datasets generated from an mTp with 3000 points, 5 clusters and $\sigma_{\boldsymbol{X}}$ varying from 1 to 100. In the formula, a_x and a_y are the bin width along the x and the $y-$axis, respectively.

2.3.3 Second-order statistics for the estimation of cluster sizes and hard core radii

In spatial ecology, the CSR hypothesis mostly comes to fail due to several reasons: on one side, changes in environmental conditions, such as in physical features of the landscape or in chemical composition of the soil, may lead to inhomogeneous patterns. On the other side, different seed dispersal mechanisms may favour the formation of clumped structures as well as dispersed one. These pattern characteristics are revealed by second-order statistics, which take into account the correlations between pair of points due to possible interactions. Below we briefly recall the most common second order statistics and their use.

The probably most used second-order statistics for homogeneous pattern is the Ripley's K-function $K(r)$ (see Chapter 1), which we recall is the expected number of points falling within a distance r from a point of the process $\boldsymbol{X}$, divided by the intensity function $\lambda_{\boldsymbol{X}}$. Since under CSR hypothesis $K_{\boldsymbol{X}}(r) = \pi \cdot r^2$ scales quadratically with the distance, it is usually substituted with the L-function $L_{\boldsymbol{X}}(r) = \sqrt{K_{\boldsymbol{X}}(r)/\pi} - r$, which takes constant zero value for the CSR model. Both Ripley's and L functions are regarded in literature as cumulative statistics (Wiegand and Moloney, 2013), which do not permit to properly infer pattern characteristics

at an arbitrary chosen scale.
This limitation can be avoided by taking the pair correlation function $g_{\boldsymbol{X}}(r) = K'_{\boldsymbol{X}}(r)/2\pi r$ (Adorisio et al., 2009; Azaele, Maritan et al., 2015). It is defined as the ratio between the density of points falling within a small ring distant r from a point of the pattern and the constant intensity function $\lambda_{\boldsymbol{X}}$ of the process, assumed to be homogeneous.
Since values of $g_{\boldsymbol{X}}(r)$ greater than 1 indicate clustering, whereas values less than 1 indicate dispersion, the point where the pair correlation function intercepts the straight line $y = 1$ gives a rough estimate of the average diameter of clusters for clumped patterns or the average hard core radius for overdispersed ones.
As pointed out in Schiffers et al., 2008 and Wiegand and Moloney, 2013, both Ripley's and the pair correlation function are *window-dependent*, in the sense that their ability in detecting the scale of clustering or dispersion depends on the window within which we arrange the data. For example, if the plot contains empty spaces, these statistics put in evidence a clustered structure which is not due to the pattern's generating process. This phenomenon is called *virtual aggregation* and it affects how these statistics perform in inhomogeneous patterns. In these latter cases, the apparent clumping nature of the pattern is due to the fact that both $K_{\boldsymbol{X}}$ and $g_{\boldsymbol{X}}$ are normalised with the empirical intensity $\hat{\lambda}_{\boldsymbol{X}}$, which is wrongly assumed as constant. Schiffers's index (see Chapter 1 and Section 1.3.1), using the derivative of the pair correlation function, is less sensitive to this phenomenon.
Here we test the ability of Knuth's method to deal with clustered, dispersed and inhomogeneous patterns and to cope with the virtual aggregation phenomenon which may arise when considering non homogeneous patterns. We have thus selected a model of clustered process, the modified Thomas process and a dispersed one with an hard core repulsion radius, which are briefly described below (see Chapter 1 for a more detailed description of such processes).
The *mTp*, which is one of the simplest variant of the Poisson cluster process (He, Legendre et al., 1997; Condit et al., 2000; Plotkin, Potts et al., 2000; Morlon et al., 2008; Azaele, Cornell et al., 2012), is one of the most used model in literature to describe the clumping mechanism of plants' species. This process, in addition to being mathematically tractable (Illian et al., 2008; Diggle, 2013) has been shown to be more efficient than others in capturing important biological curves such as the *species-area relationship* (see, e. g. Plotkin, Potts et al., 2000) or to model species occupancies at different spatial scales (Azaele, Cornell et al., 2012). Instead, as shown in Morlon et al., 2008, it results to be inadequate in reproducing the *distance-decay relationship* (see Chapter 3), thus indicating that some of the assumptions of *mTp* do not hold in nature.
The hard core process is generated from a uniform distribution with the additional constraint that if a point comes to fall within a fixed hard core distance from a pre-existing one, it is rejected. This model is of importance in ecology to describe reproductive mechanisms where the seeds are shot apart from parents to avoid species-specific predators (the so-called *Janzen-Connell effect*, see Janzen, 1970; Connell, 1971; Adorisio et al., 2009).

2.3.4 Interplay between second-order statistics and Knuth's method

Despite their ability to detect significant departures from CSR processes, the pair correlation function and the Schiffers index lack in reliability at determining the cluster size when dealing with inhomogeneous patterns being subject, the former much more than the latter, to the virtual aggregation phenomenon.
To show this, we compute the pair correlation function and the Schiffers index for three generated spatial point processes, respectively: a) an mTp, b) an overdispersed process with fixed hard core radius and c) an inhomogeneous Poisson process.
To see the virtual aggregation phenomenon, they are firstly considered within a 500×500 units window (black curves of figure 2.6), and then within a larger one, obtained from the previous by adding at its bottom an empty square box of the same area (grey curves). In the top panels of figure 2.6 we show the results for the clustered point process. This latter has been generated according to a mTp with parameters $\rho_{\boldsymbol{X}} = 2 \cdot 10^{-4}$, $\sigma_{\boldsymbol{X}} = 10$, $\mu_{\boldsymbol{X}} = 10$, where $\rho_{\boldsymbol{X}}$ is the intensity of the Poisson process from which parents are generated, $\sigma_{\boldsymbol{X}}$ is the standard deviation of the Gaussian distribution of the offspring around each parent and $\mu_{\boldsymbol{X}}$ is the mean number of offspring per parent. The two second-order statistics $g_{\boldsymbol{X}}$ and Schiffers's index well capture the clustered structure of the pattern, both in the original window and in the expanded one. In particular, they reveal an average clump diameter around 37-40 units, which overestimates the true value $2\sigma_{\boldsymbol{X}} \cdot \sqrt{\pi/2} \approx 25$ units. Looking at the grey curves, we can observe the phenomenon of the virtual aggregation. While the black pair correlation function intercepts the line $y = 1$, the grey one sees the species as clustered at any scale. By contrast, their corresponding Schiffers's indexes are much closer one another, meaning the addition of an empty box slightly affects this latter statistic. By applying Knuth's method we get a 22×20 grid which results in a clump size of 26 units circa, which is therefore the closest to the real one. Notice that Knuth's method automatically restricts to the data span V (see Section 2.1) therefore it results to be insensitive to voids in the pattern and hence does not suffer from the virtual aggregation effect.
In the middle panels of figure 2.6 we carry out the analysis for 500 points sampled from a uniform distribution with the additional constraint that if a point comes to fall within a fixed hard core distance from a pre-existing one, it is rejected. Once again all statistics are able to capture the overdispersion at small scales: they return an hard core radius around the true value of 10 units, with a big and a slight difference between black and grey curves in the pair correlation and the Schiffers index case, respectively. Applying the Knuth's method we obtain a 1×1 grid as in the CSR case: the estimated density is therefore correctly detected by the optimal histogram which sees the homogeneity of the pattern. Here we see a limitation of the Knuth's method which is unable to reveal second-order information such as the dispersion of a spatial pattern or its hard core radius.
The last considered pattern, whose results are shown in the right panels of figure 2.6, is sampled from a Poisson process with intensity $\lambda_{\boldsymbol{X}}$ increasing with the x-coordinates, thus presenting a small gradient along the axis. Here the phenomenon of virtual aggregation is again well visible looking at the graph of $g_{\boldsymbol{X}}$: the pattern

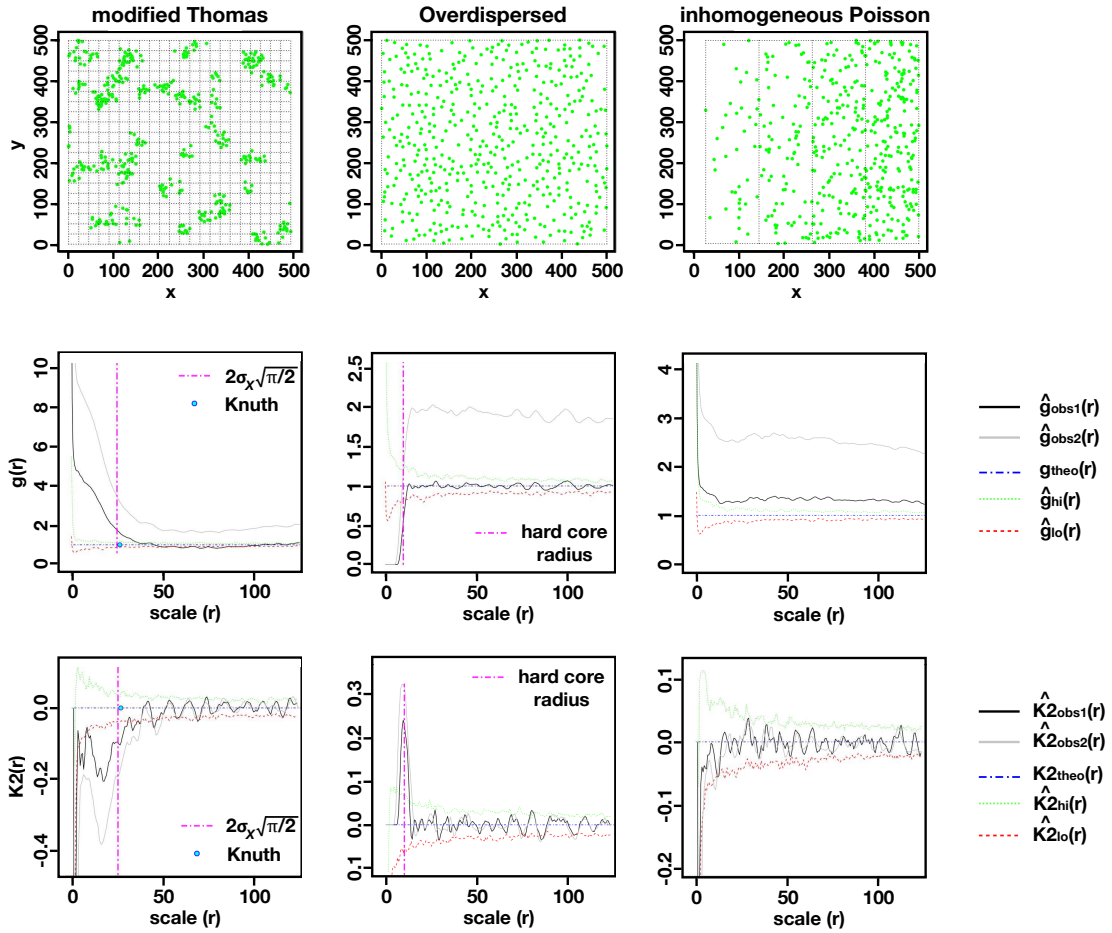


Figure 2.6: Second-order statistics for three generated pattern: a modified Thomas process ($\rho_{\boldsymbol{X}} = 2 \cdot 10^{-4}$, $\sigma_{\boldsymbol{X}} = 10$, $\mu_{\boldsymbol{X}} = 10$), an overdispersed (of hard core radius of 10 units, $N = 500$) and an inhomogeneous Poisson process ($\lambda_{\boldsymbol{X}} = 8 \cdot 10^{-6}x$). From top to bottom: point patterns with superimposed Knuth grid, pair correlation function('g' in the legend) and Schiffers's index ('K2' in the legend). Significance departures from the CSR model are tested by Monte Carlo simulations (blue lines are theoretical CSR curves, green and red are the higher and lower band of the 99% confidence envelopes, respectively). Black lines in the graphs of $\hat{g}_{\boldsymbol{X}}$ and $\hat{K}_{2,\boldsymbol{X}}$ refer to the original datasets, plotted within a 500×500 units window. Grey lines are obtained by considering the same data points but within an enlarged window, created by adding an empty 500×500 units square at the bottom of the previous window. For the mTp process we also inserted the value of Knuth binning diameter $2\sqrt{a/\pi}$. Computations of both g and Schiffers's index have been performed using the R code provided in Schiffers et al., 2008.

is detected as clustered at all scales by the statistic. Instead, the Schiffers index is not affected by this, although it cannot distinguish the process from a CSR one. By contrast, the Knuth's method arrange data in a 4×1 grid, which permits us to capture the homogeneity along the y-axis and the gradient along the x-axis.
In conclusion, combining Knuth's method to Schiffers's index leads to a better understanding of the underlying process from which a pattern is generated: the former permits to detect homogeneity against gradient in densities and gives a measure of how structured it is in the sense that small bins indicate clumping while large bin indicate Poisson-like distributions. The latter, on the other hand, allows to give quantitative information on the scale of both clumping and dispersion.

2.4 Application to the BCI ecological database

We consider an open access ecological dataset consisting of the spatial coordinates of individuals belonging to 300 different species of plants located in a 50 ha rectangle of the Barro Colorado island rainforest in Panama. Our goal is to show that the choice of modelling BCI species' distribution through a mTp, which has been proven to be efficient in capturing some important biological curves but not others (Plotkin, Potts et al., 2000; Morlon et al., 2008; Azaele, Maritan et al., 2015), in many cases is not supported by Knuth's method.
Notice that checking the goodness-of-fit of a fitted model using a minimum contrast method (Diggle, 2013; Plotkin, Potts et al., 2000) when the theoretical form of the summary function (usually Ripley's K) is not known is a difficult task. For a mTp process the form of Ripley's function is known and equals

$$K_{\boldsymbol{X}}(r)_{mTp} = \pi r^2 + \frac{1}{\rho_{\boldsymbol{X}}}\Big(1 - \exp\Big(-\frac{r^2}{(2\sigma_{\boldsymbol{X}})^2}\Big)\Big).$$

Hence the fitted pattern is already optimal with respect to a minimum contrast goodness-of-fit criterion. Nevertheless, the analysis of the difference of the optimal bin size for the real and the mTp-generated pattern may reveal a strong departure form the real data.
To find the species that are suitable to be described by a clumped pattern, we select all species with abundance between 20 and 3000 individuals (204 species) and we compute the mTp parameters $(\rho_{\boldsymbol{X}}, \sigma_{\boldsymbol{X}}, \mu_{\boldsymbol{X}})$, for each species by fitting the empirical Ripley's function (panels (a-c) of figure 2.7).
As in Morlon et al., 2008, we discard the species whose distribution is not quite distinguishable from a random one – in this case the Poisson cluster process cannot capture correctly the underlying structure of the data – according to the following criteria: i) the mean cluster diameter $\sigma_{\boldsymbol{X}}\sqrt{2\pi}$ is bigger than 500 m and ii) the number of clusters is bigger than the number of the individuals. We found that 183 of the 204 species satisfied both the criteria.
Panels (a-c) of figure 2.7 show the histograms of the mTp parameters of the selected BCI species.
For each of them we compute the quantity $\sqrt{a/\pi}$, which is the radius of the circle equivalent to a rectangular Knuth optimal bin of area a. We know that this latter

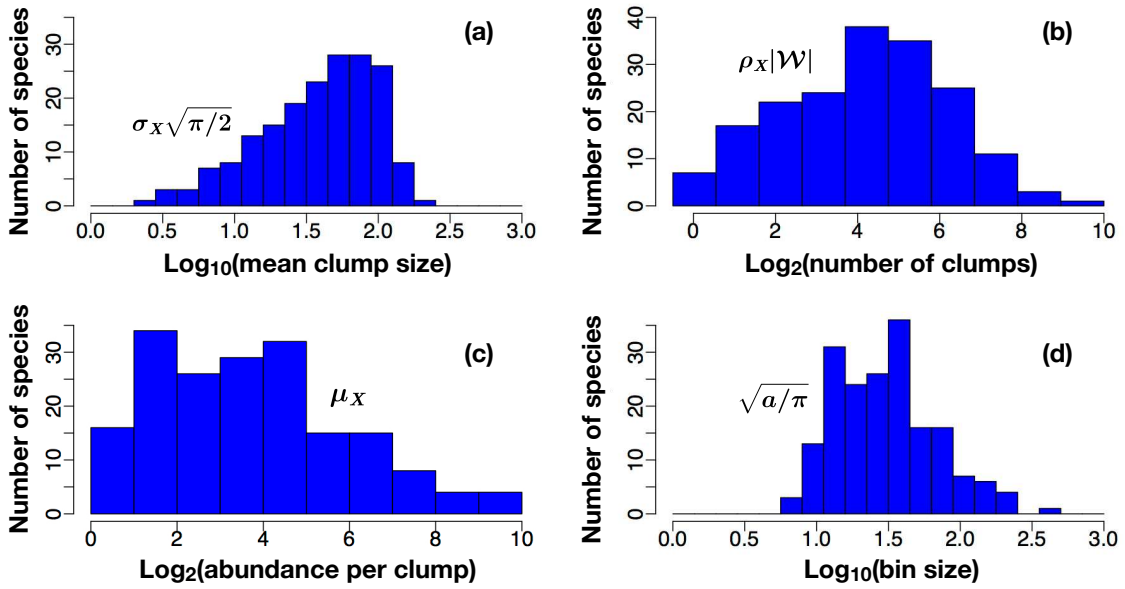


Figure 2.7: Frequency histograms of the mTp parameters fitted by the minimum contrast method and Knuth's optimal bin radius, defined as the square root of the optimal bin area over π.

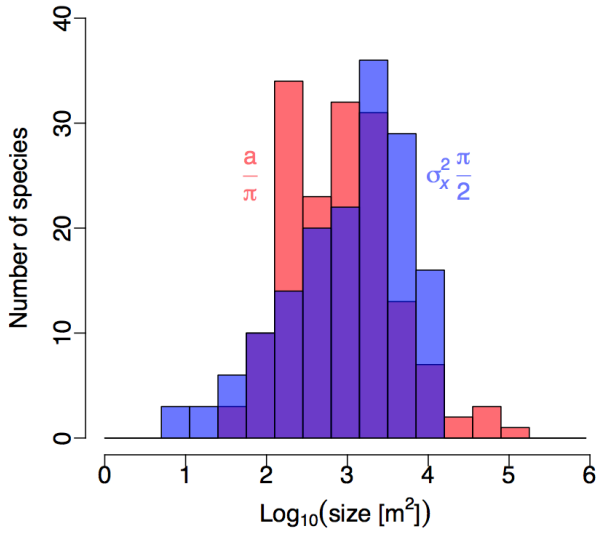


Figure 2.8: Comparison between Knuth's optimal radius and estimated clumping mTp radius for the species of BCI (for a better visualisation in the histograms we plotted their square values). The first one is obtained as $\sqrt{a/\pi}$, which represents the radius of the circle equivalent to Knuth's optimal bin. The second one is given by $\sigma_X\sqrt{\pi/2}$, which is the average clumping radius according to the radially symmetric, bivariate Gaussian density.

represents a measure of the size of the underlying minimal structure of the dataset. From (d) panel of figure 2.7 we see that there is no preferred choice for a common optimal bin area for all species, since the values of the optimal binning areas span from 100 m^2 to $5 \cdot 10^5$ m^2 circa. This is not surprising, since each species has its own distribution due to myriad of factors such as seed dispersal, gap recruitment or adaptation to the surrounding soil (Augspurger, 1984; Plotkin, Potts et al., 2000). In figure 2.8 the frequency histogram of the mTp clump area $\sigma^2_{\boldsymbol{X}} \cdot \pi/2$ is superimposed to the one of a/π. In this cumulative plot the two histograms are respectively right and left-skewed showing that globally Knuth's method assigns a finer structure to the real patterns with respect to the mTp fitting.

2.4.1 Difference index for BCI

In order to see how well the mTp reproduces the species of BCI, we generate, for each of them, a mTp counterpart as follows. Given each species's total number of individuals n_s and its mTp parameters obtained through the minimum contrast method $(\rho_{\boldsymbol{X}_s}, \sigma_{\boldsymbol{X}_s}, \mu_{\boldsymbol{X}_s})$, we simulate a mTp process of such parameters and of n_s points in the 1000×500 units window representing our BCI surveyed region (see Chapter 1, Section 1.5.2). For this test we select only species with $\sigma^2_{\boldsymbol{X}} \pi/2 < 10^4$ m^2. We thus compare Knuth's optimal bin size a for each BCI real species with the one obtained for its correspondent mTp generated counterparts.

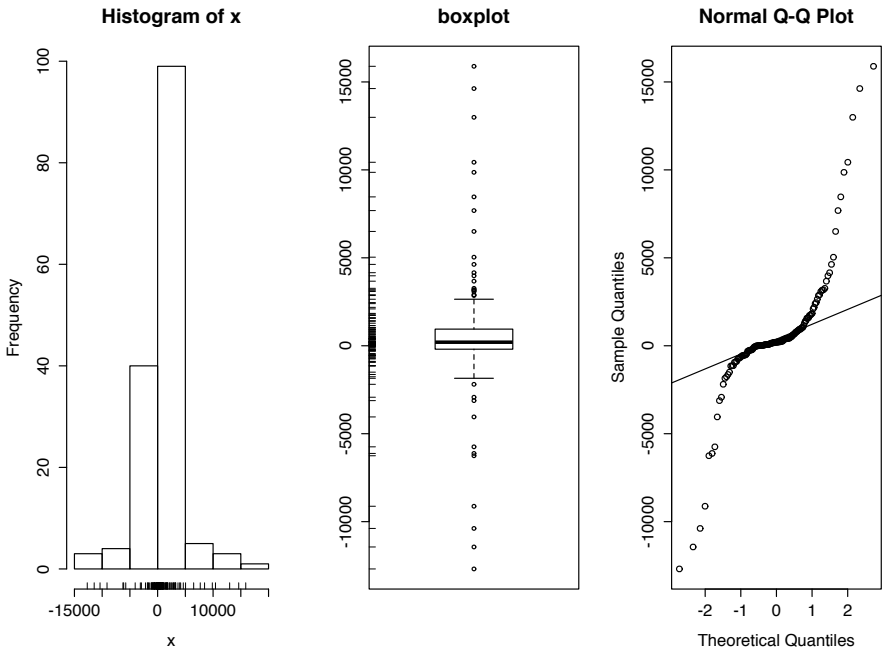


Figure 2.9: R-exploratory analysis of difference data $x = \Delta = a(mTp) - a(real)$: histogram, boxplot and relation with the quantiles of a normal distribution. The mean is 546.4, the median 202.9 and the standard deviation equals 3458.9.

In figure 2.9 we display the exploratory analysis (histogram, boxplot and QQ plot) with R-software of the resulting difference $\Delta = a(mTp) - a(real)$. We see that

the histogram is right-skewed with fat tails. For 50% of the species the difference of the bin area Δ is bigger than 2 times the smaller of the two. For these species which are in the tails of the distributions, the mTp process fails at reproducing the real pattern either because the generated clustered pattern has a coarser scale than the real one losing details of the original fine structure ($a(real) \ll a(mTP)$, right tail) or because it introduces an artificial clustered structure on a more uniform real pattern ($a(real) \gg a(mTp)$, left tail).
To have a visual inspection of the difference between real and generated species, we select three species respectively i) in the right tail, ii) in the centre and iii) in the left tail.
In figure 2.10 we display the real and the mTp generated patterns of such species. In the first case the mTp reproduces a coarser pattern than the real one while in the latter it introduces an artificial finer structure. Finally, in the second case, Knuth's method recognises as similar the two spatial structures and therefore the hypothesis that the species is distributed according to a mTp is not rejected.

2.4.2 Anisotropy index for BCI

We compute the anisotropy index $I_{an} = |a_y - a_x|/\max(a_x, a_y)$ both for the BCI real species and for their correspondent mTp generated counterparts.
In figure 2.11 we show the frequency histogram of the anisotropy index for real species superimposed onto the one for mTp generated species. In the first case there is a high number of species whose index is far from 0, meaning that their underlying density function is not recognised as isotropic by Knuth's method. In the second case, Knuth histogram is more shifted against the $y-$axis, meaning that Knuth's method sees the new generated species' process more isotropic than before. This fact suggests that a reason for which mTp process fails in capturing some important ecological curves is the fact that the hypothesis of isotropic clusters is too strong and is not supported by the real data.

2.4.3 Relation with the abundance

The relation between species abundance and degree of aggregation is still debated as an important issue of ecological theory (He, Legendre et al., 1997; Plotkin, Potts et al., 2000; Condit et al., 2000; Morlon et al., 2008). Moreover, as pointed out in Morlon et al., 2008, the correlation between these two quantities strongly depends on how they are measured. For example, for Pasoh forest, in He, Legendre et al., 1997 the proposed *Donnelly clumping index* based on nearest-neighbour distance shows a slightly positive correlation between abundance and aggregation. By contrast, the *relative neighbourhood density* Ω_{0-10} (Condit et al., 2000; Harte, Conlisk et al., 2005; Ostling et al., 2000) and the Cramer-von Mises-type k statistic (Plotkin, Potts et al., 2000) are negatively correlated to abundance.
To investigate if mTp and Knuth aggregation parameters are similarly correlated with species abundance or not, we compute on one hand the correlation between the abundance and the following mTp quantities: the mean number of parents $\rho_{\boldsymbol{X}} \cdot |\mathcal{W}|$, the mean clump radius $\sigma_{\boldsymbol{X}} \cdot \sqrt{\pi/2}$, the mean number of offspring per parent

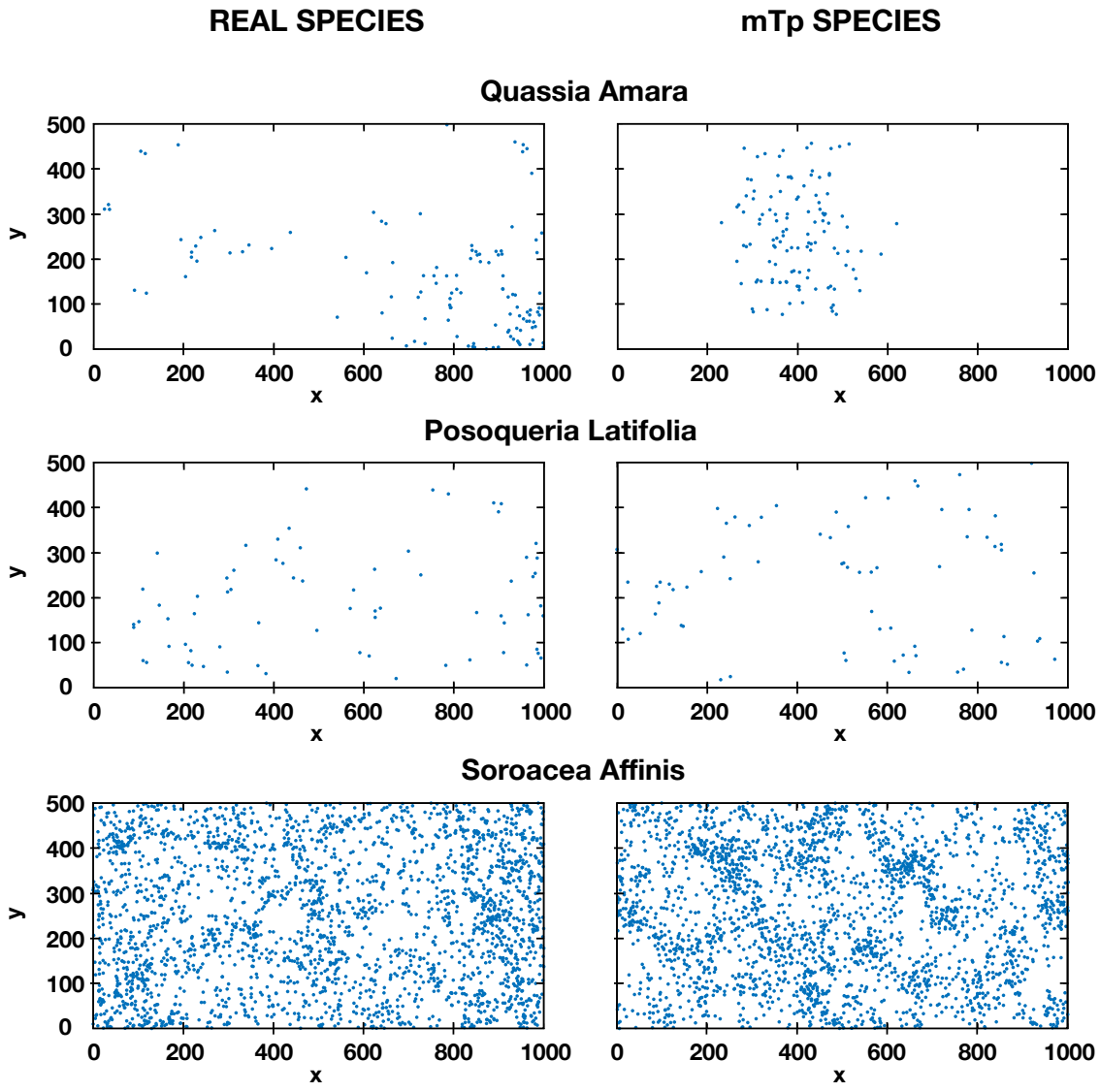


Figure 2.10: Plot of three species distribution from the BCI surveyed area (left column) and plot of the same species distribution generated according to the mTp with parameters fitted from the real data (right column). Species are selected to display the three cases : i)(*Quassia amara*, top) $a(mTp) \gg a(real)$, meaning that the Knuth's method detects a finer structure for the real species compared to the generated one; ii) (*Posoqueria latifolia*, middle) $a(mTp) \approx a(real)$ Knuth's method recognises as similar the two spatial structures and therefore the hypothesis that the species is distributed according to a mTp is not rejected; iii) (*Soroacea affinis*, bottom) $a(mTp) \ll a(real)$, the Knuth's methods detects a finer structure for the generated species with respect to the real one.

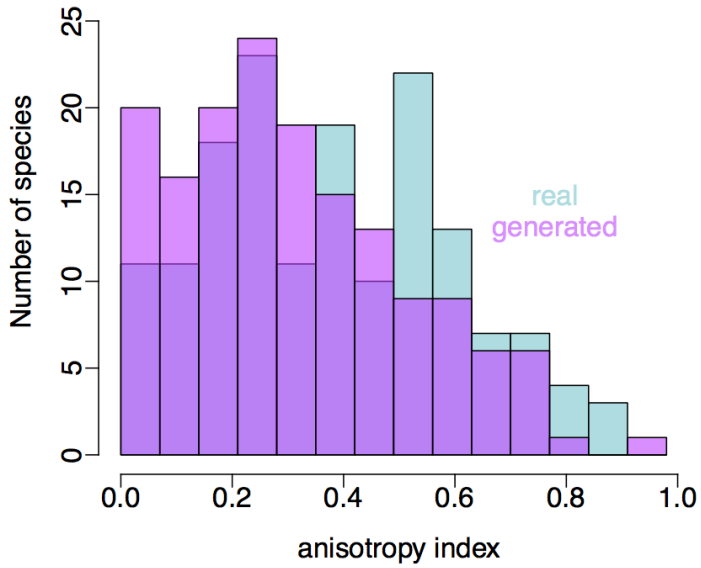


Figure 2.11: Frequency histograms of anisotropy index for the real BCI species and the ones generated by a modified Thomas process with parameters fitted by data.

μ_X and the relative neighbourhood density Ω_{0-10} (see Morlon et al., 2008) and on the other hand the correlation of species' abundance with Knuth optimal bin area and with index of anisotropy (see figure 2.12). These latter, as we have seen, give us information on how structured is the data density function and how far it is from a uniform or isotropic one.

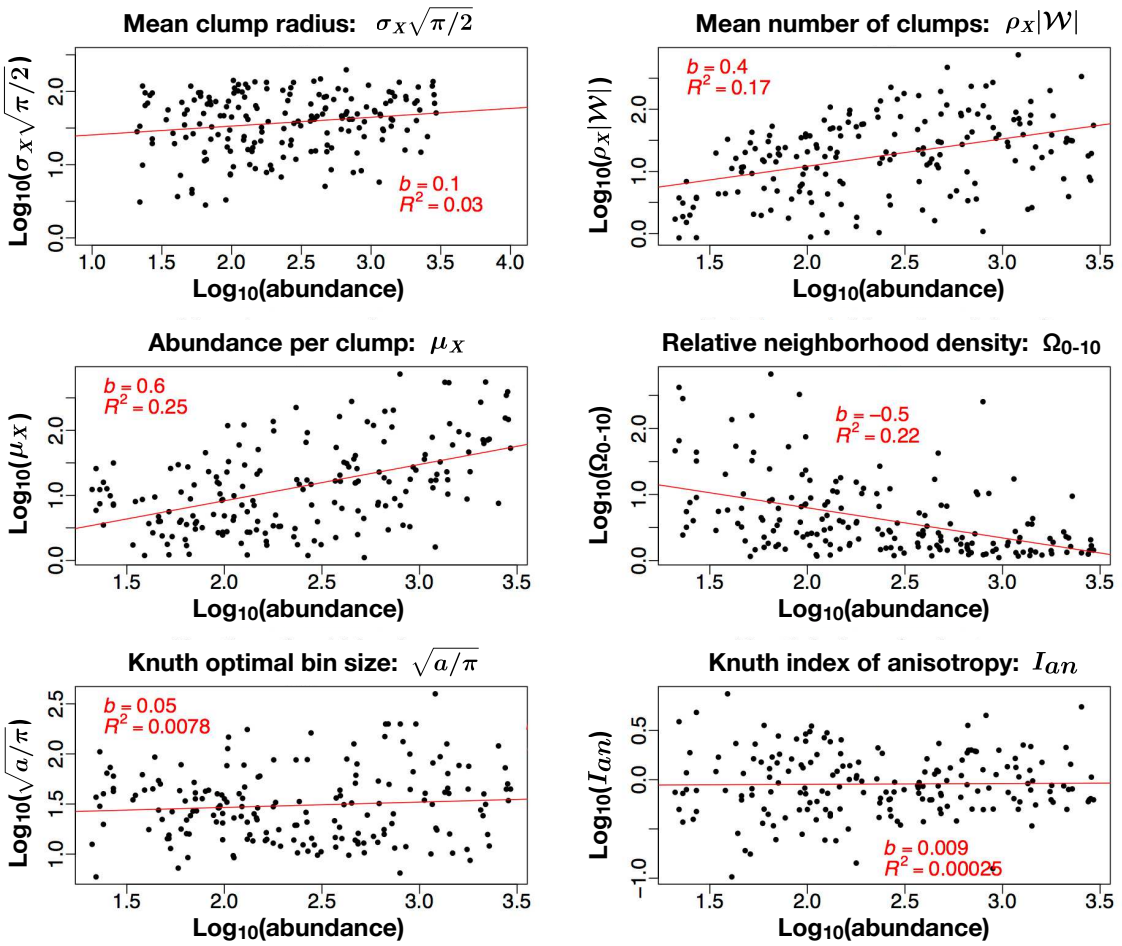


Figure 2.12: Correlation between mTp and Knuth's parameters and the abundance of a species.

From the determination coefficients, only the relative neighbourhood density shows a negative correlation with abundance, while the other mTp parameters result to be slightly positive correlated with it. This is in accordance with the literature (Morlon et al., 2008). Knuth optimal bin area and index of anisotropy are, instead, insignificantly correlated with the abundance with respect to the determination coefficient (bottom panels of figure 2.12). This is quite reasonable because Knuth optimal grid depends only on data distribution, and not on their abundance.

3

Diversity and Similarity Indexes

Estimating biodiversity of forests is a central issue in modern conservation ecology. Both from the theoretical and field application point of view it represents a daunting challenge.

The very same word *biodiversity* may, and actually does, assume many different meanings and refer to a vast number of notions depending on the subject under study, so that different additional terms have been introduced to reflect the multiple aspects of this important concept (Colwell, 2009).

Following Whittaker[1] (Whittaker, 1960), we distinguish between *alpha*, *beta* and *gamma-diversity* in relation to the scale of investigation. In particular, we use the term alpha-diversity when biodiversity is measured at the scale of a single sample, while beta-diversity or *turnover*[2] refers to the change in species composition between samples. Finally, we talk about gamma-diversity when describing the diversity of an assemblage of samples.

In the first part of this chapter we will explore in details some of these notions and we will present the most important indexes introduced in literature to measure ecological diversity in community composition. We will then study how to insert them in the context of point processes' theory. This will be useful in the next chapter, where we will focus on the more specific problem of describing the decay of similarity between two regions of a landscape as a function of the distance between them.

[1] In Whittaker, 1972 the author proposed a new terminology to refer to different diversities in space (Magurran, 2013). Indeed, he distinguished between seven spatial scales: within sample (point diversity), between samples within a defined habitat (pattern diversity), within a defined habitat (alpha-diversity), between habitats within a defined landscape (beta-diversity), within a defined landscape (gamma-diversity), between landscapes within a defined biogeographic province (delta-diversity) and within a defined biogeographic province (epsilon-diversity). Here we stick with the original definition of 1960.

[2] The term *turnover* is more often used to refer to changes in time.

3.1 The concepts of diversity and similarity

The first and most intuitive indicator of diversity is the *species richness*, term coined by McIntosh to indicate the total number of species S of a community (Magurran, 2013; McIntosh, 1967). This is the oldest mathematical descriptor, and yet the poorest, since it does not give any information of the actual composition of the community and it clearly cannot be used for comparisons between ecosystems, if not in a trivial way.
However, despite its simplicity, estimating species richness from small samples to bigger areas is not trivial at all. Indeed, lots of estimators have been introduced to tackle the problem of inferring the number of missing species which we do not see in the surveyed area, but which are present in the ecological community under study (see Chapter 5). The major problem is that usually the species estimates depend on the sampling effort Gaston, 1996. In literature, there have also been proposed other simple diversity indexes based on species richness which, apart from the observed number of species S, they also take into account the observed number of individuals N in order to reduce the dependence on the sample scale (Magurran, 2013). Two examples are the Margalef's diversity index $(S-1)/\ln N$ and (Clifford and Stephenson, 1975) and Menhinick's index $S/\sqrt{N}$. Anyway, even these latter lack in giving information about the composition of the sampling unit under study.
Complementary to the notion of species richness, are the ones of *evenness* (Heip et al., 1998; Pielou, 1969, 1975; Smith and Wilson, 1996; Gray, 2000) and *dominance* or *concentration* (Magurran, 2013; Hill, 1973; Jost, 2010b). In particular, a community has high evenness if all species are equally abundant, whereas it has high dominance if there is one or few species with population much larger than all the others. Both these concepts are therefore strictly connected to the species-abundance distribution of the community, which gives information on how similar species are in their abundances and it is thus related to Preston's concepts of commonness and rarity of species (Preston, 1948).
Any index of measure which takes into account both species richness and evenness is called an *heterogeneity measure* (Magurran, 2013; Good, 1953; Gray, 2000). They can be distinguished in two classes, depending on whether they make assumptions on the species-abundance distribution of the community (parametric indexes) or not (non-parametric indexes). Examples of parametric indexes are the α parameter of the log-series distribution (see Chapter 5) and the λ parameter of the log-normal (Taylor, 1978), this latter defined as the ratio between the observed number of species and the standard deviation of the distribution. Log-series and log-normal distributions have both been widely used in literature to describe the abundances of species (Slik et al., 2015; Ter Steege, Sabatier et al., 2017; White et al., 2012; Azaele, Cornell et al., 2012; Magurran and Henderson, 2003; Preston, 1948). Simpson's index D (Simpson, 1949) and Shannon's information index (Shannon and Weaver, 1949), are instead examples of non-parametric indexes of diversity (see section below).
In what follows we will describe some of the most important diversity indexes which have been introduced in literature to describe the alpha-diversity of a community.

3.2 Alpha-diversity indexes

As stated in the introduction, with the term alpha-diversity we mean the degree of diversity in species' composition within a finite and precise habitat, which can be a single sample unit of a larger database. In literature, lots of indexes have been introduced to measure the alpha-diversity, in order to be able to compare the composition of different sampling units, within ecological theory (Jost, 2010b; Hill, 1973; Magurran, 2013; Wolda, 1983) but also in physics and information theory (Abe and Rajagopal, 2001; Tsallis et al., 1998).
Before exploring the most famous ones, let us introduce some notations which will be useful in the following. We denote with S the number of species and with N the total number of individuals within a region $\mathcal{W} \subset \mathbb{R}^2$. Moreover, let us denote with n_s the abundance of the s^{th} species, where s runs from 1 to S. Let us then introduce the vector

$$p = (p_1, \dots, p_S) = \left(\frac{n_1}{N}, \dots, \frac{n_S}{N} \right)$$

of relative abundance of each species.
One of the most used indexes to measure the alpha-diversity of a community is Simpson's index (Simpson, 1949; Hill, 1973; Lande, 1996; Heip et al., 1998; Shimatani, 2001; Shimatani and Kubota, 2004; Legendre and Legendre, 2012; Wiegand and Moloney, 2013).

Definition 3.1. Simpson's diversity index expresses the probability that, picking at random a pair of individuals from the community, they belong to different species:

$$D = 1 - \sum_{s=1}^{S} \frac{n_s(n_s - 1)}{N(N-1)}. \tag{3.1}$$

In literature, the term *Simpson's index* is sometimes used to denote the **Simpson's dominance** or **concentration index**, which instead refers to the quantity

$$D' = \sum_{s=1}^{S} p_s^2, \tag{3.2}$$

or even the quantity

$$D'' = 1 - \sum_{s=1}^{S} p_s^2, \tag{3.3}$$

also known as **Gini-Simpson's index** (Jost, 2006).
Indeed, in his original paper, the author considered a population whose individuals are classified into S different groups (Simpson, 1949). By calling π_s the proportion of individuals belonging to the s^{th} group, he defined the quantity $\lambda = \sum_{s=1}^{S} \pi_s^2$ as a measure of the *concentration*, i.e. the similarity degree achieved with the classification. In fact, λ can assume values between $1/S$ (smallest concentration or lowest similarity, $\pi_s = 1/S$ for every s) and 1 (complete concentration or highest similarity: $\pi_s = 1$ for a unique fixed $s \in \{1, \dots, S\}$ and $\pi_t = 0$ for $t \neq s$). In particular, λ represents the probability that, picking two individuals from the community, randomly and independently one another, they belong to the same group.

Let us suppose that we have a sample of N individuals of such community. An unbiased estimator of λ is then the following quantity (Simpson, 1949)

$$l = \sum_{s=1}^{S} \frac{n_s(n_s-1)}{N(N-1)} = 1 - D,$$

where n_s is the number of individuals belonging to the s^{th} group.
The index in Definition 3.1 is therefore a natural modification of the Simpson's original concentration index measuring diversity instead of similarity and taking values between 0 (smallest diversity) and $N/S \cdot (S-1)/(N-1)$ (largest diversity). Other famous examples of alpha-diversity indexes are species richness S and Rényi's entropy (Rényi, 1961; Jost, 2006). This latter, given a real number $q \geq 0$ called the *order* of the entropy, is defined as follows

$$H_q = \frac{1}{1-q} \ln \left(\sum_{s=1}^{S} p_s^q \right). \tag{3.4}$$

The limiting case for $q \to 1$ of eq. (3.4) leads to Shannon's information (Shannon, 1948; Shannon and Weaver, 1949; Hill, 1973; Lande, 1996; Jost, 2006), another widely used non-parametric index of diversity:

$$H = -\sum_{s=1}^{S} p_s \ln p_s.$$

3.3 Binary similarity indexes

In appendix B we show how the concept of beta-diversity may be related to the one of alpha and gamma-diversity. Nevertheless, in literature there have been introduced various beta-diversity indexes to compare two different communities of an ecosystem (Rogers and Tanimoto, 1960; Clifford and Stephenson, 1975; Peters, 1968; Cheetham and Hazel, 1969; Romesburg, 1984; Wolda, 1981, 1983; Anderson et al., 2011) without involving the other two concepts. Some of them are incidence-based or *binary* (Sørensen, 1948; Jaccard, 1908; Chao et al., 2006; Lennon et al., 2001; Wolda, 1981; Hubalek, 1982; Gower, 1985; Kulczyński, 1928), thus accounting only for the presence or absence of a species; others, as the ones we have mentioned in the previous sections, are instead abundance-based (Horn, 1966; Gower, 1971; Bray and Curtis, 1957; Lance and Williams, 1967; Lande, 1996; Morisita, 1959; Renkonen, 1938), thus also looking at the relative abundance of each species in the surveyed community.
Here we introduce the two most famous (and exploited) incidence-based ones which have found large application within ecology and point process theory (Chao et al., 2006; Morlon et al., 2008; Engen et al., 2011; Réjou-Méchain and Hardy, 2011).

3.3.1 Jaccard's and Sørensen's similarity indexes

One of the most used diversity indexes to compare two communities is the Jaccard index (Jaccard, 1900, 1908; Hagmeier and Stults, 1964; Peters, 1968; Krebs, 1989).

Let A and B be two disjoint subsets of the region $\mathcal{W}$ under study. We denote with $S(A)$, $S(B)$ the number of species within the regions A and B, respectively. Let then denote with $S(A,B)$ the number of species that A and B have in common. Clearly $S(A)$, $S(B)$ and $S(A,B)$ are all less or equal than the number of species, S.

Definition 3.2. Jaccard's index of similarity $\mathrm{JAC}(A,B)$ is defined as the ratio between the species present in both the regions A and B and the total number of species in their union $A \cup B$:

$$\mathrm{JAC}(A,B) = \frac{S(A,B)}{S(A)+S(B)-S(A,B)}, \tag{3.5}$$

provided that $S(A)+S(B) > 0$.

Slightly modifying eq. (3.5), one can obtain another fundamental diversity index, firstly proposed by Sørensen (Sørensen, 1948; Sokal and Sneath, 1963; Peters, 1968; Krebs, 1989; Chao et al., 2005).

Definition 3.3. Sørensen's index of similarity $\mathrm{SØR}(A,B)$ is defined as the ratio between the common species in the two regions A and B and the mean number of species in them:

$$\mathrm{SØR}(A,B) = \frac{S(A,B)}{\frac{1}{2}[S(A)+S(B)]} = \frac{2S(A,B)}{S(A)+S(B)}, \tag{3.6}$$

provided that $S(A)+S(B) > 0$.

Let us notice that, if $S(A) = S(B)$, then $\mathrm{SØR}(A,B)$ is the number of co-present species per species.
As it is well known, the number of present or co-present species depends on the size of the regions A and B, therefore we assume, as it is generally the case, that A and B have the same size a.
In the next section we will study how to insert these two notions into the context of point processes' theory.

3.4 Similarity indexes in the context of point processes

When point process theory is applied to spatial data such as the location of trees within a given region, the key point is to think the dataset as a typical realisation of an unknown point process to be discovered. The goal is therefore to infer from the pattern under study the main characteristics of such underlying process and to estimate the associated moment functions (intensity, pair correlation function, Ripley's function and so on). These latter will help reconstruct the hidden stochastic process which best reflects and models the empirical data.
At this point it is fundamental to distinguish a point process from the data representing one of its possible realisations. For each random variable $x_{\boldsymbol{X}}$ associated to the point process $\boldsymbol{X}$, we will denote with $\overline{x_{\boldsymbol{X}}}$ the value it assumes in the realisation

under study. For example, if $\mathcal{N}_{\boldsymbol{X}}(B)$ is the counting random measure associated to the point process $\boldsymbol{X}$ evaluated at some subset B, we will denote with $\overline{\mathcal{N}_{\boldsymbol{X}}(B)}$ the actual number of data points of our realisation falling within B.
In the same spirit of the paper Shimatani, 2001, we want now to reformulate the notion of Jaccard's and Sørensen's indexes in the language of spatial point processes. Let us then suppose that our dataset consists of the locations of individuals belonging to S different species within a subregion $\mathcal{W} \subset \mathbb{R}^2$. We model the presence of such S species by a superposition of S independent and unknown spatial point processes $\boldsymbol{X}_s$, $s \in \{1, \dots, S\}$, thus considering (see Chapter 1) the following spatial point process

$$\boldsymbol{X} = \cup_{s=1}^{S} \boldsymbol{X}_s.$$

This way of superimposing two or multiple processes so that a community-level pattern can be derived has been widely used in literature (Hui and McGeoch, 2007).

3.4.1 Jaccard's and Sørensen's indexes for point processes

From Chapter 1 we know that each point process $\boldsymbol{X}_s$ of $\boldsymbol{X}$ is uniquely determined by its capacity functional $T_{\boldsymbol{X}_s}$, which gives, for every subset $B \subseteq \mathcal{W}$ the probability that at least one point of the process falls within it. We also know that to each point process $\boldsymbol{X}_s$ and to each subset B of $\mathcal{W}$ it is possible to associate two random variables named the *presence indicator* $1^B_{\boldsymbol{X}}$ and the *vacancy indicator* $v^B_{\boldsymbol{X}}$.
Let us then rewrite the Jaccard and the Sørensen index in terms of these random variables.
For the stochastic process $\boldsymbol{X} = \bigcup_{s=1}^{S} \boldsymbol{X}_s$ generating our dataset, the number of species falling within a region B is given by the random variable $S_{\boldsymbol{X}}(A) = \sum_{s=1}^{S} 1^A_{\boldsymbol{X}_s}$. Thus, from eq. (1.4) and the linearity of the expected value we have

$$\mathbb{E}[S_{\boldsymbol{X}}(A)] = \mathbb{E}\left[\sum_{s=1}^{S} 1^A_{\boldsymbol{X}_s}\right] = \sum_{s=1}^{S} \mathbb{E}[1^A_{\boldsymbol{X}_s}] = \sum_{s=1}^{S} T_{\boldsymbol{X}_s}(A).$$

Moreover, eq. (1.5) leads to

$$\mathbb{E}[S_{\boldsymbol{X}}(A,B)] = \mathbb{E}\left[\sum_{s=1}^{S} 1^A_{\boldsymbol{X}_s} 1^B_{\boldsymbol{X}_s}\right] = \sum_{s=1}^{S} \mathbb{E}[1^A_{\boldsymbol{X}_s} 1^B_{\boldsymbol{X}_s}] = \sum_{s=1}^{S} \mathbb{P}(\mathcal{N}_{\boldsymbol{X}_s}(A) > 0, \mathcal{N}_{\boldsymbol{X}_s}(B) > 0). \tag{3.7}$$

The equivalent of Jaccard's and Sørensen's similarity indexes for the regions A and B are thus given, respectively, by the random variables

$$\mathrm{JAC}_{\boldsymbol{X}}(A,B) = \frac{S_{\boldsymbol{X}}(A,B)}{S_{\boldsymbol{X}}(A) + S_{\boldsymbol{X}}(B) - S_{\boldsymbol{X}}(A,B)}$$

and

$$\mathrm{S\O R}_{\boldsymbol{X}}(A,B) = \frac{2S_{\boldsymbol{X}}(A,B)}{S_{\boldsymbol{X}}(A) + S_{\boldsymbol{X}}(B)} \tag{3.8}$$

which are the ratio of two random quantities.
The expected value of this ratio can be computed from $\mathbb{E}[S_{\boldsymbol{X}}(A)]$, $\mathbb{E}[S_{\boldsymbol{X}}(B)]$ and

$\mathbb{E}[S_{\boldsymbol{X}}(A,B)]$ using the method of statistical differentials. Indeed, for general random variables X and Y we have the following relation (see e.g. Johnson et al., 2005)

$$\mathbb{E}\left[\frac{X}{Y}\right] \approx \frac{\mathbb{E}[X]}{\mathbb{E}[Y]}\left[1+\frac{var(Y)}{\mathbb{E}[Y]^2}-\frac{cov(X,Y)}{\mathbb{E}[X]\mathbb{E}[Y]}\right] = \frac{\mathbb{E}[X]}{\mathbb{E}[Y]}(1+\phi(X,Y)) \tag{3.9}$$

Applying eq. (3.9) above to the point process formulation of the average of Jaccard's and Sørensen's similarity indexes, we obtain

$$\begin{aligned}\mathbb{E}[\mathrm{JAC}_{\boldsymbol{X}}(A,B)] &= \frac{\mathbb{E}[S_{\boldsymbol{X}}(A,B)]}{\mathbb{E}[S_{\boldsymbol{X}}(A)]+\mathbb{E}[S_{\boldsymbol{X}}(B)]-\mathbb{E}[S_{\boldsymbol{X}}(A,B)]}\cdot(1+\phi_{\mathrm{JAC}}(A,B)) \\ &= \frac{\sum_{s=1}^{S}\mathbb{P}(\mathcal{N}_{\boldsymbol{X}_s}(A)>0,\mathcal{N}_{\boldsymbol{X}_s}(B)>0)\cdot(1+\phi_{\mathrm{JAC}}(A,B))}{\sum_{s=1}^{S}T_{\boldsymbol{X}_s}(A)+\sum_{s=1}^{S}T_{\boldsymbol{X}_s}(B)-\sum_{s=1}^{S}\mathbb{P}(\mathcal{N}_{\boldsymbol{X}_s}(A)>0,\mathcal{N}_{\boldsymbol{X}_s}(B)>0)}\end{aligned} \tag{3.10}$$

and

$$\begin{aligned}\mathbb{E}[\mathrm{S\O R}_{\boldsymbol{X}}(A,B)] &= \frac{2\mathbb{E}[S_{\boldsymbol{X}}(A,B)]}{\mathbb{E}[S_{\boldsymbol{X}}(A)]+\mathbb{E}[S_{\boldsymbol{X}}(B)]}\cdot(1+\phi_{\mathrm{S\O R}}(A,B)) \\ &= \frac{2\sum_{s=1}^{S}\mathbb{P}(\mathcal{N}_{\boldsymbol{X}_s}(A)>0,\mathcal{N}_{\boldsymbol{X}_s}(B)>0)}{\sum_{s=1}^{S}T_{\boldsymbol{X}_s}(A)+\sum_{s=1}^{S}T_{\boldsymbol{X}_s}(B)}\cdot(1+\phi_{\mathrm{S\O R}}(A,B)).\end{aligned} \tag{3.11}$$

Let us now consider, as regions A and B, two infinitesimal disks, $\mathcal{B}_u$ and $\mathcal{B}_v$, centred at u and v, having equal area $\mathrm{d}u = \mathrm{d}v$ and being disjoint.
Let us denote with $n_s(u) = \mathcal{N}_{\boldsymbol{X}_s}(\mathcal{B}_u)$ and $n_s(v) = \mathcal{N}_{\boldsymbol{X}_s}(\mathcal{B}_v)$ the number of points of the process $\boldsymbol{X}_s$ contained in $\mathcal{B}_u$ and $\mathcal{B}_v$, respectively. Let then $\lambda_{\boldsymbol{X}_s}(u)$ be the intensity of $\boldsymbol{X}_s$. Generalising formula (1.13) to the case of S processes, we get that the intensity function of the superposed process $\boldsymbol{X} = \cup_s \boldsymbol{X}_s$ is

$$\lambda_{\boldsymbol{X}}(u) = \sum_s \lambda_{\boldsymbol{X}_s}(u). \tag{3.12}$$

Let then $\rho_{\boldsymbol{X}_s}(u,v)$ be the associated second moment density (see Chapter 1, Section 1.2.2) and set for simplicity's sake $\rho^{\boldsymbol{X}}(u,v) = \sum_s \rho_{\boldsymbol{X}_s}(u,v)$. The following interpretations are standard (see Chapter 1, Sections 1.2.1 and 1.2.2)

$$\begin{aligned}\lambda_{\boldsymbol{X}_s}(u)\mathrm{d}u = P(n_s(u)=1),\\ P(n_s(u)>1) = o(\mathrm{d}u)\end{aligned}$$

and

$$\begin{aligned}\rho_{\boldsymbol{X}_s}(u,v)\mathrm{d}u\mathrm{d}v = P(n_s(u)=1,n_s(v)=1),\\ P(\{n_s(u)>1\}\cup\{n_s(v)>1\}) = o(\mathrm{d}u\mathrm{d}v).\end{aligned}$$

Denoting with $S_{\boldsymbol{X}}(u)$, $S_{\boldsymbol{X}}(u)$ and $S_{\boldsymbol{X}}(u,v)$ the number of species (point processes) present in $\mathcal{B}_u$, $\mathcal{B}_v$ and in their union, respectively, and neglecting higher order terms in $\mathrm{d}x$ or $\mathrm{d}x\mathrm{d}y$, we have that the expected number of species found in $\mathcal{B}_u$, $S_{\boldsymbol{X}}(u)$, can be expressed as

$$\mathbb{E}[S_{\boldsymbol{X}}(u)] \sim \sum_{s=1}^{S} P(n_s(\mathrm{d}u)=1) = \sum_{s=1}^{S}\lambda_{\boldsymbol{X}_s}(u) = \lambda_{\boldsymbol{X}}(u)\mathrm{d}u, \tag{3.13}$$

while the average number of co-present species in the infinitesimal regions $\mathcal{B}_u$ and $\mathcal{B}_v$ around u and v, $S_{\boldsymbol{X}}(u,v)$, is

$$\mathbb{E}[S_{\boldsymbol{X}}(u,v)] \sim \sum_s P(n_s(u)=1, n_s(v)=1) = \sum_{s=1}^{S} \rho_{\boldsymbol{X}_s}(u,v)\mathrm{d}u\mathrm{d}v = \rho^{\boldsymbol{X}}(u,v)\mathrm{d}u\mathrm{d}v. \tag{3.14}$$

We remark that the number $S_{\boldsymbol{X}}(u)$ of species at u and the number $S_{\boldsymbol{X}}(u,v)$ of shared species at u and v are discrete random variables whose expected values can be described by the above formulae eqs. (3.13) and (3.14). Apart from their averages, their distribution is assumed to be unknown.
By inserting eqs. (3.13) and (3.14) into eqs. (3.10) and (3.11), we obtain the expressions of the expected values of Jaccard's and Sørensen's similarity indexes for infinitesimal regions $\mathcal{B}_u$ and $\mathcal{B}_v$

$$\mathbb{E}[\mathrm{JAC}_{\boldsymbol{X}}(u,v)] \sim \frac{\rho^{\boldsymbol{X}}(u,v)\mathrm{d}u\mathrm{d}v}{\lambda_{\boldsymbol{X}}(u)\mathrm{d}u + \lambda_{\boldsymbol{X}}(v)\mathrm{d}v - \rho^{\boldsymbol{X}}(u,v)\mathrm{d}u\mathrm{d}v} \cdot (1 + \phi_{\mathrm{JAC},\boldsymbol{X}}(u,v))$$

$$\mathbb{E}[\mathrm{S\O R}_{\boldsymbol{X}}(u,v)] \sim \frac{2\rho^{\boldsymbol{X}}(u,v)\mathrm{d}u\mathrm{d}v}{\lambda_{\boldsymbol{X}}(u)\mathrm{d}u + \lambda_{\boldsymbol{X}}(v)\mathrm{d}v} \cdot (1 + \phi_{\mathrm{S\O R},\boldsymbol{X}}(u,v)). \tag{3.15}$$

Under the additional hypothesis of $\boldsymbol{X}_s$ homogeneous for every $s \in \{1,\dots,S\}$ and setting $\mathrm{d}v = \mathrm{d}u$ we can approximate the average of Jaccard's and Sørensen's indexes as follows:

$$\begin{aligned}
\mathbb{E}[\mathrm{JAC}_{\boldsymbol{X}}(u,v)] &\sim \frac{\rho^{\boldsymbol{X}}(u,v)\mathrm{d}u\mathrm{d}u}{\lambda_{\boldsymbol{X}}\mathrm{d}u + \lambda_{\boldsymbol{X}}\mathrm{d}u - \rho^{\boldsymbol{X}_s}(u,v)\mathrm{d}u\mathrm{d}u} \cdot (1 + \phi_{\mathrm{JAC},\boldsymbol{X}}(u,v)) \\
&= \frac{\sum_{s=1}^{S} \rho_{\boldsymbol{X}_s}(u,v)\mathrm{d}u}{2\lambda_{\boldsymbol{X}} - \sum_{s=1}^{S} \rho_{\boldsymbol{X}_s}(u,v)\mathrm{d}u} \cdot (1 + \phi_{\mathrm{JAC},\boldsymbol{X}}(u,v)) \\
&= \frac{\sum_{s=1}^{S} g_{\boldsymbol{X}_s}(u,v)\lambda^2_{\boldsymbol{X}_s}\mathrm{d}u}{2\lambda_{\boldsymbol{X}} - \sum_{s=1}^{S} g_{\boldsymbol{X}_s}(u,v)\lambda^2_{\boldsymbol{X}_s}\mathrm{d}u} \cdot (1 + \phi_{\mathrm{JAC},\boldsymbol{X}}(u,v))
\end{aligned}$$

$$\begin{aligned}
\mathbb{E}[\mathrm{S\O R}_{\boldsymbol{X}}(u,v)] &\sim \frac{2\rho^{\boldsymbol{X}}(u,v)\mathrm{d}u\mathrm{d}u}{\lambda_{\boldsymbol{X}}\mathrm{d}u + \lambda_{\boldsymbol{X}}\mathrm{d}u} \cdot (1 + \phi_{\mathrm{S\O R},\boldsymbol{X}}(u,v)) \\
&= \frac{\sum_{s=1}^{S} \rho_{\boldsymbol{X}_s}(u,v)\mathrm{d}u}{\lambda_{\boldsymbol{X}}} \cdot (1 + \phi_{\mathrm{S\O R},\boldsymbol{X}}(u,v)) \\
&= \frac{\sum_{s=1}^{S} g_{\boldsymbol{X}_s}(u,v)\lambda^2_{\boldsymbol{X}_s}\mathrm{d}u}{\lambda_{\boldsymbol{X}}} \cdot (1 + \phi_{\mathrm{S\O R},\boldsymbol{X}}(u,v)),
\end{aligned} \tag{3.16}$$

where we have used the definition of the pair correlation function $g_{\boldsymbol{X}_s}(u,v) = \frac{\rho_{\boldsymbol{X}_s}(u,v)}{\lambda_{\boldsymbol{X}_s}(u)\lambda_{\boldsymbol{X}_s}(v)}$ (see Chapter 1, Section 1.2.2).
Let us focus on Sørensen's similarity index, which will be further studied in the following chapter.
By generalising expression eq. (1.17) of the second moment density $\rho_{\boldsymbol{X}}(u,v)$ for the superposition of two point processes to the case of S homogeneous processes, we get:

$$\rho_{\boldsymbol{X}}(u,v) = \sum_{s=1}^{S} \rho_{\boldsymbol{X}_s}(u,v) + \sum_{\substack{s,t=1 \\ s\neq t}}^{S} \lambda_{\boldsymbol{X}_s}\lambda_{\boldsymbol{X}_t},$$

from which we obtain a generalisation of the pair correlation function expressed by eq. (1.18):

$$g_{\boldsymbol{X}}(u,v) = \frac{\sum_{s=1}^{S} \rho_{\boldsymbol{X}_s}(u,v) + \sum_{\substack{s,t=1 \\ s\neq t}}^{S} \lambda_{\boldsymbol{X}_s}\lambda_{\boldsymbol{X}_t}}{\left(\sum_{s=1}^{S} \lambda_{\boldsymbol{X}_s}\right)^2} = \frac{\sum_{s=1}^{S} \rho_{\boldsymbol{X}_s}(u,v) + \sum_{\substack{s,t=1 \\ s\neq t}}^{S} \lambda_{\boldsymbol{X}_s}\lambda_{\boldsymbol{X}_t}}{\lambda_{\boldsymbol{X}}^2}. \tag{3.17}$$

We can now rewrite the above equation as follows

$$\begin{aligned} g_{\boldsymbol{X}}(u,v) &= \frac{\sum_{s=1}^{S} \rho_{\boldsymbol{X}_s}(u,v) + \sum_{\substack{s,t=1 \\ s\neq t}}^{S} \lambda_{\boldsymbol{X}_s}\lambda_{\boldsymbol{X}_t}}{\lambda_{\boldsymbol{X}}^2} \\ &= \frac{\sum_{s=1}^{S} g_{\boldsymbol{X}_s}(u,v)\lambda_{\boldsymbol{X}_s}^2 + \sum_{\substack{s,t=1 \\ s\neq t}}^{S} \lambda_{\boldsymbol{X}_s}\lambda_{\boldsymbol{X}_t}}{\lambda_{\boldsymbol{X}}^2} \\ &= \frac{\sum_{s=1}^{S} g_{\boldsymbol{X}_s}(u,v)\lambda_{\boldsymbol{X}_s}^2}{\lambda_{\boldsymbol{X}}^2} + \frac{\sum_{\substack{s,t=1 \\ s\neq t}}^{S} \lambda_{\boldsymbol{X}_s}\lambda_{\boldsymbol{X}_t}}{\lambda_{\boldsymbol{X}}^2} \\ &= \frac{1}{\lambda_{\boldsymbol{X}}} \frac{\sum_{s=1}^{S} g_{\boldsymbol{X}_s}(u,v)\lambda_{\boldsymbol{X}_s}^2}{\lambda_{\boldsymbol{X}}} + \frac{(\sum_{s=1}^{S} \lambda_{\boldsymbol{X}_s})^2 - \sum_{s=1}^{S} \lambda_{\boldsymbol{X}_s}^2}{\lambda_{\boldsymbol{X}}^2} \\ &= \frac{1}{\lambda_{\boldsymbol{X}} \mathrm{d}x} \frac{\mathbb{E}[\mathrm{S\O R}_{\boldsymbol{X}}(u,v)]}{(1+\phi_{\mathrm{S\O R},\boldsymbol{X}}(u,v))} + 1 - \frac{\sum_{s=1}^{S} \lambda_{\boldsymbol{X}_s}^2}{\lambda_{\boldsymbol{X}}^2}. \end{aligned} \tag{3.18}$$

Thus, we have the following

Proposition 3.4. *Let $\boldsymbol{X}$ be the superposition of S homogeneous point processes $\boldsymbol{X}_s$ defined in $\mathcal{W} \subset \mathbb{R}^2$ and having intensity functions $\lambda_{\boldsymbol{X}_s}$, $s = 1, \dots, S$. Let then $\lambda_{\boldsymbol{X}}$ and $g_{\boldsymbol{X}}(\cdot,\cdot)$ be the intensity function and the pair correlation function of $\boldsymbol{X}$ given by eqs. (3.12) and (3.17), respectively. Then, the expected value of the Sørensen index at locations u and v in $\mathcal{W}$ is given by*

$$\mathbb{E}[S\O R_{\boldsymbol{X}}(u,v)] = \lambda_{\boldsymbol{X}} du \left(g_{\boldsymbol{X}}(u,v) - 1 + \frac{\sum_{s=1}^{S} \lambda_{\boldsymbol{X}_s}^2}{\lambda_{\boldsymbol{X}}^2} \right) (1 + \phi_{S\O R,\boldsymbol{X}}(u,v)). \tag{3.19}$$

If $\boldsymbol{X}_s$ is also isotropic for every $s \in \{1, \dots, S\}$, then both Sørensen's index and the pair correlation function result to depend only on the distance r between the two points u and v.
Let us compute Sørensen's index for the two motion-invariant processes introduced in Chapter 1.

Example 3.5 (Homogeneous Poisson Process)**.** Let us assume that each species' process $\boldsymbol{X}_s$ is a homogeneous Poisson process of intensity $\lambda_{\boldsymbol{X}_s}$. From eq. (1.23) we know that the pair correlation function of each process is constantly equal to 1. Therefore, substituting into eq. (3.19) we get the expected value of Sørensen's index

for the superposed process $\boldsymbol{X} = \bigcup_s \boldsymbol{X}_s$, which is a function only of the distance r because of the motion invariant property of the Poisson process:

$$\begin{aligned}\mathbb{E}[\text{SØR}_{\boldsymbol{X}}(r)] &= \lambda_{\boldsymbol{X}} \mathrm{d}u \left(1 - 1 + \frac{\sum_{s=1}^{S} \lambda_{\boldsymbol{X}_s}^2}{\lambda_{\boldsymbol{X}}^2}\right)(1 + \phi_{\text{SØR},\boldsymbol{X}}(u,v)) \\ &= \lambda_{\boldsymbol{X}} \mathrm{d}u \sum_{s=1}^{S} p_s^2 (1 + \phi_{\text{SØR},\boldsymbol{X}}(u,v)).\end{aligned}$$

where we have set $p_s = \frac{\lambda_{\boldsymbol{X}_s}}{\lambda_{\boldsymbol{X}}}$.

Example 3.6 (Neyman-Scott process)**.** Let us assume that, for every index s, the process $\boldsymbol{X}_s$ is a Neyman-Scott process of intensity $\lambda_{\boldsymbol{X}_s} = \rho_{\boldsymbol{X}_s}\mu_{\boldsymbol{X}_s}$, where $\rho_{\boldsymbol{X}_s}$ is the density of clusters and $\mu_{\boldsymbol{X}_s}$ the average number of points per cluster. In this case $p_s = \rho_{\boldsymbol{X}_s}\mu_{\boldsymbol{X}_s}/\lambda_{\boldsymbol{X}}$.
We restrict to the processes defined in Section 1.5.2:

- Matérn cluster process of parameters $(\rho_{\boldsymbol{X}_s}, \mu_{\boldsymbol{X}_s}, R)$:

$$\mathbb{E}[\text{SØR}_{\boldsymbol{X}}(r)] = \lambda_{\boldsymbol{X}} \mathrm{d}x \left(\sum_{s=1}^{S} [1 + (\rho_{\boldsymbol{X}_s} 2\pi r)^{-1} h^{\text{pol}}_{\boldsymbol{X}_s}(r)] p_s^2\right)(1 + \phi_{\text{SØR},\boldsymbol{X}}(u,v)),$$

  with $h^{\text{pol}}_{\boldsymbol{X}_s}(r)$ given by eq. (1.35).

- Modified Thomas process of parameters $(\rho_{\boldsymbol{X}_s}, \mu_{\boldsymbol{X}_s}, \sigma_{\boldsymbol{X}_s})$:

$$\mathbb{E}[\text{SØR}_{\boldsymbol{X}}(r)] = \lambda_{\boldsymbol{X}} \mathrm{d}x \left(\sum_{s=1}^{S} \left[1 + (\rho_{\boldsymbol{X}_s} 4\pi\sigma^2_{\boldsymbol{X}_s})^{-1} \exp\left(-\frac{r^2}{4\sigma^2_{\boldsymbol{X}_s}}\right)\right] p_s^2\right)(1 + \phi_{\text{SØR},\boldsymbol{X}}(u,v)).$$

- Exponential cluster process of parameters $(\rho_{\boldsymbol{X}_s}, \mu_{\boldsymbol{X}_s}, \beta_{\boldsymbol{X}_s})$:

$$\mathbb{E}[\text{SØR}_{\boldsymbol{X}}(r)] = \lambda_{\boldsymbol{X}} \mathrm{d}x \left(\sum_{s=1}^{S} \left[1 + \frac{1}{\rho_{\boldsymbol{X}_s}} \frac{\beta^4_{\boldsymbol{X}_s} r^2}{16\pi} K_2(\beta_{\boldsymbol{X}_s} r)\right] p_s^2\right)(1 + \phi_{\text{SØR},\boldsymbol{X}}(u,v)).$$

- Cauchy cluster process of parameters $(\rho_{\boldsymbol{X}_s}, \mu_{\boldsymbol{X}_s}, b_{\boldsymbol{X}_s})$:

$$\mathbb{E}[\text{SØR}_{\boldsymbol{X}}(r)] = \lambda_{\boldsymbol{X}} \mathrm{d}x \left(\sum_{s=1}^{S} \left[1 + \frac{1}{\rho_{\boldsymbol{X}_s}} \frac{b^2_{\boldsymbol{X}_s}}{\pi(4 + r^2/b^2)^{3/2}}\right] p_s^2\right)(1 + \phi_{\text{SØR},\boldsymbol{X}}(u,v)).$$

3.4.2 Spatial-dependent alpha and beta-diversity

Shimatani (Shimatani, 2001; Shimatani and Kubota, 2004) extended the notions of Simpson's diversity index and introduced, in the context of point processes, two functions α and β measuring, respectively, the alpha and beta-diversity within a landscape. He defined $\alpha(r)$ as the probability that two points of the process within distance less or equal than r, belong to different species. $\beta(r)$ expresses, instead, the conditional probability that, given two points of the process distant r apart, are not conspecific.

Let us begin, by convenience, with the beta-diversity.
We start by computing the probability that, given two regions A and B occupied by the process $\boldsymbol{X}$, the same species s is present in both of them, for $s \in \{1, \dots, S\}$ fixed. By probability theory, we have that

$$\begin{aligned}\mathbb{P}(\mathcal{N}_{\boldsymbol{X}_s}(A) > 0, \mathcal{N}_{\boldsymbol{X}_s}(B) > 0 | \mathcal{N}_{\boldsymbol{X}}(A) > 0, \mathcal{N}_{\boldsymbol{X}}(B) > 0) = \\ = \frac{\mathbb{P}(\mathcal{N}_{\boldsymbol{X}_s}(A) > 0, \mathcal{N}_{\boldsymbol{X}_s}(B) > 0, \mathcal{N}_{\boldsymbol{X}}(A) > 0, \mathcal{N}_{\boldsymbol{X}}(B) > 0)}{\mathbb{P}(\mathcal{N}_{\boldsymbol{X}}(A) > 0, \mathcal{N}_{\boldsymbol{X}}(B) > 0)} \\ = \frac{\mathbb{P}(\mathcal{N}_{\boldsymbol{X}_s}(A) > 0, \mathcal{N}_{\boldsymbol{X}_s}(B) > 0)}{\mathbb{P}(\mathcal{N}_{\boldsymbol{X}}(A) > 0, \mathcal{N}_{\boldsymbol{X}}(B) > 0)}.\end{aligned}$$

By summing up over $s \in \{1, \dots, S\}$ we get the probability that, given that the two regions are occupied by points of $\boldsymbol{X}$, there is at least one species which is present in both of them. This sum, by eq. (3.7), is equal to the mean value of $S_{\boldsymbol{X}}(A, B)$. We can now define the β-diversity index of the process $\boldsymbol{X}$ between the two regions A and B:

$$\beta_{\boldsymbol{X}}(A, B) = 1 - \frac{\mathbb{E}[S_{\boldsymbol{X}}(A, B)]}{\mathbb{P}(\mathcal{N}_{\boldsymbol{X}}(A) > 0, \mathcal{N}_{\boldsymbol{X}}(B) > 0)}. \tag{3.20}$$

It expresses the conditional probability that regions A and B do not share any species, given that both of them contain points of $\boldsymbol{X}$.
If we assume, as before, that A and B are infinitesimally small balls within $\mathcal{W}$ with sizes $\mathrm{d}u$ and $\mathrm{d}v$ and centred in u and v, respectively, we can rewrite the ratio in eq. (3.20) as

$$\begin{aligned}\frac{\mathbb{E}[S_{\boldsymbol{X}}(A, B)]}{\mathbb{P}(\mathcal{N}_{\boldsymbol{X}}(A) > 0, \mathcal{N}_{\boldsymbol{X}}(B) > 0)} &= \frac{\sum_{s=1}^{S} \mathbb{P}(\mathcal{N}_{\boldsymbol{X}_s}(A) > 0, \mathcal{N}_{\boldsymbol{X}_s}(B) > 0)}{\mathbb{P}(\mathcal{N}_{\boldsymbol{X}}(A) > 0, \mathcal{N}_{\boldsymbol{X}}(B) > 0)} \\ &= \lim_{\substack{\mathrm{d}u \to 0 \\ \mathrm{d}v \to 0}} \frac{\sum_{s=1}^{S} \rho_{\boldsymbol{X}_s}(u, v)\mathrm{d}u\mathrm{d}v}{\rho_{\boldsymbol{X}}(u, v)\mathrm{d}u\mathrm{d}v} \\ &= \frac{\sum_{s=1}^{S} \rho_{\boldsymbol{X}_s}(u, v)}{\rho_{\boldsymbol{X}}(u, v)}\end{aligned}$$

Thus, in this case, the β-diversity function of the process $\boldsymbol{X}$ between the two regions A and B is given by

$$\begin{aligned}\beta_{\boldsymbol{X}}(A, B) &= 1 - \frac{\sum_{s=1}^{S} \rho_{\boldsymbol{X}_s}(u, v)}{\rho_{\boldsymbol{X}}(u, v)} \\ &= 1 - \frac{\sum_{s=1}^{S} \rho_{\boldsymbol{X}_s}(u, v)}{\sum_{s=1}^{S} \rho_{\boldsymbol{X}_s}(u, v) + \sum_{\substack{s,t=1 \\ s \neq t}}^{S} \lambda_{\boldsymbol{X}_s}(u)\lambda_{\boldsymbol{X}_t}(v)}\end{aligned}$$

Lastly, if we add the hypothesis of $\boldsymbol{X}_s$ isotropic and homogeneous for every $s \in \{1, \dots, S\}$, the beta-diversity index becomes a function of the distance r between

the two considered regions A and B:

$$\begin{aligned}\beta_{\boldsymbol{X}}(r) &= 1 - \frac{\sum_{s=1}^{S} \rho_{\boldsymbol{X}_s}(r)}{\sum_{s=1}^{S} \rho_{\boldsymbol{X}_s}(r) + \sum_{\substack{s,t=1 \\ s\neq t}}^{S} \lambda_{\boldsymbol{X}_s}\lambda_{\boldsymbol{X}_t}} \\ &1 - \frac{\sum_{s=1}^{S} g_{\boldsymbol{X}_s}(r)\lambda_{\boldsymbol{X}_s}^2}{\sum_{s=1}^{S} g_{\boldsymbol{X}_s}(r)\lambda_{\boldsymbol{X}_s}^2 + \sum_{\substack{s,t=1 \\ s\neq t}}^{S} \lambda_{\boldsymbol{X}_s}\lambda_{\boldsymbol{X}_t}} \\ &= 1 - \frac{\sum_{s=1}^{S} g_{\boldsymbol{X}_s}(r)\lambda_{\boldsymbol{X}_s}^2}{g_{\boldsymbol{X}}(r)\lambda_{\boldsymbol{X}}^2} \end{aligned} \tag{3.21}$$

Last equality of eq. (3.21) is exactly the definition of the function $\beta(r)$ given by Shimatani (Shimatani, 2001).
Let us now compute the alpha-diversity, which is the probability that, given two points of the process $\boldsymbol{X}$ at distance no more than r, they belong to different species. In order to compute it, we firstly need to express the probability that, given such two points, they belong to the same species $s \in \{1, \dots, S\}$. Since we assume that all the processes $\boldsymbol{X}_s$ are motion invariant, also this probability depends solely on the distance of the two points and it is independent on their exact locations in $\mathcal{W}$.
Therefore, by arbitrarily fixing the position of the first point at x, what we are looking for is given by

$$\mathbb{P}(x \in \boldsymbol{X}_s, \mathcal{N}_{\boldsymbol{X}_s}(\mathcal{B}(x,r)) > 1 | x \in \boldsymbol{X})$$

or, equivalently

$$\mathbb{P}(x \in \boldsymbol{X}_s | x \in \boldsymbol{X}) \mathbb{P}(\mathcal{N}_{\boldsymbol{X}_s}(\mathcal{B}(x,r)) > 1 | x \in \boldsymbol{X}_s).$$

Let us firstly focus on the left factor:

$$\begin{aligned}\mathbb{P}(x \in \boldsymbol{X}_s | x \in \boldsymbol{X}) &= \frac{\mathbb{P}(x \in \boldsymbol{X}_s, x \in \boldsymbol{X})}{\mathbb{P}(x \in \boldsymbol{X})} = \frac{\mathbb{P}(x \in \boldsymbol{X}_s)}{\mathbb{P}(x \in \boldsymbol{X})} \\ &= \lim_{\mathrm{r}_x \to 0} \frac{\mathbb{P}(\mathcal{N}_{\boldsymbol{X}_s}(\mathcal{B}(x, r_x)) > 0)}{\mathbb{P}(\mathcal{N}_{\boldsymbol{X}}(\mathcal{B}(x, r_x)) > 0)} \\ &= \lim_{\mathrm{r}_x \to 0} \frac{\lambda_{\boldsymbol{X}_s} \mathrm{d}x}{\lambda_{\boldsymbol{X}} \mathrm{d}x} \\ &= \frac{\lambda_{\boldsymbol{X}_s}}{\lambda_{\boldsymbol{X}}}. \end{aligned}$$

It remains to compute the probability $\mathbb{P}(\mathcal{N}_{\boldsymbol{X}_s}(\mathcal{B}(x,r)) > 1 | x \in \boldsymbol{X}_s)$ that, given that the point at x belongs to species s, there is at least another point of the same species within $\mathcal{B}(x,r)$. This is equivalent to the ratio between the mean number of extra-points in $\mathcal{B}(x,r)$ belonging to $\boldsymbol{X}_s$ and the ones of $\boldsymbol{X}$:

$$\begin{aligned}\mathbb{P}(\mathcal{N}_{\boldsymbol{X}_s}(\mathcal{B}(x,r)) > 1 | x \in \boldsymbol{X}_s) &= \frac{\mathbb{E}[\mathcal{N}_{\boldsymbol{X}_s}(\mathcal{B}(x,r)) | x \in \boldsymbol{X}_s] - 1}{\mathbb{E}[\mathcal{N}_{\boldsymbol{X}}(\mathcal{B}(x,r)) | x \in \boldsymbol{X}] - 1} \\ &= \frac{\lambda_{\boldsymbol{X}_s} K_{\boldsymbol{X}_s}(r)}{\lambda_{\boldsymbol{X}} K_{\boldsymbol{X}_s}(r)}. \end{aligned}$$

Therefore, the probability that, given two points of $\boldsymbol{X}$ at distance no more than r, they both belong to $\boldsymbol{X}_s$ is given by

$$\mathbb{P}(x \in \boldsymbol{X}_s, \mathcal{N}_{\boldsymbol{X}_s}(\mathcal{B}(x,r)) > 1 | x \in \boldsymbol{X}) = \frac{\lambda_{\boldsymbol{X}_s}}{\lambda_{\boldsymbol{X}}} \frac{\lambda_{\boldsymbol{X}_s} K_{\boldsymbol{X}_s}(r)}{\lambda_{\boldsymbol{X}} K_{\boldsymbol{X}}(r)} = \frac{\lambda^2_{\boldsymbol{X}_s} K_{\boldsymbol{X}_s}(r)}{\lambda^2_{\boldsymbol{X}} K_{\boldsymbol{X}}(r)}.$$

By summing up over all $s \in \{1, \ldots, S\}$ we get the probability that given two points of $\boldsymbol{X}$ at distance no more than r, they belong to the same species. We are thus ready to write the alpha-diversity function

$$\alpha(r) = 1 - \sum_{s=1}^{S} \frac{\lambda^2_{\boldsymbol{X}_s} K_{\boldsymbol{X}_s}(r)}{\lambda^2_{\boldsymbol{X}} K_{\boldsymbol{X}}(r)}.$$

4

Decay of Similarity

In this chapter we focus on the study of the decay of similarity between two regions of a landscape as a function of the distance between them (Tovo and Favretti, 2017).
Since the pioneering work of Whittaker (Whittaker, 1972) and Preston (Preston, 1962a,b), a number of diversity indexes have been introduced in literature and their effectiveness has been tested against field data, with various degrees of success. To specify the intuitive concept of similarity we will adopt Sørensen's[1] similarity index (see Chapter 3) and its associated spatial density. Equally used in literature is the notion, complementary to the concept of similarity, of *species' turnover* or beta-diversity, which, as we have seen in the previous chapter, is the change in species' composition between two plots as a function of the distance between them. Even stated in these terms, this more restricted problem is hard to reduce to a mathematical model since on real landscapes many drivers of diversity are acting at the same time and may contribute with different intensity depending on the spatial scales (Soininen et al., 2007): at a continental scale climatic factor may dominate, at a smaller scale orographic factors may create specific environmental gradients due to the change in altitude or to the orientation of valleys. At any scale, the effect of past transformations of the environment may have shaped the territory with dispersal barriers or niches. The heterogeneity of these factors may have hampered the construction of an all-compassing mathematical model and, effectively, a relatively small (compared the the huge number of articles dedicated to biodiversity issues) number of works are available on the specific problem of finding the function that best describes the change in species' composition with the distance.

[1] We are aware that each diversity index has its own field of applicability and it is more or less biased, hence necessarily our analytical treatment of a decay of similarity function based on a specific index will suffer from the same limitations of the index itself, but we are confident that the same procedure can be applied to other indexes.

4.1 Similarity decay functions

4.1.1 Sørensen index's spatial density

We begin by recalling the definition of Sørensen's similarity index. Let us consider a flat region W with no environmental gradients and two disjoint ($A \cap B = \emptyset$) subregions A and B of $\mathcal{W}$ having area a. Let $S(A)$ and $S(B)$ be the number of different species present respectively in A and B and let us denote with $S(A, B)$ the number of co-present species in A and B. Provided that $S(A) + S(B) > 0$, the Sørensen similarity between the two regions A and B is the symmetric function $0 \leq \text{SØR}(A, B) \leq 1$

$$\text{SØR}(A, B) = \frac{S(A, B)}{\frac{1}{2}(S(A) + S(B))}. \tag{4.1}$$

As in Chapter 3, let us consider S independent and homogeneous point processes representing the S species of our ecological community residing in the region $\mathcal{W} \subset \mathbb{R}^2$.
Let us introduce the quantity $\chi(A, B)$, denoting the number of co-present species per species and per unit of survey area, i.e. the spatial density of Sørensen's similarity:

$$\chi(A, B) = \frac{\text{SØR}(A, B)}{a}.$$

Let us now consider, as regions A and B, two infinitesimal disjoint disks of $\mathcal{W}$, $\mathcal{B}_u$ and $\mathcal{B}_v$, centred in u and v, respectively, and having equal area $\mathrm{d}u = \mathrm{d}v$. From Proposition 3.4, we know that the expected value of Sørensen's similarity index can be formulated, in the context of point processes, as

$$\mathbb{E}[\text{SØR}_{\boldsymbol{X}}(u, v)] = \lambda_{\boldsymbol{X}}\mathrm{d}u\left(g_{\boldsymbol{X}}(u, v) - 1 + \frac{\sum_{s=1}^{S}\lambda_{\boldsymbol{X}_s}^2}{\lambda_{\boldsymbol{X}}^2}\right)(1 + \phi_{\text{SØR}}(u, v)), \tag{4.2}$$

where $\boldsymbol{X} = \cup_s \boldsymbol{X}_s$ is the superposed process, $\lambda_{\boldsymbol{X}} = \sum_s \lambda_{\boldsymbol{X}_s}$ the sum of the (constant) intensities of the processes $\boldsymbol{X}_s$, $s = 1, \ldots, S$, (i.e. the intensity of the superposed process $\boldsymbol{X}$) and $g_{\boldsymbol{X}}(\cdot, \cdot)$ is the pair correlation function of $\boldsymbol{X}$ given by eq. (3.17).
Considering the spatial density associated to the quantity in eq. (4.2)

$$\chi_{\boldsymbol{X}}(u, v)\mathrm{d}u = \mathbb{E}[\text{SØR}(u, v)], \tag{4.3}$$

we have the following form for the similarity decay function (Tovo and Favretti, 2017)

$$\chi_{\boldsymbol{X}}(u, v) = \lambda_{\boldsymbol{X}}\left(g_{\boldsymbol{X}}(u, v) - 1 + \frac{\sum_{s=1}^{S}\lambda_{\boldsymbol{X}_s}^2}{\lambda_{\boldsymbol{X}}^2}\right)(1 + \phi_{\text{SØR},\boldsymbol{X}}(u, v)). \tag{4.4}$$

The factor $1 + \phi_{\text{SØR},\boldsymbol{X}}(u, v)$ is hard to compute analytically. Let us thus analyse the simplest case of $\phi_{\text{SØR},\boldsymbol{X}}(u, v) = 0$. In Section 4.2 we will then show that, for our applications, we can make such an assumption.

4.1.2 Similarity decay under $\phi_{\mathbf{S\O R},\boldsymbol{X}} = 0$, stationarity and isotropy hypotheses

Let us consider the special case where $\phi_{\mathrm{S\O R},\boldsymbol{X}}(u,v)$ equals 0 for all points u and v in our study region $\mathcal{W}$.
If, for $s = 1,\ldots,S$, $g_{\boldsymbol{X}_s}(u,v)$ is a function of the distance $r = |u-v|$ (i.e. if each process $\boldsymbol{X}$ is isotropic), then also the pair correlation function of the superposed process $\boldsymbol{X}$ depends only on r. Under these assumptions, eq. (4.4) above becomes distance dependent

$$\begin{aligned}\chi_{\boldsymbol{X}}(r) &= \lambda_{\boldsymbol{X}}\left(g_{\boldsymbol{X}}(r) - 1 + \frac{\sum_{s=1}^{S}\lambda_{\boldsymbol{X}_s}^2}{\lambda_{\boldsymbol{X}}^2}\right)\\ &= \lambda_{\boldsymbol{X}}\left(g_{\boldsymbol{X}}(r) - 1 + \sum_{s=1}^{S}p_s^2\right)\\ &= \lambda_{\boldsymbol{X}}(g_{\boldsymbol{X}}(r)-1) + \lambda_{\boldsymbol{X}}\sum_{s=1}^{S}p_s^2,\end{aligned} \tag{4.5}$$

where we have set $p_s = \frac{\lambda_{\boldsymbol{X}_s}}{\lambda_{\boldsymbol{X}}}$. Note that the quantity $\sum_s p_s^2$ in eq. (4.5) above is Simpson's dominance index $D'_{\boldsymbol{X}}$ for the point process $\boldsymbol{X}$ (see Chapter 3).
We thus have that the Sørensen similarity decay function coincides with the pair correlation function of the superposed process up to a change of scale.[2]
Unlike the original Sørensen's index of eq. (4.1), which is an incidence-based index (i.e. giving equal weight to abundant or rare species), our *distance-dependent* Sørensen's similarity function, which is a probabilistic quantity, takes into account the relative intensity of species p_s. As noted in (Chao et al., 2005), incidence-based indexes are generally biased downward, underestimating similarity especially when species richness is large or sample size is small.
Moreover, from eq. (3.16), we also have the alternative formulation[3] (Tovo and Favretti, 2017)

$$\chi_{\boldsymbol{X}}(r) = \lambda_{\boldsymbol{X}}\sum_{s} g_{\boldsymbol{X}_s}(r)p_s^2, \tag{4.6}$$

which shows that the similarity decay function we propose is dominated by the most abundant species, a feature which has been previously recognised in the literature (see Morlon et al., 2008), but that is rigorously motivated in our formula.
A useful property of the formulae (4.5) and (4.6) above is that they are independent of the size of the plots. This is the consequence of our strategy of considering

[2] If all the species have the same clustering, i.e. are described by the same pair correlation function $g_{\boldsymbol{X}_s}(\cdot) = g(\cdot)$, $s \in \{1,\ldots,S\}$, one has that $\chi_{\boldsymbol{X}}(r) = \lambda_{\boldsymbol{X}} g_{\boldsymbol{X}}(r) D'_{\boldsymbol{X}}$, while if they are equally abundant ($\lambda_{\boldsymbol{X}_s} = \lambda$), then $\chi_{\boldsymbol{X}}(r) = \lambda_{\boldsymbol{X}} S^{-1}\sum_s g_{\boldsymbol{X}_s}(r)$.

[3] Let us remark the our similarity decay function can be related to both the distance-dependent Simpson's-inspired index of Shimatani (Shimatani, 2001), $\beta_{\boldsymbol{X}}(r) = 1 - g_{\boldsymbol{x}}(r)^{-1}\sum_s g_{\boldsymbol{X}_s}(r)p_s^2$, which gives the conditional probability that two individuals at distance r belong to different species given that there are two individuals of $\boldsymbol{X}$ at distance r, and to a spatial point process formulation of the co-dominance index of Chave and Leigh (Chave and Leigh, 2002), $F_{\boldsymbol{X}}(r) = 1 - \beta_{\boldsymbol{X}}(r)$, which gives the probability of the complementary event (see Tovo and Favretti, 2017).
Indeed, we have the additional identities $\chi_{\boldsymbol{X}}(r) = \lambda_{\boldsymbol{X}} g_{\boldsymbol{X}}(r)[1-\beta_{\boldsymbol{X}}(r)] = \lambda_{\boldsymbol{X}} g_{\boldsymbol{X}}(r) F_{\boldsymbol{X}}(r)$ and $\beta_{\boldsymbol{X}}(r) g_{\boldsymbol{X}}(r) = D''_{\boldsymbol{X}}$, where $D''_{\boldsymbol{X}}$ is Gini-Simpson's index for the point process $\boldsymbol{X}$.

virtually infinitesimal plots and the spatial density of Sørensen's similarity.
In a sense, our approach is orthogonal to the one adopted in the works Morlon et al., 2008 and Plotkin, Chave et al., 2002 where the emphasis is first on the development of an analytical formula for the similarity between plots of finite area (even if the assumption of relatively small sample size and relatively large distances is made in Morlon et al., 2008) and then on dealing with the problem of determining how the model parameters (e.g. the negative binomial clumping parameter in Plotkin, Chave et al., 2002) vary with the area under study. See Section 4.5 below for an overview on this problem. We think that our approach, even if not all-compassing, is more straightforward and easier to test against field data.
A crucial point however is how to deal with the assumption of infinitesimal plot size in the design and application of the statistical estimator of the similarity (see eq. (4.13) below) which can be defined only for finite cell size. This point will be thoroughly discussed in Section 4.2 below.

4.1.3 Complete spatial randomness case

Let us consider the similarity decay function of the last equality of eq. (4.5).
If the complete spatial randomness hypothesis (CSR) holds for every species, then $g_{\boldsymbol{X}_s}(r) \equiv 1$ for every $s \in \{1, \dots, S\}$ and $r \in (0, \infty)$. Thus, Sørensen's spatial density becomes distance-independent and constant:

$$\chi_{\boldsymbol{X},\infty} = \lambda_{\boldsymbol{X}} \sum_{s=1}^{S} p_s^2 = \lambda_{\boldsymbol{X}}(1 - D''_{\boldsymbol{X}}), \tag{4.7}$$

where $D''_{\boldsymbol{X}} = 1 - \sum_s p_s^2 = 1 - D'_{\boldsymbol{X}}$, is the Gini-Simpson index (see Chapter 3). Hence $\chi_{\boldsymbol{X},\infty}$ is the product of the probability $\lambda_{\boldsymbol{X}}$ of finding an individual times the probability $1 - D''_{\boldsymbol{X}}$ of finding two individuals belonging to the same species.
Note that, with this definition eq. (4.5) for a general spatial point process becomes

$$\chi_{\boldsymbol{X}}(r) = \lambda_{\boldsymbol{X}}(g_{\boldsymbol{X}}(r) - 1) + \chi_{\boldsymbol{X},\infty}. \tag{4.8}$$

Therefore, since the pair correlation function $g_{\boldsymbol{X}}(r)$ tends to unity as r goes to infinity, $\chi_{\boldsymbol{X}}(r)$ tends to χ_∞ asymptotically.
Formula (4.8) sets the range of applicability of our point process formulation of Sørensen's index: it measures the average change of species' composition at a scale which is comparable with the largest of the correlation lengths $r^c_{\boldsymbol{X}_s} = \min\{r : g_{\boldsymbol{X}_s}(r') = 1 \text{ for } r' > r\}$ of the different species s ($g_{\boldsymbol{X}}(r) = 1$ for $r > r^c_{\boldsymbol{X}} = \max_s r^c_{\boldsymbol{X}_s}$).
At a broader scale, the spatial point process description of the landscape's composition becomes in a sense trivial, because all the species appear to be randomly distributed (therefore satisfying the CSR hypothesis).
Accordingly, the asymptotic value $\chi_{\boldsymbol{X},\infty}$ does not depend on the clustering properties of the various species but only on their abundances. This feature may be useful for establishing a test of the theory independent of the chosen cluster model. Using the standard estimator for the intensity $\hat{\lambda}_{\boldsymbol{X}_s} = n_s(\mathcal{W})/|\mathcal{W}|$, where $|\mathcal{W}|$ is the size of the whole study region $\mathcal{W}$ and $n_s(\mathcal{W}) = \overline{\mathcal{N}_{\boldsymbol{X}_s}(\mathcal{W})}$ is the number of points of the s species within $\mathcal{W}$, we derive the following estimator for the asymptotic value $\chi_{\boldsymbol{X},\infty}$

in eq. (4.7)

$$\hat{\chi}_{\boldsymbol{X},\infty} = \frac{1}{|\mathcal{W}|}\frac{\sum_{s=1}^{S} n_s(\mathcal{W})^2}{\sum_{s=1}^{S} n_s(\mathcal{W})}. \tag{4.9}$$

Let us suppose that the species-abundance distribution (SAD) $\phi(n)$ of the region $\mathcal{W}$ under study is known by other means, independently of the assumptions made for the spatial process. Then, eq. (4.9) can be rewritten as follows

$$\hat{\chi}_{\boldsymbol{X},\infty} = \frac{1}{|\mathcal{W}|}\frac{\mathbb{E}_\phi[n^2]}{\mathbb{E}_\phi[n]}. \tag{4.10}$$

In this case, $\hat{\chi}_{\boldsymbol{X},\infty}$ depends on the ratio between the second and first moment of n.

4.1.4 Analytical formula for finite-size cells under the CSR hypothesis

For later use, we derive here an analytical formula for the similarity between two regions of finite and equal area a under the CSR hypothesis, i.e. for a superposition of Poisson point processes $\boldsymbol{X}_S$ of intensities $\lambda_{\boldsymbol{X}_s}$ (see e.g. Diggle, 2013). Under this assumption, as we have seen in Chapter 1, every point is uncorrelated to the others, hence the spatial point pattern is independent of the distance between its points. This framework is thus intended to describe the similarity between two plots of finite area very far away so that the species' compositions are virtually independent. This is the approach taken in Plotkin, Chave et al., 2002.
In this case, for two regions $A, B \subset \mathcal{W}$ of equal area a, we have that – see eqs. (3.13) and (3.14) –

$$\mathbb{E}[S_{\boldsymbol{X}}(A)] = \sum_{s=1}^{S} P(\mathcal{N}_{\boldsymbol{X}}(A) \geq 1) = \sum_{s=1}^{S} 1 - e^{-\lambda_{\boldsymbol{X}_s} a} = \mathbb{E}[S_{\boldsymbol{X}}(B)],$$

and, since $P(\mathcal{N}_{\boldsymbol{X}_s}(A) \geq 1, \mathcal{N}_{\boldsymbol{X}_s}(B) \geq 1) = P(\mathcal{N}_{\boldsymbol{X}_s}(A) \geq 1)P(\mathcal{N}_{\boldsymbol{X}_s}(B) \geq 1)$,

$$\mathbb{E}[S_{\boldsymbol{X}}(A,B)] = \sum_{s} P(\mathcal{N}_{\boldsymbol{X}_s}(A) \geq 1, \mathcal{N}_{\boldsymbol{X}_s}(B) \geq 1) = \sum_{s} (1 - e^{-\lambda_{\boldsymbol{X}_s} a})^2.$$

Therefore

$$\chi^{\text{CSR}}_{\boldsymbol{X}}(a) = \frac{1}{a}\frac{\mathbb{E}[S_{\boldsymbol{X}}(A,B)]}{\mathbb{E}[S_{\boldsymbol{X}}(A)]} = \frac{1}{a}\frac{\sum_{s=1}^{S}(1 - e^{-\lambda_{\boldsymbol{X}_s} a})^2}{\sum_{s=1}^{S} 1 - e^{-\lambda_{\boldsymbol{X}_s} a}}.$$

Using the standard estimation for $\lambda_{\boldsymbol{X}_s}$, $\hat{\lambda}_{\boldsymbol{X}_s} = n_s(\mathcal{W})/|\mathcal{W}|$, we get (see pink solid lines in figure 4.3)

$$\hat{\chi}^{\text{CSR}}_{\boldsymbol{X}}(a) = \frac{1}{a}\frac{\sum_{s=1}^{S}\left(1 - e^{-\frac{n_s(\mathcal{W})}{|\mathcal{W}|}a}\right)^2}{\sum_{s=1}^{S} 1 - e^{-\frac{n_s(\mathcal{W})}{|\mathcal{W}|}a}}. \tag{4.11}$$

Eq. (4.11) gives the predicted asymptotic value of the similarity between two patches of finite area a far away, given the value of $n_s(\mathcal{W})$.
This quantity can be considered as the discrete analogous of equation (10) in (Plotkin, Chave et al., 2002) under the CSR hypothesis (see also (Morlon et al., 2008), Supporting Information F2).

Using the Taylor expansion of the exponential function $e^x = 1+x+o(x^2)$ for $x \to 0$, when $a \to 0$ the above formula reduces to eq. (4.9):

$$\lim_{a\to 0} \hat{\chi}_{\boldsymbol{X}}^{\mathrm{CSR}}(a) = \frac{1}{a} \frac{\sum_{s=1}^{s} \left(1 - 1 + \frac{a}{|W|} n_s(\mathcal{W})\right)^2}{\sum_{s=1}^{S} 1 - 1 + \frac{a}{|W|} n_s(\mathcal{W})} = \frac{1}{|W|} \frac{\sum_{s=1}^{S} n_s(\mathcal{W})^2}{\sum_{s=1}^{S} n_s(\mathcal{W})} = \hat{\chi}_{\boldsymbol{X},\infty}. \quad (4.12)$$

4.2 Estimators for $\chi_{\boldsymbol{X}}$ and $g_{\boldsymbol{X}}$

4.2.1 Direct estimators for Sørensen's spatial density

Let $S = \overline{S_{\boldsymbol{X}}(\mathcal{W})}$ be the total number of species and $N = \overline{\mathcal{N}_{\boldsymbol{X}}(\mathcal{W})}$ the total number of individuals in our study region W, supposed to be a rectangular window of side lengths l_x and l_y. For each species s, the coordinates x_i^s, $i \in \{1, \dots, n_s(\mathcal{W})\}$ of all its $n_s(\mathcal{W}) = \overline{\mathcal{N}_{\boldsymbol{X}_s}(\mathcal{W})}$ individuals falling within $\mathcal{W}$ are known.
In order to estimate the Sørensen similarity decay function defined by eq. (4.3), we first divide our region $\mathcal{W}$ in cells of area a and call $C = |\mathcal{W}|/a$ the total number of cells. Let c_i be the centre coordinates of cell $i = 1, \dots, C$. Then, calling K_r the number of cells having distance r from each other, we give the following estimator (Tovo and Favretti, 2017)

$$\hat{\chi}_{\boldsymbol{X}}(r; a) = \frac{2}{aK_r} \sum_{\substack{i,j=1 \\ j\neq i}}^{C} \frac{S(i,j)}{S(i)+S(j)} \mathrm{II}(\| c_i - c_j \| = r), \quad (4.13)$$

where $S(i) = \overline{S_{\boldsymbol{X}}(i)}$ is the number of species in cell i, $S(i,j) = \overline{S_{\boldsymbol{X}}(j)}$ is the number of co-present species in cells i and j, and $\mathrm{II}(\cdot)$ is the indicator function:

$$\mathrm{II}(\mathcal{C}) = \begin{cases} 1 & \text{if condition } \mathcal{C} \text{ holds} \\ 0 & \text{otherwise.} \end{cases}$$

Let us remark that, with the estimator 4.13, we are considering the spatial average over all cells in $\mathcal{W}$ having distance r, while point process theory considers the average, for fixed cells located in x and y, over many realisations of the point process. This latter cannot be computed, since, of course, we are given a single realisation of the process, that is the 'real' forest. However, the two kind of averages (respectively over space and over realisations) are equivalent provided that the stationarity and isotropy hypotheses are realistic assumptions for our forest.
Denoting with $X = S(i,j)$ and with $Y = S(i) + S(j)$, we have that eq. (4.13) will give an estimate of $\mathbb{E}[X/Y]$. The estimator for the ratio $\mathbb{E}[X]/\mathbb{E}[Y]$ can, instead, be obtained through the following formula

$$\hat{\chi}'_{\boldsymbol{X}}(r; a) = \frac{2}{a} \frac{\frac{1}{K_r} \sum_{\substack{i,j=1 \\ j\neq i}}^{C} S(i,j) \mathrm{II}(\| c_i - c_j \| = r)}{\frac{1}{K_r} \sum_{\substack{i,j=1 \\ j\neq i}}^{C} [S(i)+S(j)] \mathrm{II}(\| c_i - c_j \| = r)}. \quad (4.14)$$

We already noticed (see Chapter 3) that the following relation holds:

$$\chi_{\boldsymbol{X}}(u,v) = \mathbb{E}\left[\frac{S_{\boldsymbol{X}}(u,v)}{S_{\boldsymbol{X}}(u)+S_{\boldsymbol{X}}(v)}\right] = \frac{\mathbb{E}[S_{\boldsymbol{X}}(u,v)]}{\mathbb{E}[S_{\boldsymbol{X}}(u)+S_{\boldsymbol{X}}(v)]}(1+\phi_{\text{SØR},\boldsymbol{X}}(u,v)). \tag{4.15}$$

Thus, if we compute $\mathbb{E}[X/Y]$ and $\mathbb{E}[X]/\mathbb{E}[Y]$ through estimators eqs. (4.13) and (4.14), we can also compute $\phi_{\text{SØR},\boldsymbol{X}}(u,v)$.
Let us note that, under the hypotheses of stationarity and isotropy under which we are working, the average $\mathbb{E}[S_{\boldsymbol{X}}(u,v)]$ only depends on the modulus $r = |u-v|$ (indeed, $\rho_{\boldsymbol{X}}(u,v) = \rho_{\boldsymbol{X}}(r)$), while both $\mathbb{E}[S_{\boldsymbol{X}}(u)]$ and $\mathbb{E}[S_{\boldsymbol{X}}(v)]$ are constantly equal to $\lambda_{\boldsymbol{X}}dx$. Therefore, also $\phi_{\text{SØR},\boldsymbol{X}}(u,v)$ can be written as a function of r, $\phi_{\text{SØR},\boldsymbol{X}}(r)$. Using eqs. (4.13) and (4.14), we have found that, for our empirical data, $\phi_{\text{SØR},\boldsymbol{X}}(r) < 0.05$ at any distance r considered, leading to a relative error of less than 5% between these two estimators (see figure 4.1), $\hat{\chi}_{\boldsymbol{X}}(r;a)$ and $\hat{\chi}'_{\boldsymbol{X}}(r;a)$. Therefore, in the following we will approximate $\phi_{\text{SØR},\boldsymbol{X}}(r) \approx 0$.

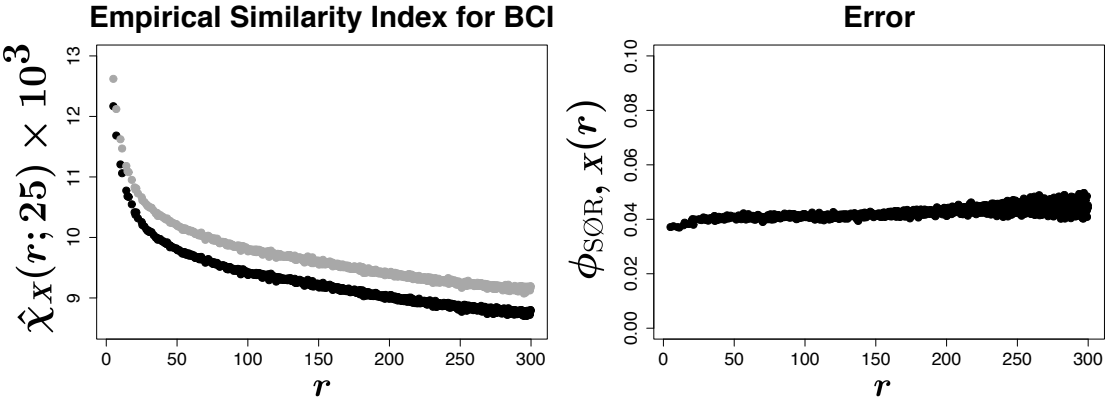


Figure 4.1: Comparison between empirical estimators of $\chi_{\boldsymbol{X}}$. We have computed the similarity index for the BCI dataset considering species with more than 200 individuals through eq. (4.13) (black curve on the left panel) and we compare it with the one obtained through eq. (4.14) (grey curve). In both cases a has been set equal to 25. On the left we plot the values of $\phi_{\text{SØR},\boldsymbol{X}}(r)$ at any distance. It results always smaller than 0.05, leading to a relative error between the two estimators of less than 5%. In the rest of this chapter we will therefore approximate $\phi_{\text{SØR},\boldsymbol{X}}(r) \approx 0$.

4.2.2 Estimator for $\chi_{\boldsymbol{X}}$ based on the estimator for $g_{\boldsymbol{X}}$

Following Stoyan and Stoyan, 1994, we estimate the empirical pair correlation function of the s species as follows

$$\hat{g}_{\boldsymbol{X}_s}(r) = \frac{1}{\hat{\lambda}_{\boldsymbol{X}_s} n_s(\mathcal{W})} \sum_{\substack{j,k=1 \\ k\neq j}}^{n_s(\mathcal{W})} \frac{w(\parallel x_j^s - x_k^s \parallel - r)}{2\pi r B(r)}, \tag{4.16}$$

where $\hat{\lambda}_{\boldsymbol{X}_s} = n_s(\mathcal{W})/|\mathcal{W}|$ is the unbiased estimator of the density of individuals of the s species, $\parallel \cdot \parallel$ denotes the Euclidean distance on the plane, $B(r) = 1 - r(2l_x +$

$2l_y - r)/l_x l_y \pi$ is the edge corrector function and $w(\cdot)$ is the Epanechnikov kernel, defined as

$$w(z) = \begin{cases} \frac{3}{4\delta}\left(1 - \frac{z^2}{\delta^2}\right) & \text{for } |z| < \delta \\ 0 & \text{for } |z| \geq \delta, \end{cases}$$

with $\delta = 0.2/\sqrt{\hat{\lambda}_{\boldsymbol{X}_s}}$.
The estimator for the pair correlation function $g_{\boldsymbol{X}}(\cdot)$ of the superposed process $\boldsymbol{X}$ can thus be computed using the following relation (see eq. (3.18))

$$g_{\boldsymbol{X}}(r) = \frac{1}{\lambda_{\boldsymbol{X}}^2}\Big[\sum_{s=1}^{S} \lambda_{\boldsymbol{X}_s}^2 g_{\boldsymbol{X}_s}(r) + \sum_{\substack{s,t=1 \\ s \neq t}} \lambda_{\boldsymbol{X}_s}\lambda_{\boldsymbol{X}_t}\Big],$$

which yields

$$\hat{g}_{\boldsymbol{X}}(r) = \frac{1}{\hat{\lambda}_{\boldsymbol{X}}^2}\Big[\sum_{s=1}^{S} \hat{\lambda}_{\boldsymbol{X}_s}^2 \hat{g}_{\boldsymbol{X}_s}(r) + \sum_{\substack{s,t=1 \\ s \neq t}}^{S} \hat{\lambda}_{\boldsymbol{X}_s}\hat{\lambda}_{\boldsymbol{X}_t}\Big], \tag{4.17}$$

where $\hat{\lambda}_{\boldsymbol{X}} = N/|\mathcal{W}|$ is the unbiased estimator of the total density of individuals. Therefore, an indirect estimator of $\chi_{\boldsymbol{X}}(\cdot)$ can be obtained from plugging eq. (4.17) above and (4.9) into eq. (4.8):

$$\hat{\chi}_{\boldsymbol{X}}(r) = \hat{\lambda}(\hat{g}_{\boldsymbol{X}}(r) - 1) + \hat{\chi}_{\boldsymbol{X},\infty}. \tag{4.18}$$

4.2.3 Scaling for finite-size cells

Note that the statistical estimator $\hat{\chi}_{\boldsymbol{X}}(r; a)$ given by eq. (4.13) for $\chi_{\boldsymbol{X}}(r)$ depends on an extra parameter, the cell size a.
Since our analytical formula (4.5) is designed for an ideally infinitesimal area, we are faced with the problem of coupling the results of the direct estimator $\hat{\chi}_{\boldsymbol{X}}(r; a)$ relative to finite area cells to the output of eq. (4.18), where the finiteness is only taken into account by the Epanechnikov kernel. Below we show how to properly rescale the decay curves of $\hat{\chi}_{\boldsymbol{X}}(r; a)$ to different areas.
In Section 4.1.4 we have derived the analytical formula (4.11) for the similarity between two regions of equal area under the complete spatial randomness hypothesis. Out of this special case, it is hard to guess how the output of the similarity estimator (4.13) depends on the cell size a.
Here we test the incidence of the cell size on the output of similarity estimator (4.13) by superimposing five different grids onto the BCI 50ha plot, with square cells of area 1, 4, 25, 100 and 625 square meters respectively.
In figure 4.2, we can see that the choice of the cell size a strongly influences the curves $\hat{\chi}_{\boldsymbol{X}}(r; a)$ although the general trend results stable. However, as shown in the right panels of the same figure, the curves $\hat{\chi}_{\boldsymbol{X}}(r; a)$ divided by their respective value at the largest considered distance, $\hat{\chi}(r_{\max}; a)$, are approximatively independent on the cell size. For us $r_{\max}$ is the maximal available distance in the study area given by the shorter side of the rectangular study area, that is $r_{\max} = 500$ m. In the following, we set $\hat{\chi}_{\boldsymbol{X}}(r_{\max}; a) = \hat{\chi}_{\boldsymbol{X},r_{\max}}(a)$.

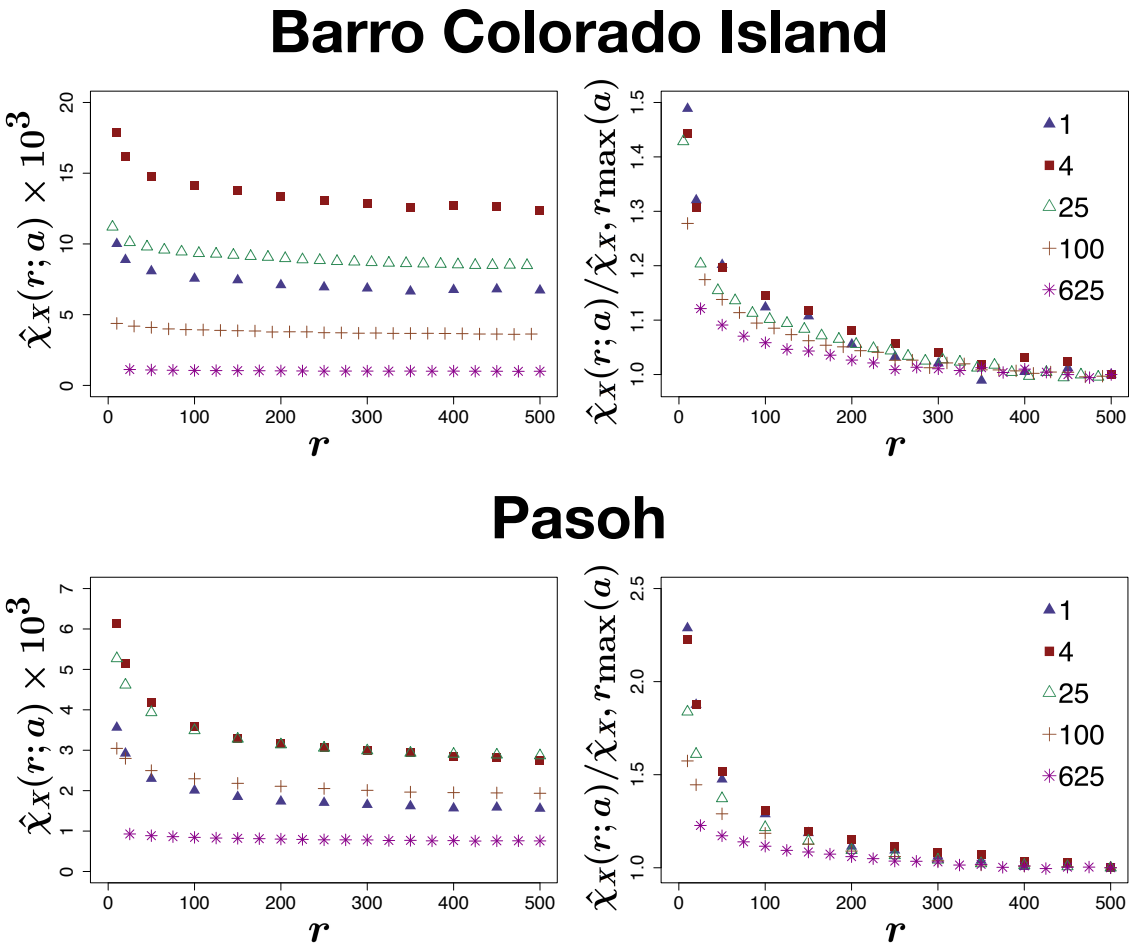


Figure 4.2: Sensitivity of Sørensen similarity decay function on cell size. We have computed the similarity index for the BCI (top panels) and Pasoh (bottom panels) datasets considering species with more than 200 and 100 individuals, respectively. We have superimposed to the 1000x500 observation window different regular square grids and estimated the corresponding Sørensen spatial density $\hat{\chi}_{\boldsymbol{X}}(r;a)$ via eq. (4.13). In figure, different colours represent different cell sizes, as in the legend. On the left panels: the choice of the cell size strongly influences the result, by "shifting the $\hat{\chi}_{\boldsymbol{X}}(r;a)$ curves along the y-axis". On the right panel: we divide each curve by its empirical value at the maximum considered distance, $\hat{\chi}_{\boldsymbol{X},r_{\max}}(a)$. The resulting curve can be considered approximatively independent of the cell size and the error decreases with the distance.

Therefore, from the right panels of figure 4.2, we can experimentally deduce that, at any distance r, the following relations holds

$$\frac{\hat{\chi}_{\boldsymbol{X}}(r;a)}{\hat{\chi}_{\boldsymbol{X},r_{\max}}(a)} \approx \hat{f}_{\boldsymbol{X}}(r),$$

Thus, the ratio between the empirical similarity decay function and its last-computed value is independent of the cell size. Hence, for two cell sizes a and b, we have that

$$\hat{\chi}_{\boldsymbol{X}}(r;b) \approx \hat{f}_{\boldsymbol{X}}(r)\hat{\chi}_{r_{\max}}(b) \approx \hat{\chi}(r;a)\frac{\hat{\chi}_{\boldsymbol{X},r_{\max}}(b)}{\hat{\chi}_{\boldsymbol{X},r_{\max}}(a)}. \tag{4.19}$$

In the limit for $b \to 0$, we may derive the statistical estimator for an infinitesimal cell size, given the estimator for a finite cell size $\hat{\chi}_{\boldsymbol{X}}(r;a)$

$$\hat{\chi}_{\boldsymbol{X}}(r;0) \approx \hat{\chi}_{\boldsymbol{X}}(r;a)\frac{\hat{\chi}_{\boldsymbol{X},r_{\max}}(0)}{\hat{\chi}_{\boldsymbol{X},r_{\max}}(a)} \approx \hat{\chi}_{\boldsymbol{X}}(r;a)\hat{\gamma}_{\boldsymbol{X}}(a), \tag{4.20}$$

where we have set $\hat{\chi}_{\boldsymbol{X},r_{\max}}(0) \equiv \hat{\chi}_{\boldsymbol{X},\infty}$ and

$$\hat{\gamma}_{\boldsymbol{X}}(a) = \hat{\chi}_{\boldsymbol{X},\infty}/\hat{\chi}_{\boldsymbol{X},r_{\max}}(a). \tag{4.21}$$

Note that the scaling factor $\hat{\gamma}_{\boldsymbol{X}}(a)$ is the ratio of two similarities between plots very far away, where we may assume that the CSR hypothesis holds. In Section 4.1.4, we have derived the analytical function $\chi^{\mathrm{CSR}}_{\boldsymbol{X}}(a)$ of the similarity under CSR hypothesis and its limit for $a \to 0$, $\hat{\chi}_{\boldsymbol{X},\infty}$ (see eq. (4.11)).
We wish to test if the assumption $\hat{\chi}_{\boldsymbol{X},r_{\max}}(a) = \hat{\chi}^{\mathrm{CSR}}_{\boldsymbol{X}}(a)$ does hold for our study area, so that we may compute analytically the scaling factor as

$$\gamma_{\boldsymbol{X}}(a) = \hat{\chi}_{\boldsymbol{X},\infty}/\hat{\chi}^{\mathrm{CSR}}_{\boldsymbol{X}}(a). \tag{4.22}$$

Results are contained in figure 4.3, where we plot the function $\hat{\chi}^{\mathrm{CSR}}_{\boldsymbol{X}}(a)$ defined by eq. (4.11), for a up to 625 square meters (solid pink lines) and the empirical values of $\hat{\chi}_{\boldsymbol{X},r_{\max}}(a)$ computed through the estimator (4.13) for different cell-sizes (colored symbols). There is a very good agreement between these two quantities, which improves with the cell size a (the off-curve point $a = 1$ square meters is probably due to the finite diameter of the plants) meaning that, at least for $a \geq 25$ square meters, the CSR hypothesis holds for distances $r \sim r_{\max} = 500$ meters where correlations between points become negligible. Note that the curve $\hat{\chi}^{\mathrm{CSR}}_{\boldsymbol{X}}(a)$ (pink solid lines) tends asymptotically to $\hat{\chi}_{\boldsymbol{X},\infty}$ (light blue lines) for $a \to 0$ as prescribed by eq. (4.12). With the scaling formula (4.20), we can now rescale the output of the direct statistical estimator $\hat{\chi}_{\boldsymbol{X}}(r;a)$ for a finite cell size a to an infinitesimal cell size in order to compare it with the output of the indirect estimator (4.18) based on the pair correlation function $g_{\boldsymbol{X}}$ of the superposed process.
Note that the rescaling (both upscaling or downscaling) can be done also between two empirical curves $\hat{\chi}_{\boldsymbol{X}}(r;a)$ and $\hat{\chi}_{\boldsymbol{X}}(r;b)$ through eq. (4.19), using the empirical scaling factor $\hat{\gamma}_{\boldsymbol{X}}(b,a) = \hat{\chi}_{\boldsymbol{X},r_{\max}}(b)/\hat{\chi}_{\boldsymbol{X},r_{\max}}(a)$ or the analytical one $\gamma_{\boldsymbol{X}}(b,a) = \chi^{\mathrm{CSR}}_{\boldsymbol{X}}(b)/\chi^{\mathrm{CSR}}_{\boldsymbol{X}}(a)$.

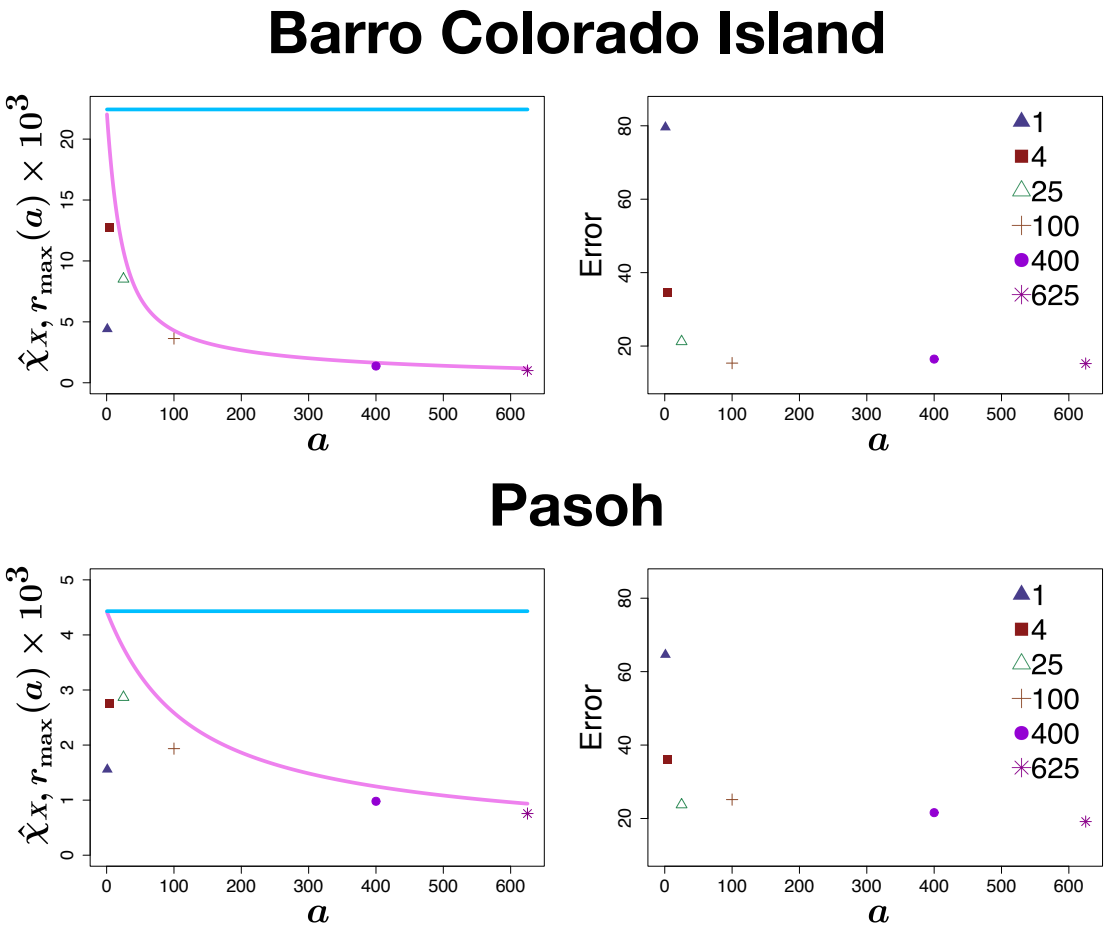


Figure 4.3: χ_∞ for finite-size cells. On the left: the pink lines represent the asymptotic value of the theoretical similarity index, $\hat{\chi}^{\mathrm{CSR}}_{\boldsymbol{X}}(a)$ as a function of the cell-area – see eq. (4.11) –. Coloured symbols are the empirical values $\hat{\chi}_{\boldsymbol{X},r_{\max}}(a)$ computed via eq. (4.13) for cell sizes of 1, 4, 25, 100, 400 and 625 square meters, respectively, and considering only species with more than 200 individuals. The straight light blue line represents the value of $\hat{\chi}_{\boldsymbol{X},\infty}$ for infinitesimal cells computed through the abundances (see eq. (4.9)). On the right: relative percentage error between the empirical values and the theoretical ones.

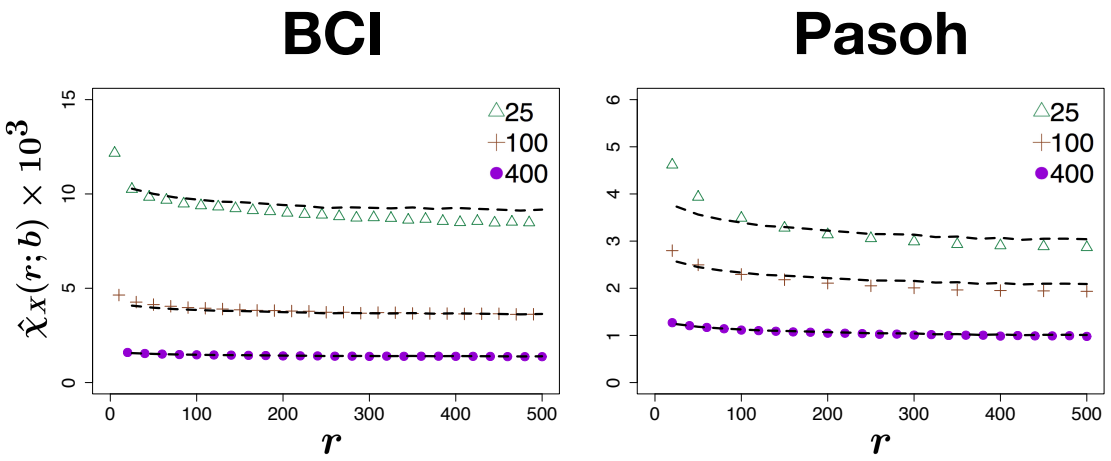


Figure 4.4: Comparison between rescaled $\chi_{\boldsymbol{X}}$ estimators for BCI and Pasoh datasets. We tested the goodness of the rescaling eq. (4.19) by plotting the estimator (4.13) for $\hat{\chi}_{\boldsymbol{X}}(r;b)$ with $b = 25$ (triangles), $b = 100$ (crosses) and $b = 400$ (dots) against the rescaled one $\hat{\chi}_{\boldsymbol{X}}(r;a)\gamma_{\boldsymbol{X}}(b,a)$ for $a = 625$ (dashed line). The difference between the two curves increases with $|b-a|$, but there is always an excellent fit.

In figure 4.4 we downscale the similarity decay function $\hat{\chi}_{\boldsymbol{X}}(r;a)$ estimated via eq. (4.13) with $a = 625$ to $\hat{\chi}_{\boldsymbol{X}}(r;b) = \hat{\chi}_{\boldsymbol{X}}(r,a)\gamma_{\boldsymbol{X}}(b,a)$ for $b = 25, 100, 400$ (dashed lines) and compare the curves with the original ones $\hat{\chi}_{\boldsymbol{X}}(r;b)$ for the same values of b (colored symbols). As expected, for such fairly large areas, all the rescaled estimators are in good agreement. In the sequel we will use the empirical scaling factor $\hat{\gamma}_{\boldsymbol{X}}(a)$ for smaller plots (1 to 4 square meters).

4.2.4 Preliminary test on computer-generated forests

It is evident from eq. (4.6) or eq. (4.8) that our Sørensen's similarity decay function depends on the relative abundances of species and on their pair correlation functions. These latter, in turn, depend essentially on the clustering of the individuals. Therefore, the crucial point for the description of real data patterns is the choice of the spatial point process' cluster model. In this chapter we limit ourselves to the Neyman-Scott (NS) processes introduced in Chapter 1.
Let us recall some notions on these latter. Considering S NS processes $\boldsymbol{X}_s$ of parameters $(\rho_{\boldsymbol{X}_s}, \mu_{\boldsymbol{X}_s}, \gamma_{\boldsymbol{X}_s})$, the Sørensen's similarity decay function of the process resulting from their superposition is given by plugging eq. (1.26) applied to each $\boldsymbol{X}_s$, $s \in \{1, \ldots, S\}$, into eq. (4.6)

$$\chi_{\boldsymbol{X}}(r) = \chi_{\boldsymbol{X},\infty} + \sum_{s=1}^{S} \mu_{\boldsymbol{X}_s} f_{\gamma_{\boldsymbol{X}_s}}(r) p_{\boldsymbol{X}_s}, \tag{4.23}$$

where $f_{\gamma_{\boldsymbol{X}_s}}$ is the convolution of the two-dimensional dispersal kernel $d_{\gamma_{\boldsymbol{X}_s}}(r)$. In the general case, the form of this latter function reflects the cluster characteristics and may have short or long tails. The most used are Gaussian (single or mixture),

inverse power or exponential (see Tanaka et al., 2008). Since $\chi_{\boldsymbol{X}}(r)$ tends to $\chi_{\boldsymbol{X},\infty}$, necessarily

$$\lim_{r\to\infty} f_{\gamma_{\boldsymbol{X}_s}}(r) = 0,$$

while the limit of $f_{\gamma_{\boldsymbol{X}_s}}(r)$ for $r \to 0^+$ may be finite or infinite (in this latter case the function is said to have a pole at 0).
The choice of the dispersal kernel $d_{\gamma_{\boldsymbol{X}_s}}(r)$ is a crucial point also in seed dispersal studies (see e.g. Clarke and Lidgard, 2000; Chave and Leigh, 2002 for an overview of the problem). The validation of the dispersal kernel choice is particularly difficult for tropical rainforests where the tree crowns overlap and therefore it is impossible to identify the seed shadow dispersed by a single tree.
The theory exposed above, establishing a link between the form of the dispersal kernel – see Chapter 1 – and the form of the resulting similarity decay function, could help to select the form of the dispersal kernel from the empirical similarity curve as an inverse problem. In particular, it would be interesting to determine the dispersal kernel that gives rise to a compound exponential decay function, as predicted by Hubbell's neutral theory.
In Section 4.3, we will apply the theory exposed above to real forests. In particular, we will compare the above analytical curve eq. (4.23) for $\chi_{\boldsymbol{X}}(r)$ with the empirical ones coming from field data to select the best cluster model and determine the corresponding cluster parameters.
Here we test the validity of the estimator (4.13), rescaled through eq. (4.20), on four artificial forests generated according to different point processes (see Chapter 1): a Poisson one and three Neyman-Scott processes (exponential, Gaussian(modified Thomas) and Cauchy).
In all cases, we consider a square window of side 500 meters and we generate a forest consisting of 50 species having abundances distributed according to a normal distribution of mean and standard deviation equal to 1000 and 300 individuals, respectively.
For the exponential and modified Thomas cluster processes, to each species is assigned a random average clumping radius $r_{\boldsymbol{X}}$ drawn from a normal distribution of mean 20 and standard deviation 5. These values determine the cluster parameters $\beta = 2/r_{\boldsymbol{X}}$ and $\sigma = r_{\boldsymbol{X}}\sqrt{2/\pi}$ (see Chapter 1). Since the average radius is not well defined for a Cauchy cluster process, in this case we arbitrarily set the cluster parameter b equal to $r_{\boldsymbol{X}}/2$.
Once generated the four forests, we estimate the similarity decay function for infinitesimal area using the rescaled estimator $\hat{\chi}_{\boldsymbol{X}}(r;b)\hat{\gamma}_{\boldsymbol{X}}(b)$ for $b = 25,\ 100,\ 200$ and 625. We will then compare the empirical curves with the theoretical similarity decay functions given by eq. (4.7) for the Poisson process and eq. (4.23) for the Neyman-Scott processes. Results are displayed in figure 4.5. We can see that, with the exception of the Poisson case, where for all cell sizes b the empirical curves are constantly around the theoretical value of $\hat{\chi}_{\boldsymbol{X},\infty}$, in the other three cases the estimated curves tend from below to the theoretical ones as b approaches zero, in accordance with the theory.

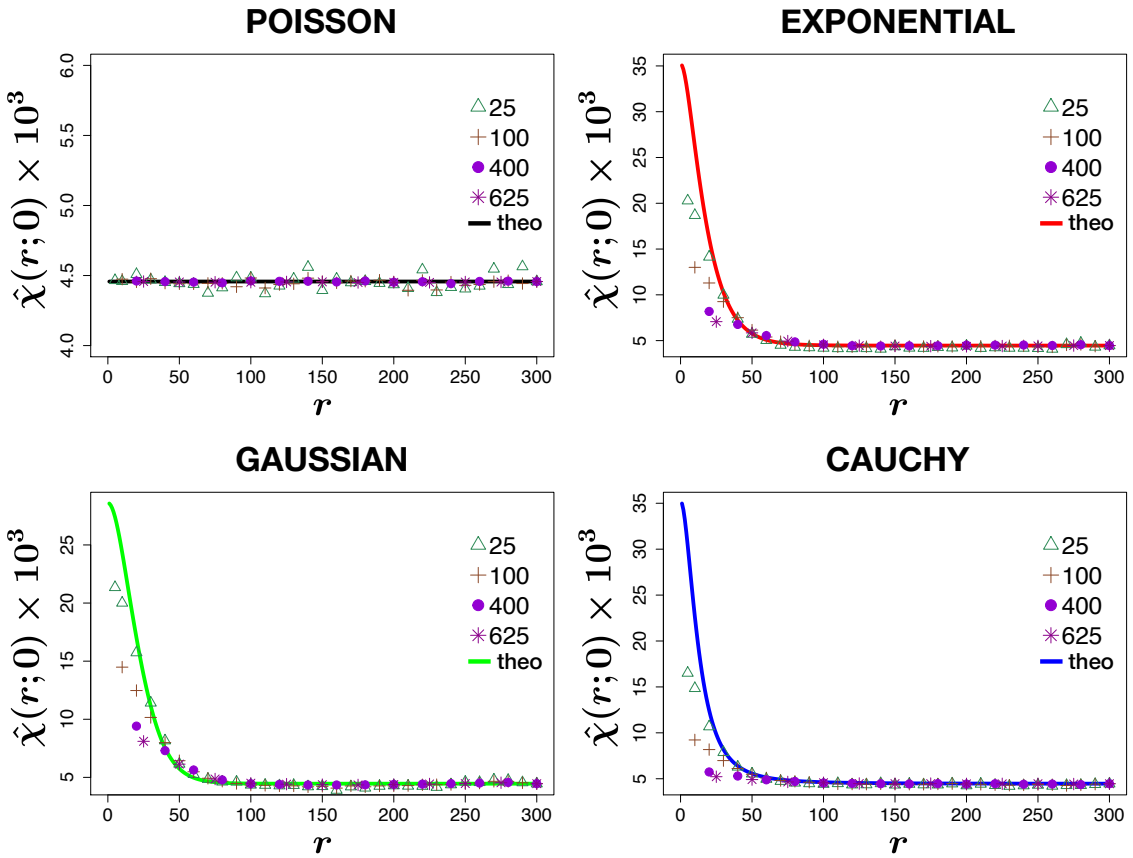


Figure 4.5: Test of the estimator (4.20) on four artificial forests generated according to different point processes: a Poisson one and three cluster processes (exponential, modified Thomas and Cauchy). We compare $\hat{\chi}_{\boldsymbol{X}}(r;b)\hat{\gamma}_{\boldsymbol{X}}(b)$ for $b = 25$ (triangles), 100 (crosses), 400 (dots) and 625 (stars) with the theoretical similarity decay function (solid curves) given by eq. (4.7) for the Poisson process and by eq. (4.23) for the Neyman-Scott processes. The agreement between the empirical and theoretical curves increases as b decreases, but it is always very good.

4.3 Test of the estimators on BCI ecological dataset

We test our analytical formula eq. (4.8) and its related estimators (4.13) and (4.18) on the Barro Colorado Island ecological dataset (BCI) consisting of the spatial coordinates of 222602 individuals belonging to 301 different species of plants within a 50ha rainforest plot.

4.3.1 Species selection and sub-sampling

Species selection. To see if neglecting the scarcely abundant species strongly affects the similarity index for BCI, we superimpose a 5x5 grid onto the plot and compare our distance-dependent Sørensen index computed via eq. (4.13) taking into account species having least abundance of 0, 20, 100, 200, 300 and 500 individuals, which represent, respectively, the 100%, 73%, 49%, 36%, 30% and 24% of the total biodiversity and which account for the 100%, 99%, 98%, 96%, 93% and 90% of the total number of individuals in the 50ha plot. From our analytical formula, we know that the similarity is affected only by the most abundant species.
On the left panel of figure 4.6 we plot the obtained empirical curves, while in the right plot the corresponding relative percentage error with respect to the Sørensen index computed for the whole forest. By selecting the species with a population of at least 200 individuals (107 species), we get a curve which differs from the similarity curve of the whole BCI for about 5%, a reasonable restriction for our goal.

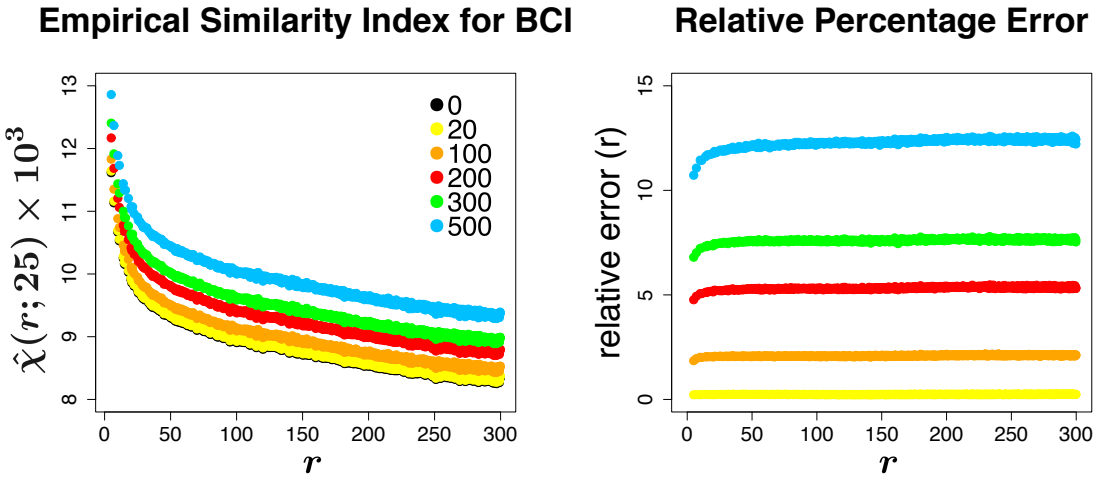


Figure 4.6: Sensitivity of Sørensen's similarity decay function on the abundances. We have already noticed that our distance-dependent Sørensen index is dominated by the most abundant species. On the left panel, we show the different curves $\hat{\chi}_{\boldsymbol{X}}(r;25)$ obtained via estimator (4.13) by taking into account only species having least abundance as indicated in the legend. On the right panel, we insert the relative percentage error with respect to the black curve, which is the similarity index for the whole BCI forest.

Sub-sampling. As a last preliminary test, we check whether sub-sampling affects the estimate of our Sørensen's similarity decay function. Again, we superimpose a 5x5 grid on our study region and we consider three scales of sub-sampling by randomly taking the following percentages of the 20·000 available cells: 50%, 25% and 5%. In figure 4.7, coloured points are the result of averaging over 10 trials for each sub-scale. We can observe that, although lower percentages affect the curve by significantly increasing the fluctuations, the general trend of the curve is very well preserved for a sub-sampling of up to 5% of the data.

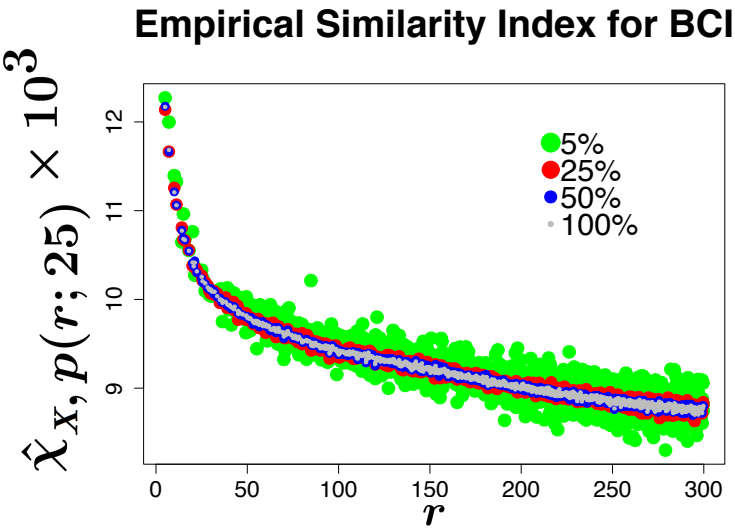


Figure 4.7: Sensitivity of Sørensen's similarity decay function on sub-sampling. We investigate the effect of sub-sampling in computing the empirical Sørensen index estimated through eq. (4.13). We first superimpose a 5x5 grid on the window, corresponding to $C = 20{\cdot}000$ square cells. For every percentage p shown in the legend, we randomly choose pC cells and compute the distance-dependent similarity index between them, $\hat{\chi}_{\boldsymbol{X},p(r;25)}$. Points are the mean on 10 trials. As expected, the lower the percentage of considered cells, the more fluctuations affect the curve, although the trend results very stable under sub-sampling.

4.3.2 Comparison of direct and indirect similarity estimators

We can test the validity of our similarity decay model (formulae (4.6) and (4.8))

$$\chi_{\boldsymbol{X}}(r) = \lambda_{\boldsymbol{X}}(g_{\boldsymbol{X}}(r) - 1) + \chi_{\boldsymbol{X},\infty}$$

by estimating independently its left and right-hand sides. For $\chi_{\boldsymbol{X}}(r)$ we use the estimator $\hat{\chi}_{\boldsymbol{X}}(r;a)\hat{\gamma}_{\boldsymbol{X}}(a)$ given by eq. (4.13) multiplied by the finite cell size scaling factor $\hat{\gamma}_{\boldsymbol{X}}(a)$ defined in eq. (4.21), whereas for the right-hand side we use the indirect estimator (4.18). Results are displayed in figure 4.8. We can see that there is a very good agreement between the two estimates if the smallest cell sizes are used ($a = 1$ or 4 square meters), which are the closest to the theoretical hypothesis of

infinitesimal cell size.
Moreover, we see that the estimated decay curves tends to the analytical one monotonically from below. Therefore the analytical curve sets an upper bound for the density of similarity. Let us also note that, when using the smallest cell sizes, the curve displays a tri-phasic behaviour with a steep initial descent, a linear descent in the middle and a hollow tail. This behaviour is not captured with coarser cell sizes. We will discuss a possible explication of this phenomenon in Section 4.3.3.
We stress that the basic estimator (4.13) is independent of any assumptions on the clustering of the individuals, which are not known a priori, while these assumptions are contained in the indirect estimator (4.18). The good agreement between the two estimates supports the conclusion that, at the considered scale $0 < r < 500$ m, clustering is a main driver of the species turnover. Also, at contrast with Hubbell's neutral theory (Hubbell, 2001b), we find that rare species do not contribute significantly to the species' turnover even at local scales.

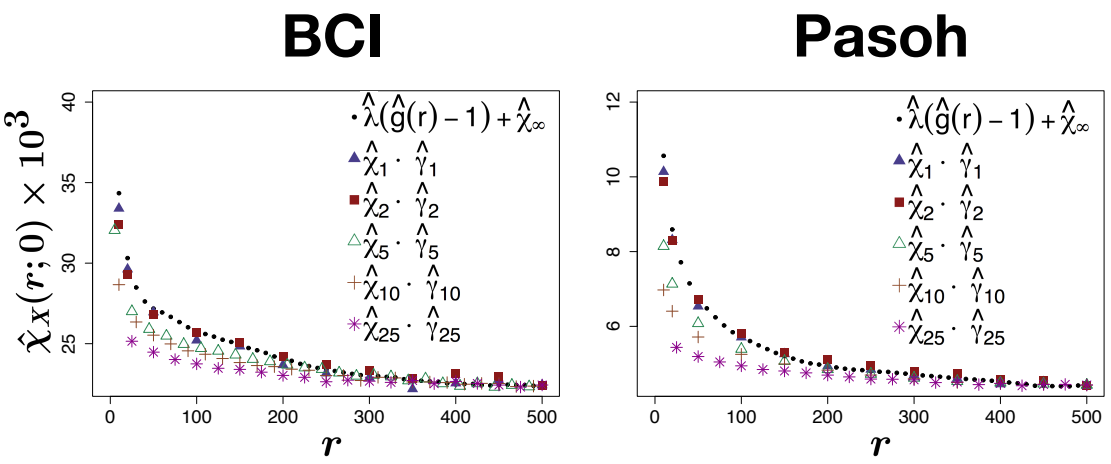


Figure 4.8: Similarity decay function for BCI and Pasoh forests. Similarity index for BCI and Pasoh datasets computed via eq. (4.18) (black dots) compared with the estimator (4.20) for different cell sizes a (coloured symbols). The agreement between the two estimators increases as the cell size decreases, in accordance with the theory.

4.3.3 Comparing estimated and theoretical similarity functions

Let us now focus on modelling BCI species through the three NS point processes described in Section 4.2.4: exponential, Gaussian and Cauchy cluster processes.
For each species s, the first step is to estimate the set of parameters $(\rho_{\boldsymbol{X}_s}, \mu_{\boldsymbol{X}_s}, \gamma_{\boldsymbol{X}_s})$ which best describe its pattern. We do this by the method of minimum contrast (see Chapter 1, Section 1.6.2), which relies on the minimisation of the following integral

$$\int_0^{r\max} \left(\hat{g}_{\boldsymbol{X}_s}(r)^{\frac{1}{4}} - g_{\boldsymbol{X}_s}(r)^{\frac{1}{4}}\right)^2 \mathrm{d}r.$$

where $\hat{g}_{\boldsymbol{X}_s}$ is the empirical pair correlation function estimated according to eq. (4.16) and $r_{\max}$ is the maximum considered distance, which we set equal to 500 meters in our estimation. In figure 4.9 we insert the values of parameters fitted according to the three models for BCI species.

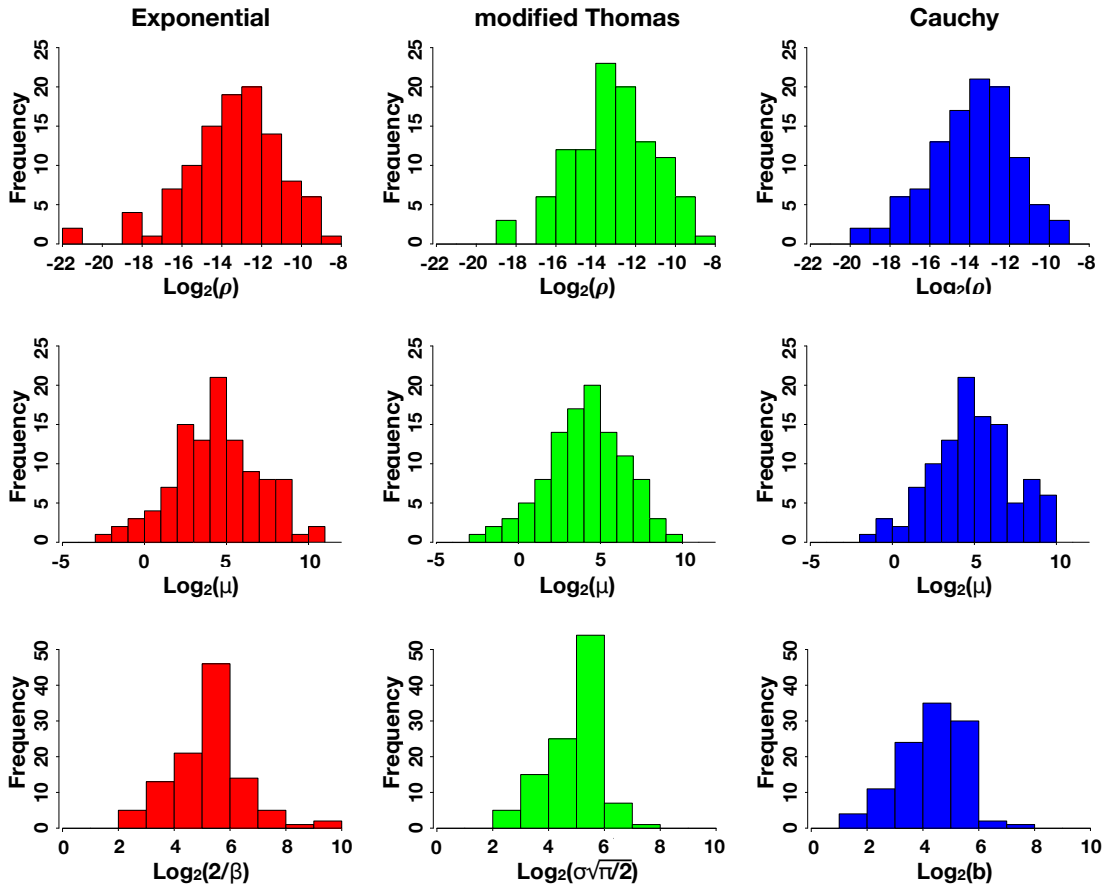


Figure 4.9: Model parameters for BCI forests. Fitted model parameters via the method of minimum contrast for BCI species according to the exponential process (left column), modified Thomas process (centre column) and Cauchy process (right column). From top to bottom: frequency histogram of density of cluster $\rho_{\boldsymbol{X}_s}$, number of offspring per parent $\mu_{\boldsymbol{X}_s}$ and clustering parameters $\beta_{\boldsymbol{X}_s}$, $\sigma_{\boldsymbol{X}_s}$ and $b_{\boldsymbol{X}_s}$, respectively. We remark that for the first two models the mean radius of cluster is well-defined and equals $2/\beta_{\boldsymbol{X}_s}$ and $\sigma_{\boldsymbol{X}_s}\sqrt{\pi/2}$ (whose histograms are shown above), while for the Cauchy process such quantity is not defined (thus above we show the histogram of $\log_2 b_{\boldsymbol{X}_s}$). Similar histograms have been obtained for Pasoh forest.

By inserting the fitted parameters of each model into the corresponding formulation of eq. (4.23), we can get the predicted theoretical similarity curve for the BCI and Pasoh forests (see figure 4.10). We find a good agreement between the models' predictions and the empirical data for all cluster types. Nevertheless, the Cauchy cluster results to give the best fitting for most of the species' pair correlation function, as we can see from figure 4.11, where we have compared the goodness of fit between the model according to the χ^2 test.

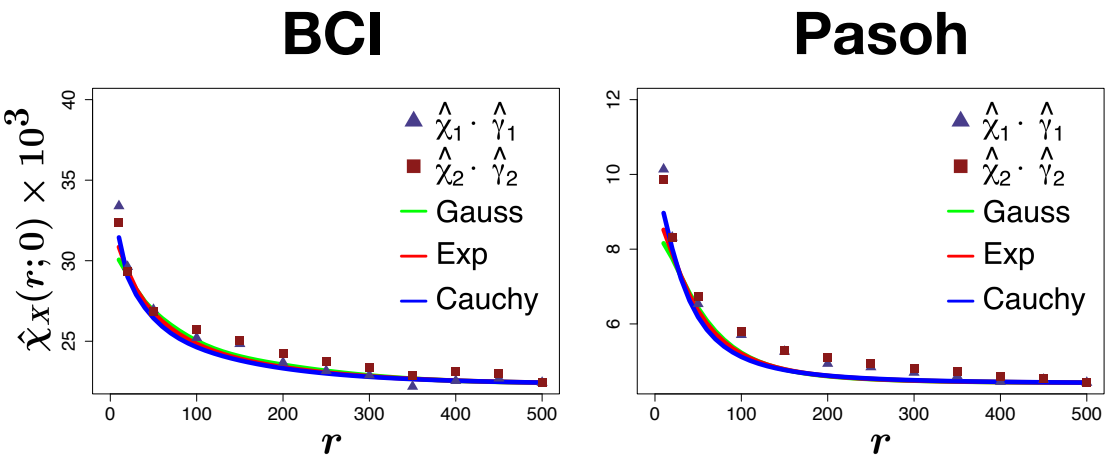


Figure 4.10: Sørensen index for BCI. Comparison between the empirical distance-dependent Sørensen index $\hat{\chi}_{\boldsymbol{X}}(r;0) = \hat{\chi}_{\boldsymbol{X}}(r;a)\hat{\gamma}_{\boldsymbol{X}}(a)$ (where $\hat{\chi}_{\boldsymbol{X}}(r;a)$ has been computed via (4.13) and $\hat{\gamma}_{\boldsymbol{X}}(a) = \hat{\chi}_{\boldsymbol{X},\infty}/\hat{\chi}_{\boldsymbol{X},r_{\max}}(a)$) for square cells of area 1 (triangles) and 4 (squares) square meters and the exact functional form of $\chi_{\boldsymbol{X}}(r)$ by the three cluster models using eq. (4.23). We find a good agreement between model prediction and empirical data for all cluster type.

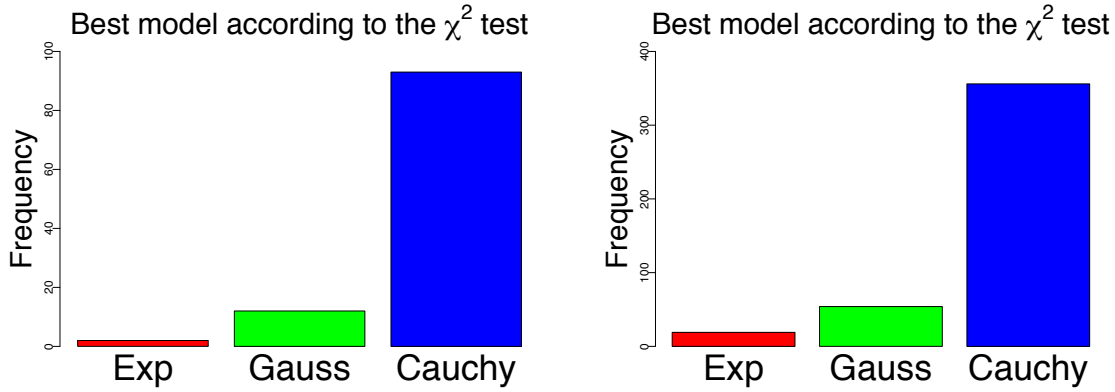


Figure 4.11: Comparison between models for BCI and Pasoh forests. We compute the goodness of each model in fitting the pair correlation function of real species through the χ^2 test. For each species, we then select the best fitting model and we group data into frequency histograms. We find that for the overwhelming majority of the species in both the BCI and Pasoh databases, the Cauchy model results to better fit the empirical pair correlation function, followed by the modified Thomas and the exponential one.

4.4 Impact of the stationarity and isotropy hypotheses

Our analytical similarity decay function (4.8) has been derived using the point process framework and it is based on the following assumptions: 1) we compare the

species' composition of two regions of *infinitesimal size* and 2) the point process is *stationary and isotropic* (i.e. translation and rotation invariant) and *homogeneous*, i.e. the *intensity* $\lambda_{\boldsymbol{X}}$, which is the density of individuals, is constant (see Section 4.1.1).
In the previous sections, we have already discussed the implications of using a statistical estimator based on finite-size cells and how to compare it with theoretical formulae developed for infinitesimal sizes. Here we would like to discuss the impact of hypothesis 2).
To investigate this, we generate three artificial forests as follows: for each BCI and Pasoh species having more than 200 and 100 individuals, respectively, we generate a Neyman-Scott *homogeneous, stationary and isotropic* cluster process within the 50ha plot having the same number of individuals as the original species and according to the three different cluster types parameters (see Chapter 1, Section 1.5.2). We then compute the empirical Sørensen similarity decay function $\hat{\chi}_{\boldsymbol{X}}(r;0) = \hat{\chi}_{\boldsymbol{X}}(r;a)\hat{\gamma}_{\boldsymbol{X}}(a)$, with $a = 2$ and 25 for the superposed process and compare it with the theoretical one (see eq. (4.23)). Results are displayed in figure 4.12 and figure 4.13.
For the forests generated according to the three cluster model hypothesis 2) holds. Accordingly, the empirical similarity decay function of the artificial forest with a cell area of 25m^2 (left column), does not display the linear descent in the middle part, which is on the contrary present in the empirical curves of the BCI and Pasoh data. At smaller cell area scale (4m^2, right column), we found mixed results: the linear descent is present in the Gaussian dispersal kernel, is absent in the exponential one and is very hollow in the Cauchy kernel. This is a consequence of the different average cluster radius of the three models, which is infinite for the Cauchy cluster.
To give a hint for a possible explanation for the linear descent phenomenon, in figure 4.14 we show the pattern of two different species of the BCI forest, one showing an homogeneous behaviour (left column, first panel) and one affected by anisotropy and non-stationarity (right column, first panel). For both species, bottom panels show the analytical curves of the pair correlation function (coloured lines, see eq. (1.26)), with parameters obtained by the minimum contrast method against the empirical pair correlation estimated via eq. (4.17). In contrast to the first species, where all three models are able to capture the empirical curve, for the second species the fit is much worse and, more importantly, the curve shows a hollow part in the middle due to overdispersion at that scale. Since the pair correlation function of the superposed process is the sum weighted by the abundances of the pair correlation function of the different species, we may think that the linear descent of the curve in its middle part is an average behaviour due to the fact that for some species the patterns are not homogeneous nor translation or rotation invariant. We finally test the ability of the three studied cluster models to capture another important macro-pattern in ecological theory, which is the species-area relationship (SAR), giving the mean number of species as a function of the surveyed area. The modified Thomas process has already been noticed to be able to reproduce it with high fidelity (Plotkin, Potts et al., 2000; Morlon et al., 2008). Here we wish to check whether also the exponential and the Cauchy processes have a similar performance. In order to estimate the SAR from our data, we divide the 50ha plot into cells of different areas a and we average among the number of species falling within each of

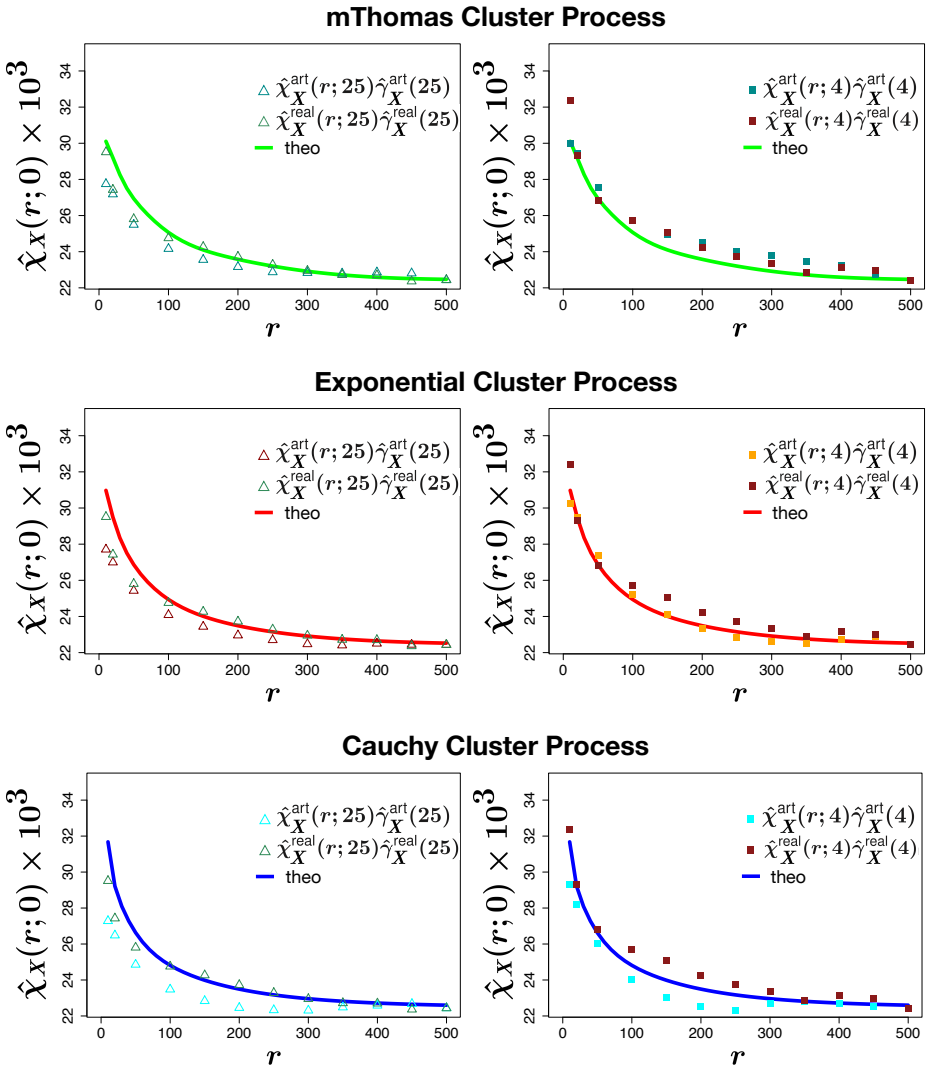


Figure 4.12: Empirical Sørensen's similarity decay function for BCI artificial forests. We generate three artificial forests as follows: for each species of BCI having more than 200 individuals, we generate a Neyman-Scott process (modified Thomas in the top panels, exponential in the middle panels and Cauchy in the bottom panels) within the 50 ha plot having the same number of individuals as the original species. We then compute the empirical Similarity index for the new generated superposed process $\hat{\chi}_{\boldsymbol{X}}^{\text{art}}(r;0) = \hat{\chi}_{\boldsymbol{X}}^{\text{art}}(r;a)\hat{\gamma}_{\boldsymbol{X}}(a)^{\text{art}}$ (the superscript "art" stands for artificial forest) and compare it with the theoretical one (see eq. (4.23)) and the empirical one for the real BCI $\hat{\chi}_{\boldsymbol{X}}^{\text{real}}(r;0) = \hat{\chi}_{\boldsymbol{X}}^{\text{real}}(r;a)\hat{\gamma}_{\boldsymbol{X}}(a)^{\text{real}}$ (the superscript "real" stands for real forest), when using 25 square meters cell area (left column) and 4 square meters cell area (right column).

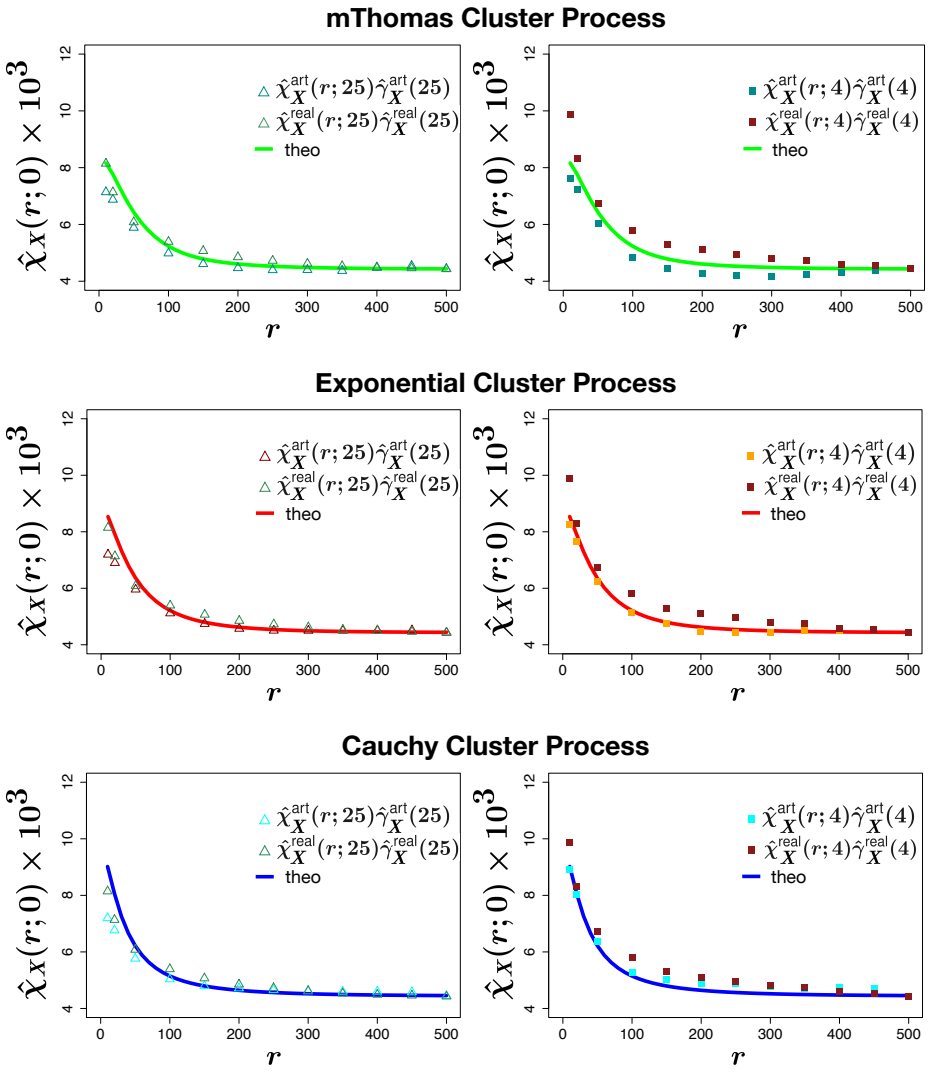


Figure 4.13: Empirical Sørensen's similarity decay function for Pasoh artificial forests. We generate three artificial forests as follows: for each species of Pasoh having more than 100 individuals, we generate a Neyman-Scott process (modified Thomas in the top panels, exponential in the middle panels and Cauchy in the bottom panels) within the 50 ha plot having the same number of individuals as the original species. We then compute the empirical similarity index for the new generated superposed process $\hat{\chi}_{\boldsymbol{X}}^{\text{art}}(r;0) = \hat{\chi}_{\boldsymbol{X}}^{\text{art}}(r;a)\hat{\gamma}_{\boldsymbol{X}}(a)^{\text{art}}$ (the superscript "art" stands for artificial forest) and compare it with the theoretical one (see eq. (4.23)) and the empirical one for the real Pasoh $\hat{\chi}_{\boldsymbol{X}}^{\text{real}}(r;0) = \hat{\chi}_{\boldsymbol{X}}^{\text{real}}(r;a)\hat{\gamma}_{\boldsymbol{X}}(a)^{\text{real}}$ (the superscript "real" stands for real forest), when using 25 square meters cell area (left column) and 4 square meters cell area (right column).

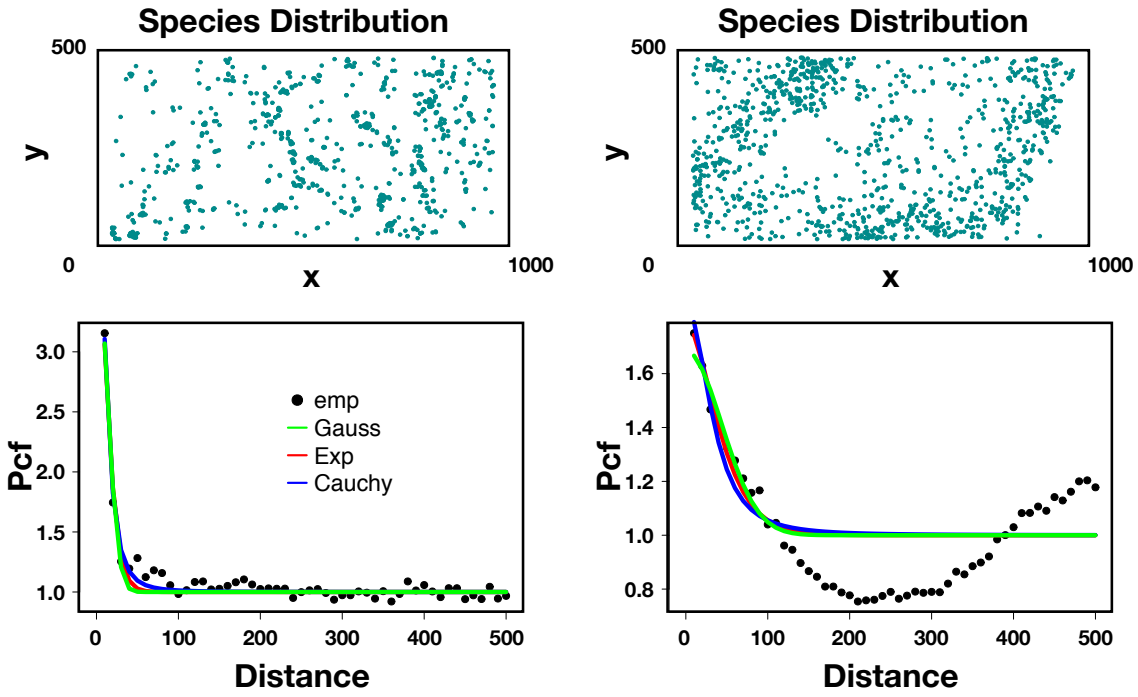


Figure 4.14: Pair Correlation Function for two species. On the top: two species distributions within the 1000x500 surveyed area of the BCI. We can notice that the species on the right panel shows a non-isotropic nor homogeneous pattern, resulting in contrast with our model hypothesis. Such behaviour does not characterise the species on the left. On the bottom: empirical pair correlation function computed via eq. (4.16) (black dots) and the analytical one (solid lines) computed through eq. (1.26) with parameters fitted by minimum contrast method.

them. In figure 4.15 we compare the SAR obtained for real BCI and Pasoh data with those of the artificial forests generated according to the exponential, Gaussian and Cauchy cluster processes. Coloured points and bars represent mean and standard errors for the empirical SARs, respectively. For each considered area a, we find a good agreement between both the real forest and those generated with our models.

4.5 A synopsis on similarity decay functions

In this section we give an account of the various approaches to the problem of describing the decay in similarity with the distance that we have found in literature. In their pioneering paper (Nekola and White, 1999), Nekola and White studied North America boreal spruce forests using data from 34 nine hectare plots distributed from Newfoundland to Alaska. The similarity was computed using Jaccard's index and species were subdivided in homogeneous classes in terms of growth or dispersal form. Linear regression was used to calculate the decay rate of the logarithm of the similarity against linear distance. This implies an *exponential* rate of the distance decay, with different exponents for various classes.
In Hubbell's neutral theory of ecology (Hubbell, 2001b) the similarity decay is also

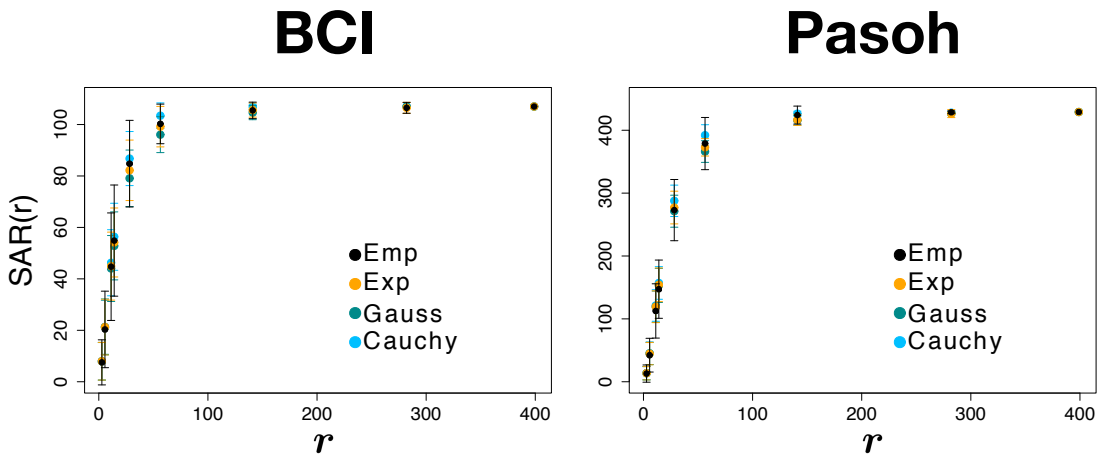


Figure 4.15: Species-area relationship for BCI and Pasoh. Comparison between the empirical SAR of the real BCI and Pasoh forests and the ones obtained for the three artificial forests generated as described in figures 4.12 and 4.13. At each scale r, the 1000×500 window plot has been divided into C_r cells of side $\sqrt{r^2\pi}$. Black points and bars are mean value and three times the standard error, respectively, of the number of different species falling within each cell. Coloured points and bars refer to the empirical SAR computed for three artificial forests with the same species-abundance distribution as for the BCI/Pasoh but generated according to the exponential, modified Thomas and Cauchy processes. We find a good agreement between artificial and real forests, meaning that all the models are able to capture this macro-ecological pattern.

considered for an artificial community. The form of the decay function is a *compound exponential*, i.e. a linear combination of exponentials with different exponents. The steeper decay rate is the contribution due to the rare species, which are also confined to restricted areas, while the tail of the curve has a lower decay due to the abundant and widespread species with lower turnover. The overall decay is steeper and the overall similarity is lower if a smaller grain size (i.e. plot size) is used. This latter aspect can be discussed only qualitatively within the theory. However, the dependence of the decay curve on the size of the plot is an unavoidable consequence of the very definition of similarity, which is area-dependent. This renders more difficult the comparison of different graphs realised with diverse grain sizes and extents, and its potential impact on the conclusions drawn from these data have been recalled in various works (Steinbauer et al., 2012; Palmer and White, 1994).
In Morlon et al., 2008, an analytical model for the similarity decay function is presented, extending a spatially implicit model contained in Plotkin, Chave et al., 2002. This spatially explicit model considers two small regions A and B of area a at distance r and is based on the conditional probability

$$P(n_s(B) \geq 1 | n_s(A) \geq 1, r),$$

which gives a distance dependent similarity index. In (Morlon et al., 2008), the probabilities are computed for a specific spatial point process, a modified Thomas cluster process of parameters $(\rho_{\boldsymbol{X}_s}, \mu_{\boldsymbol{X}_s}, \sigma_{\boldsymbol{X}_s})$. The resulting formula of the Sørensen similarity, for a discrete number of species S, is the following (see Morlon et al., 2008, Supporting Information F2)

$$\text{SØR}(r;a) = \frac{\sum_{s=1}^{S} \Big(1 - \exp(-ac_{\boldsymbol{X}_s}(a)n_s)\Big)\Big(1 - \exp(-ac_{\boldsymbol{X}_s}(a)n_s g_{\boldsymbol{X}_s}(r))\Big)}{\sum_{s=1}^{S} \Big(1 - \exp(-ac_{\boldsymbol{X}_s}(a)n_s)\Big)}, \quad (4.24)$$

where $g_{\boldsymbol{X}_s}(r)$ is the pair correlation function of the modified Thomas process

$$g_{\boldsymbol{X}_s}(r) = 1 + \frac{e^{-\frac{r^2}{4\sigma^2_{\boldsymbol{X}_s}}}}{4\pi\rho_{\boldsymbol{X}_s}\sigma^2_{\boldsymbol{X}_s}}.$$

The term $c_{\boldsymbol{X}_s}(a)$ in eq. (4.24) is an area-dependent correcting factor for the clustering of individuals having dimension $area^{-1}$:

$$c_{\boldsymbol{X}_s}(A) = \frac{1}{\mu_{\boldsymbol{X}_s} A} \int_W \Big(1 - e - \int_A d_{\gamma_{\boldsymbol{X}_s}}(u - v)du\Big)dv,$$

where $d_{\gamma_{\boldsymbol{X}_s}}$ is the Gaussian dispersal kernel function (see eq. (1.36)). For randomly distributed individuals $c_{\boldsymbol{X}_s}(a) = 1$, while $c_{\boldsymbol{X}_s}(a)$ tends to zero for highly clustered patterns. Note that, for a random pattern, also $g_{\boldsymbol{X}_s}(r) = 1$ for every species s. Therefore, in this case, dividing eq. (4.24) by the area a, we obtain eq. (4.11). For the continuous case the formula is more involved. However, when the term $ac_{\boldsymbol{X}_s}(a)n_s$ is small, keeping only the leading term in the Taylor expansion of the exponential as done before, we find that the Sørensen similarity for very small plots is (this derivation is ours)

$$\text{SØR}(r;a) = a\frac{\sum_{s=1}^{S} g_{\boldsymbol{X}_s}(r)c^2_{\boldsymbol{X}_s}(a)n_s^2}{\sum_{s=1}^{S} c_{\boldsymbol{X}_s}(a)n_s(a)},$$

which again, under the complete spatial random placement hypothesis, reduces to eq. (4.12) when divided by a. There are probably other approaches to the problem of determining the form of the similarity decay function of which we are not aware of, but, as far as we know, it seems to us that formulating the theory for the similarity between plots of finite area produces very complex formulae in which the dependence on the area is not easy to investigate. We are convinced that the formulation presented in this chapter based on infinitesimal plots offers a clearer picture of the problem.

Part II

From Local to Global: The Problem of Upscaling

> *"Wonderful!" I cried in astonishment. "It is incredible that a man can count, at a glance, all the branches in a tree, all the flowers in a garden. That skill can bring immense riches to anyone."*

Malba Tahan, *The Man Who Counted*

5

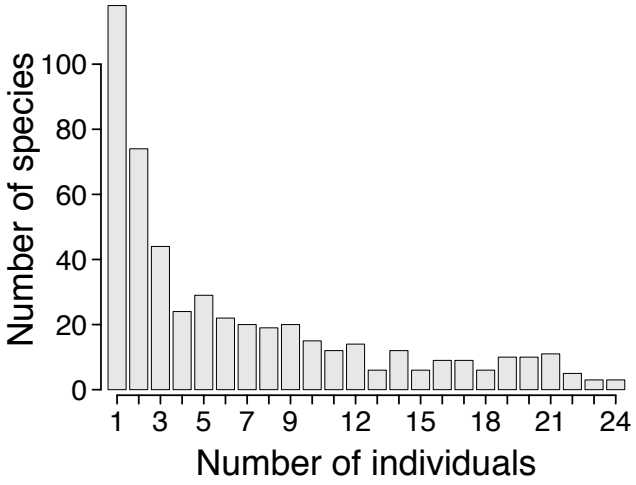
Upscaling Species Richness and Abundances

5.1 The problem of inferring biodiversity

Up to now we have only worked in our relatively small surveyed sample of rainforest. But what can we say of the rest of it? Is there a way to extrapolate information on bigger scales?

The problem of inferring total biodiversity when only scattered samples are observed is a long-story problem. In the early 1940s, the British chemist and naturalist Alexander Steven Corbet spent two years in Malaya to trap butterflies (Corbet, 1941). For every species he saw, he noted down how many individuals of that species he trapped (see figure 5.1). When Corbet returned to England, he showed the table to its colleague Ronald Aylmer Fisher and asked him how many new species he would trap if he returned to Malaya for another couple of years. The father of statistics was only the first mathematician to tackle the problem of species estimation, which since then has found large applications in different scientific fields, from ecology (Colwell and Coddington, 1994; Bunge and Fitzpatrick, 1993; Chao and Bunge, 2002) to bioscience (Locey and Lennon, 2016; Hughes et al., 2001; Ionita-Laza et al., 2009), leading to the development of a myriad of estimators (Good and Toulmin, 1956; Orlitsky et al., 2016; Chao and Chiu, 2016; Kunin et al., 2017).

In this chapter we will focus on the problem of upscaling tree species richness and abundances when only scattered samples are available (see figure 5.2). Tropical forests have long been recognised as one of the largest pools of biodiversity (Crowther et al., 2015). In fact, more than two-fifths of the number of worldwide trees can be found either in tropical or sub-tropical forests, but only $\approx 0.000067\%$ of species identities are known. Global patterns of empirical abundance distributions show that tropical forests vary in their absolute number of species but display surprising similarities in the distribution of individuals across species (Volkov, Banavar,



N	S
1	118
2	74
3	44
4	24
⋮	⋮
22	5
23	3
24	3

Figure 5.1 & Table 5.1: Species-abundance distribution of the butterflies trapped by Corbet in Malaya.

He et al., 2005; McGill et al., 2007; Suweis et al., 2012). For practical reasons, biodiversity is typically measured or monitored at fine spatial scales. However, important drivers of ecological change tend to act at large scales (Alonso et al., 2008; Bertuzzo, Carrara et al., 2016). Conservation issues, for example, apply to diversity at global, national or regional scales. Extrapolating species richness from the local to the whole forest scale is not straightforward. Indeed, a vast number of different biodiversity estimators have been developed under different statistical sampling frameworks (Bunge and Fitzpatrick, 1993; Brose et al., 2003; Mao and Colwell, 2005; Wang and Lindsay, 2005; Bunge, Woodard et al., 2012), but most of them have been designed for local/regional-scale extrapolations, and they tend to be sensitive to the spatial distribution of trees (Plotkin, Potts et al., 2000; Carrara et al., 2012; Azaele, Suweis et al., 2016), sample coverage and sampling methods (Chao, Colwell et al., 2009).
Here we introduce an analytical framework that provides robust and accurate estimates of species richness and abundances in biodiversity-rich ecosystems, as confirmed by tests performed on both in silico-generated and real forests (Tovo, Suweis et al., 2017). Our theoretical method confirms that the vast majority of species in our 15 analysed forests are rare or hyper-rare and suggests that this may be a signature of critical-like behaviour, characteristic of species-rich ecosystems, which can provide a buffer against extinction.

5.2 Negative binomial SAD upscaling method

In this section, we describe in detail our theoretical framework developed for upscaling biodiversity (Tovo, Suweis et al., 2017).
A common statistical tool used to describe the commonness and rarity of species in an ecological community is the species-abundance distribution (SAD), which is a list of species present within a region along with the number of individuals per species

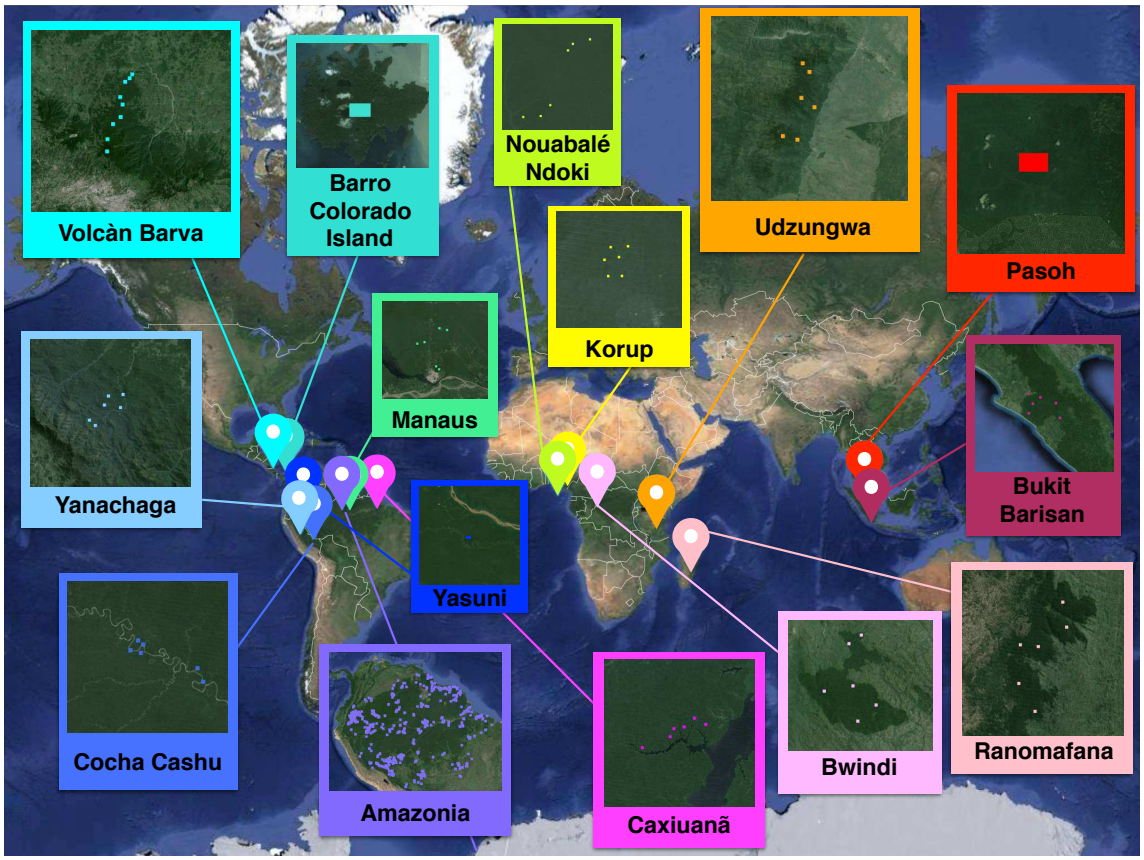


Figure 5.2: The challenge of estimating global tropical species richness. A map depicting the 15 forests in our dataset for which the coordinates of each subplot (squares) are known. Our goal is to deduce the species richness and abundances of each entire forest on the basis of the very limited knowledge coming from the scattered samples in the marked dots.

(MacArthur, 1960; Magurran, 2013). Typically, the SAD is measured at local scales (e.g., in quadrats or transects), in which the identities of all the individuals living in the area are known. The sampled SAD can be fit to a given functional form at that scale. However, that form may change at different spatial scales (see figure 5.3), thus hindering analytical treatment (Azaele, Maritan et al., 2015).
When upscaling, we are interested in the SAD and in the total number of species, S, at the scale of the whole forest area A. We will denote as $P(n|1)$ the probability that a species has exactly n individuals – also known as relative species abundance (RSA) in theoretical ecology – at the whole forest scale (here 1 refers to the whole forest). Note that $P(n|1)$ should be defined only for $n \geq 1$, because S is the total number of species actually present in the forest, thus each having at least one individual.
Here the SAD is postulated to have a negative binomial functional form (NB) (He and Hubbell, 2003; He and Gaston, 2003), $\mathcal{P}(n|r,\xi)$ with parameters (r,ξ) (r is known as the clustering coefficient):

$$P(n|1) = c(r,\xi)\mathcal{P}(n|r,\xi) \tag{5.1}$$

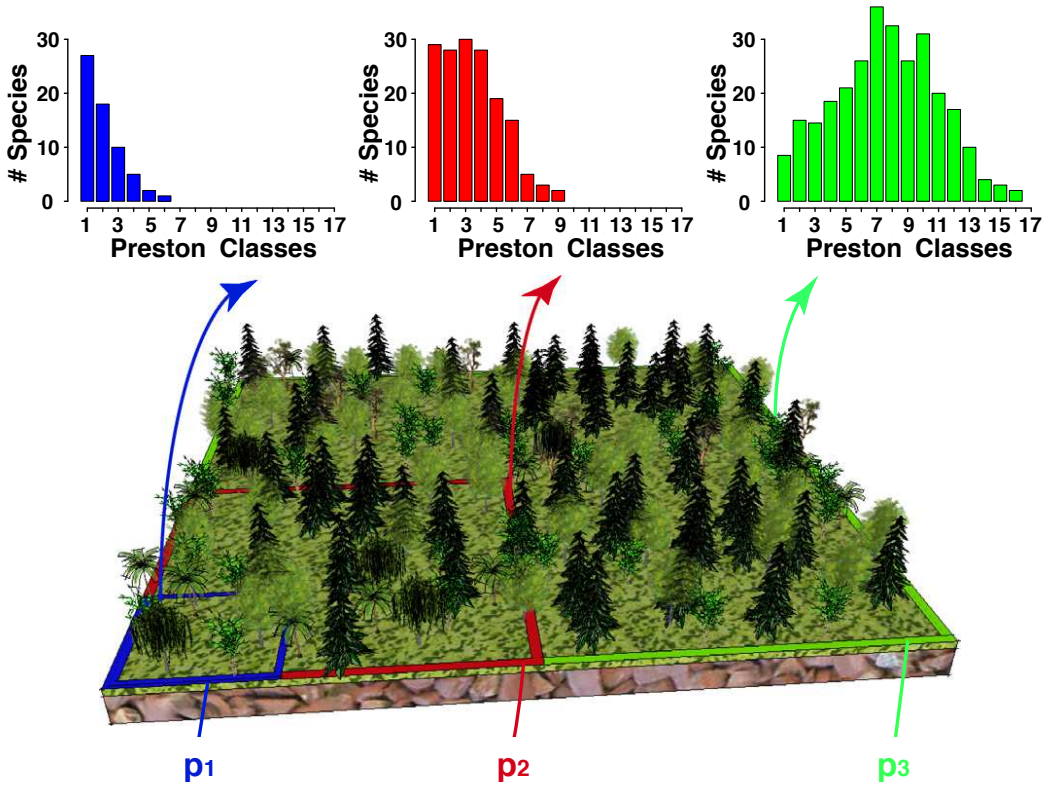


Figure 5.3: Species-abundance distribution at different spatial scales for Barro Colorado Island, in Panama. The form of the SAD changes with the scale. It displays a monotonic decreasing behaviour when only a small portion of the whole forest is sampled, while it exhibits an internal mode when larger areas are surveyed.

with

$$\mathcal{P}(n|r,\xi) = \binom{n+r-1}{n}\xi^n(1-\xi)^r, \qquad c(r,\xi) = \frac{1}{1-(1-\xi)^r},$$

where $c(r,\xi)$ is the normalisation constant, determined by imposing

$$\sum_{n=1}^{\infty} P(n|1) = 1.$$

In Section 5.2 the sum starts from $n = 1$ because we are taking into account only species with non-zero abundance, as already mentioned. Note that the classic negative binomial $\mathcal{P}(n|r,\xi)$ is instead normalised for $n \geq 0$.
Let us now consider a sub-sample of area a of the whole forest and define $p = a/A$ the scale of the sample, that is the fraction of the sampled forest. The first step is to compute the RSA of the sub-sample.
We will assume that this latter is not affected by spatial correlations due to both interspecific and intraspecific interactions. This hypothesis is well satisfied as we will show in Section 5.5.2 using *in silico* generated forests with various degrees of spatial correlations. Under this hypothesis, the conditional probability that a species has k individuals in the smaller area, $a = pA$, given that it has total abundance n in the

whole forest of area A is given by the binomial distribution

$$\mathcal{P}_{binom}(k|n,p) = \begin{cases} \binom{n}{k} p^k (1-p)^{n-k} & k = 0, \dots, n \\ 0 & k > n. \end{cases}$$

Indeed, in absence of spatial correlations, the probability that one of the species's individuals will fall within a is exactly p.
We now prove the following key result.

Proposition 5.1 (Self-Similarity Property for the NB distribution)**.** *Let $P(n|1) = c(r,\xi)\mathcal{P}(n|r,\xi)$ be the RSA at the whole forest scale and let us denote with $\mathcal{P}(k|n,p)$, sampling probability at a sub-scale $p \in (0,1)$, i.e. the conditional probability that a species has abundance k in the sample p, given that it has n individuals in A.*
If $\mathcal{P}(k|n,p) = \mathcal{P}_{binom}(k|n,p)$ is binomially distributed, then the RSA at the sample scale p, $\mathcal{P}_{sub}(k|p)$, is again a negative binomial, for $k \geq 1$, with rescaled parameter ξ and the same r:

$$\mathcal{P}_{sub}(k|p) = \begin{cases} c(r,\xi) \cdot \mathcal{P}(k|r,\hat{\xi}_p) & k \geq 1 \\ 1 - c(r,\xi)/c(r,\hat{\xi}_p) & k = 0 \end{cases}$$

with

$$\hat{\xi}_p = \frac{p\xi}{1-\xi(1-p)}. \tag{5.2}$$

Proof. The probability, $\mathcal{P}_{sub}(k|p)$, of finding a species with a population of k individuals, $k \geq 0$, in the sub-plot of area $a = pA$ is

$$\begin{aligned} k \geq 1: \quad \mathcal{P}_{sub}(k|p) &= \sum_{n \geq k} \mathcal{P}_{binom}(k|n,p) P(n|1) \\ &= \sum_{n \geq k} \binom{n}{k} p^k (1-p)^{n-k} \cdot c(r,\xi) \binom{n+r-1}{n} \xi^n (1-\xi)^r \\ &= c(r,\xi) \binom{k+r-1}{k} \Big(\frac{p\xi}{1-\xi(1-p)}\Big)^k \Big(\frac{1-\xi}{1-\xi(1-p)}\Big)^r \\ &= c(r,\xi) \binom{k+r-1}{k} \hat{\xi}_p^k (1-\hat{\xi}_p)^r = c(r,\xi) \cdot \mathcal{P}(k|r,\hat{\xi}_p), \end{aligned} \tag{5.3}$$

$$\begin{aligned} k = 0: \quad \mathcal{P}_{sub}(0|p) &= 1 - \sum_{k \geq 1} \mathcal{P}_{sub}(k|p) \\ &= 1 - \sum_{k=1}^{\infty} c(r,\xi) \cdot \binom{k+r-1}{k} \hat{\xi}_p^k (1-\hat{\xi}_p)^r \\ &= 1 - c(r,\xi) \cdot \sum_{k=1}^{\infty} \mathcal{P}(k|r,\hat{\xi}_p) = 1 - \frac{c(r,\xi)}{c(r,\hat{\xi}_p)}, \end{aligned} \tag{5.4}$$

where we have inserted the explicit relation eq. (5.2) for $\hat{\xi}_p$ in the penultimate equality of eq. (5.3). □

Recall that our method uses only the information we can infer from a sub-sample at some scale p^*. Therefore, we only have information on the abundances of the $S^* \leq S$ species present in the surveyed area. By denoting the number of species of abundance k at scale p^* by $S^*(k)$, we get, for $k \geq 1$

$$\begin{aligned}\frac{S^*(k)}{S^*} \equiv P(k|p^*) &= \frac{\mathcal{P}_{sub}(k|p^*)}{\sum_{k'\geq 1}\mathcal{P}_{sub}(k'|p^*)} \\ &= \frac{\mathcal{P}(k|r,\hat{\xi}_{p^*})}{\sum_{k'\geq 1}\mathcal{P}(k'|r,\hat{\xi}_{p^*})} \\ &= c(r,\hat{\xi}_{p^*})\cdot\mathcal{P}(k|r,\hat{\xi}_{p^*}) \end{aligned} \tag{5.5}$$

which, due to eq. (5.1), is a NB normalised for $k \geq 1$, whereas $\mathcal{P}(k|r,\hat{\xi}_{p^*})$ is normalised for $k \geq 0$. We have therefore obtained the key result that starting with a NB distribution for the RSA at the global scale, the RSA at smaller scales is also distributed according to a negative binomial with the same clustering coefficient r and a rescaled parameter $\hat{\xi}_{p^*}$ depending on both ξ and p^*. A RSA with the property of having the same functional form at different scales is said to be *form-invariant*. By fitting the RSA of the data at the sampling scale p^* we can thus find both the parameters r and $\hat{\xi}_{p^*}$ and, by inverting eq. (5.2), we can get the value of ξ:

$$\xi = \frac{\hat{\xi}_{p^*}}{p^* + \hat{\xi}_{p^*}(1-p^*)}. \tag{5.6}$$

Using eq. (5.2) to eliminate ξ from the last equation, one gets the following relation for the parameter ξ at the two scales p and p^*

$$\hat{\xi}_p = \frac{p\hat{\xi}_{p^*}}{p^* + \hat{\xi}_{p^*}(p-p^*)} \equiv U(p,p^*|\hat{\xi}_{p^*}). \tag{5.7}$$

from which, of course, one can recover both eqs. (5.2) and (5.6) where one has to use that $\xi \equiv \hat{\xi}_{p=1}$.
We now wish to determine the relationship between the total number of species at the whole scale $p = 1$, S, and the total number of species surveyed at a local scale p, S_p. For the sampling scale p^*, in the following, we will use the notation $S^* \equiv S_{p^*}$. Note that

$$\mathcal{P}_{sub}(k=0|p^*) = (S-S^*)/S \tag{5.8}$$

$$\mathcal{P}_{sub}(k|p^*) = S^*(k)/S. \tag{5.9}$$

Using eq. (5.4), the total number of species in the whole forest, in terms of the data on the surveyed sub-plot is given by

$$\begin{aligned} S &= \frac{S^*}{1-\mathcal{P}_{sub}(k=0|p^*)} \\ &= S^*\frac{1-(1-\xi)^r}{1-(1-\hat{\xi}_{p^*})^r}, \end{aligned} \tag{5.10}$$

where ξ is given by eq. (5.6).
We choose the negative binomial distribution in eq. (5.1) as the SAD. Apart from its simplicity and versatility, we choose this form for our analysis for four reasons.

1. The negative binomial, depending on its parameters displays both log-series like behaviour or an interior mode (see Section 5.4 and figure 5.9), i.e. it can accommodate different SAD shapes. Therefore we can use the same SAD function to reproduce different ecosystems' SAD, as those we observed in our dataset. Even more generally, if more complex SAD are encountered, a virtually perfect fit of data is still possible by using linear combinations of negative binomials – a case for which our framework still works.

2. The NB distribution arises naturally as the steady-state SAD of an ecosystem that undergoes simple birth and death dynamics, with an effective birth rate accounting for the effects of immigration events and/or intraspecific interactions (Volkov, Banavar, He et al., 2005; Azaele, Suweis et al., 2016), and under the neutral hypothesis that individuals are demographically identical (see appendix C for a brief review on Hubbell's neutral theory).

3. In the limit of $r \to 0$, the NB in becomes the well-known Fisher's log-series (LS), which has been widely used to describe the patterns of abundance in ecological communities (White et al., 2012; Slik et al., 2015; Ter Steege, Sabatier et al., 2017; Harte, 2011; Harte, Smith et al., 2009; Kitzes and Harte, 2015). Of course, because of the flexibility of choosing r to be non-zero, the NB distribution is always more versatile than the LS. The SAD, especially at large scales or with increasing sampling effort (Chisholm, 2007), often displays an interior mode that cannot be captured by a LS distribution (see figure 5.3).

4. Finally and importantly, if one chooses two contiguous patches with NB as SADs characterised by the same parameters r and $\xi \equiv \xi_{1/2}$ and combines the two, remarkably, the resulting larger patch is also characterised by a NB distribution with the same scale-invariant value of r and a new scale-dependent parameter, ξ, given by the analytical expression in eq. (5.6) above with $p = 1/2$. This special form-invariant property of the NB distribution, albeit with a scale-dependent parameter, makes it particularly well suited for our extrapolation studies.

In the following sections we will explore these properties in more details.

5.2.1 Flexibility of negative binomial distribution in describing empirical SADs

Our method can be generalised to a linear combination of two NBs with the same parameter ξ and different parameters r_1 and r_2. For example, this result is particularly useful when dealing with data which present unusual behaviours which cannot be captured by a single NB distribution (see figure 5.4). Indeed, one finds that in this case the predicted biodiversity is given by

$$S = S^* \frac{\lambda[1-(1-\xi)^{r_1}] + (1-\lambda)[1-(1-\xi)^{r_2}]}{\lambda[1-(1-\hat{\xi}_{p^*})^{r_1}] + (1-\lambda)[1-(1-\hat{\xi}_{p^*})^{r_2}]},$$

where $\lambda \in (0,1)$ is the coefficient of the linear combination of the two negative binomials. The parameter ξ is given by eq. (5.6) whereas the parameters $r_1,\ r_2,\ \lambda$ and $\hat{\xi}_{p^*}$ are obtained by the best fit of the RSA of the surveyed area at scale p^* using the linear combination

$$\lambda c(r_1, \hat{\xi}_{p^*}) \cdot \mathcal{P}(k|r_1, \hat{\xi}_{p^*}) + (1-\lambda) c(r_2, \hat{\xi}_{p^*}) \cdot \mathcal{P}(k|r_2, \hat{\xi}_{p^*}). \tag{5.11}$$

In principle, one could use a generic combination of an arbitrary number $m \in \mathbb{N}$

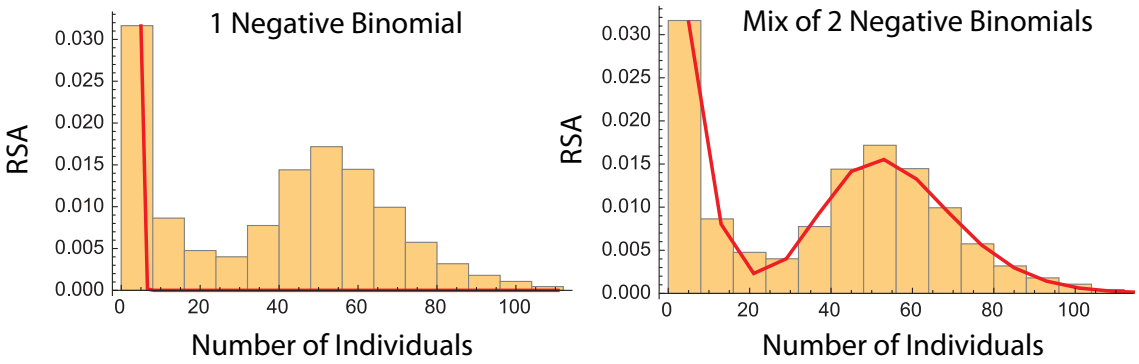


Figure 5.4: Fit of a RSA consisting of a mixture of a log-series and a log-normal distributions. On the left the RSA is fitted through a negative binomial, which cannot capture the unusual behaviour of the distribution. On the right is shown an improved fit with a combination of two negative binomials with the same parameter ξ and different clustering coefficients like in eq. (5.12).

of negative binomials to better fit the distribution (see figure 5.5). In this case one has the upscaling formula

$$S = S^* \frac{\sum_{i=1}^{m} \lambda_i [1-(1-\xi)^{r_i}]}{\sum_{i=1}^{m} \lambda_i [1-(1-\hat{\xi}_{p^*})^{r_i}]},$$

where $\lambda_i \in (0,1),\ \sum_{i=1}^{m} \lambda_i = 1$ are the coefficients of the linear combination of m negative binomials. Again, the parameter ξ is given by eq. (5.6) whereas the parameters $r_i,\ \lambda_i,\ i = 1, \ldots, m$ and $\hat{\xi}_{p^*}$ are obtained by the best fit of the RSA of the surveyed area at scale p^* using the linear combination

$$\sum_{i=1}^{m} \lambda_i c(r_i, \hat{\xi}_{p^*}) \mathcal{P}(k|r_i, \hat{\xi}_{p^*}). \tag{5.12}$$

For our analysis, we make the parsimonious choice of a single NB function because it suffices to approximately describe the available tropical forest data.

5.2.2 Stochastic model leading to a negative binomial and a log-series SAD

As already mentioned, the NB distribution can be derived from first principles on the basis of biological processes (Volkov et al., 2007; Azaele and Peruzzo, 2016). Let us assume that our ecological community consists of S species independent

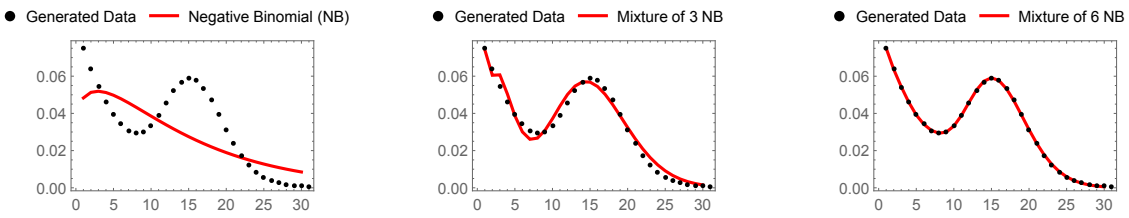


Figure 5.5: Fitting of synthetic data generated from a mixture of discrete distributions (a binomial distribution of parameters $r = 40$ and $\xi = 0.8$, a geometric distribution of parameter $\mu = 0.15$ and a Poisson distribution with parameter $\lambda = 15$) with one, three and six negative binomials. As shown, in this latter case we obtain a perfect fit of the data.

one another, so that no interspecific interactions occur. Let then $\mathcal{P}_{n,s}(t)$ be the probability that, at time t, species s has exactly n individuals, where $s \in \{1, \dots, S\}$ is the species label. We assume that the population dynamics of each species is governed by two terms, $b_{n,s}$ and $d_{n,s}$, which are the birth and death rates for species s with n individuals. The master equation regulating the evolution of $\mathcal{P}_{n,s}(t)$ for $n \geq 0$ is then

$$\frac{\partial}{\partial t}\mathcal{P}_{n,s}(t) = \mathcal{P}_{n-1,s}(t)b_{n-1,s} + \mathcal{P}_{n+1,s}(t)d_{n+1,s} - \mathcal{P}_{n,s}(t)b_{n,s} - \mathcal{P}_{n,s}(t)d_{n,s}. \tag{5.13}$$

The above equation is also valid for $n = 0$ and $n = 1$ if we set $b_{-1,s} = d_{0,s} = 0$ (reflecting boundary conditions). The steady-state solution is, for $n > 0$,

$$\mathcal{P}_{n,s} = P_{0,s} \prod_{i=0}^{n-1} \frac{b_{i,s}}{d_{i+1,s}}. \tag{5.14}$$

The term $P_{0,s}$ is a normalisation factor which can be found by imposing $\sum_{n=1}^{\infty} \mathcal{P}_{n,s} = 1$.

Let us assume that the birth term in the above equation depends on a density-independent term b_s, which is the per-capita birth rate, and on the term r_s, which takes into account immigration events or intraspecific interactions:

$$b_{n,s} = b_s(n + r_s).$$

Analogously, let us suppose that the death term depends on a density-independent term d_s, which is the per-capita death rate:

$$d_{n,s} = d_s n. \tag{5.15}$$

These suppositions are reasonable in ecology. By substituting in eq. (5.14) and setting $\xi_s = b_s/d_s$, we obtain

$$\mathcal{P}_{n,s} = P_{0,s} \binom{n + r_s - 1}{n} \xi_s^n.$$

The normalisation constant can be found by imposing

$$1 = \sum_{n=1}^{\infty} \mathcal{P}_{n,s} = P_{0,s} \sum_{n=0}^{\infty} \binom{n + r_s - 1}{n} \xi_s^n = P_{0,s}[1 - (1 - \xi_s)^{r_s}](1 - \xi_s)^{-r_s}.$$

Therefore, the probability that the s^{th} species has n individuals at equilibrium is given by a negative binomial with parameters (r_s, ξ_s) and normalised for non-zero abundances:

$$\mathcal{P}_{n,s} = \frac{1}{1-(1-\xi_s)^{r_s}} \binom{n+r_s-1}{n} \xi_s^n (1-\xi_s)^{r_s}. \tag{5.16}$$

Under the neutral hypothesis, in which all species are considered to be demographically equivalent (in the sense that each individual has the same probability of giving birth, dying, speciating and migrating), we can remove the species index s from the above equation, thus obtaining a negative binomially distributed RSA for the ecosystem under study (see eq. (5.1)).[1]

Let us notice that, with a different choice of the birth rate, one can also obtain, as stationary solution of the master equation eq. (5.13), another important RSA distribution called the Fisher log-series, named after the father of statistics who discovered it experimentally in 1943 while studying Corbet's tables of butterflies.

Indeed, let us now assumes that the population dynamics in the community are governed by ecological drift and random speciation instead of migration from meta-community. Then one can set the birth rate equal to

$$b_{n,s} = b_s n + \delta_{n,0}\nu. \tag{5.17}$$

Adding the additional reflecting boundary condition $b_{0,s} = \nu$, one has that the birth rate now accounts for reproduction and speciation. In particular, the ν parameter ensures that, whenever the species goes extinct, the community is always populated by one individual. Then, by substituting eqs. (5.15) and (5.17) into eq. (5.14) and setting $x_s = b_s/d_s$, one finds the following stationary solution:

$$\mathcal{P}_{n,s} = P_{0,s} \frac{\nu}{b_s} \frac{x_s^n}{n}.$$

The normalisation constant $P_{0,s}$ is again determined by imposing

$$1 = \sum_{n=1}^{\infty} \mathcal{P}_{n,s} = P_{0,s} \frac{\nu}{b_s} \sum_{n=0}^{\infty} \frac{x_s^n}{n} = P_{0,s} \frac{\nu}{b_s} [-\log(1-x_s)],$$

which leads to

$$\mathcal{P}_{n,s} = -\frac{1}{\log(1-x_s)} \frac{x_s^n}{n}. \tag{5.18}$$

Again, if we assume that all species are demographically identical, we can drop the s index from both $\mathcal{P}_{n,s}$ and x_s.

From the RSA (given by either eq. (5.16) or eq. (5.18)) one can find the corresponding SAD as follows

$$\mathrm{SAD}(n) = \mathbb{E}[\phi_n] = \sum_{n=1}^{\infty} \mathcal{P}_{n,s} = S\mathcal{P}_n, \tag{5.19}$$

where ϕ_n denotes the number of species in the community having abundance n and where we have dropped, once again, the species label s because of the neutrality

[1] The continuum version of the NB, i.e., the gamma distribution, is also the stationary state of a model that captures the temporal turnover of species (Bertuzzo, Suweis et al., 2011), an important aspect of tropical tree dynamics (Azaele, Pigolotti et al., 2006).

hypothesis.
Therefore, one gets

$$\mathrm{SAD}(n) = \begin{cases} S \cdot c(r,\xi)\binom{n+r-1}{n}\xi^n(1-\xi)^r & \text{NB RSA} \\ S \cdot \alpha(x)\dfrac{x^n}{n} & \text{LS RSA,} \end{cases}$$

where we set $\alpha(x) = -\dfrac{1}{\log(1-x)}$.
In the following section we will see how our framework can also be applied when assuming a log-series RSA at the scale of the whole forest.

5.2.3 Log-series SAD upscaling method

As already observed, the log-series distribution is a special case of the negative binomial obtainable as the limiting case of eq. (5.1) when r goes to zero:

$$\lim_{r\to 0} c(r,\xi)\mathcal{P}(n|r,\xi) = \lim_{r\to 0}\frac{(1-\xi)^r}{1-(1-\xi)^r}\binom{n+r-1}{n}\xi^n = \frac{\xi^n}{-n\ln(1-\xi)}, \tag{5.20}$$

where we have used the fact that

$$\binom{n+r-1}{n} = \frac{\Gamma(n+r)}{\Gamma(n+1)\Gamma(r)} \overset{r\approx 0}{\approx} \frac{r}{n+1}.$$

Let us note that eq. (5.20) is eq. (5.21) with $x = \xi$.
Therefore, it should not surprise that all the results holding for the negative binomial distribution also hold for the log-series one, which has been the basis for many different upscaling methods Slik et al., 2015; Ter Steege, Sabatier et al., 2017; Harte, Smith et al., 2009.
Here we see how our framework can be applied also when the forest SAD at the global scale $p = 1$ is modelled through a log-series.
Let us then suppose that the RSA at the global scale is distributed according to a log-series with parameter x:

$$P(n|1) = P^{LS}(n|x) = \alpha(x)\frac{x^n}{n}, \qquad \alpha(x) = -(\log(1-x))^{-1}, \tag{5.21}$$

where $\alpha(x)$ is the normalisation constant.
Again, assuming that no spatial correlations affect the sub-sample RSA, one finds that also the log-series do satisfy the similarity property.

Proposition 5.2 (Self-Similarity Property for the LS distribution)**.** *Let $P(n|1) = \alpha(x)\mathcal{P}^{LS}(n|x)$ be the RSA at the whole forest scale and let us denote with $\mathcal{P}(k|n,p)$, sampling probability at a sub-scale $p \in (0,1)$, i.e. the conditional probability that a species has abundance k in the sample p, given that it has n individuals in A.*
If $\mathcal{P}(k|n,p) = \mathcal{P}_{binom}(k|n,p)$ is binomially distributed, then the RSA at the sample scale p, $\mathcal{P}^{LS}_{sub}(k|p)$, is again a log-series, for $k \geq 1$, with rescaled parameter x:

$$\mathcal{P}^{LS}_{sub}(k|p) = \begin{cases} \alpha(x)\cdot\mathcal{P}^{LS}(k|\hat{x}_p) & k \geq 1 \\ 1-\alpha(x)/\alpha(\hat{x}_p) & k = 0 \end{cases}$$

with

$$\hat{x}_p = \frac{px}{1-x(1-p)}. \tag{5.22}$$

Proof. The probability, $\mathcal{P}^{LS}_{sub}(k|p)$, of finding a species with population $k \geq 0$ in the sub-plot of area $a = pA$ is

$$\begin{aligned} k \geq 1: \quad \mathcal{P}^{LS}_{sub}(k|p) &= \sum_{n\geq k} \mathcal{P}_{binom}(k|n,p) P(n|1) \\ &= \sum_{n\geq k} \binom{n}{k} p^k (1-p)^{n-k} \cdot \alpha(x) \frac{x^n}{n} \\ &= \alpha(x) \Big(\frac{px}{1-x(1-p)} \Big)^k \frac{1}{k} \\ &= \alpha(x) \frac{\hat{x}_p^k}{k} = \alpha(x) \cdot \mathcal{P}^{LS}(k|\hat{x}_p), \end{aligned} \tag{5.23}$$

$$\begin{aligned} k = 0: \quad \mathcal{P}^{LS}_{sub}(0|p) &= 1 - \sum_{k\geq 1} \mathcal{P}^{LS}_{sub}(k|p) \\ &= 1 - \sum_{k=1}^{\infty} \alpha(x) \cdot \frac{\hat{x}_p^k}{k} \\ &= 1 - \alpha(x) \cdot \sum_{k=1}^{\infty} \mathcal{P}^{LS}(k|\hat{x}_p) = 1 - \frac{\alpha(x)}{\alpha(\hat{x}_p)}, \end{aligned} \tag{5.24}$$

where we have inserted the explicit relation (5.22) for $\hat{x}_p$ in the penultimate equality of eq. (5.23). □

Let us notice that the relation eq. (5.22) is the same as eq. (5.2). Thus the analogue of eq. (5.6),

$$x = \frac{\hat{x}_p}{p + \hat{x}_p(1-p)}, \tag{5.25}$$

and eq. (5.7) also holds in this case.
The RSA, $P(k|p)$, is obtained as in eq. (5.5) and it is given by

$$P(k|p) = \frac{\mathcal{P}^{LS}_{sub}(k|p)}{\sum_{k'\geq 1} \mathcal{P}^{LS}_{sub}(k'|p)} = \alpha(\hat{x}_p) \frac{\hat{x}_p^k}{k} = P^{LS}(n|\hat{x}_p). \tag{5.26}$$

As for the negative binomial, also the Fisher log-series is scale invariant.
The number of species with population $k \geq 1$, $S_p(k)$, in the sub-sample of area $a = pA$ is given by

$$S_p(k) \equiv S\mathcal{P}_{sub}(k|p) = S\alpha(x) \frac{\hat{x}_p^k}{k} = \hat{\alpha} \frac{\hat{x}_p^k}{k}, \tag{5.27}$$

where we have gathered both the constants S and $\alpha(x)$ into a unique term $\hat{\alpha}$, which does not depend on the scale p. Again when referring to the sampling scale p^*, we will use the shorthand notation $S^*(k) \equiv S_{p^*}(k)$.

Then the total number of species S^* and the total abundance N^* at the scale p^* are given, respectively, by Fisher et al., 1943

$$S^* = \sum_{k=1}^{\infty} S^*(k) = -\hat{\alpha}\log(1-\hat{x}_{p^*}) \tag{5.28}$$

$$N^* = \sum_{k=1}^{\infty} kS^*(k) = \hat{\alpha}\frac{\hat{x}_{p^*}}{1-\hat{x}_{p^*}}. \tag{5.29}$$

From the sample, because S^* and N^* are known, we can get the $\hat{\alpha}$ parameter by solving the following equation:

$$N^* - \hat{\alpha}\Big(\exp\Big(\frac{S^*}{\hat{\alpha}}\Big) - 1\Big) = 0, \tag{5.30}$$

which has been obtained by inserting the expression for $\hat{x}_{p^*}$ from eq. (5.28) into eq. (5.29).
We now wish to infer information at the global scale $p = 1$ from the information we have at the sampling scale $p = p^*$. We know from previous considerations that the $\hat{\alpha}$ parameter is scale-independent. Therefore, we have the following analogous relations for S and N:

$$S = -\hat{\alpha}\log(1-x)$$
$$N = \hat{\alpha}\frac{x}{1-x},$$

from which we obtain

$$S = \hat{\alpha}\log\Big(1+\frac{N}{\hat{\alpha}}\Big), \qquad \hat{\alpha} = S\alpha(x). \tag{5.31}$$

In order to deduce the biodiversity S at the global scale, we first require an estimate of the total abundance N. Here we set $N = N^*/p^*$. This is consistent with our theoretical framework that assumes a form-invariant RSA. In fact, one can prove that the mean total abundance scales linearly with the area when one assumes a LS-distributed RSA at the global scale:

$$\mathbb{E}(N^*) = \sum_{k=1}^{\infty} kS^*(k) = \sum_{k=1}^{\infty} k\hat{\alpha}\frac{\hat{x}_{p^*}^k}{k} = \alpha\frac{\hat{x}_{p^*}}{1-\hat{x}_{p^*}} = \hat{\alpha}\frac{px}{1-x} = p^*\mathbb{E}(N),$$

where we have used eq. (5.22). The very same result can be obtained if one assumes the RSA distributed as a negative binomial:

$$\begin{aligned}\mathbb{E}(N^*) &= \sum_{k=1}^{\infty} kSc(r,\xi)\binom{k+r-1}{k}\hat{\xi}_{p^*}^k(1-\hat{\xi}_{p^*})^r \\ &= Sc(r,\xi)r\frac{\hat{\xi}_{p^*}}{1-\hat{\xi}_{p^*}} = Sc(r,\xi)r\frac{p\xi}{1-\xi} \\ &= p\mathbb{E}(N).\end{aligned}$$

Another way to infer the total biodiversity at the global scale is by using, as for the NB method, the following relation

$$S = \frac{S^*}{\sum_{k=1}^{\infty}\mathcal{P}(k|\hat{x}_{p^*})} = S^*\cdot\frac{\log(1-x)}{\log(1-\hat{x}_{p^*})}. \tag{5.32}$$

In this case, we do not need an estimate of the total number of individuals N within the area A. We have applied both the methods to extract biodiversity in our empirical forests to verify that the predictions are essentially the same (Tovo, Suweis et al., 2017). In figure 5.6 we show a schematic presentation of our upscaling framework.

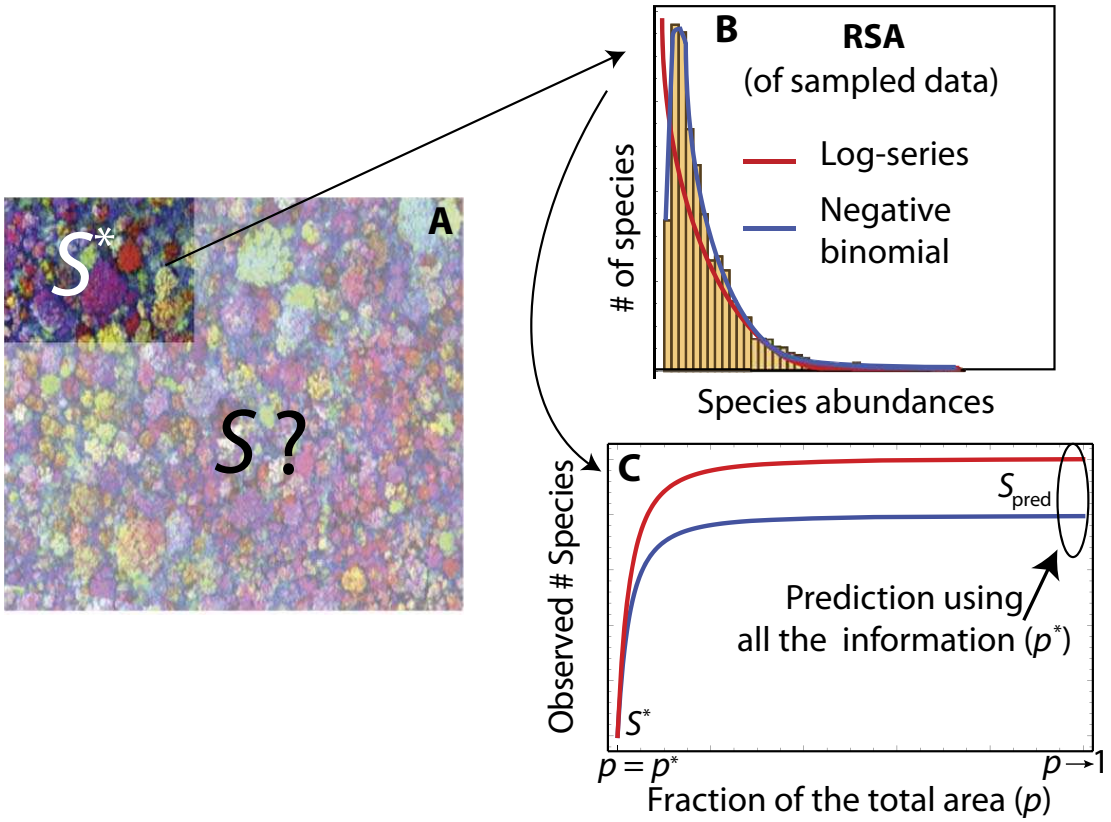


Figure 5.6: Schematic presentation of our theoretical upscaling framework. It consists of the following three steps. A) Sampling: sample a fraction p^* of the whole forest and then obtain the vector $\mathbf{n}_{p^*} = \{n_1, n_2, \ldots, n_{S^*}\}$ of the abundances of the S^* observed species. B) Fitting: use a log-series or a linear combination of a suitable number of negative binomial distributions with the same $\hat{\xi}_{p^*}$ and different values of r to perform a best fit (maximum likelihood) of the empirical SAD. C) Upscaling: using the best fit parameters obtained in (B) and using our upscaling eqs. (5.1), (5.6) and (5.10), predict the biodiversity S_{pred} and the SAD of the whole forest.

5.3 Assumptions of the upscaling framework

In our analysis, we assume that the probability that an individual tree falls within a given region is proportional to the region's area $a = pA$. This allows us to use the formalism introduced in the previous section. We refer to this assumption as the *mean field hypothesis*.

A consequence of this latter is that when we wish to sample a fraction p of an area A where every individual has been catalogued into a list according to the species it belongs to, this is equivalent to sampling the same fraction p of the individuals on this list. This is the only unbiased procedure one can utilise when neither spatial coordinates of the individuals nor spatial correlations are available.
In order for this hypothesis to be satisfied, one must first check if the region under study does not present strong inhomogeneities and anisotropies (Azaele, Suweis et al., 2016; Tovo, Formentin et al., 2016; Muneepeerakul et al., 2008) – otherwise some species may tend to inhabit specific habitats of the region and therefore the assumption of a homogeneous spatial distribution of the individuals may fail. When extrapolating information to larger scales which present environmental inhomogeneities, we need a large number of randomly located samples in order to cover all the possible habitats, as emphasised in Slik et al., 2015.
It may also not be possible to neglect spatial correlations since they could have a strong influence on the spatial distribution of the individuals. For example, we test the influence of spatial correlations between individuals on empirical singleton curves for the French BBS dataset of 2010, which records the occupancy number of 246 species in 1096 cells located all around France. At variance with the case of tropical forests, here the curves obtained by considering or neglecting spatial effects are quite different especially for scales $\lesssim 60\%$. This discrepancy suggests that space cannot be neglected and thus it must be taken into account when analysing those kinds of datasets (see figure 5.8).

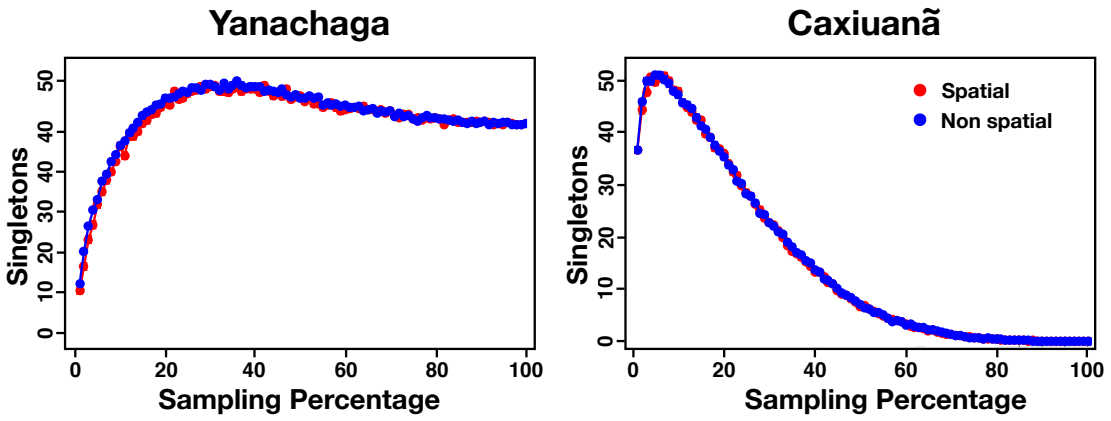


Figure 5.7: Test of the influence of eventual spatial aggregation between individuals on empirical singleton curves for two tropical forests. Spatial-dependent curves (the red ones) are then obtained by randomly choosing a fraction p of the cells and counting how many singletons are observed. Non-spatial curves (the blue ones) are obtained by randomly choosing the same fraction p of the individuals from their list and, once grouped according to the species they belong to, by counting the resulting number of species with one individual. Error bars and data points in the graphs refer to meaning among 100 trials for each percentage of sampling. The two curves are practically equivalent, meaning that they are not affected by spatial interactions.

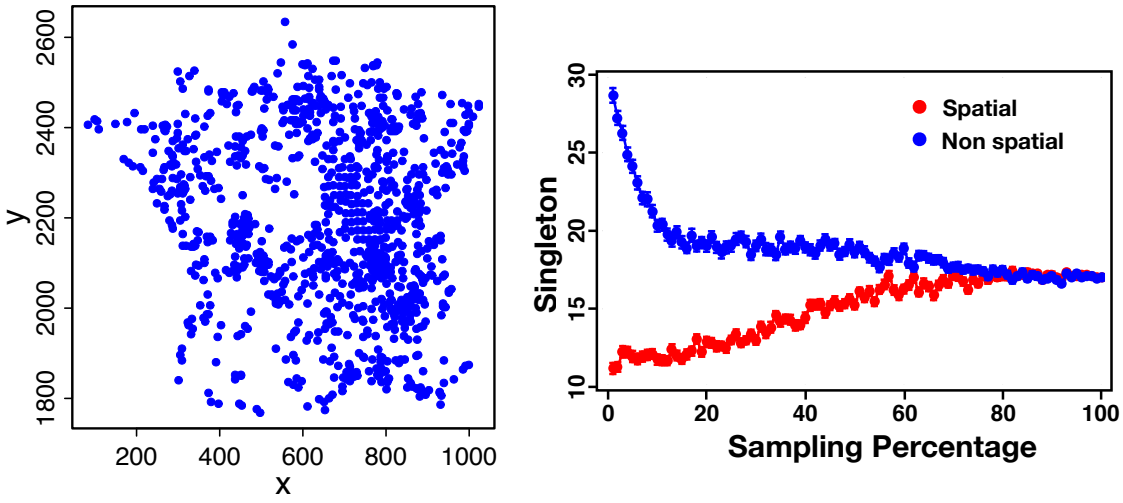


Figure 5.8: Influence of spatial aggregation between individuals on empirical singleton curves for the French BBS dataset of 2010, consisting of the occupation number of birds in 1096 cells located all around France (on the left). At contrast with the case of tropical forests, here the curves obtained by considering or not the space (red and blue curves, respectively) are very different one from the other for most of the considered spatial scales. See figure 5.7 for a detailed description of the sampling methods.

5.4 Limitation of the LS upscaling methods

The LS method suffers from some important limitations. The first, already noted by several groups (Chave, 2004; Magurran, 2005; Chave, Alonso et al., 2006; Volkov et al., 2007; Magurran, 2013; Matthews and Whittaker, 2014; Azaele, Maritan et al., 2015; Azaele, Suweis et al., 2016), is that in many cases the log-series distribution is not flexible enough to describe the distinct observed RSA patterns: unimodal distributions are the norm, rather than the exception in tropical forests. Indeed, owing to partial sampling, the empirical SAD of a small sample of a forest will likely show a monotonic decreasing behaviour, because such samples contain many rare species with just few individuals. However, a relatively larger sample may exhibit an internal mode, because relatively rare species are not found as the sampling effort increases (this happens, e.g., if the SAD at the whole forest scale is well described by a log-normal). Both situations are well captured by the NB distribution, whose functional form can accommodate both shapes, depending on the value of its different parameters (see figure 5.9). When extrapolating to larger spatial scales (up-scaling), a single NB distribution (eq. (5.1)) retains the same value of the parameter r – so we say that r is scale invariant –, whereas the parameter ξ depends on the sampling scale. The same holds true for a linear combination of NB distributions with different values of r and the same ξ. This fact is reflected in the better performance of the NB method in predicting the biodiversity at larger scales in both artificial forests and in empirical tests (see Sections 5.5 and 5.6). Moreover, to ascertain that the increased reliability of the NB method with respect to the LS method is not due to

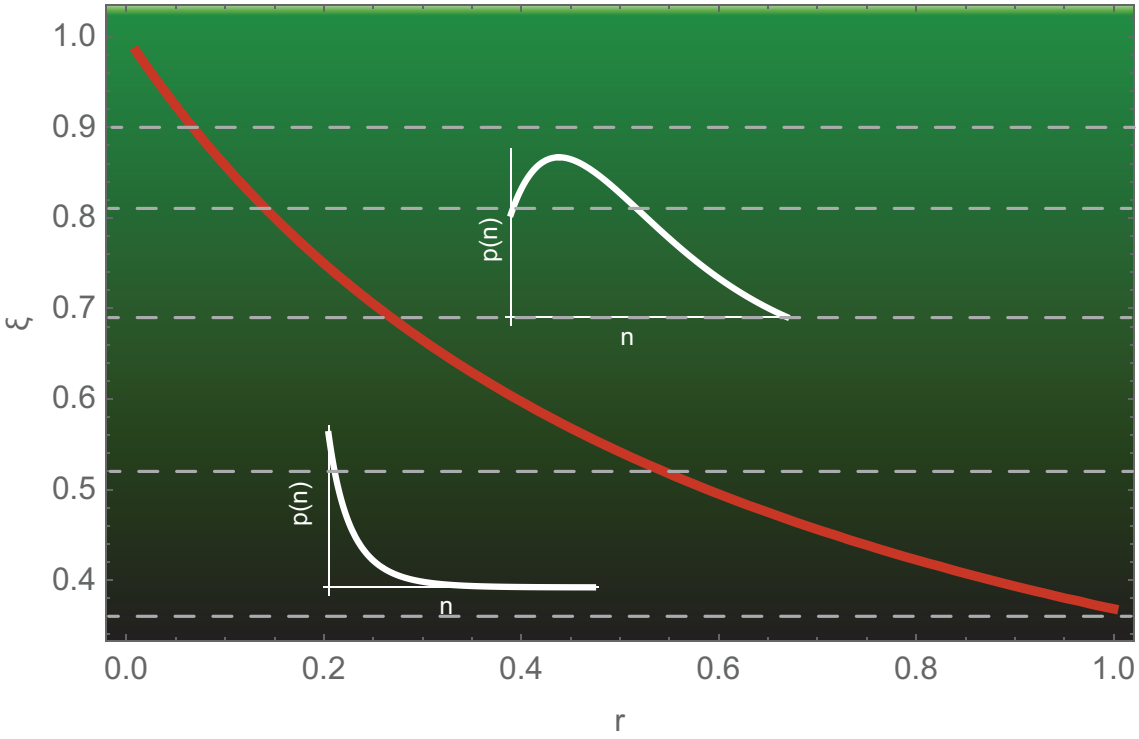


Figure 5.9: Versatility of the NB distribution. The NB distribution is a two-parameter distribution that shows self-similarity and can display both monotonic log-series-like behaviour (in the limit $r \to 0$, the NB tends to the LS distribution) and a unimodal shape, as a function of the scaling parameter ξ. The red line represents the analytical threshold separating these two cases (here the number of individuals in the x-axis are grouped in Preston classes, see Preston, 1948). The SAD, especially at large scales or with increasing sampling effort (Chisholm, 2007), often displays an interior mode that cannot be captured by the LS distribution but can be described by the NB. An example is shown of how the parameter ξ of the NB increases as the area of the forest doubles. Starting from ξ=0.36, as the area doubles, the ξ value moves to the value corresponding to the successive (dashed) horizontal line in the upward direction.

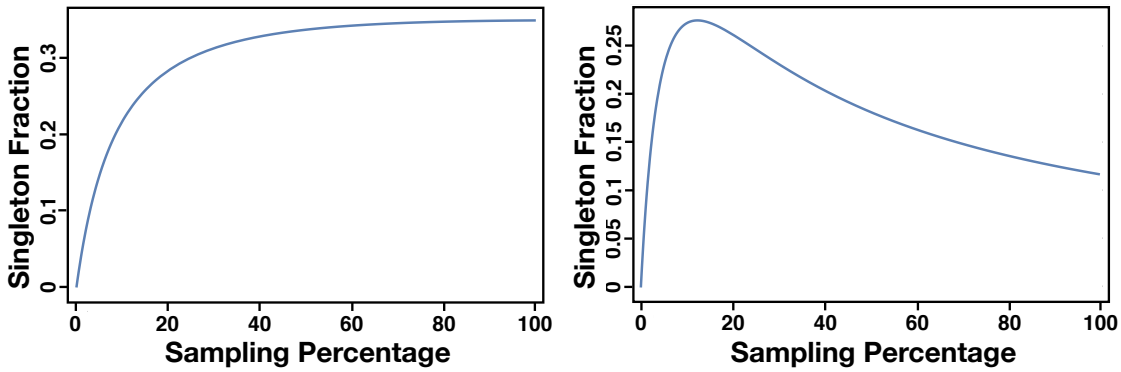


Figure 5.10: Assuming that the global RSA is distributed according to a negative binomial, we can compute the probability that a species comprises a single individual at the scale p by using eq. (5.34). Left panel: singleton fraction as a function of the percentage of sampled area for a global RSA with parameters $r = 0.1$ and $\xi = 0.9$. Right panel: singleton fraction at different scales p for a global RSA with parameters $r = 0.9$ and $\xi = 0.9$. In contrast with the log-series case, the curve does not necessarily increase monotonically.

the introduction of the additional parameter r, we have used the Akaike information criterion, which shows that the NB is the preferred model for all tropical forests in our dataset except one for which r is very close to zero.
There are two other important limitations that we describe below in detail.

5.4.1 Lack of flexibility of the LS in describing the singleton curve

Using the theoretical framework described above we can determine the number of singletons in a sub-plot whose area is a fraction p of the whole forest's area. The LS method predicts that the number of singletons is given by (see eq. (5.27))[2]:

$$S_p(1) \equiv S\mathcal{P}^{LS}_{sub}(k=1|p) = S\alpha(x)\hat{x}_p = S\alpha(x)\frac{px}{1-x(1-p)} \tag{5.33}$$

This is a monotonically increasing function of p, since S and $\alpha(x)$ are positive constants depending only on the composition of the forest at the global scale and $x \in (0,1)$. In contrast, the number of singletons predicted by our approach, using a single NB, is given by (see eq. (5.8)):

$$S_p(1) \equiv S\mathcal{P}_{sub}(k=1|p) = Sc(r,\xi)r\hat{\xi}_p(1-\hat{\xi}_p)^r \tag{5.34}$$

This, in contrast with eq. (5.33), is not necessarily an increasing function of the sampled area, as we can see in figure 5.10, but it depends on the values of the parameters. The negative binomial distribution is therefore more flexible.

[2]Note that in eq. (5.27) we have used the notation $S^*(k)$ instead of $S_p(k)$ used here.

5.4.2 Dependence of Fisher's α from the sampling scale

Slik et al. (Slik et al., 2015) showed that Fisher's α, that they deduced from three surveyed macro-regions using eq. (5.30), displays an asymptotic behaviour and they use the corresponding asymptotic value as a reliable estimate for Fisher's α at the global scale. This asymptotic α could be an artefact as its behaviour is affected by having sampled too low a percentage of the area.
To prove this, we compute Fisher's α for the Amazonian dataset at different scales using the same eq. (5.30) and the empirical values of N^* and S^* (first panel of figure 5.11). In particular, because no explicit spatial data are available, but just the RSA of the 4962 recorded species, mean values and error bars at each scale refer to 100 samples and the corresponding fraction of individuals, randomly picked among all the surveyed populations (see Section 5.3 for an assessment of the spatial effects). At small scales (up to $\sim 10\%$), we can observe the same increasing behaviour as for Slik's curves (see left bottom panel of figure 5.11). Nevertheless, when the sampling percentage increases, the α-curve starts to slowly decrease. This means that in some intermediate range, as the sampled area increases, singletons disappear (because other individuals of the same species are found) at a rate faster than that at which new singletons are found. After this regime, the number of singletons reaches an asymptotic value. This phenomenon is even more evident in other cases, such as the Caxiuanã forest (second column of figure 5.11).
The choice of the value of the parameter α strongly affects the predictions of both the number of species and singletons at the global scale, since both estimates are proportional to α itself (see eqs. (5.31) and (5.33)). In table 5.2 we display the number of singletons inferred with NB and LS methods for all the forests in our dataset. Except for few cases, where the results are comparable, usually the number of singletons predicted by the LS method is much larger than the one inferred by the NB approach.

5.5 Tests on computer-simulated forests

In order to compare the LS upscaling method with our framework based on the NB distribution, we generate various kinds of artificial forests with and without spatial correlations.
We have already noticed that our theoretical framework holds exactly when species are spatially uncorrelated. However, as we will see in the next sections, our *in silico* experiments indicate that the framework is robust even in the presence of spatial correlations and for different sampling methods.

5.5.1 Artificial forests without spatial correlations

In this case, the forests are obtained by drawing 5000 species from three of the commonly used RSA for modelling tropical forest abundances: a log-series, a negative binomial and a log-normal (LN) distribution. This latter, originally proposed by Preston (Preston, 1948), has provided a reasonably good fit to the SAD of several tropical forests (Magurran and Henderson, 2003; Azaele, Suweis et al., 2016).

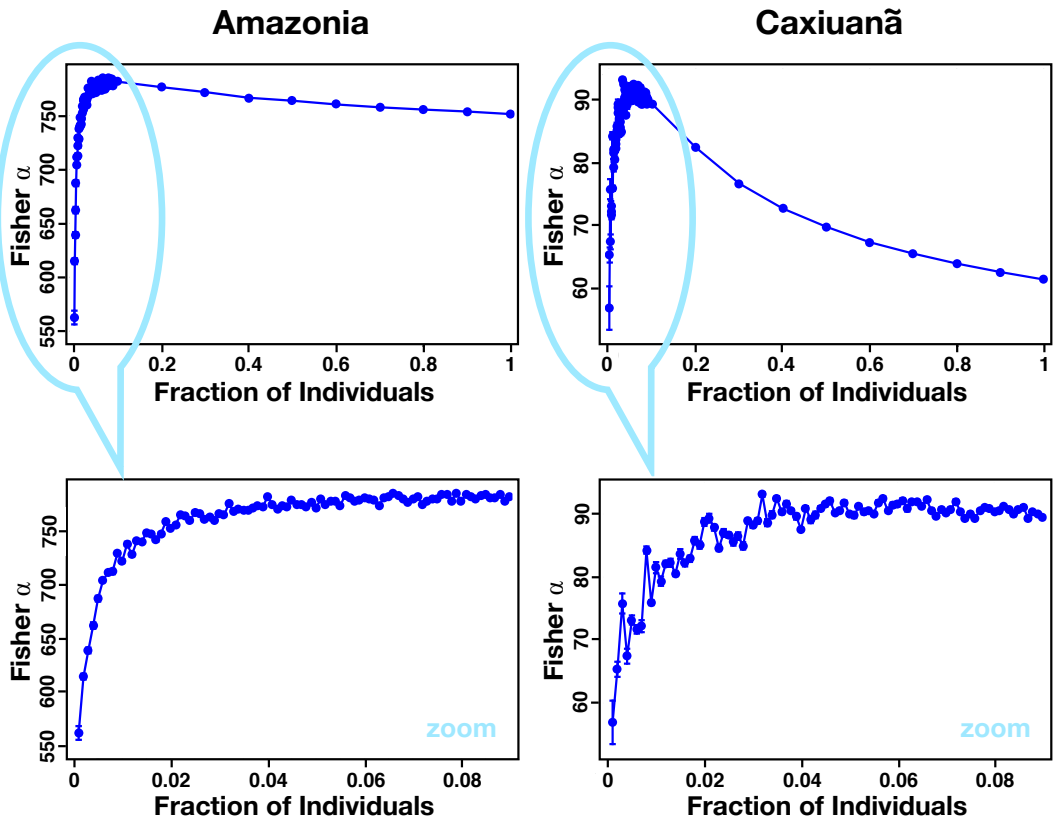


Figure 5.11: On the top: Fisher's α for two different rainforests, Amazonia and Caxiuanã. There is no asymptotic limit of Fisher's α for $p < 1$. On the bottom: we zoom the Fisher's α for only the smallest scales. We can see that in this case an apparent asymptote is reached. Nevertheless, this is not the real asymptotic Fisher's α. Mean values and error bars at each scale refer to 100 samples and the corresponding fraction of individuals, randomly picked among all the surveyed population.

Table 5.2: Predicted number of singletons in the whole area of each tropical forest obtained by applying our method (NB column). In the last column, we show the results of the LS method. The NB method yields lower results to the LS method, and it does not need an estimate of N, the total number of trees.

Forest	Observed Singletons	Method NB	Method LS
AMAZONIA	645	581	751
BARRO COLORADO NATURE MONUMENT	17	16	34
BUKIT BARISAN	13	2	62
BWINDI IMPENETRABLE FOREST	3	1	19
CAXIUANÃ	1	1	61
COCHA CASHU MANU NATIONAL PARK	12	3	94
KORUP NATIONAL PARK	0	1	37
MANAUS	11	0	175
NOUABALÉ NDOKI	0	0	18
PASOH FOREST RESERVE	94	30	118
RANOMAFANA	3	2	40
UDZUNGWA MOUNTAIN NATIONAL PARK	3	1	15
VOLCÀN BARVA	5	1	59
YANACHAGA CHEMILLÉN NATIONAL PARK	52	58	58
YASUNI NATIONAL PARK	7	4	97

We thus tested the two methods predicting the total biodiversity starting from different spatial scales, $p = 1\%$ and $p = 5\%$. In the mean field hypothesis, sampling the fraction p of the whole forest area is equivalent to randomly sample a fraction p of the individuals. Thus, at each scale, we create a list of all the forest individuals and we randomly choose a fraction p of them. Then we count the number of different species they belong to (our p^*) and we apply both LS and NB upscaling frameworks. We stress that, when fitting the sample of the simulated forest with the log-series (LS method), we use as the number of individuals N of the whole area its exact value (that we know as we generate the forest). We do this to favourably bias the chances of success of the LS method. The results are reported in table 5.3.
We find that the NB method, even using a single negative binomial, works well in all cases, while the LS method overestimates the biodiversity when the generated forest has a RSA which is not a log-series. Therefore, the NB method is more flexible and robust even when a negative binomial distribution is not the RSA of the whole forest, while it is as efficient as the LS method when applied to a log-series forest. Indeed, the best fit of the RSA with a negative binomial has led to an $r \approx 0$, so that the NB distribution is very close to a log-series.

5.5.2 Artificial forests with spatial correlations

To test the robustness of our method with respect to spatial correlations and sampling methods, we distribute the individuals of a NB (4974 species), a LS (5000

Table 5.3: We compare our method with LS method for three *in-silico* test forests generated according to different RSA distributions: log-series, log-normal and negative binomial. The datasets consists of $S = 5000$ species and we perform the analysis by considering two different spatial scales (1% and 5%). At each scale, we randomly choose the corresponding fraction of the forest individuals. Our NB method outperforms the LS one for both the forests generated through a LN and NB distribution. Moreover, in the case of the forest generated using the log-series distribution, the NB method is just as efficient as the LS one. Indeed, the best fit of the SAD with a negative binomial leads to an r parameter very close to zero, so that the NB distribution is effectively converging to a log-series, as we know it should from analytical computations.

Percentage	**Method**	**Empirical S**	**Forest RSA distribution**		
			LS	**LN**	**NB**
p=1%	LS	5000	4972	10068	9860
	NB	5000	4972	6274	4996
p=5%	LS	5000	4975	5745	7466
	NB	5000	4975	4945	5001

species) and a LN (5000 species) forests according to two modified Thomas processes (see Tovo, Formentin et al., 2016; Plotkin, Potts et al., 2000; Azaele, Cornell et al., 2012 and Chapter 1). We recall that this process can be simulated by first distributing the parents' locations (clusters' centres) according to a Poisson process with intensity $\rho_{\boldsymbol{X}}$. Given then the total number of individuals to be placed within the area of the sample, we randomly assign each of them to one of the previously generated parents. We thus place the offspring at a position drawn from a two-dimensional Gaussian distribution centred at the location of the parent and with standard deviation $\sigma_{\boldsymbol{X}}$. We impose toroidal boundary conditions in order to minimise finite-size effects for the whole synthetic forest. Finally, the parents are removed from the dataset, leaving just the offspring at their locations. Modified Thomas cluster models have reproduced empirical species-area curves with high fidelity (Plotkin, Potts et al., 2000; Plotkin, Chave et al., 2002).
We set the density of clusters $\rho_{\boldsymbol{X}} = 6 \cdot 10^{-5}$ and we choose two clumping parameters $\sigma_{\boldsymbol{X}} = 15$ and 200 in order to compare the performance of the methods for different degrees of spatial correlations. The area of the global region is chosen with the same density of individuals per unit area N/A as for the Amazonia forest. We then infer the number of species in the whole area by sampling a percentage $p^* = 1\%$ and 5% of it.
For the NB forest, we consider two different sampling methods: a first one where we survey non-overlapping 1-ha plots at randomly chosen locations within the available area and a second one where we collect data within a unique plot of the same total desired area. In figure 5.12, we show a schematic presentation of the datasets generated according to different clumping parameters and of the different sampling methods.
We find that the NB method works well in all cases and its results are robust with

respect both to the sampling method and the presence of spatial correlations (see table 5.4). In contrast, the LS method does not give reliable results, since it much overestimates the empirical number of species. This is because the basic hypothesis of a log-series RSA does not hold and, as we noticed in Section 5.5.1, the LS method is very sensitive to the empirical forest RSA. We thus compare the two

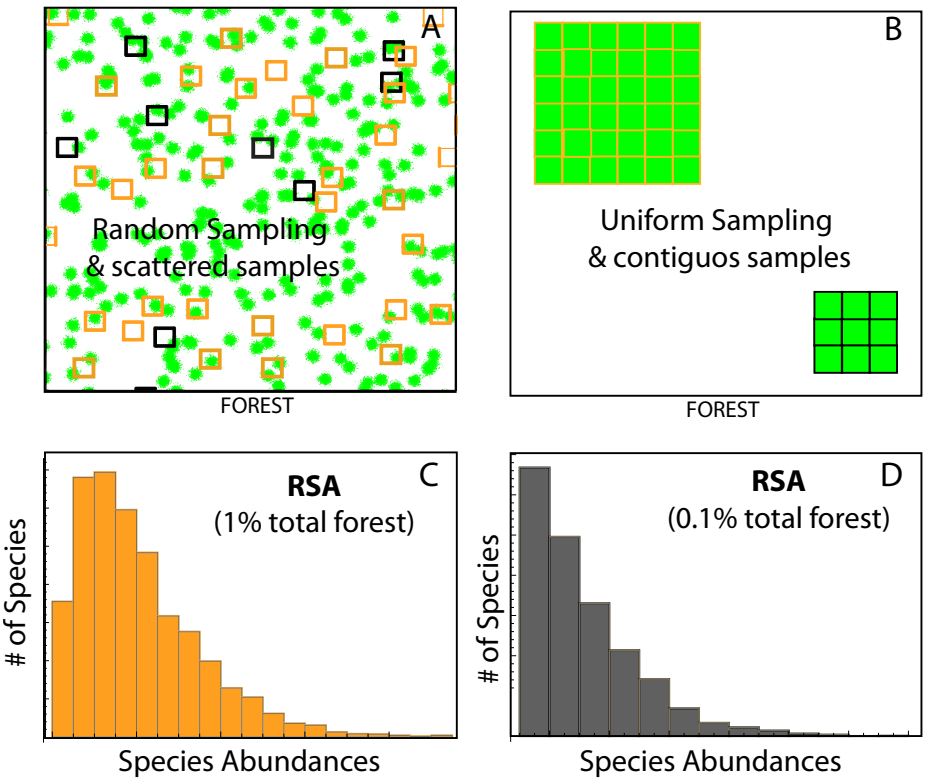


Figure 5.12: Robustness of the Method. A) We test the robustness of the NB method with respect to different spatial correlations and sampling methods. We distribute the individuals of an "artificial" forest on an area A according to two modified Thomas processes with the same density of clusters $\rho_{\boldsymbol{X}} = 6 \cdot 10^{-5}$, two different clumping parameters $\sigma_{\boldsymbol{X}} = 15,\ 200$ and different RSAs. In A)-B) green dots are plants' individuals which are either highly (A) or lowly clumped (B). We then wish to infer the number of species in the whole area by sampling a fraction p^* of it. We consider two different sampling methods: a first one where we survey non-overlapping plots at randomly chosen locations within the available area (left panel) and a second one where we collect data within a unique plot of the same desired area (right panel). In the figure, orange squares correspond to $p^* = 0.01$ sampling, while black squares represent $p^* = 0.001$ (i.e. 1% and 0.1% respectively). C-D) RSA of the species sampled at the 1% and 0.1% scale. We note that the RSA in (D) does not exhibit a mode due to the effect of the veil-line (Chisholm, 2007): the rarest species in the 1% case are not sampled in the 0.1%, leading to a mode of the observed distribution in the 1% case and not in the 0.1% case.

methods for a LS forest with a lowly-clustered distribution of individuals and the random sampling method and we find that, as expected, in this case the LS method

Table 5.4: Prediction of the total number of species obtained by applying both NB and LS methods to a NB forest consisting of 4974 species and whose individuals are distributed according to two different modified Thomas processes with the same density of clusters $\rho_{\boldsymbol{X}} = 6 \cdot 10^{-5}$ and different clumping parameters $\sigma_{\boldsymbol{X}}$: 15 (high-clustered forest) and 200 (low-clustered forest). Mean values and related standard errors on 100 trials are reported for each percentage of sampling. The NB method works well in all cases and its results are robust with respect both to the sampling method and the presence of spatial correlations. In contrast, the LS method does not give reliable results, because the basic hypothesis of a log-series RSA does not hold.

p=1% Forest Type	Samples Type	Empirical S	NB Method	LS Method
High-clustered	random	4974	4973±5	9823±20
	increasing-area	4974	4961±5	9918±35
Low-clustered	random	4974	4970±4	9834±5
	increasing-area	4974	4968±4	9876±21
p=5% Forest Type	**Samples Type**	**Empirical S**	**NB Method**	**LS Method**
High-clustered	random	4974	4974±1	7448±7
	increasing-area	4974	4981±1	7567±26
Low-clustered	random	4974	4975±1	7440±1
	increasing-area	4974	4975±1	7550±26

performs very well, predicting a species richness of 4930 against the true value of 5000 (error $\sim$ 1.3%). The very same result is obtained by using the NB method. Indeed, as in the test without spatial correlations, the best fit of the RSA with a negative binomial has led to an $r \sim 10^{-5}$, so that the predicted RSA at the whole forest scale is very close to the one predicted with the LS method. We have predicted the species richness also with the Chao estimator Chao_{wor} based on sampling without replacement (see Section 5.6 and table 5.7 for a detailed description of this method). We find that Chao's method underestimate the number of species giving a prediction of 3878 (error $\sim$ 22%). Indeed, previous results have shown (Ter Steege, Sabatier et al., 2017) that the Chao estimator for upscaling species richness based on sampling with replacement performs poorly in hyper-diverse communities with many rare species. Here we find that the very same result holds for the estimator based on sampling without replacement, an assumption consistent with the way empirical forests are sampled.

We finally test the robustness of the NB framework with respect to different clumping parameters of the generating modified Thomas process for the LN forest (see figure 5.13 and table 5.5).

Again, we sample non-overlapping 1 unit plots at randomly chosen locations covering only a small fraction, $p^* = 5\%$, of the area and attempt to predict S using only this partial information. We perform the estimation of the total species richness of

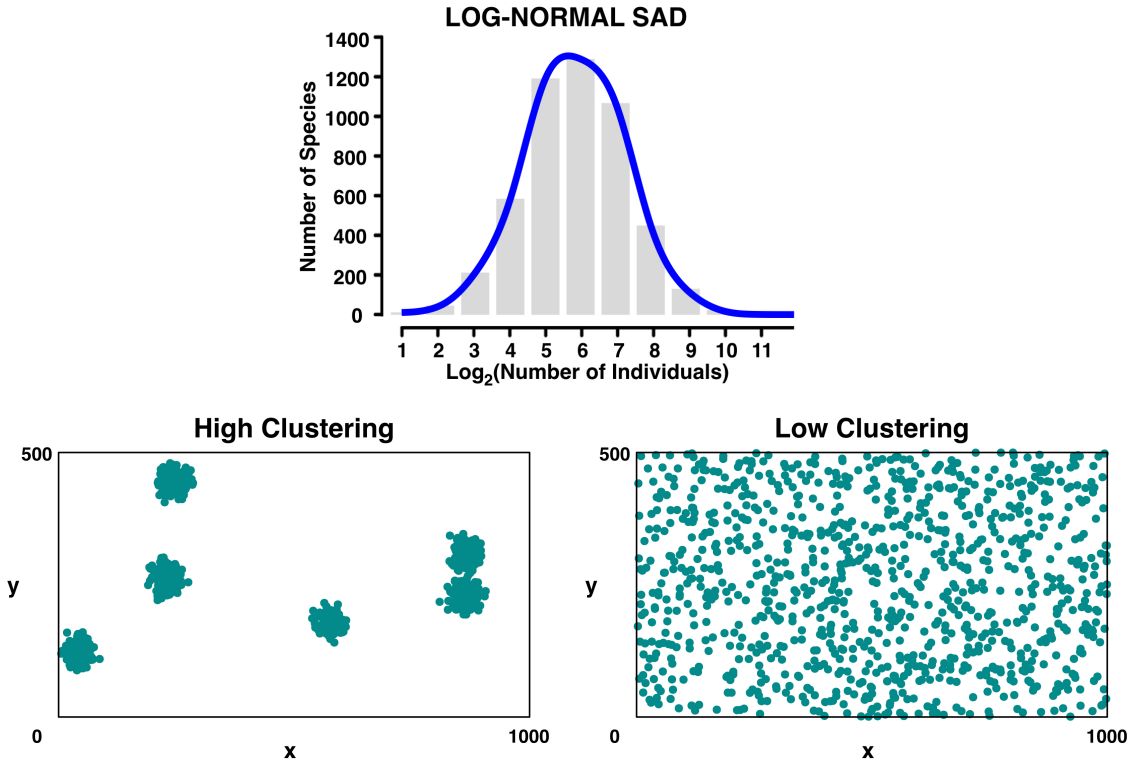


Figure 5.13: On the top: species-abundance distribution of the log-normal generated forest, consisting of 5000 species. On the bottom: each species of the *in-silico* forest is distributed according to a modified Thomas process with two different clustering coefficient, as for the LS and the NB forests' cases (15 units on the left panel and 200 units on the right panel).

the computer-generated forest by using a single negative binomial distribution or a linear combination of two negative binomial distributions, the LS method and the Chao_{wor} estimator. Notice that using a higher number of parameters as for the two negative binomial case introduces numerical complexity, so that the fit must be supervised. For this reason, the fitting algorithm works slowly and we did not perform a complete statistical analysis of the prediction errors in this case. The results in table 5.5 for the LN computer-simulated forest are therefore obtained conducting the analysis only on one single sample.
For both clustering regimes, the prediction of the number of species using the NB framework with just one negative binomial is already very good (error $< 2\%$). The linear combination of two NB increases the accuracy of the prediction at the whole forest scale $p = 1$ (with two parameters we obtain an error $< 0.2\%$). Chao's method gives results comparable to those with one negative binomial (error $< 2\%$), while underestimating the true number of species instead of overestimating it. In contrast, the LS method strongly overestimates the number of species (error $> 56\%$). We thus find that even though the original forest has a log-normal SAD entangled with spatial correlations, a single NB or a linear combination of two NBs lead to surprisingly good predictions and systematically outperform the LS method.

Table 5.5: Prediction of the total number of species obtained by applying both NB and LS methods to a LN forest whose individuals are distributed according to two different modified Thomas processes with the same density of clusters $\rho_{\boldsymbol{X}} = 6{\cdot}10^{-5}$ and different clump sizes $\sigma_{\boldsymbol{X}}$: 15 (high-clustered forest) and 200 (low-clustered forest). Results refer to a single sample consisting of non-overlapping 1 unit plots at randomly chosen locations covering a fraction $p^* = 5\%$ of the area. The prediction of the number of species using the NB framework with just one negative binomial is already quite good (error $< 2\%$). The introduction of two additional fitting parameters, necessary when using a linear combination of two negative binomials improves the estimates (error $< 0.2\%$). In contrast, the LS method overestimates the number of species (error $> 56\%$). In the last row prediction with Chao_{wor} method are also shown for comparison. We refer to table 5.7 for a detailed description of this method.

Method	**Forest type**	**Empirical S**	**Predicted S_{pred}**
LS	Low-clustered	5000	9036
	High-clustered	5000	7838
one NB	Low-clustered	5000	5067
	High-clustered	5000	5095
two NB	Low-clustered	5000	5011
	High-clustered	5000	4995
Chao_{wor}	Low-clustered	5000	4931
	High-clustered	5000	4938

Table 5.6: Number of observed species S^* and singletons in the 15 forests in our dataset. In the last column we insert the sample scale p^* (multiplied by 100), representing the fraction of surveyed area of each forest on which we have information.

Forest	S^*	Singletons	$p^* \cdot 100$
AMAZONIA	4962	645	0.00016
BARRO COLORADO NATURE MONUMENT	301	17	3.20513
BUKIT BARISAN	340	13	0.00169
BWINDI IMPENETRABLE FOREST	128	3	0.01813
CAXIUANÃ	386	1	0.01818
COCHA CASHU MANU NATIONAL PARK	489	12	0.00035
KORUP NATIONAL PARK	226	0	0.00473
MANAUS	946	11	0.06000
NOUABALÉ NDOKI	110	0	0.00143
PASOH FOREST RESERVE	927	94	0.35714
RANOMAFANA	269	3	0.01463
UDZUNGWA MOUNTAIN NATIONAL PARK	109	3	0.00302
VOLCÀN BARVA	392	5	0.02025
YANACHAGA CHEMILLÉN NATIONAL PARK	209	52	0.00372
YASUNI NATIONAL PARK	481	7	0.61100

5.6 Tests on empirical data

In order to test the accuracy of our method on more realistic distributions of trees (e.g. habitat heterogeneity, species spatial distributions, etc.), we use sub-samples taken from empirical forest data. We use a global-scale compilation of 1248 local sites collected over 15 forests around the planet on different tropical field stations of the equatorial zone[3]. The number of observed species and singletons for each forest are reported in table 5.6.

For each forest, we sub-sample a small fraction $p < p^*$ of its individuals and apply our framework to infer the number of species at the corresponding largest empirically-observable scale p^*. Moreover, we compare our results to those obtained with other methods to upscale species richness and abundances, previously proposed in the literature (Chao, 2005; Slik et al., 2015; Chao and Chiu, 2016) and summarised in table 5.7. We find that our method outperforms the one of Chao denoted with Chao_{wor} (Chao, 2005; Chao and Chiu, 2016) – which typically underestimates the forest species richness – for Amazonia (see first panel of figure 5.14), Pasoh and Yasuni. For the remaining forests, the NB method perform better than the LS

[3]All data are publicly available. The Pasoh and Barro Colorado Island datasets are provided by the Center of Tropical Research Science of the Smithsonian Tropical Research Institute (http://www.ctfs.si.edu/site). The Amazonian dataset comes from the paper Ter Steege, Pitman et al., 2013 (http://science.sciencemag.org/content/342/6156/1243092/tab-figures-data). All other data are provided by the Tropical Ecology Assessment and Monitoring (TEAM) Network of Conservation International (see http://www.teamnetwork.org/data/use). We conducte our analysis by considering all provided species, with no restriction based on dbh (saplings included). Following the analysis in Slik et al., 2015, we removed only individuals whose taxa are classified as unknown.

Table 5.7: Summary table of the most popular biodiversity estimators.

Estimator	Predicted S	Details
NB	$S^* \dfrac{1-(1-\xi)^r}{1-(1-\hat{\xi}_{p^*})^r}$	$(\xi,\ r)$ NB parameters at the global scale $p=1$ $(\hat{\xi}_{p^*},\ r)$ NB parameters at the sample scale p^*
LS	$\hat{\alpha}\log\left(1+\frac{N^*}{\hat{\alpha}}\right)$	$\hat{\alpha}$ s.t. $N^* - \hat{\alpha}(e^{S^*/\hat{\alpha}}-1)=0$ N^* = observed individuals at the sample scale p^*
Chao_1	$S^* + \begin{cases} \dfrac{f_1^2}{2f_2} & f_2>0 \\ \dfrac{f_1(f_1-1)}{2} & f_2=0 \end{cases}$	f_n = number of species with n individuals at the sample scale p^*
Chao_{bc}	$S^* + \dfrac{N^*}{N^*-1}\dfrac{f_1(f_1-1)}{2(f_2-1)}$	
iChao_1	$S_{Chao_1} + \dfrac{N^*-3}{N^*}\dfrac{f_3}{4f_4}\max\left(f_1 - \dfrac{(N^*-3)f_2f_3}{(N^*-1)2f_4}, 0\right)$	$S_{Chao_1} = S$ predicted by Chao_1 method
Chao_{wor}	$S^* + \dfrac{f_1^2}{\frac{N^*}{N^*-1}2f_2 + \frac{p^*}{1-p^*}f_1}$	
Jacknife_1	$S^* + \dfrac{N^*-1}{N^*}f_1$	
Jacknife_2	$S^* + \dfrac{2N^*-3}{N^*}f_1 - \dfrac{(N^*-2)^2}{N^*(N^*-1)}f_2$	
Turing	$S^*_{abun} + \dfrac{S^*_{rare}}{\hat{C}_{rare}}$	$S^*_{abun} = \sum_{n>10} f_n$ $S^*_{rare} = \sum_{n=1}^{10} f_n$ $\hat{C}_{rare} = 1 - f_1/\sum_{n=1}^{10} nf_n$
ACE	$S_{Turing} + \dfrac{f_1}{\hat{C}_{rare}}\hat{\gamma}^2_{rare}$	$\hat{\gamma}^2_{rare} = \max\{\gamma - 1, 0\}$ where $\gamma = \dfrac{f_1}{\hat{C}_{rare}}\dfrac{\sum_{n=1}^{10} n(n-1)f_n}{(\sum_{n=1}^{10} nf_n)(\sum_{n=1}^{10} nf_n - 1)}$

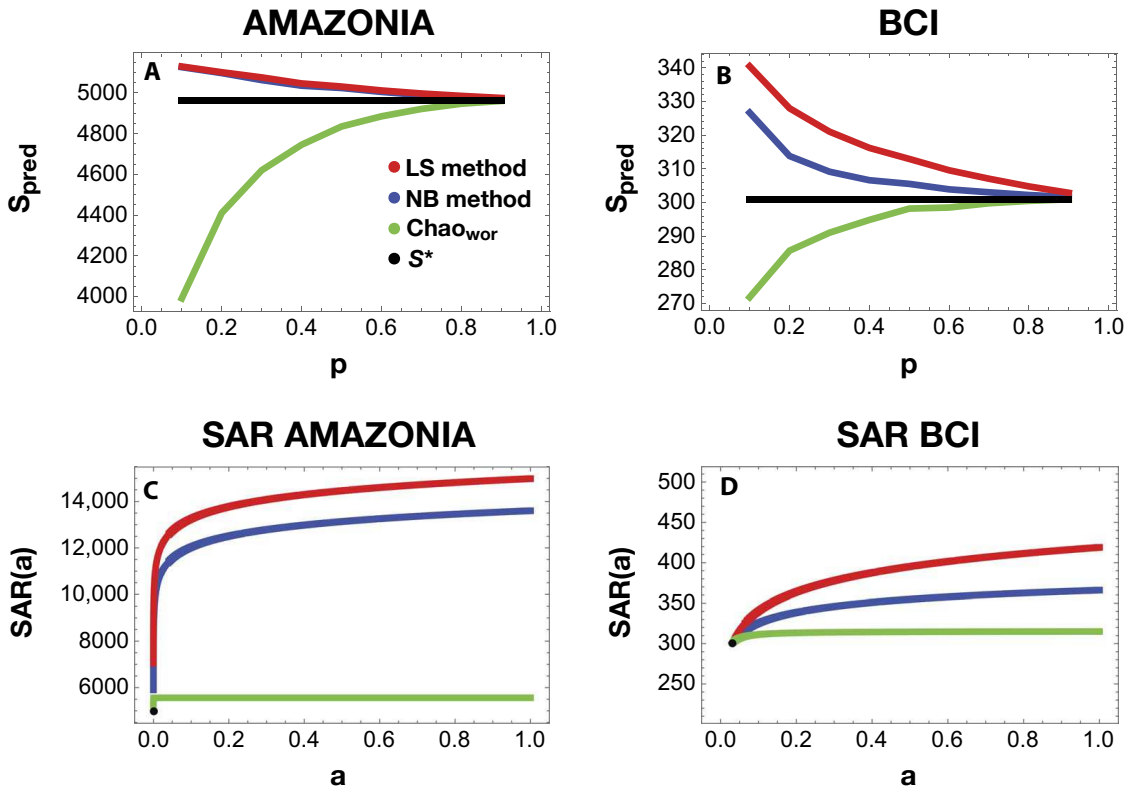


Figure 5.14: Comparison between NB, LS and Chao estimators. Top panels: predictions at different sub-scales of the number of species (the number corresponding to $p^* = 1$ is represented as a constant black line) of the LS method (red line), the NB method (blue line) and the method of Chao_{wor} (green line) for Amazonia (A) and BCI (B) forests. The first two methods perform better for the Amazonian forest, where the number of singletons, on which Chao's estimate is based, is high at every sub-scale but not enough to compensate the difference $S_{p^*} - S_p$ at small scales (see text). In contrast, for the BCI forest, both the NB and the Chao methods give comparable predictions, because here the number of singletons is very small as is the difference between S_{p^*} and S_p. Bottom panels: Amazonia (C) and BCI (D) species-area relationship (SAR), i.e. the predicted number of species at different normalised area a ($p^* < a < 1$) predicted by the three methods. In the figures, the black dots are the number of species observed at the sample scale p^*. In contrast with the canonical SAR obtained with the NB and LS method, Chao's prediction remains constant over a large part of the upscaling area range.

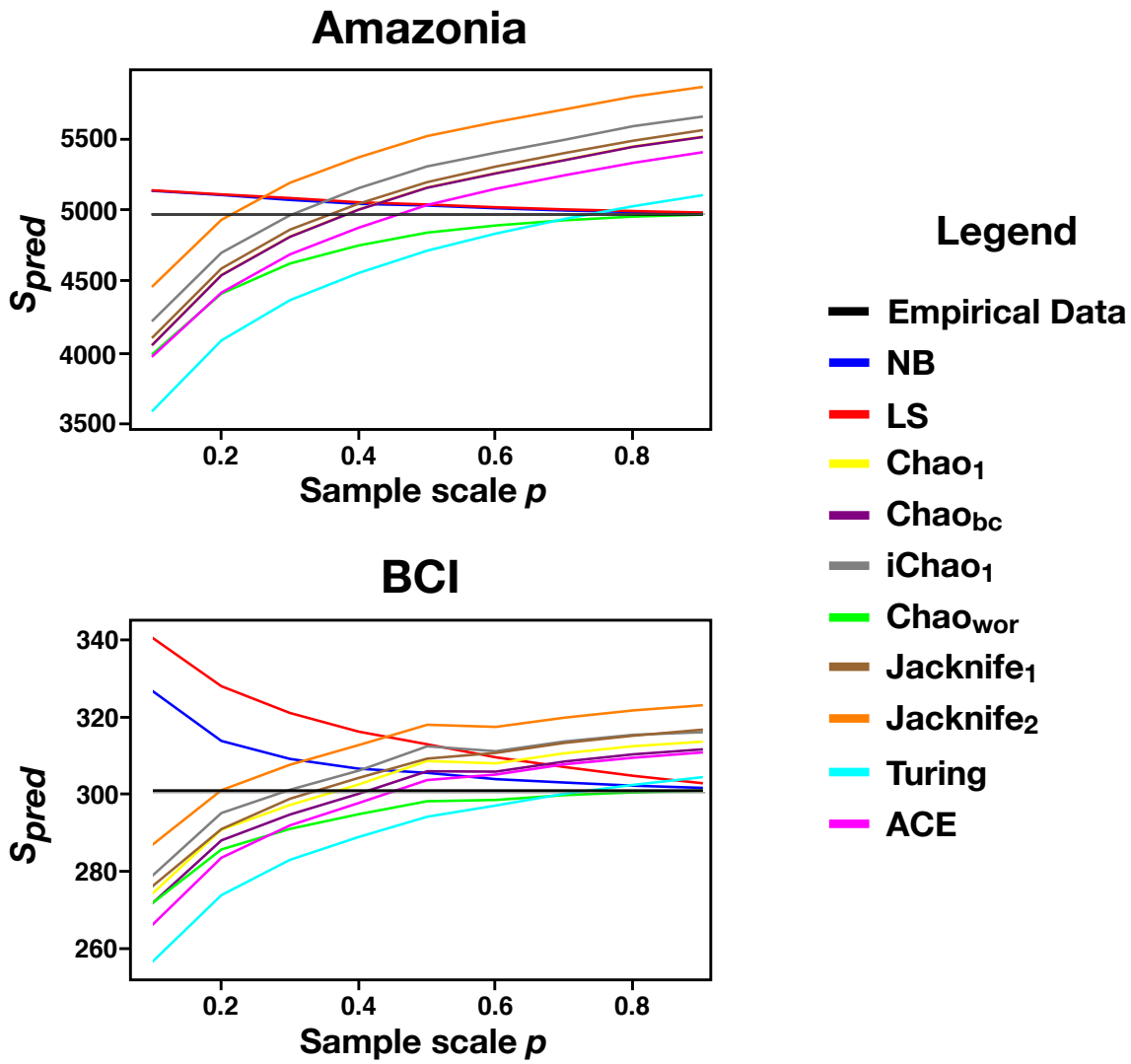


Figure 5.15: Comparison between biodiversity estimators for Amazonia and BCI forests Predicted biodiversity at the sample scale $p^* = 1$ from subsamples at scales $p < p^*$ with the most popular estimators summarised in table 5.7. While the NB, LS and Chao$_{wor}$ methods do converge at S_{p^*} as p goes to p^*, all the others have a monotonically increasing behaviour which, at some scale, overstep the true value of S^* due to the independence, in their predictions, of the scale p.

method, which overestimates the number of species at p^*, and it is comparable to Chao's estimator. However, we remark that the accuracy in Chao's predictions is due to the following fact. In such forests, the difference between S_{p^*} - the number of observed species at the sampled scale p^* - and S_p - the one at sub-sample scale p - goes to zero very fast as p approaches p^*. At the same time, the number of singletons quickly decreases to very small values (even zero). In these cases, Chao's predictions, based on singleton and doubleton species conservatively gives as output the number of species at the observation scale itself, i.e. $S_{pred} \approx S_{p^*}$ This limitation is evident in figure 5.14, which shows the tropical forest species-area relationship (SAR, i.e. the number of observable species as a function of the fraction of the sampled area a, $(p^* \leq a \leq 1)$) predicted by the three methods. While LS and NB show the expected qualitative behaviour observed in real ecosystems, the method of Chao saturates almost immediately at $a \approx p^*$, which is clearly an artefact of the method. Indeed for this range of p^*, the SAR predicted by Chao can be approximated as $S^p_{pred} \approx S_{p^*} + \frac{(\#\text{ singletons})^2}{2\ \#\text{ doubletons}}$. The same results were also observed when using Chao's estimator based on sampling with replacement (Ter Steege, Sabatier et al., 2017).
We finally find that other methods (Chao, 2005; Chao and Chiu, 2016) do not converge to S_{p^*} as $p \to p^*$, i.e., they do not have an explicit dependence on the surveyed area, rather they give an upscaled biodiversity estimates only based on the number of singletons or doubletons (see figure 5.15). Therefore we exclude these predictors from our following analysis.

5.6.1 Comparison with Harte's method

Another very popular upscaling procedure was proposed by Harte (Harte, Zillio et al., 2008; Harte, Smith et al., 2009; Kitzes and Harte, 2015) on the basis of *maximum entropy principle* (MaxEnt). Here we briefly describe Harte's method and we then compare its performance with respect to our framework.
Harte considered a system described by four state variables: the whole forest area A, the total number of species S, the total number of individuals N and the total metabolic rate E. Then he defined the joint probability distribution that a species has n individuals and that one of its individuals, chosen at random, has metabolic rate ϵ, $R(n, \epsilon)$ and he maximised its information entropy under three constraints: the normalisation condition and the constraints on the mean number of individual per species (equal to N/S) and on the mean energy per species (equal to E/S). Through the Langrange multipliers method, he found the following expression for $R(n, \epsilon)$:

$$R(n, \epsilon) = \frac{e^{-(\lambda_1 + \lambda_2 \epsilon)n}}{Z(\lambda_1, \lambda_2)}, \tag{5.35}$$

where $Z(\lambda_1, \lambda_2)$ is the partition function and λ_1 and λ_2 are the multipliers associated to the constraints on N/S and E/S. Imposing these latter and under some particular assumptions (Harte, Zillio et al., 2008), one can find the following relations for λ_1

and λ_2:

$$\frac{S}{N}\sum_{n=1}^{N} e^{-\lambda_1 n} = \sum_{n=1}^{N}\frac{e^{-\lambda_1 n}}{n} \tag{5.36}$$

$$\lambda_2 = \frac{S}{E} \tag{5.37}$$

By substituting eqs. (5.36) and (5.37) into eq. (5.35) and integrating this latter over the metabolic rate, one gets that the species-abundance distribution is given by a log-series (truncated at N)

$$P^{Harte}(n|1) = \lambda\frac{e^{-\lambda_1 n}}{n}, \tag{5.38}$$

where λ is the normalisation constant satisfying the following equation

$$\lambda = \frac{N}{S}\frac{(1-e^{-\lambda_1})}{e^{-\lambda_1} - e^{-\lambda_1(N+1)}}.$$

With MaxEnt Harte also obtained a form for the spatial abundance distribution $\mathcal{P}(k|n,p)$ describing the probability that a species has k individuals in a sample covering a fraction p of the total area A, given that it has total abundance n in the whole forest area. Under the constraints due to the normalisation and on the mean number of individuals in the sample (which is np if one assumes that the total number of individuals scales linearly with the surveyed area), Harte found that $\mathcal{P}(k|n,p)$ is given by a geometric distribution truncated at n

$$\mathcal{P}(k|n,p) = \mathcal{P}_{geom}(k|n,p) = \lambda' e^{-\lambda_3 k},$$

where λ' is the normalisation constant and λ_3 is the Lagrange multiplier associated to the constraint on the mean number of individuals at p. Under some particular assumptions (Harte, Zillio et al., 2008), Harte found the following relation

$$np = \frac{\sum_{k=1}^{n} n e^{-\lambda_3 k}}{\sum_{k=1}^{n} e^{-\lambda_3 k}} = \frac{1}{1-e^{-\lambda_3(n+1)}}\left[\frac{e^{-\lambda_3}}{1-e^{-\lambda_3}} - e^{-\lambda_3(n+1)}\left(n + \frac{1}{1-e^{-\lambda_3}}\right)\right] \tag{5.39}$$

from which one can numerically compute the value of λ_3.
He thus found its key expression for the species-area relationship

$$\begin{aligned} S_p &= S\sum_{n=1}^{N}(1-\mathcal{P}_{geom}(0|n,p))P^{Harte}(n|1) \\ &= -S\sum_{n=1}^{N}\left[1-\left(\sum_{k=0}^{n} e^{-\lambda_3 k}\right)^{-1}\right]\frac{1}{\log(\lambda_1)}\frac{e^{-\lambda_1 n}}{n}. \end{aligned} \tag{5.40}$$

Let us see how Harte exploited this result to upscale species richness from a sample of area $a/2$ to a double area a. For this special case, $\mathcal{P}_{geom}(0|n,p)$ has the simple form (Harte, Smith et al., 2009)

$$\mathcal{P}_{geom}(0|n,p) = \frac{1}{n+1},$$

since, from eq. (5.39), the λ_3 parameter equals zero.
Then, by inverting eq. (5.40) and performing the computations, the total number of species in $a/2$ is given by

$$S(a) = S(a/2)e^{-\lambda_1(a)} - N(a)\frac{1 - e^{-\lambda_1(a)}}{e^{-\lambda_1(a)} - e^{-\lambda_1(a)(N(a)+1)}}\left(1 - \frac{e^{-\lambda_1(a)N(a)}}{N(a)+1}\right), \qquad (5.41)$$

where we explicated the dependence of the number of individuals, the number of species and the Lagrange multiplier λ_1 on the area.
Since, by hypothesis, the number of individuals in a is given by $N(a) = 2N(a/2)$, eq. (5.41) contains only two unknowns, the total number of species in a, $S(a)$, and the value of the Lagrange multiplier λ_1. One can therefore numerically solve eq. (5.41) together with eq. (5.36) and obtain the species richness in the area a. By iterating the procedure, one can upscale the biodiversity up to areas which are powers of two of the anchor area p^*A.
We apply Harte's procedure[4] on four empirical forests (see table 5.8). For each of them, we sub-sample a fraction $p = 0.1$ of the individuals and predict the species richness at the $p = 1$ scale, where the true value of S^* is known (second column of table 5.8). Because Harte's upscaling procedure only allows one to scale up by successive factors of two (Harte, Smith et al., 2009), we cannot obtain an estimate at $p = 1$. Last two columns of table 5.8 refer to the predictions at $p = 0.8$ and $p = 1.6$, which represent, respectively, a lower and an upper bound for the species richness at the desired scale. For the first three forests, Harte's method does not perform as well as the others, with a typical error around 20%. For the last forest, the performance is comparable to Chao_{wor}, while being a bit worse than both the NB and LS methods. These two latter methods yield the very same results because the best fitting of the empirical SAD with a negative binomial resulted in an r parameter very close to zero. The SAD, in this case, does in fact resemble a log-series (hypothesis on which Harte's method is based).
We finally compare the NB, LS, Chao_{wor} and Harte methods on BCI empirical data, when a contiguous area is sampled (see table 5.9). More precisely, we sub-sample a fraction $p = 0.25$ and $p = 0.5$ of the individuals and predict the species richness at the $p = 1$ scale. In both cases NB method outperforms Harte's one, whose estimates are comparable to those of the LS method. This is in accordance with theoretical expectations, since both LS and Harte's procedure are based on the assumption of a log-series SAD.

5.6.2 Self-consistency test

To check the self-consistency of our framework, we run the following test on each empirical forest. We generate the corresponding forest at $p = 1$ according to the RSA and the number of species predicted by our method at the global scale, S_{pred}. We then sample $N_{p*} = p^*N$ individuals, where N is the total abundance of the whole synthetic forest, and measure the number of different species (S_{p*}) to which they belong. In summary, from the predicted RSA at the global scale, we can reproduce,

[4] We resort to the Python code provided as Supporting Information in Kitzes and Harte, 2015.

Table 5.8: Comparison between NB, LS, Chao_{wor} and Harte methods on empirical data. For each tropical forest, we sub-sample a fraction $p = 0.1$ of the individuals and predict the species richness at the whole forest scale ($p = 1$), where the true value of S^* is known (second column). For Harte's method two estimates are given, since its iterative method only permits to upscale at scales which are power of two with respect to the anchor scale Harte, Smith et al., 2009. Here we show the predictions at $p = 0.8$ and $p = 1.6$.

FOREST	**True S***	**NB**		**LS**		**Chao_{wor}**		**Harte**	
		S_{pred}	**% error**	**S_{pred}**	**% error**	**S_{pred}**	**% error**	**S_{pred}**	**% error**
BCI	301	327	8.6	341	13.3	272	9.6	382/430	26.9/42.9
PASOH FOREST RESERVE	927	910	1.8	1049	13.2	805	13.2	1192/1362	28.6/49.9
AMAZONIA	4962	5127	3.3	5130	3.4	3991	19.6	6060/7107	22.1/43.2
YANACHAGA	209	182	12.9	182	12.9	148	29.2	241/320	15.3/53.1

Table 5.9: Comparison between NB, LS, Chao_{wor} and Harte's methods on BCI empirical dataset. We consider two sub-samples consisting of contiguous fractions $p = 0.25$ and $p = 0.5$ of the surveyed area.

Sample scale $p < p^*$	**True S***	**NB**		**LS**		**Chao_{wor}**		**Harte**	
		S_{pred}	**% error**	**S_{pred}**	**% error**	**S_{pred}**	**% error**	**S_{pred}**	**% error**
0.25	301	310	3.0	325	8.0	287	4.7	333	10.6
0.5	301	306	1.7	313	4.0	298	1.0	315	4.7

by sub-sampling, the empirical values of the number of species, S_{p*}, and the number of individuals, N_{p*}, at the scale p^*. For each forest, we run the test 100 times In figure 5.16 we insert the histogram we have obtained for two of our forests (see Tovo, Suweis et al., 2017, Supplementary Material, for the histograms of all the other forests). For all the forests, the red lines representing the empirical values of S^* and N^* in our dataset turn out to be *typical* values.

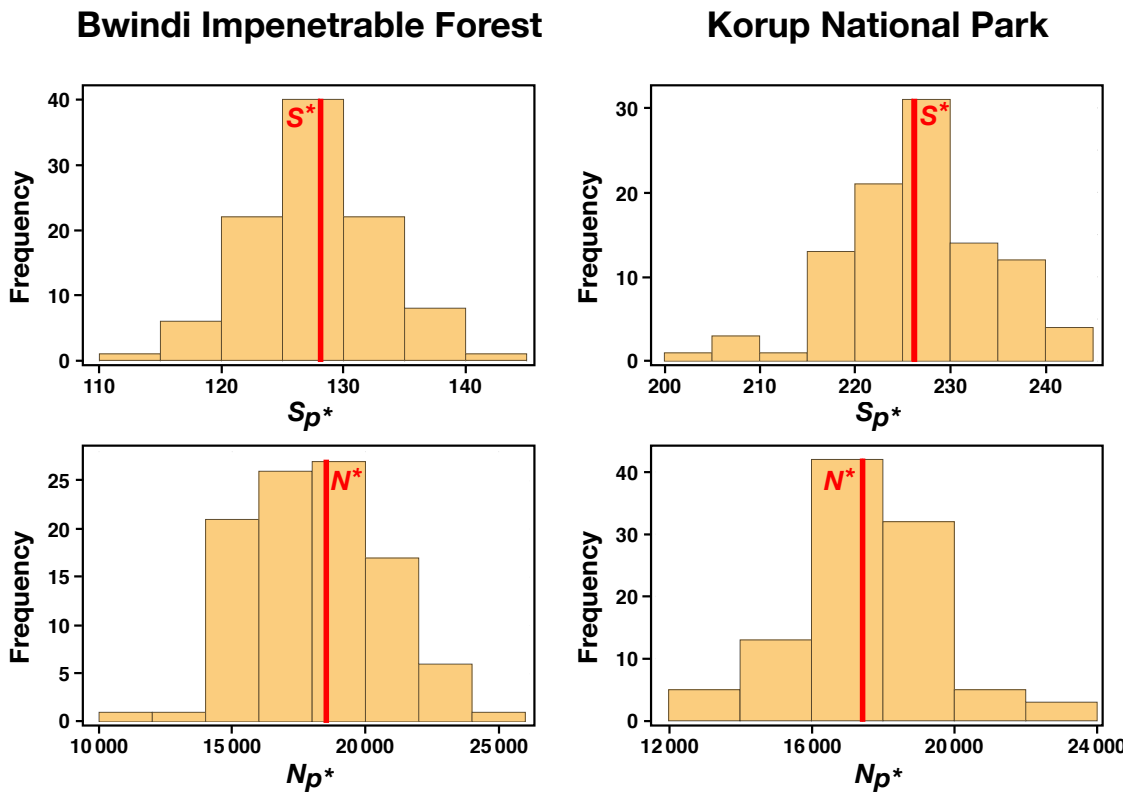


Figure 5.16: Self consistency test of our framework Starting with the RSA and the number of species at the global scale predicted by our method, we generate an artificial forest. We then sample a fraction p^* of the area and measure the number of different species (S_{p^*}) and the number of individuals (N_{p^*}) at the scale p^*. For each RSA of an empirical forest, we have run this test 100 times producing the histograms depicted above. The red lines represent the empirical value S^* and N^* of S_{p^*} and N_{p^*} in our dataset.

5.7 Biodiversity estimates in tropical forests

After testing our model on controlled computer generated data and real forest sub-samples, we apply our framework to predict the species richness and abundances of tropical forest data. Because of the good agreement between NB predictions and the true species richness in the tested forests, we choose to work with a single NB. Indeed, such a form can be derived from basic ecological processes (Volkov et al.,

2007; Azaele, Suweis et al., 2016) and it also permits an exact analytical treatment of the upscaling protocol. Although in few cases using more than one NB improves the accuracy of the predictions, in general it increases the likelihood that the empirical data are over-fitted at the sampled scale. We therefore attempt to predict, through the NB method with a single negative binomial, the species richness at the whole forest scale ($p = 1$) for each of the 15 tropical forests around the equatorial zone, and we compare our predictions with those of previous results based on the LS distribution (Ter Steege, Pitman et al., 2013; Slik et al., 2015) and with those obtained with the method of Chao. We find that the LS method systematically leads to higher estimates of the number of rare species and consequently of the forest species richness at the largest scale (see table 5.10). Only for the Yanachaga Chimillen National Park, the two estimates with NB and LS are essentially the same. The discrepancies in the estimate increase to approximately 10% for Amazonia and Barro Colorado (BCI), reach $30 - 40\%$ for Pasoh and Bukit Barisan and range between 72% and 152% for the remaining 10 forests. In contrast, Chao's method predicts a much smaller number of species at the whole forest scale. The errors in our estimates are also given in table 5.10.

5.8 Estimation of the critical p^*: how much remains to be sampled?

Using our results on the upscaled forest biodiversity, it is possible to estimate the percentage of the forest that still needs to be sampled in order to have an estimation error around 5%. We proceed as follows:

1. employing our estimation of the RSA parameters and of the total number S of the species at the global scale, we generate the predicted forest;

2. we sample the global forest at larger and larger scales p, extracting for each of them 100 samples consisting of $N_p = pN$ randomly chosen individuals;

3. we apply our method to each sample obtaining an estimation S_{pred} of S;

4. we compute for each scale mean values μ and standard deviations σ of the 100 obtained relative errors $(S_{pred} - S)/S$;

5. we select the scale at which 95% of the samples lead to an error less or around 5% with respect to the true value of S.

Interestingly, we find that for some forests (BCI, Caxiuanã, Manaus, Volcàn Barva and Yasuni), the present sampling effort may be sufficiently informative and representative to characterise the biodiversity of the whole forest. In contrast, we propose an estimate of the further sampling required for all the other forests (see table 5.10). Amazonia, for example, would need approximately twice as much the current amount of sampling, Cocha and Nouabalé approximately ten times, and Bwindi, Udzungwa and Yanachaga several hundred times the current sampling (see third column of table 5.11 showing the ratio between the predicted needed sampling

Table 5.10: Predicting Biodiversity in Tropical Forests. Predicted total number of species, S_{pred}, at the whole forest scale (corresponding to $p = 1$) for each of the 15 tropical forests in our database. Predictions are determined by using information on the sampled scale p^* (fourth column), where we observe N^* trees belonging to S^* species (second and third columns). In the fifth column we show the predictions obtained by using the NB framework with a single negative binomial for fitting the sampled SAD. Standard errors are computed by propagating the errors in the fitting parameters of the SAD (obtained by the bootstrapping method) and of S^*. This latter is determined as follows: for each dataset, we create the corresponding predicted forest at the scale $p = 1$ by generating S_{pred} numbers distributed according to a negative binomial with parameters (r, ξ). We then sample a fraction p^* of the list of individuals, as in the original data, and we count the number of observed species. The last two columns show the predictions of the LS and Chao's methods.

Forest	**Local scale information**			**Global scale predictions**		
	$\mathbf{S^*}$	$\mathbf{N^*}$	$\mathbf{100 \cdot ^*}$	$\mathbf{S_{pred}}$ **(NB)**	$\mathbf{S_{pred}}$ **(LS)**	$\mathbf{S_{pred}}$ **(Chao)**
AMAZONIA	4962	553949	0.00016	13602 ± 711	14984	5561
BARRO COLORADO NATURE MONUMENT	301	222602	3.20513	366 ± 15	419	315
BUKIT BARISAN	340	14974	0.00169	471 ± 40	1020	346
BWINDI IMPENETRABLE FOREST	128	18490	0.01813	163 ± 15	288	129
CAXIUANÃ	386	32701	0.01818	437 ± 14	915	386
COCHA CASHU MANU NATIONAL PARK	489	16640	0.00035	731 ± 63	1674	501
KORUP NATIONAL PARK	226	17427	0.00473	282 ± 23	591	226
MANAUS	946	38933	0.06000	1016 ± 14	2242	956
NOUABALÉ NDOKI	110	7196	0.00143	125 ± 8	316	110
PASOH FOREST RESERVE	927	310520	0.35714	1193 ± 36	1590	1049
RANOMAFANA	269	34580	0.01463	336 ± 22	620	269
UDZUNGWA MOUNTAIN NATIONAL PARK	109	18447	0.00302	146 ± 20	269	114
VOLCÀN BARVA	392	44439	0.02025	448 ± 16	895	395
YANACHAGA CHEMILLÉN NATIONAL PARK	209	2041	0.00372	802 ± 211	802	259
YASUNI	481	13817	0.61100	565 ± 20	974	484

and the actual one).
In figure 5.17, we plot these values against the percentage of hyper-rare species for each forest in log-log scale (see table 5.12). Intuitively, the higher the number of the rare species of a forest, the bigger the percentage one should sample in order to get an estimate of the total number of species within a given error. Indeed, we can observe a slight increasing trend in the data points. If we exclude the Amazonia dataset, which is clearly an outlier, we get a correlation coefficient of 0.5. If we also exclude the three forests of Bwindi, Udzungwa and Yanachaga, for which we would need a few hundred times the actual sampling to have an estimation precision around 5%, the correlation coefficient rises up to 0.8.

Table 5.11: Sampling targets for forest percentage cover. Using our results on upscaled forest species richness, it is possible to estimate the fraction $\mathbf{p_{pred}}$ of the forest that must be sampled to achieve an estimation error of approximately 5% with certainty of 95%. We derive these values by creating the predicted forest at the whole forest scale (we generate S_{pred} numbers according to a negative binomial with parameters r and ξ) and sampling it at increasingly larger scales until the desired accuracy in the estimation of the global species richness is reached. The last column indicates how much extra sampling is needed (if the number is greater than 1) to reach 5% precision.

Forest	$\mathbf{100 \cdot p_{pred}}$	$\mathbf{p_{pred}/p^*}$
AMAZONIA	0.0003	1.875
BARRO COLORADO NATURE MONUMENT	3	1
BUKIT BARISAN	0.05	18
BWINDI IMPENETRABLE FOREST	5	386
CAXIUANÃ	0.01	0.55
COCHA CASHU MANU NATIONAL PARK	0.003	8.57
KORUP NATIONAL PARK	0.02	1.06
MANAUS	0.02	0.17
NOUABALÉ NDOKI	0.015	10.5
PASOH FOREST RESERVE	0.5	1.4
RANOMAFANA	0.1	6.84
UDZUNGWA MOUNTAIN NATIONAL PARK	1.5	497
VOLCÀN BARVA	0.02	0.25
YANACHAGA CHEMILLÉN NATIONAL PARK	1	269
YASUNI	0.3	0.49

5.9 Fisher's paradox

5.9.1 The emergence of hyper-rarity

We also estimate the number of hyper-rare species, defined as species with fewer than 1000 individuals, and the number of hyper-dominant species, defined as the

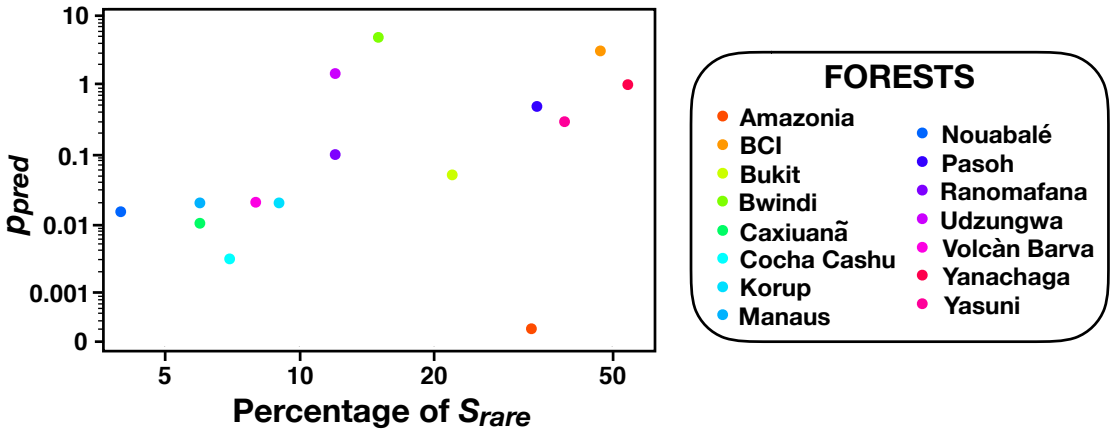


Figure 5.17: Plot, in logarithmic scale, of the scale p_{pred} one ought to sample to have an estimate precision around 5% against the predicted percentage of hyper-rare species, i.e. species with less than 1000 individuals at the global scale. Data points show a slight increasing trend with few outliers.

most abundant species contributing approximately 50% to the total number of individuals of the forest (Ter Steege, Pitman et al., 2013) (see table 5.12). Our analysis shows that hyper-rarity, as also suggested by previous works (Ter Steege, Pitman et al., 2013; Slik et al., 2015), is a recurrent pattern in large scale tropical forests, which may contribute to the fact that these tropical forests are biodiversity hotspots (Myers et al., 2000) (see also discussion below). Focusing on Amazonia, we predict that roughly 4500 Amazon tree species are hyper-rare. If they could be found and identified, this would automatically qualify them for inclusion in the IUCN Red List of Threatened Species. In fact the NB upscaling for all the Amazon forest predicts that half the total number of trees belong to just 300 hyper-dominant species, while 33% of the 13602 tree species are hyper-rare. In this way, ecologists would have an estimate of how many Amazon tree species face the most severe threats of extinction. Such rare species in the Amazon forest (and our planet's biodiversity) are like dark matter in cosmology, which accounts for much of the universe. Nevertheless, in most of the forests, we obtain a smaller number of hyper-rare species and a higher number of hyper-dominant ones with respect to previous estimates (Ter Steege, Pitman et al., 2013; Slik et al., 2015). This result is in agreement with the tests we have performed both *in-silico* and on empirical forest data. We believe that this is due to the fact that the asymptotic value of Fisher's α in the LS method is strongly biased when a very small fraction of the forest is sampled (typically $< 1\%$) (see Section 5.4).

As well as being a crucial and practical measure of fragile biodiversity in conservation ecology, hyper-rarity is also an important theoretically intriguing and open question that goes under the name of "Fisher's paradox" (Hubbell, 2015; Ter Steege, Sabatier et al., 2017). In fact, we still do not know why there is such a huge separation of population size scales between rare and hyper-dominant species. Our framework provides a possible interpretation for this phenomenon and suggests that hyper-rarity could be a manifestation of criticality in tropical forests (Stanley, 1999;

Zillio, Banavar et al., 2008; Bak, 2013).

5.9.2 The concept of criticality

In order to explain the concept of *criticality*, it is useful to resort to a model borrowed from physics. In statistical mechanics, the Ising model helps understand ferromagnetic behaviours. Let us thus consider the standard Ising model (Newman and Barkema, 1999; Nishimori, 2001) on a two-dimensional square lattice (a bar magnet), where at each of its N sites (the single atoms) we position a spin variable σ_i (the magnetic dipole moment) which takes binary values: either $+1$ or -1.The *magnetisation* $M = \mathbb{E}[\sigma_i] = \frac{1}{N}\sum_i^N \sigma_i$ provides a measure of how ordered is the system under study. In fact, we see that the magnetisation vanishes whenever the number of spins pointing up is the same of the ones pointing down. This is considered the configuration of maximum disorder of the system. Contrarily, the higher the number of up (down) spins, the higher (lower) becomes M. The maximum absolute value it can reaches is $+1$, which corresponds to the configuration where all the spins point towards the same direction, so that the system is in the most ordered state.
Let us then assign to each pair of sites (i, j) an interaction energy $-J\sigma_i\sigma_j$, with $J > 0$. We have that such a quantity is equal to $-J$ if two spins are oriented in the same direction, and $+J$ otherwise. Finally, let us suppose that no external field affect the system (corresponding to set the Zeeman energy h equal to zero).
Given a system configuration $(\sigma_1, \ldots, \sigma_N)$ on the lattice, its energy is given by the Hamiltonian function, defined as

$$H(\sigma_1, \ldots, \sigma_N) = -\sum_{\substack{i,j=1 \\ i \neq j}}^{N} J\sigma_i\sigma_j.$$

Since the constant J is set to be positive, the pairs of spins oriented in the same direction contribute negatively to the energy, while the ones in opposite directions contribute positively. Therefore the former configuration, being a minimum of the system energy, is favoured in order to reach the thermodynamic equilibrium. This positive interaction is thus called *ferromagnetic*, since it yields to a system configuration in which all spins are oriented in the same direction.
Following statistical mechanics prescription, the probability that the system will be found in a particular configuration $(\sigma_1, \ldots, \sigma_N)$ in thermal equilibrium with a reservoir at temperature T is given by the Boltzmann distribution

$$P(\sigma_1, \ldots, \sigma_N) = \frac{1}{Z(T)} e^{-\frac{H(\sigma_1, \ldots, \sigma_N)}{kT}}, \tag{5.42}$$

depending on both the temperature T and the energy $H(\cdot)$ (Gibbs, 1902). The normalisation contant Z on the denominator of eq. (5.42) is the partition function while k is Boltzmann's constant, equal to $1.38 \times 10^{-23} JK^{-1}$.
It is well known that, at low temperatures, the absolute value of the magnetisation M is very close to either $+1$ or -1, since almost all the spins are oriented in the same direction. The system is therefore in the ferromagnetic state, so that bar magnet

is capable of attracting M thumbtacks (Stanley, 1999). Let us assume the starting M equals $+1$. As the temperature increases, the order parameter M decreases, and more and more clusters of sign -1 appear. By further heating the system, small clusters having positive spins start to grow inside the previously formed ones, so that a fractal structure emerges. Its typical size is called *correlation length*.
At a certain temperature T_c, called *critical point* or *Curie temperature*, the system undergoes a transition phase into a paramagnetic state, where the magnetisation is zero and the bar magnet looses its capability of attracting thumbtacks. In the proximity of this critical temperature, the correlation length diverges and the system appears the same at different scales, i.e., it is self-similar (Stanley, 1999). This critical phenomena are usually investigated by the mathematical technique of the renormalisation group, which let us describe the behaviour of the system at different scales in terms of equations for the model parameters. The scale-invariance property confers to the system an acute sensitivity to certain types of external perturbations or disturbances whose effects are realised at long distances. In this way the system can optimally adapt to any new environmental conditions by fine-tuning different properties closely related to the magnetisation, such as the magnetic susceptibility or the constant-field specific heat, both diverging in correspondence of the Curie temperature.
Another intriguing property is that, in the correspondence of the critical point, the diverging quantities of squared magnetisation, susceptibility and specific heat for the bar magnet are all characterised by being related to the quantity $(T - T_c)/T_c$ through power laws, with critical exponents connected by the so-called *scaling laws* (Stanley, 1999). Surprisingly, such exponents take the same values for seemingly very different physical systems and are thus divided into *universal* classes (Lee and Yang, 1952).
The same critical phenomenon can be observed, for example, in the liquid-vapour system, which, in the vicinity of its critical point, is characterised by density fluctuations that become very large, with droplets of water and bubbles of gas of all sizes thoroughly interspersed, and it undergoes a phase transition into a totally new state, called *supercritical fluid*, with physical properties inherited from both the vapour and liquid states. Also in this case, the properties connected to the density, such as the viscosity, the relative permittivity and the solvent strength are all characterised by huge gradients in the proximity of the critical point, so that the system can fine-tune them in order to easily adapt whenever even small changes in the environmental variables occur.

5.9.3 Forests are in the vicinity of a critical point

The fitted parameters of the NB distributions that provided the best predictions of the upscaled species richness in all the considered tropical forests fall within a tiny region of parameter space: $0 < r < 0.7$ and $\xi \approx 1$. This result is somewhat surprising, because there are neither theoretical nor biological reasons why tropical forests should have their parameters localised within such a narrow region, especially when considering that they are in completely different geographical regions with differing evolutionary histories. However, a closer look at the NB distribution reveals

that, in this region of parameter space, the relative fluctuation of abundances is maximised.
In fact, let $P(n|1) = c(r,\xi)\mathcal{P}(n|r,\xi) = \binom{n+r-1}{n}\xi^n(1-\xi)^r/[1-(1-\xi)^r]$ be the RSA in the NB hypothesis (see eq. (5.1)) and let us compute its first two moments, which we denote by $\mathbb{E}[n]$ and $\mathbb{E}[n^2]$ respectively. These can be calculated to give

$$\mathbb{E}[n] = \sum_{n=1}^{\infty} nP(n|1) = \frac{\xi r}{(1-\xi)(1-(1-\xi)^r)}$$

and

$$\mathbb{E}[n^2] = \sum_{n=1}^{\infty} n^2 P(n|1) = \frac{\xi r(1+\xi r)}{(1-\xi)^2(1-(1-\xi)^r)}.$$

Then the relative fluctuation in abundances $F(\xi, r)$, is given by

$$\begin{aligned} F(\xi,r) &= \frac{\mathbb{E}[(n- < n >)^2]}{\mathbb{E}[n]^2} = \frac{\mathbb{E}[n^2] - \mathbb{E}[n]^2}{\mathbb{E}[n]^2} \\ &= \frac{(1-(1-\xi)^r)(1+\xi r)}{\xi r} - 1. \end{aligned}$$

Let us notice that the function $F(\xi, r)$ diverges as $-\log(1-\xi)$ as $r \to 0$.
In the left panel of figure 5.18, we insert the values of $\sqrt{F(\xi,r)}$ versus those of $\mathbb{E}[n]$ obtained for the 15 analysed rainforests. In the right panel we show the contour plot of $\sqrt{F(\xi,r)}$, where the black points correspond to the parameter values obtained for the 15 forests.
The above computation on the form of the NB distribution reveals that the square root of the relative fluctuation of abundances, $\sqrt{F(\xi,r)}$, diverges as $\xi \to 1$ and $r \to 0$. Thus, parameter values in the vicinity of this region allow an ecosystem to have the highest heterogeneity in its abundance distribution.
The observed large abundance fluctuations suggest that tropical forests may be critical systems and may be relatively reactive to disturbances (Hidalgo, Grilli, Suweis, Muñoz et al., 2014; Hidalgo, Grilli, Suweis, Maritan et al., 2016) and able to adapt optimally to new external conditions/constraints.
Under a given set of environmental conditions, only a few species are best at exploiting the limited available resources (Grilli et al., 2013). Because of environmental fluctuations, these conditions may not continue to remain advantageous for the existing very few abundant species. However, a large pool of species may serve as a reservoir of new opportunities and responses and as a buffer against newly changed conditions (Grilli et al., 2013). According to this view, hyper-rarity is essential for an ecosystem to maintain its functions and react promptly to changes: rare species may provide the key to an ecosystem's future (Hull et al., 2015).

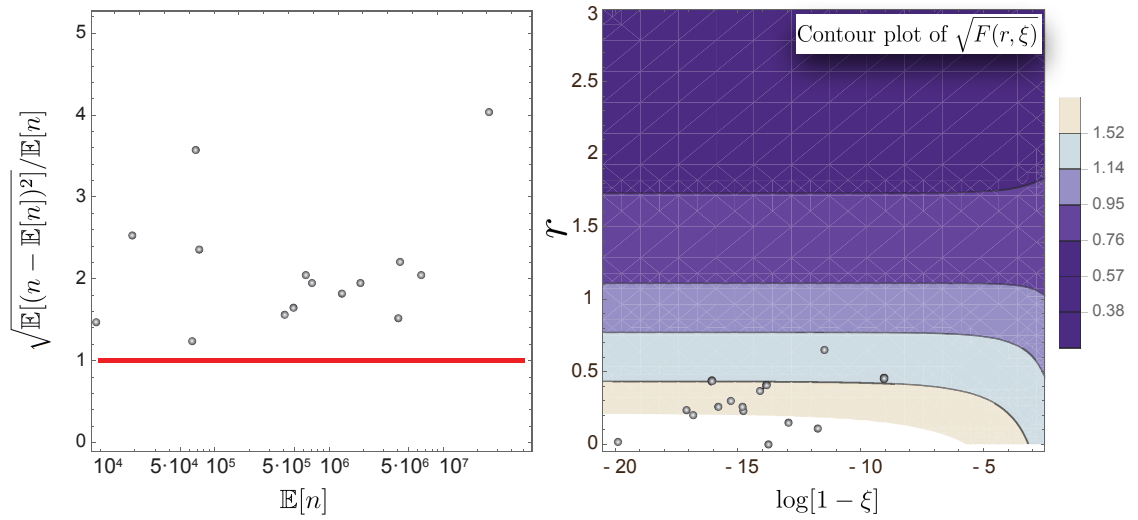


Figure 5.18: Tropical forests are poised in the vicinity of criticality. A) Plot of the square root of the relative fluctuations in species' abundances, $\sqrt{\mathbb{E}[(n-\mathbb{E}[n])^2]}/\mathbb{E}[n]$, in linear scale versus the average abundance $\mathbb{E}[n]$ in logarithmic scale. The black points denote the predicted values of the fluctuations for each of the 15 tropical forests in our database at the whole forest scale, and the red curve is the line of equation $y = 1$. All values are located above this line, thus indicating that the relative fluctuation in abundances are considerable for all the forests. B) Contour plot of the square root of the relative fluctuation in abundances $F(\xi, r)$ for a negative binomial SAD. The black points represent the pairs $(r, \log[1-\xi])$, where r and ξ are the predicted parameters for each forest of our dataset after up-scaling at the whole forest scale. These points are all located in the region of the parameter space around which the function $F(\xi, r)$ diverges, i.e., $\xi \approx 1$ and $0 < r < 0.7$.

Table 5.12: Fisher's paradox (Hubbell, 2015). Hyper-rare species (defined as species with fewer than 1000 individuals (Ter Steege, Pitman et al., 2013; Slik et al., 2015)) and hyper-dominant species (the most abundant species, accounting for $\approx 50\%$ of the total number of individuals) percentages are predicted in the whole area of each tropical forest obtained by applying both the NB and LS methods. Interestingly, we find that by using our NB method, the number of hyper-rare species in most of the forests is drastically reduced with respect to the LS method, thus suggesting that the extremely high value of hyper-rare species predicted in previous studies (Ter Steege, Pitman et al., 2013; Slik et al., 2015) is an artefact of the LS method. Nevertheless, we find that the hyper-rarity phenomenon is a genuine emergent pattern in tropical forests.

Forest	**Hyper Rare**		**Hyper Dominant**	
	NB Method	LS Method	NB Method	LS Method
AMAZONIA	33%	37%	2.2%	2.0%
BARRO COLORADO NATURE MONUMENT	47%	60%	5.5%	4.8%
BUKIT BARISAN	22%	46%	7.9%	1.9%
BWINDI IMPENETRABLE FOREST	15%	48%	7.4%	3.5%
CAXIUANÃ	6%	49%	10.3%	3.2%
COCHA CASHU MANU NATIONAL PARK	7%	41 %	8.4%	2.5%
KORUP NATIONAL PARK	9%	51%	9.3%	3.1%
MANAUS	6%	59%	14.5%	2.8%
NOUABALÉ NDOKI	4%	43%	11.2%	2.4%
PASOH FOREST RESERVE	34%	55%	6.5%	3.1%
RANOMAFANA	12%	49%	7.5%	2.7%
UDZUNGWA MOUNTAIN NATIONAL PARK	12%	48%	6.3%	3.0%
VOLCÀN BARVA	8%	52%	10.5%	2.5%
YANACHAGA CHEMILLÉN NATIONAL PARK	54%	56%	3.0%	2.7%
YASUNI NATIONAL PARK	39%	74%	11.6%	4.4%

Conclusions

To summarise, throughout this thesis we have studied models for biodiversity and issues in statistical analysis of ecological databases.
In particular, we have seen that spatial point processes are a powerful statistical tool for the description of patterns in tropical rainforests and we have developed a novel method for upscaling biodiversity from scattered samples to larger scales.
The main directions of this research work can be resumed in the following points.

Knuth's optimal binning method

- We studied how Knuth's optimal binning method, based on Bayes's Theorem and maximum a-posteriori estimation, allows to infer the least biased estimate of the underlying density function of a point pattern, without needing any a priori assumption about the phenomena that generated the data. Moreover, the optimal bin size sets the most informative scale at which to observe the data.

- We showed how to use the Knuth optimal bin size and shape to estimate the intensity of a spatial process and infer characteristic spatial features as anisotropy and clusterisation.

- Tested against both kernel method and non-kernel methods it resulted to be more efficient in detecting CSR processes and in avoiding noisy fluctuations due to the sampling process. Moreover, since it is based on a maximisation procedure, it does not contain adjustable parameters and it is not subject to the virtual aggregation phenomenon.

- When used in conjunction with Schiffer's index, it allows to infer qualitative and quantitative information on both first and second-order statistics.

- We tested our findings on the BCI ecological dataset to have information about distribution of the size of cluster-like structure of plants, anisotropy of plant distribution and existence of uniformly distributed species. We found evidence that the choice of modelling a species' distribution through a modified Thomas process, which has been proven to be efficient in capturing some important biological curves but not others, is not always supported by Knuth's method.

- We found that the cluster size measured by optimal bin area is insignificantly correlated with the abundance of a species.

Globally, Knuth's method can be used as a trusted tool for the preliminary statistical analysis of a spatial dataset. It provides a reliable method to test whether an hypothesis made on the underlying process of a pattern is justified or not. We are confident that this survey of Knuth algorithm' performance can be of help for the scientific community.
A possible deepening, for example, could be the application of this method to different datasets, maybe consisting of temporal positions instead of spatial ones. Indeed, the same tendency to form clusters has been observed also in human activities, such as email sending. Tests on these kind of databases could further confirm the reliability of our results.
The present analysis should help inform future investigations of temporal or spatial features of different complex systems in ecology and human dynamics (Simini et al., 2011; Formentin et al., 2014; Sanlı and Lambiotte, 2015).

Similarity decay function

- We investigated the role of spatial clustering in shaping the curve of species' turnover with the distance.
- We derived an analytical formula for the average decay in similarity with the distance between two relatively small plots. A peculiar trait of our approach is the use of the spatial density of the similarity with respect to the area.
- We found that the decay function of the similarity density is essentially given by the pair correlation function of the whole forest and that it is determined by the most abundant species. Our formula thus establishes a link between a very important concept in quantitative ecology with a widely used concept in the statistical description of a general particle system. Moreover, this hollow curve tends to an asymptotic value which is determined by the relative abundances.
- To test the analytical theory against real data, we designed a statistical estimator for the similarity based on presence-absence counts on plots of finite size and on an area-scaling factor. We were able to interface our analytical theory, which refers to plots of infinitesimal size, with the estimator, designed for finite-sized sample data.
- We tested our findings on the extent of the study area of BCI and Pasoh forests obtaining a very good agreement with the empirical data.

The limiting hypothesis of relatively small size of the plots with respect to the distance between them is present in all other works we examined dealing with this problem, and it is not easy to manage. We think that in our approach the dependence on the area is easier to control since it is transferred to the statistical estimator, which is flexible enough. At larger scales, if other drivers of biodiversity other than clustering of individuals are acting (climatic, orographic or other shaping factors), our model can not be directly applied. If no strong environmental inhomogeneities are encountered, it could be effectively employed for larger portions of rainforests,

where a complete survey of all individuals is impossible, since it performs well with respect to random sub-sampling.
Finally, another possible generalisation could be to extend the model by incorporating the covariance between two species' point processes in order to take into account also intraspecific biotic interactions. Such integration has been shown to yield surprising but important insights on species' persistence and coexistence (see Hui, Fox et al., 2017).

Negative binomial upscaling method

- We presented a theoretical framework to upscale species richness and abundances in tropical forests from a limited number of samples. The advantage of our method mainly relies on two properties.
- First, it is flexible. The negative binomial, depending on the value of its parameters, may display both a log-series like behaviour or an interior mode and it is therefore able to describe different SAD shapes. Thus we can use the same functional form to reproduce different ecosystems' SAD, as those observed in our datasets. In contrast, a log-series SAD predicts a very specific form for the SAD that is not flexible enough to describe any SAD with an interior mode. Furthermore, our approach, relying on an appropriate linear combination of negative binomials, can basically accommodate any type of complex SAD functional form.
- Second, the negative binomial (or a combination of them) is self-similar under different sampling intensities. This is the key feature that allows us to obtain an easy analytical formula to upscale the SAD from the sample scale to any arbitrary one.
- Tested on both computer-generated and real forests, our method outperformed other popular methods proposed in literature.
- In Harte, Smith et al., 2009, despite the flexibility of the approach, the upscaling can be performed only by numerically solving a pair of analytical equations. In Zillio and He, 2010 the authors propose an iterative method to estimate the species richness and the abundance distribution. Again, this method is flexible, but no analytical treatment can be performed.
- In our framework we only need the fraction of sampled area with respect to the whole forest, while, in other approaches, additional information on the upscaled forest is required (e.g. the number of individuals of the most abundant species (Borda-de-Água et al., 2012)).
- Our method allows to compute how much further sampling of the different forests would be needed to have estimates within a fixed acceptable error.
- It confirms that the vast majority of species in the analysed forests are rare or hyper-rare and suggests that this may be a signature of critical-like behaviour, which can provide a buffer against extinction.

The two fundamental properties of the NB allow our method to be applied on statistical upscaling problems beyond forest ecology. A possible application is, for example, in the field of meta-genomics. Using recently developed DNA sequencing machines, it is possible to obtain the total genomic DNA directly from a macro fauna or flora environmental sample (i.e., a macro-biome). This meta-genomic (gene of genes) approach together with taxonomic classification algorithms (Menzel et al., 2016) allows a characterisation of the biodiversity of the samples (typically prokaryotes). However, SAD curves built in this way describe the biodiversity only very locally (the scale of the given environmental sample). Nevertheless, assuming well mixed communities and finding an appropriate combination of negative binomials fitting the observed SAD, we can use our framework to upscale micro-biome SAD at a larger scale (e.g. the whole gut), as would be measured if it were possible to survey the entire environment. It can also be applied to immunology for finding the number of TCR clonotypes in a human body. These examples show the promising generality of our approach and open the possibility of new applications of the upscaling framework to other taxa or type of systems.

Appendices

A

Knuth's applications

A.1 Application of Knuth's method

We test Knuth's method both on one and two dimensional generated datasets. In order to do this we resort to MATLAB software and Knuth's *OPTBINS binning package v1.0*.

A.1.1 One-dimensional datasets

We apply Knuth's method to three sets of data, each consisting of 1000 unidimensional points sampled from known probability density functions. The results are shown in figure A.1.
For all sets of data we plot the logarithm of the relative posterior probability given by eq. (2.7), whose maximum gives the optimal number of bins $\hat{M}$. We then arrange the data in an histogram having $\hat{M}$ bins which represents our estimation of the piece-wise constant density function of our data. To highlight the fact that the method is able to avoid sampling fluctuations while it captures the main characteristics of the underlying density function, we plot, under each optimal histogram, the one we would get by using 100 bins instead of $\hat{M}$, which therefore better shows the spatial distributions of the data points.
Let us notice that the optimal binning numbers that we obtain in the three considered cases depend on the particular realisation of the process from which the data under analysis are sampled. Anyway, as we will show in Section A.1.2, Knuth's method is very stable, so that the results it gives slightly differ from a sample to another one.
The first case we consider is the sampling of 1000 data points from a uniform density. From the top panels of figure A.1, we see that the relative log posterior reaches its maximum at $\hat{M} = 1$. Knuth's histogram therefore coincides perfectly with the underlying density function.

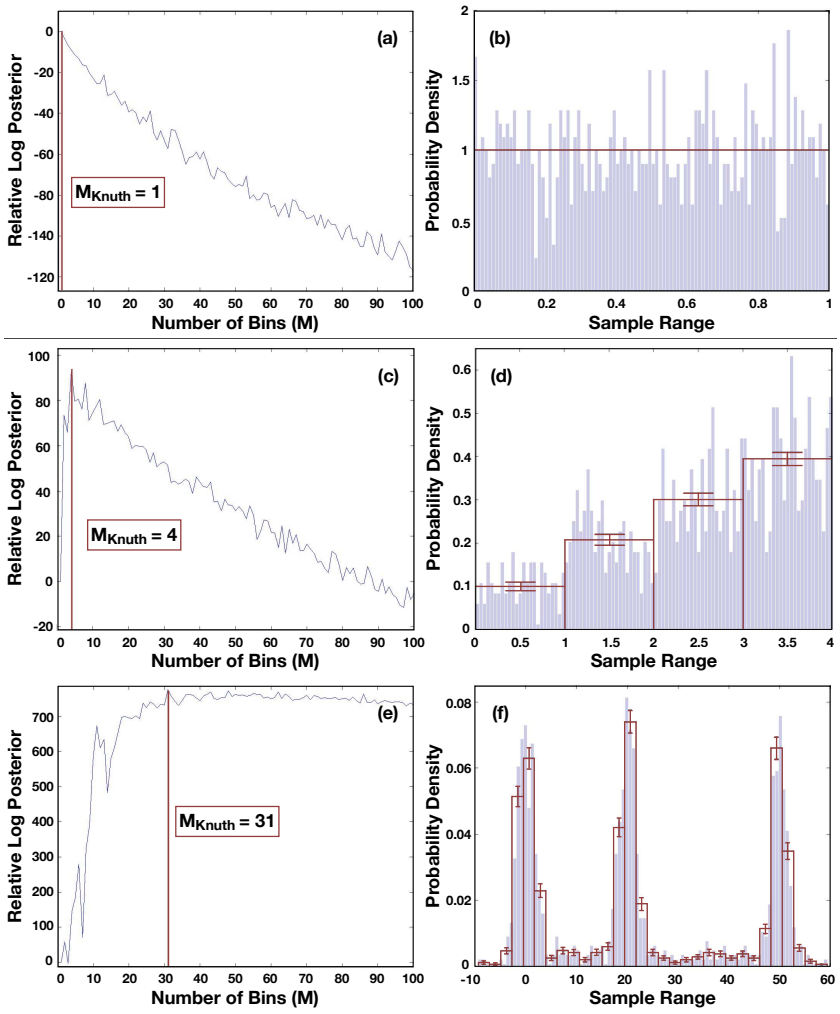


Figure A.1: Application of Knuth's method to three datasets sampled from known distributions: uniform (a,b), four-steps (c,d) and three Gaussian over a uniform background (e-f). On the left column we plot the graph of the relative log posterior and the point $\hat{M}$ where it attains the maximum (Knuth optimal binning number). On the right column we arrange the data in an histogram with that number of bins. To highlight the fact that the method is able to avoid sampling fluctuations we plot, under each optimal histogram, the one we would get by using 100 bins instead of $\hat{M}$, which therefore better shows the spatial distributions of the data sets.

In the second example, the set of data is sampled from a four-steps density function, another case for which other used binning methods cannot work (Stone, 1984; Freedman and Diaconis, 1981). Again our estimation does match the true density, since the logarithm of the relative posterior peaks at four bins.
In the last row of figure A.1 we show the results we obtain by applying the method to a set of data sampled from a density function consisting of three Gaussian peaks over a uniform background. As we can see from the left panel, the logarithm of the relative posterior probability peaks at 31 bins, which capture very well the structure of the underlying density function.
We notice that in this last case the algorithm gives an optimal binning number higher than the ones we have obtained in the previous examples. This is due to the fact that, since we are imposing that the bins have equal width v, we need a high number of bins in order to capture the real structure of the underlying density function and to identify the three peaks, in spite of the uniform background, usually arranged within only one bin, as we have seen in the first example.
This fact suggests that the number of bins given by the method could be a good parameter to distinguish clustered data from uniform ones, since the more structured is a species, the higher the number of bins we need to capture it correctly.
Moreover, we can notice that the logarithm of the relative posterior decreases much slower than the previous ones, so that each value $M \geq \hat{M}$ could be taken as the number of bins of our histogram in order to faithfully approximate the underlying density function. This is a general feature that we can observe when we are dealing with data distributions characterised by the presence of clusters, such as the three Gaussian peaks of the last example.

A.1.2 Two-dimensional datasets

Since our goal is to apply Knuth's method on spatial ecological datasets, we also test it with data sampled from two known bi-dimensional density functions: a uniform and a bivariate normal distribution.
From Section 2.1, we know that for a one-dimensional set of data uniformly distributed, the answer Knuth's algorithm gives is $\hat{M} = 1$, which correctly captures the underlying distribution. We wish to see if the answer for the two-dimensional datasets is the same. Moreover, in order to test the stability of the method, we generate 200 datasets consisting each of 1000 points within a 500×500 units sampled from a uniform density function. We then compute, for each generated dataset, Knuth's optimal binning number.
The results are shown in figure A.2, where we can see that for almost all of the 200 tests we have performed the optimal binning number results to be $\hat{M} = [1\ 1]$, leading to a perfect estimate of the true underlying density function from which the data were sampled.
Thus, Knuth's method results to be very stable with respect to the data generating distribution and not sensitive to its particular realisations. We now test Knuth's method to a dataset consisting of 10000 points distributed according to a bivariate normal distribution with zero correlation coefficient and the same standard deviation along the axes: $\sigma_X = \sigma_Y$. Our results are presented in figure A.3b.

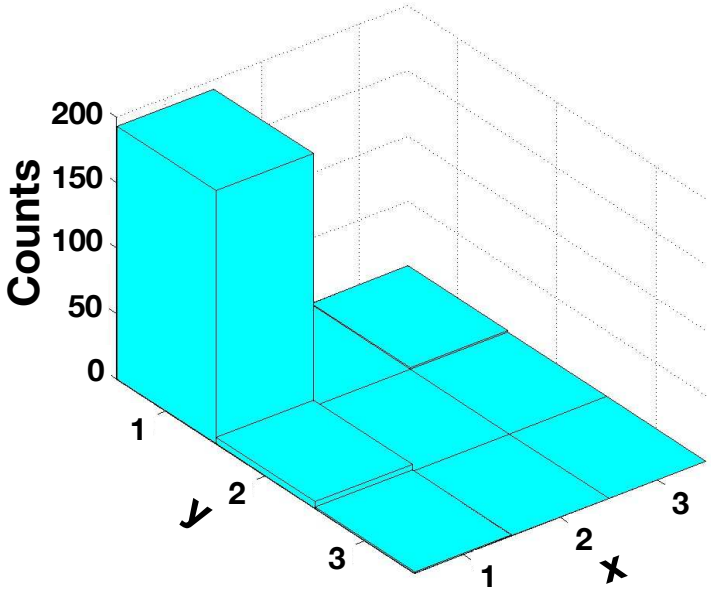
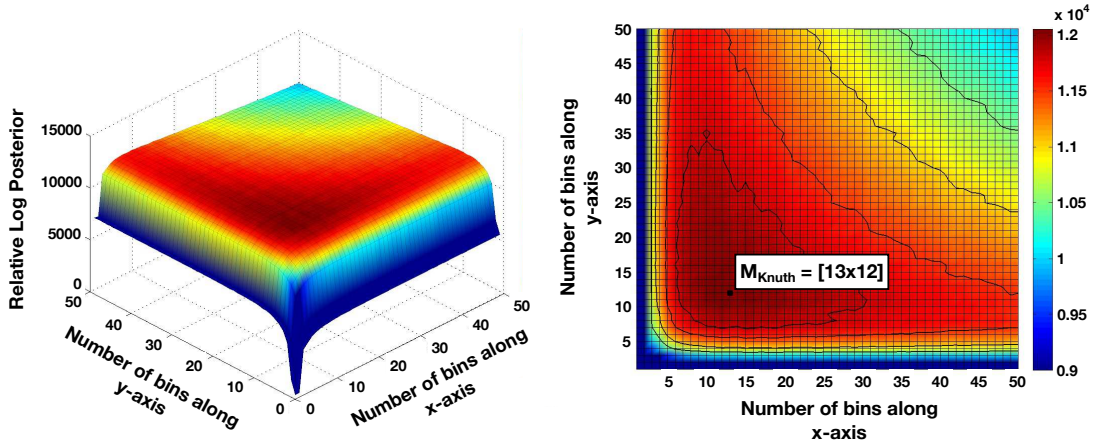


Figure A.2: Histogram of the optimal binning number obtained with Knuth's algorithm tested 200 times on a uniformly distributed dataset

In panel (a), we plot the 3D graph of the logarithm of the relative posterior probability, while in panel (b) we insert the corresponding contour plot. Here we also point out Knuth's optimal number of bins where the maximum is reached, which is 13×12 (black square in the graphic).
In the two graphics below we build the histogram representing our piece-wise constant estimate of the true density function and the spatial data distribution into a 50×50 bins histogram to show how Knuth's method is able to detect the major structure of the dataset without being affected by sampling noise.

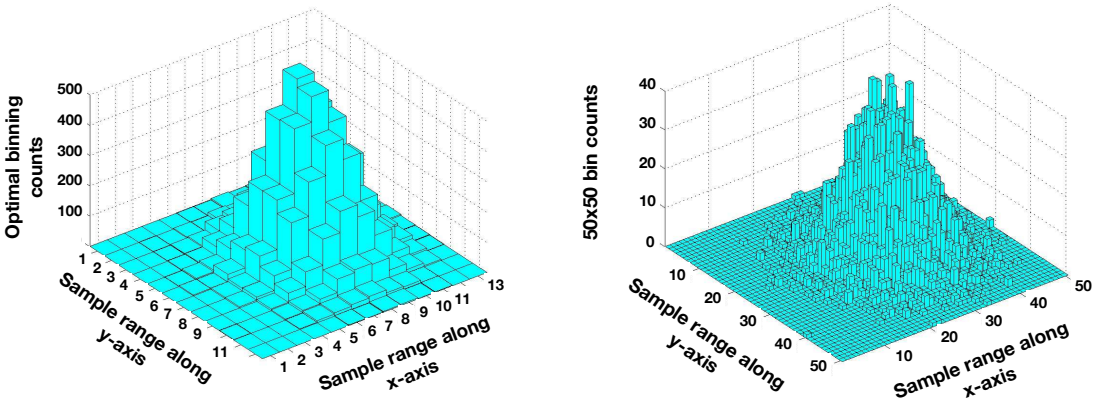
A.2 Comparison with Stone's non-kernel method

We now introduce a non-kernel method developed by Charles J. Stone in 1984 (Stone, 1984) to select the optimal number of bins which gives the best asymptotically approximation of the data density function, assumed to be known. Actually, we will see that it is not necessary to have its analytical formula, but only to know that it satisfies a particular condition. This means that the real drawback of the method is the fact that the answer it gives is optimal only when the number of data tends to infinity, while the datasets which we are usually dealing with are finite.
We will compare Knuth's and Stone's method to show that, even if the true density function is known a priori, Knuth's binning rule is more efficient in avoiding sample fluctuations.



(a) Plot of the logarithm of the relative posterior

(b) Contour Plot



(c) Histogram with Knuth's optimal number of bins

(d) Spatial data distribution into 50×50 bins

Figure A.3: Results for 10000 data sampled from a two-dimensional Gaussian density function. As for the one-dimensional case, Knuth's method correctly estimates the underlying density function from which the data were sampled avoiding noisy fluctuations due to the sampling, which are, at contrast, visible when setting a higher number of bins (see last panel).

A.2.1 Stone's optimal selection rule

As before, let us assume that we have a set of N data $\underline{d}_i \in \mathbb{R}^n$, $i = 1, \ldots, N$, sampled from a known probability density function p.
Let us denote with $\underline{a} = (a_1, ..., a_n) \in \mathbb{R}^n$ the point in the space where we wish to start building our histogram and with $\underline{b} = (b_1, ..., b_n) \in \mathbb{R}^n_+$ the dimension of the n-dimensional bin. Our histogram will therefore depends on the pair of vectors $(\underline{a}, \underline{b})$. We wish now to compute the value of the piece-wise constant probability density described by the histogram. In particular, let us denote with $\underline{k} = (k_1, ...k_n) \in \mathbb{Z}^n$ the integer coordinates, starting from $\underline{a}$ of the $\underline{k}^{th}$ bin. This latter will thus occupy,

in $\mathbb{R}^n$, the place

$$I_{(\underline{a},\underline{b}),\underline{k}} = (a_1 + (k_1 - 1)b_1,\ a_1 + k_1 b_1) \times ... \\ \times (a_n + (k_n - 1)b_n,\ a_n + k_n b_n).$$

We have obtained a partition of $\mathbb{R}^n$ into identical bins of dimension

$$v_{(\underline{a},\underline{b})} = \prod_{i=1}^{n} b_i.$$

The probability mass of the $\underline{k}^{th}$ bin, i.e. the volume of the column over it, is given by the empirical distribution

$$\pi_{(\underline{a},\underline{b}),\underline{k}} = \frac{1}{N} |\{i | 1 \leq i \leq N,\ \underline{d}_i \in I_{(\underline{a},\underline{b}),\underline{k}}\}|.$$

From this we can compute the height of the column, which is simply

$$h_{(\underline{a},\underline{b}),\underline{k}} = \frac{\pi_{(\underline{a},\underline{b}),\underline{k}}}{v_{(\underline{a},\underline{b})}}.$$

The probability density function of the histogram is therefore given by

$$h_{(\underline{a},\underline{b})}(x) = \sum_{\underline{k}} h_{(\underline{a},\underline{b}),\underline{k}} \chi_{I_{(\underline{a},\underline{b}),\underline{k}}}(x),$$

where $\chi_{I_{(\underline{a},\underline{b}),\underline{k}}}$ is the characteristic function of $I_{(\underline{a},\underline{b}),\underline{k}}$, defined as

$$\chi_{I_{(\underline{a},\underline{b}),\underline{k}}}(x) = \begin{cases} 0 & \text{if } x \notin I_{(\underline{a},\underline{b}),\underline{k}} \\ 1 & \text{if } x \in I_{(\underline{a},\underline{b}),\underline{k}} \end{cases}$$

Similarly we can define the probability mass as follows:

$$\pi_{(\underline{a},\underline{b})}(x) = \sum_{\underline{k}} \pi_{(\underline{a},\underline{b}),\underline{k}} \chi_{I_{(\underline{a},\underline{b}),\underline{k}}}.$$

In order to best approximate the density function from which the data were sampled, Stone's idea relies on minimising the integrated squared error of $h_{(\underline{a},\underline{b})}$

$$E_{(\underline{a},\underline{b})} = \int_{\mathbb{R}^n} (h_{(\underline{a},\underline{b})}(\underline{x}) - p(\underline{x}))^2 d\underline{x}. \tag{A.1}$$

Under some conditions (see Stone, 1984), minimising eq. (A.1) is asymptotically equivalent to minimising the following quantity

$$K_{(\underline{a},\underline{b})} = \frac{1}{v_{(\underline{a},\underline{b})}} \left(\frac{2}{N} - \sum_{\underline{k}} \pi^2_{(\underline{a},\underline{b}),\underline{k}} \right).$$

A.2.2 Comparison for one-dimensional datasets

Let us firstly remark that the condition above is satisfied, for example, if there is some non-empty open subset of $\mathbb{R}^n$ on which the derivative of p exists, it is continuous and non-zero (Stone, 1984). As a consequence, if we are dealing with uniform or step density functions, such as we did in the previous section with Knuth's method, we cannot apply Stone's optimal selection rule, which has therefore a smaller range of applications. This is a drawback especially when working with a real dataset sampled from an unknown density function, since we cannot exclude a priori such densities.
We compare the two methods in the one-dimensional case using L^1 and L^2 distance method (Gomes-Gonçalves et al., 2014), defined, respectively, by

$$L^1 = \int_{-\infty}^{x_0} |p(x)| \mathrm{d}x + \sum_{k=0}^{M-1} \int_{x_k}^{x_{k+1}} |p(x) - h_{k+1}| \mathrm{d}x + \int_{x_M}^{+\infty} |p(x)| \mathrm{d}x$$

and

$$L^2 = \sqrt{\int_{-\infty}^{x_0} |p(x)|^2 \mathrm{d}x + \sum_{k=0}^{M-1} \int_{x_k}^{x_{k+1}} |p(x) - h_{k+1}|^2 \mathrm{d}x + \int_{x_M}^{+\infty} |p(x)|^2 \mathrm{d}x},$$

where $x_0, \ldots, x_M$ are the edges of the M bins of the histogram, h_k is the height of the k^{th} bin with edges (x_{k-1}, x_k) and $p(x)$ is the value at x of the true density function from which data are sampled.
We generate 50 datasets each consisting of 1000 one-dimensional points sampled from a standard Gaussian density function $\mathcal{N}(0, 1)$. For each test, we compute Knuth's and Stone's optimal binning numbers and then we construct the correspondent histograms. We finally compare the L^1 and L^2 distances between the underlying density function and its piece-wise constant estimates.
In panel (a) of figure A.4 we plot the ratio between Stone's and Knuth's L^1 distance obtained for each test, while in panel (b) figure A.4 we plot the logarithm of the ratio between Stone's and Knuth's L^2 distance.
As we can see, while the two methods are practically equivalent in the first case, Knuth's L^2 distance is about ten times smaller than Stone's one, due to the fact that this latter method is more sensitive to sample fluctuations. The result may be surprising, since Stone's optimal selection rule relies on minimising exactly the L^2 distance between the underlying density function and the one obtained by the histogram. The key point is that the rule only assures that this minimum is reached when the number of data tends to infinity. In fact, it is in this asymptotic case that the sampling fluctuations diminish and thus the histogram recalls quite faithfully the distribution from which the data are sampled.
For the sake of completeness, we also insert the graphics we obtain testing the two methods on one dataset such as the ones described above (panels (c-d) of figure A.4). From bottom panels of figure A.4, it is clear that while Knuth's histogram correctly captures the main characteristics of the underlying probability density function, Stone's rule is more sensitive to noisy fluctuations due to sampling, which results in a higher optimal binning number.

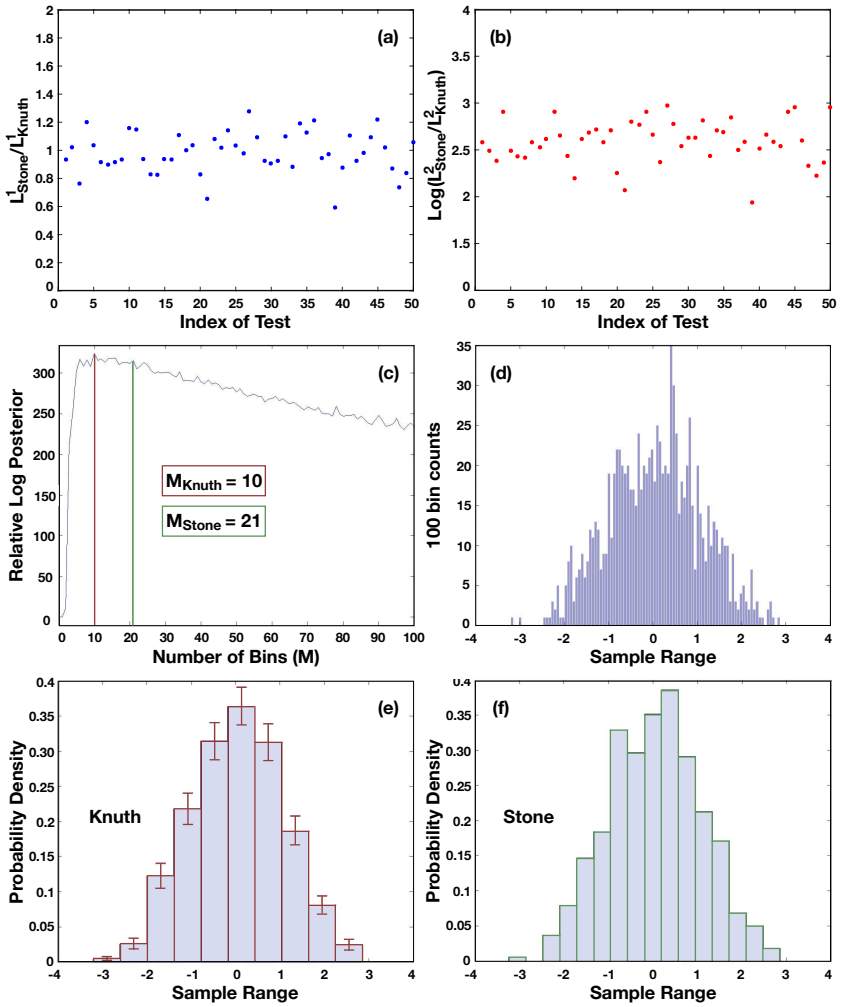


Figure A.4: L^1 and L^2- comparisons between Stone's and Knuth's methods on 50 datasets generated from a standard Gaussian distribution $\mathcal{N}(0,1)$. While the two methods give comparable results according to the L^1 test, Knuth's L^2 distance between the normal density and its estimate through the optimal histogram results ten times smaller with respect to the one of Stone. In (c-f) we see an example of Knuth's and Stone's answers to the same dataset. As we can see from the histograms, Knuth's method is able to avoid the noisy fluctuations due to the sampling, contrarily to Stone's rule. Knuth's method is therefore more efficient in the sense that Stone's histogram results often sub-optimal since it highlights random variations that are not representative of p.

B

Alpha, beta and gamma-diversity

B.1 From diversity indexes to diversity: Hill's numbers

In Section 3.2 we have introduced the most important alpha-diversity indexes proposed in literature. All of them have the property of being monotonic functions of $\sum_{s=1}^{S} p_s^q$ (or limits of such functions, as in the case of Shannon's information), for a given q called the *order* of the index. Nevertheless, they differ, as we will later see, in their sensitivity to common and rare species. Moreover, they may give counter-intuitive answers when we consider the row index computed for a given community as its diversity value (Jost, 2007). For example, let us consider a community $\mathcal{C}_1$ having S_1 equally-common species of abundance n. Now, let us add other S_2 species to $\mathcal{C}_1$, each of which having $n/2$ individuals. We call $\mathcal{C}_2$ this new community. It would be natural to say that $\mathcal{C}_2$ is more diverse than $\mathcal{C}_1$. Moreover, if $S_2 = S_1 = S$, we would spontaneously declare that the second community is twice as diverse than the first. We call this the *doubling property* of the diversity (Hill, 1973; Jost, 2006). One can think to the second community as obtained from the first by splitting each species into two equal groups (e.g. males and females).
But let us compute the indexes of similarity introduced in Chapter 3 for both $\mathcal{C}_1$ and $\mathcal{C}_2$.
Clearly, the relative abundance vector of $\mathcal{C}_1$ is $p^{(1)} = (p_1^{(1)}, \dots, p_S^{(1)}) = (1/S, \dots, 1/S)$.
Analogously, for $\mathcal{C}_2$ we have that $p^{(2)} = (p_1^{(2)}, \dots, p_{2S}^{(2)}) = (1/2S, \dots, 1/2S)$.
Let $S = 500$ and $n = 2$ (easiest case where all species are doubletons for $\mathcal{C}_1$ and

singletons for $\mathcal{C}_2$). Simspon's diversity index applied to $\mathcal{C}_1$ and $\mathcal{C}_2$ leads to

$$D^{(1)} = 1 - \sum_{s=1}^{S} \frac{n(n-1)}{N(N-1)} = 1 - \frac{1}{2S-1} = 0.999, \tag{B.1}$$

$$D^{(2)} = 1 - \sum_{s=1}^{2S} \frac{\frac{n}{2}(\frac{n}{2}-1)}{N(N-1)} = 1. \tag{B.2}$$

Therefore, both the communities have practically the same diversity according to the index.
By applying Simpson's dominance index we get an even odder result:

$$D'^{(1)} = \sum_{s=1}^{S} \left(\frac{1}{S}\right)^2 = \frac{1}{S} = 0.002 \tag{B.3}$$

$$D'^{(2)} = \sum_{s=1}^{2S} \left(\frac{1}{2S}\right)^2 = \frac{1}{2S} = 0.001. \tag{B.4}$$

The second community results to be half as diverse as the first, which is absolutely absurd at first sight if we take those values as measure of the communities' diversity. At contrast, it is perfectly reasonable if we remember that D' expresses the probability that two randomly taken individuals belong to the same species, which is clearly bigger for the the first community.
Let us continuing computing Gini-Simpson's and Shannon's information index for $\mathcal{C}_1$ and $\mathcal{C}_2$

$$D''^{(1)} = 1 - D'^{(1)} = 0.998 \tag{B.5}$$

$$D''^{(2)} = 1 - D'^{(2)} = 0.999 \tag{B.6}$$

$$H^{(1)} = -\sum_{s=1}^{S} p_s^{(1)} \ln p_s^{(1)} = \ln S = 6.215 \tag{B.7}$$

$$H^{(2)} = -\sum_{s=1}^{2S} p_s^{(2)} \ln p_s^{(2)} = \ln 2S = 6.908 \tag{B.8}$$

In both cases, the ratio between the diversity index for the first community and the one for the second community is much higher than the expected (99% for Gini-Simpson and 90% for Shannon).
How can we then reconcile our intuitive idea of the concept of diversity with the results given by measuring it through all these different indexes?
Following Hill's idea (Hill, 1973), Jost (Jost, 2006, 2007) proved that a reconciliation is possible if one does pay attention on the difference between the concept of *diversity* and the one of *diversity index*. Such notions had been widely confused before by ecologists, even leading some authors to conclude that diversity was a non-concept (Hurlbert, 1971). Moreover, he showed that all such indexes can be written with a unifying notation.
The basic intuition of Jost was the following: the only index which seems to respect our intuition for the considered example is the species richness. Therefore, in the case of an equally-distributed community, its diversity should be equal to the number of its species.

Here is the key point then. A diversity index of order q should be thought as an equivalence relation which groups communities in classes (Jost, 2006). The representative of each class is a number ${}^qD_\alpha$ called the *effective number of species* (MacArthur, 1965) which is the number of equally-common species that a community must have in order to share the value of the diversity index with all the other components of such class. We add a subscript α to highlight the fact we are dealing with alpha-diversity.
To better understand this tricky point, let us denote with I_q any diversity index of order q having the property of being a continuous and monotonic function of $\sum_{s=1}^{S} p_s^q$. Let us assume that, for a given community with S species and vector of abundances $(p_1, \dots, p_S)$, the index takes value equal to x: $I_q(\sum_{s=1}^{S} p_s^q) = x$. To find the effective number of species of such community, i.e. the representative of its equivalence class according to I_q, we must find a number ${}^qD_\alpha$ of equally-common species having the same value of the diversity index. Therefore, we have to impose that the following relation holds:

$$x = I_q\left(\sum_{s=1}^{{}^qD_\alpha} \frac{1}{({}^qD_\alpha)^q}\right).$$

Now, since I_q is assumed to be continuous and monotonic, it is also invertible, so that we can write

$$I_q^{-1}(x) = \sum_{s=1}^{{}^qD_\alpha} \frac{1}{{}^qD_\alpha^q} = \frac{1}{({}^qD_\alpha)^q} \cdot {}^qD_\alpha = \frac{1}{({}^qD_\alpha)^{q-1}},$$

from which we get

$${}^qD_\alpha = [I_q^{-1}(x)]^{\frac{1}{1-q}} = \left[I_q^{-1} I_q\left(\sum_{s=1}^{S} p_s^q\right)\right]^{\frac{1}{1-q}} = \left(\sum_{s=1}^{S} p_s^q\right)^{\frac{1}{1-q}}. \tag{B.9}$$

Since the above expression has been obtained without any hypothesis on either I_q or q, it must hold for any diversity index and any order. Therefore, it is a unifying expression for all diversity indexes. As in Jost, 2006, we will denote the number defined in B.9 with ${}^qD_\alpha$, making explicit the diversity order q. Let us remark that such quantity had been previously introduce by Hill (Hill, 1973), so that the quantity $(\sum_{s=1}^{S} p_s^q)^{\frac{1}{1-q}}$ is also known in literature as *Hill's number* of order q (Heip et al., 1998; Magurran, 1988). Let us see an example.

Example B.1 (Gini-Simpson's index)**.** Let us see how to pass from the diversity of a community to its measure through the Gini-Simpson's index defined in 3.3, which we have denoted with D''. In order not to confuse it with the diversity of the community, we will change its notation in this example into $I_{2,Gini-Simp}$, in coherence with the notation adopted above. Let thus S be the number of species of the community and $(p_1, \dots, p_S)$ their abundance vector. According to 3.3, Gini-Simpson's index for the community is given by

$$I_{2,Gini-Simp}\left(\sum_{s=1}^{S} p_s^2\right) = 1 - \sum_{s=1}^{S} p_s^2$$

To find the corresponding diversity in term of number of equivalent species, we set

$$1 - \sum_{s=1}^{S} p_s^2 = 1 - \sum_{s=1}^{{}^2D_\alpha} \left(\frac{1}{{}^2D_\alpha}\right)^2 = 1 - \frac{1}{{}^2D_\alpha}.$$

Therefore, we have that

$${}^2D_\alpha = \frac{1}{\sum_{s=1}^{S} p_s^2}.$$

We also remark that the diversity defined in B.9 satisfies the intuitive property called the *doubling property* described in the previous section. Let us consider, as before, two communities $\mathcal{C}_1$ and $\mathcal{C}_2$ with total number of species S and $2S$ and with vector of abundances $(p_1, p_2, \dots, p_S)$ and $(p_1/2, p_1/2, p_2/2, p_2/2, \dots, p_S/2, p_S/2)$, respectively. Since it is reasonable to assert that the second community is twice as diverse as the first one, then the effective number of species of the second should be twice the number of the first, regardless of the diversity order q under study. Let us see that this intuition holds. Let us denote with ${}^qD_\alpha^{(1)}$ and ${}^qD_\alpha^{(2)}$ the diversities of order q of the two communities. Then we have the following relation:

$$\begin{aligned}
{}^qD_\alpha^{(2)} &= \left(\sum_{t=1}^{2S} p_t^q\right)^{\frac{1}{1-q}} = \left(2\sum_{s=1}^{S}\left(\frac{p_s}{2}\right)^q\right)^{\frac{1}{1-q}} \\
&= \left(2^{1-q}\sum_{s=1}^{S} p_s^q\right)^{\frac{1}{1-q}} = 2\left(\sum_{s=1}^{S} p_s^q\right)^{\frac{1}{1-q}} \\
&= 2\ {}^qD_\alpha^{(1)}.
\end{aligned}$$

We thus have the reconciliation we were looking for. Indeed, let us fix $S = 500$ and compute the diversity index for $\mathcal{C}_1$ and $\mathcal{C}_2$ through eq. (B.9) for all the introduced indexes. The indexes of Simpson's dominance and of Gini-Simpson are of order 2. Thus the diversity associated to them are equal and can be computed as follows

$${}^2D_\alpha^{(1)}\left(\sum_{s=1}^{S} p_s^2\right) = \frac{1}{\sum_{s=1}^{S} p_s^2} = S = 500$$

$${}^2D_\alpha^{(2)}\left(\sum_{s=1}^{2S} p_s^2\right) = \frac{1}{\sum_{s=1}^{2S} p_s^2} = 2S = 1000.$$

The effective number of species of the communities are equal to their actual number and therefore, as we expect by definition of the diversity ${}^qD_\alpha$. Thus, the diversity of $\mathcal{C}_2$ is twice the one of $\mathcal{C}_1$.
Shannon's information is, instead, of order 1, so we have to compute ${}^1D^{(1)}$ and ${}^1D^{(2)}$:

$${}^1D_\alpha^{(1)}\left(\sum_{s=1}^{S} p_s\right) = \lim_{q\to 1}\left(\sum_{s=1}^{S} p_s\right)^{\frac{1}{1-q}} = \exp\left(-\sum_{s=1}^{S} p_s \ln p_s\right) = \exp(\ln S) = S = 500,$$

$${}^1D_\alpha^{(2)}\left(\sum_{s=1}^{2S} p_s\right) = \lim_{q\to 1}\left(\sum_{s=1}^{2S} p_s\right)^{\frac{1}{1-q}} = \exp\left(-\sum_{s=1}^{2S} p_s \ln p_s\right) = \exp(\ln 2S) = 2S = 1000.$$

Again we have obtained the wished result.
In this trivial case, the diversity of all order q are equal since the assumption of an equally-distributed community. In general, all the indexes of the same order lead to the same value of the diversity, while differing with respect to the measure computed via indexes of other order.
For example, consider the diversities of order 0, 1 and 2. They can all be seen as mean values of the relative abundances p_s weighted by these latter themselves. Indeed, let us call $w_s = p_s$ and consider it as the weight associated to the s^{th} relative abundance. It can be shown that the following relations hold (Hill, 1973)

$$\frac{1}{{}^0D_\alpha(\sum_s^S p_s^0)} = \frac{w_1 + \cdots + w_S}{\frac{w_1}{p_1} + \cdots + \frac{w_S}{p_S}} \tag{B.10}$$

$$\frac{1}{{}^1D_\alpha(\sum_s^S p_s)} = \sqrt[w_1+\cdots+w_S]{p_1^{w_1} + \cdots + p_s^{w_s}} \tag{B.11}$$

$$\frac{1}{{}^2D_\alpha(\sum_s^S p_s^2)} = \frac{w_1 p_1 + \cdots + w_S p_S}{w_1 + \cdots + w_S}. \tag{B.12}$$

Thus, the inverse of ${}^0D$, ${}^1D$ and ${}^2D$ are, respectively, the harmonic, the geometric and the arithmetic mean of the relative abundances weighted by the vector $(w_1, \ldots, w_s) = (p_1, \ldots, p_S)$. Since the geometric mean is always bigger than the harmonic and smaller than the arithmetic one, we have that, in general

$${}^0D_\alpha \ < \ {}^1D_\alpha \ < \ {}^2D_\alpha.$$

We point out that the order q of a diversity index represents how much it is influenced by the presence of common or rare species: the higher the value of q, the more sensitive is the index with respect to common species, while the lower q, the more it is affected by rarest species (Hill, 1973; Jost, 2006; Tsallis, 2001; Keylock, 2005). The threshold case $q = 1$ is not affected by commonness and rarity making no favours in relation to the abundance. This characteristic of Shannon's information index has made it particularly suitable to measure diversity in different scientific fields (Jost, 2007).

B.2 Partition of the gamma-diversity of a community

It is frequent in ecology to work with database consisting on a high number of sample units scattered within a large landscape or region covered by different habitat, each one having a particular alpha-diversity measured with an index as the ones introduced in the above section. According to Whittaker's inventory diversity scheme (Whittaker, 1972), in this case we speak about gamma-diversity of the pooled communities.
In order to define the total diversity of order q of the set of samples, ${}^qD_\gamma$, we first

need some notations.
Let U be the number of sampling units of the landscape and S be the total number of species who can be found within them. We denote with $p_{s,u}$ the fraction of individuals belonging to the s^{th} species within the u^{th} sample. Let us assign a weight w_u to each sampling unit according to an arbitrary fixed criterion which can takes into account the importance or the sample size of the units. We then define the weighted sum $\overline{p_s} = \sum_{u=1}^{U} w_u p_{s,u}$. We are now ready to give the definition of the gamma-diversity of order q of the pooled communities D_γ:

$$ {}^qD_\gamma = \left(\sum_{s=1}^{S} \overline{p_s}^q \right)^{\frac{1}{1-q}}. $$

In order to measure it, we can resort to any index I_q as the ones introduced in the above section with the additional hypothesis of concavity (Lewontin, 1972; Lande, 1996; Jost, 2006). Both Gini-Simpson's index and Shannon's information satisfy this property (Rényi, 1961; Aczél and Daróczy, 1975). This ulterior condition guarantees that the value of the gamma-diversity index computed for the pooled communities is greater than the weighted average of the alpha-diversity index computed for its components:

$$ I_q\left(\sum_{s=1}^{S} \overline{p_s}^q \right) = I_q\left(\sum_{s=1}^{S} \left(\sum_{u=1}^{U} w_u p_{s,u} \right)^q \right) \geq \sum_{u=1}^{U} w_u I_q\left(\sum_{s=1}^{S} p_{s,u}^q \right). $$

The problem of partitioning biodiversity has been deeply studied in literature (Jost, 2007, 2010a; Veech and Crist, 2010a,b; Baselga, 2010; Tuomisto, 2010; Buzas and Hayek, 1996; Ricotta, 2010, 2005; Lewontin, 1972).
Whittaker noticed that the diversity of a community should follow a *multiplicative law*, where ${}^qD_\gamma$ is given by the product of two independent terms, one due to the average value of any alpha-diversity index of order q over the sampling units and another one due to the beta-diversity among them (Whittaker, 1972):

$$ {}^qD_\gamma = {}^q D_\alpha \cdot {}^qD_\beta. \tag{B.13} $$

As for ${}^qD_{beta}$, Jost gave an interpretation also for the beta-diversity in terms of effective number of species: indeed, it represents the number of distinct communities (with no common species between them) of which the ecosystem consists.
Thanks to the relation between ${}^qD$ and I_q found by Jost, it is then possible, starting from eq. (B.13), to deduce all the other laws found in literature for the diversity indexes introduced in Chapter 3, as the additive-law of Shannon's information or the multiplicative law of species richness (Jost, 2007).
Let us continue example B.1 to see how to obtain the rule for Gini-Simpson's index from eq. (B.13).

Example B.2 (Gini-Simpson's index)**.** Let $I_{2,Gini-Simp}$ denote Gini-Simpson's index. We wish to find the relation between $I^{\alpha}_{2,Gini-Simp}$, $I^{\beta}_{2,Gini-Simp}$ and $I^{\gamma}_{2,Gini-Simp}$, which represent, respectively, the average value of the alpha-diversity index over a set of U sampling units, the beta-diversity index taking into account their differences in species' composition and the gamma-diversity index for the pooled communities

(we drop the dependence on the relative abundances vector by simplicity's sake). We know from example B.1 that the following relation holds

$$^2D = \frac{1}{\sum_{s=1}^{S} p_s^2} = \frac{1}{1 - I_{2,Gini-Simp}}.$$

Therefore, from the multiplicative law of the diversity, we have that

$$^2D_\gamma = {}^2D_\alpha \cdot {}^2D_\beta$$

from which we get

$$I^\gamma_{2,Gini-Simp} = I^\alpha_{2,Gini-Simp} + I^\beta_{2,Gini-Simp} - I^\alpha_{2,Gini-Simp} \cdot I^\beta_{2,Gini-Simp}.$$

We have obtained the additive-multiplicative law between Gini-Simpson's indexes measuring the alpha, beta and gamma diversities.
We can now resort to this result to deduce the beta-diversity index of the community under study.
Let us firstly compute the weighted average of Gini-Simpson's indexes when U communities compose the landscape under study:

$$I^\alpha_{2,Gini-Simp} = \sum_{u=1}^{U} w_u I_{2,Gini-Simp}\left(\sum_{s=1}^{S}(p_s^{(u)})^2\right) = \sum_{u=1}^{U} w_u\left(1 - \sum_{s=1}^{S}(p_s^{(u)})^2\right).$$

The gamma-diversity index of the pooled communities can be similarly computed:

$$I^\gamma_{2,Gini-Simp} = I_{2,Gini-Simp}\left(\sum_{s=1}^{S}\overline{p_s}^2\right) = 1 - \sum_{s=1}^{S}\overline{p_s}^2 = 1 - \sum_{s=1}^{S}\left(\sum_{u=1}^{U} w_u p_{s,u}\right)^2.$$

We can therefore compute the beta-diversity between communities by

$$I^\beta_{2,Gini-Simp} = \frac{I^\gamma_{2,Gini-Simp}}{I^\alpha_{2,Gini-Simp}} = \frac{1 - \sum_{s=1}^{S}\left(\sum_{u=1}^{U} w_u p_{s,u}\right)^2}{\sum_{u=1}^{U} w_u\left(1 - \sum_{s=1}^{S}(p_s^{(u)})^2\right)}.$$

C

From Darwin to Hubbell: an historical review

Here we propose a brief historical synopsis about the fundamental scientific steps which led from *The Origin of Species* of Darwin to the *The Unified Neutral Theory of Biodiversity and Biogeography* of Hubbell (see figure C.1).

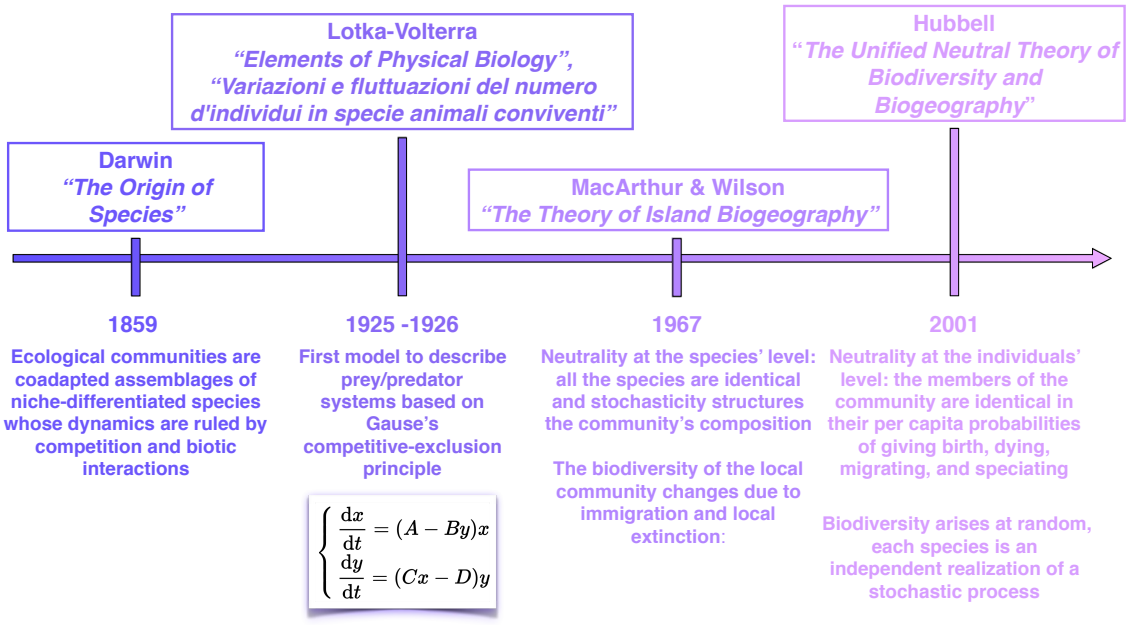


Figure C.1: Fundamental steps from Darwin's Evolution Theory and Hubbell's Neutral Theory.

Darwin's Theory of Evolution The English naturalist Charles R. Darwin was the first scientist aiming at understanding what are the ecological mechanisms on work under the recurrent patterns which emerge across wide scales of time and space, when investigating living ecosystems. Back in 1859, he formulated the hypothesis that ecological communities are the result of a long history of mutual interactions and competition between species, whose individuals have been struggling for their own life (Darwin, 1859). The species currently present have prevailed over the others thanks to their ability to evolve so to better adapt in a well defined niche of the community. This individuals' characteristic confers to the system a great resistance to external perturbations or disturbances and guarantees a stable equilibrium between its members.

The model of Lotka and Volterra The first models aiming at describing the dynamics of two interacting species (prey vs predator or resource vs consumer) were developed, within only a year one from the other and yet independently, by two mathematicians, the American Alfred J. Lotka (Lotka, 1956) and the Italian Vito Volterra (Volterra, 1927). Both their models are based on the competitive-exclusion principle, also known as Gause's law, which states that if two different (not interbreeding) and sympatric species compete for the same limiting resources of an ecological niche, necessarily the one which has even a slight advantage over the other will leads either to the extinction of the other species or to its evolutionary or behavioural shift towards a different ecological niche, where it may dominate over a third, more disadvantaged, species (Hardin, 1960). Let us denote with $x(t)$ the number of prey at time t and with $y(t)$ the number of predators at the same time. Let us also make the following assumptions:

- there are no other prey to be hunt by the predators
- the prey can always count on a limitless food supply in the territory
- each time that a prey encounter a predator, this latter eats it
- the rate of change of the population of each species is proportional to its size.
- no other environmental or biotic factor affects the dynamics during the process.

Translating the above assumptions into a mathematical framework, we find that the dynamic of Lotka and Volterra's model is determined by the following pair of non-linear differential equations of the first order:

$$\begin{cases} \dfrac{\mathrm{d}x}{\mathrm{d}t} = (A - By)x \\ \dfrac{\mathrm{d}y}{\mathrm{d}t} = (Cx - D)y. \end{cases} \tag{C.1}$$

Parameters A, B, C and D in eq. (C.1) are positive constants describing the interactions between the two species and thus regulating their evolution. In particular, in the first equation, A is the population growth rate of the prey, which are assumed to reproduce exponentially if they were not subject to predation and B is the rate

of predation, assumed to be proportional to the number of times a prey encounter a predator (thus to the product of both their populations). In the second equation, C is the population growth rate of the predators, again proportional to the number of time they encounter a prey and D is the population loss rate of the predators, which would exponentially die or migrate if there were no prey to eat in the territory under study.
Since we are using differential equations to describe the system's dynamics, we know that there exists a deterministic and continuous solution of eq. (C.1). If none of the parameters are zero, one finds a periodic solution, where the number of the predators grows as long as there is plentiful prey to hunt, whereas it starts diminishing as soon as their appetite cannot be satisfied for lack of food supply. At the same time, as the predators either migrate or die, the number of prey starts to grow since they now rarely come across their hunters. When their number is huge, the predators begin to thrive once again and the cycle restarts.
Let us also remark that the system has two equilibrium points:

- a saddle point at $x = y = 0$ (both the species are extinct). Since it is unstable, even if both the species are close to disappear, they can still recover, so that the simultaneous extinction phenomenon is rare to happen
- an elliptic point for $x = D/C$ and $y = A/B$, with periodic solutions oscillating around it. This corresponds to the situation where, at each time unit, there is an equal number of prey which are born and eaten by the predators. This number of prey represents the critical supply food threshold that keeps stationary the population of both the prey and the predators.

The idea of MacArthur and Wilson The mainstream perspective derived by Darwin's idea, constituted a solid milestone among the scientific community until the publication, in 1967, of the textbook *The Theory of Island Biogeography* by MacArthur and Wilson, which brought into light a totally new world view on the nature of the ecological communities (MacArthur and Wilson, 2016).
Before going to this important step of our historical journey, we will need to distinguish between two scales of ecological biodiversity.
At a small scale, here we refer to the *ecological community* as a group of trophically similar, sympatric species competing in a local area for the same resources.
At a global scale, we will instead refer to the *metacommunity* as all trophically similar species in a collection of local communities. Unlike species in the local community, metacommunity species may not compete because of separation in space or time.
As an example, one can think as the Soberania National Park, which extends along the shores of the Panama Canal and comprises of approximatively 200 square kilometres as our metacommunity.
Then, as an ecological community, one can consider the Barro Colorado Island, which is located in the middle of Panama Canal and with an area of approximatively 15.6 square kilometres (see figure C.2). Actually, what one has to deal with, are data concerning only a portion of the ecological community. As for the Barro Colorado Island, we have information on all the tree individuals falling within 50

hectares located in the middle of the island.

Figure C.2: Example of the three basic spatial scales of an ecosystem: within the metacommunity (Soberania National Park) species may not interact with each other because of their relative distance, while at the community scale (BCI forest) they are in competition for the same local resources (light, soil, etc.). Finally, the available data are usually compound of one or more sampling units scattered within the ecological community and giving information on the individuals an their species at a local scale.

Differently from Lotka and Volterra's model, we now consider species within the same trophic level, i.e. demanding for the same limiting resources (for plants, these are represented by sunlight, minerals, water, etc.), but not directly representing a danger one to the other.
In their book, MacArthur and Wilson considered, as a metacommunity, an archipelago where the number of species present in one of its islands (the local community) changes due to two factors leading towards opposite directions: immigration from the other islands and local extinction. According to their model, the current species' composition of the local community is not the result of an evolutive history where competition and adaptive behaviours have let some species prevail over some others, but they have been only determined by random dispersal and stochastic local extinction. In their model, the state variable is only the total number of present species, regardless of their label. In this sense, all the species are considered ecologically equivalent, having the same probability of migrate or extinguish, neglecting whether or not they have been showed to actually be the best competitors on the different ecological niche of the island. It was in this context that American ecologist Stephen P. Hubbell proposed, for the first time, the term *neutral* (Hubbell, 2001a). We underline that, despite the appearances, the model proposed by MacArthur and Wilson does not deny the obvious existence of niche-differentiated species. Indeed, while Darwin's theory involve a period of time covering many millennia, the new

model considers a much smaller time span (even if still long), where the competitive aspects of evolution can be assumed to have less influence on the system's dynamics.

Hubbell's Neutral Theory As highlighted by Hubbell in his masterpiece *The Unified Neutral Theory of Biodiversity and Biogeography* of 2001 (Hubbell, 2001a), the concept of neutrality among species preposed by MacArthur and Wilson, even if it laid the foundations of a new fruitful theory, was still lacking in some important aspects of ecology. For example, it does not consider, as a driving factor, the phenomenon of speciation which is observed in real ecosystems. Moreover, since it only looks at the species richness, it cannot predict other fundamental macro-patterns like the species-abundance distribution or the species-area relationship.
The brilliant idea of Hubbell was to transfer the neutrality concept from the species level to the individual level. The term *neutral* here means that all the individuals of the metacommunity are demographically equivalent, in the sense that they have the same per capita probabilities of giving birth, dying, migrating, and speciating. Thus, the presence of a species instead of another one is not due to the biological differences between their members. This implies each species follows a random walk being an independent realisation of a stochastic process. Ecological drift, random migration, and random speciation are the only forces structuring the community's composition and biodiversity.
Let us study in details Hubbell's model.
Let $\mathcal{P}^M_{n,s}(t)$ be the probability that a species s, $s \in \{1, \dots, S^M\}$, has n individuals at time t in the metacommunity, for example the Soberania National Park in Panama. We make the following assumptions:

- *Independence*: there are no interspecific interactions between the individuals
- *Neutrality*: all the individuals are demographically equivalent
- *Stochasticity*: the population dynamics are governed by birth and death processes, where the birth rate also account for speciation

Then, each species is an independent realisation of a stochastic process ruled by two terms, the birth rate $b^M_{n,s} = b^M_n$ and the death rate $d^M_{n,s} = d^M_n$.
The master equation regulating the evolution of $\mathcal{P}^M_{n,s}(t) = \mathcal{P}^M_n(t)$ for $n \geq 0$ is then

$$\frac{\partial}{\partial t}\mathcal{P}^M_n(t) = \mathcal{P}^M_{n-1}(t)b^M_{n-1} + \mathcal{P}^M_{n+1}(t)d^M_{n+1} - \mathcal{P}^M_n(t)b^M_n - \mathcal{P}^M_n(t)d^M_n. \tag{C.2}$$

yielding, for $b^M_{-1} = d^M_0 = 0$, the following stationary solution

$$\mathcal{P}^M_n = \mathcal{P}^M_0 \prod_{i=0}^{n-1} \frac{b_i}{d_{i+1}},$$

where $\mathcal{P}^M_0$ is the normalisation condition guaranteeing that the abundance probabilities $\mathcal{P}^M_n$ sum up to 1.
In Hubbell's model, the birth rate at the metacommunity level is given $b^M_n =$

$b^M n + \delta_{n,0}\nu$, where the ν parameter accounts for speciation. eq. (C.2) then becomes the Fisher log-series (see Chapter 5, Section 5.2.2):

$$\mathcal{P}_n^M = \mathcal{P}_0^M \frac{\nu}{b^M} \frac{x^n}{n},$$

where $x = b^M/d^M$.
Let us denote with ϕ_n^M the random number of species in the metacommunity having a population of n individuals. Its average value gives the SAD of the metacommunity

$$\mathbb{E}[\phi_n] = \sum_{k=1}^{S^M} \mathcal{P}_{n,s}^M = S^M \mathcal{P}_n^M = \theta \frac{x^n}{n},$$

where $\theta = S^M P_0^M \nu / b^M$ is called the *biodiversity parameter*. Finally, let us denote with N^M the expected total number of individuals in the metacommunity, given by

$$N^M = \sum_{n=1}^{\infty} n\mathbb{E}[\phi_n^M] = \frac{\theta x}{1-x}.$$

Let us note that, in order for N^M to be finite, the x parameter should be strictly less than 1.
Let us now change spatial scale by considering a local ecological community within the metacommunity (e.g. the Barro Colorado Island). Moreover, let us add the following hypotheses:

- the evolutionary timescale for the local community is much faster than for the metacommunity whose equilibrium distribution is thus considered negligibly modified
- the population of the local community has reached the saturation level, N^L
- the speciation phenomenon does not affect the dynamics during the process
- the surrounding metacommunity is treated as a permanent source pool of species, which can thus emigrate to the local community

Therefore, the driving factors for the local community population are ecological drift and random migration from the metacommunity.
In figure C.3 we insert a schematic plot of how the model works. Let us analyse it step by step:

1. *Birth of an individual and death of another*: with probability $1-m$, randomly pick two individuals from the local community (the BCI) and look at their species: if they belong to different species, choose at random one of them and replace it with the offspring of the other; if they are conspecific, goes directly to the following step.

2. *Death of an individual and immigration of another*: with probability m, randomly choose a single individual from the local community. Then, pick a species from the surrounding metacommunity with a probability proportional to its abundance and replace the previously chosen individual with the offspring of such species.

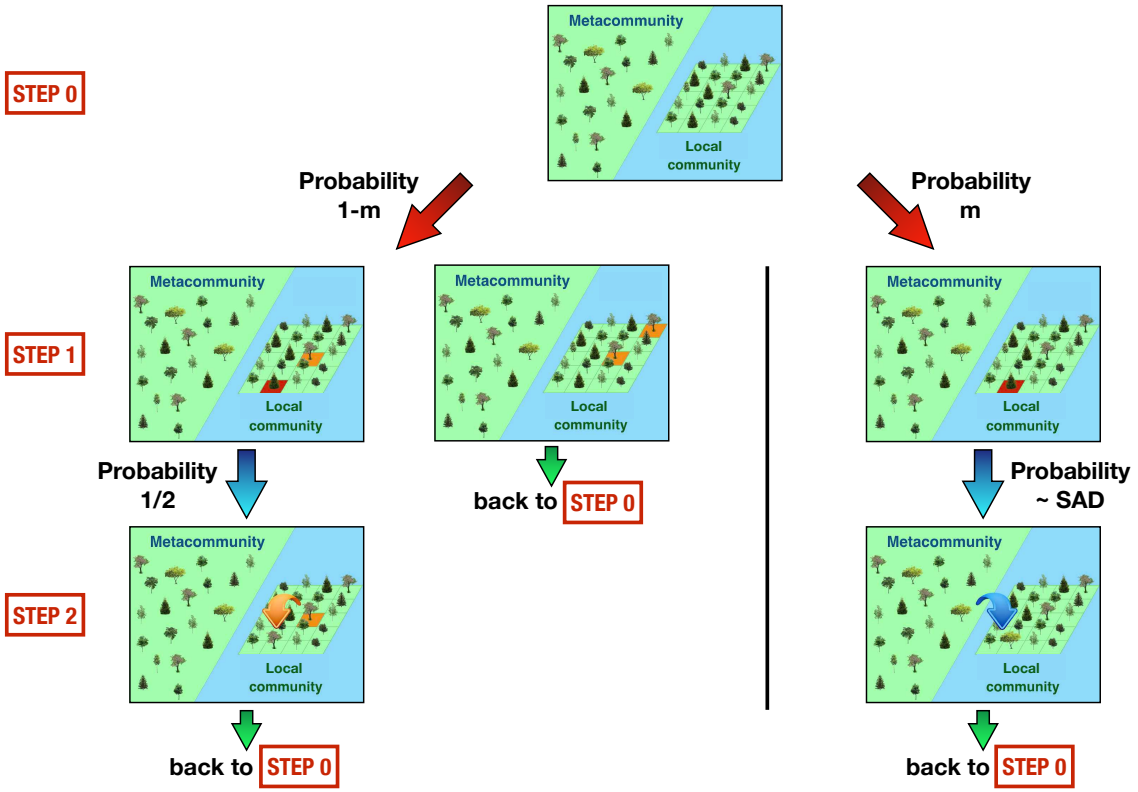


Figure C.3: Scheme of Hubbell's neutral model. At step zero we have a local community (the Barro Colorado Island in Panama, for example) surrounded by a metacommunity (the Soberania National Park) which represents a limitless reservoir of species. At step one we have two possibilities, one chosen with probability equal to $1-m$ and one with probability m. In the first case, two randomly chosen individuals of the local community are picked: if they are conspecific, we just go back to step zero, otherwise, we proceed to step two, where with equal probability, we remove one of the two individuals and we replace it with a daughter seed of the other. If the second action is chosen at the beginning, we randomly pick an individual and we proceed to step two, which consists in randomly picking a species from the metacommunity, with a probability proportional to its numerosity, and to substitute the individual chosen in step one by an individual of this second species.

The dynamics can be simulated on the computer and the two steps repeated until the equilibrium is reached.
In literature, several variations of the model have been proposed, some of which lead to mathematically exact predictions for the macro-ecological patterns of interest (Rosindell et al., 2011). An example is the model proposed in Volkov et al., 2003, where Hubbell's rules are translated into the following mathematical equations for the birth and death rates for the k^{th} species:

$$b^L_{n,k} = (1-m)\frac{n}{N^L}\frac{N^L-n}{N^L-1} + m\frac{n^M_k}{N^M}\Big(1-\frac{n}{N^L}\Big)$$
$$d^L_{n,k} = (1-m)\frac{n}{N^M}\frac{N^L-n}{N^L-1} + m\Big(1-\frac{n^M_k}{N^M}\Big)\frac{n}{N^L},$$

where n^M_k is the number of individuals belonging to species k within the metacommunity. Then, by performing the computations, one finds the following expression for the SAD of the local community,

$$\mathbb{E}[\phi^L_n] = \theta\binom{N^L}{n}\frac{\Gamma(\gamma)}{\Gamma(N^L+\gamma)}\int_0^\gamma \mathrm{e}^{-y\theta/\gamma}\frac{\Gamma(n+y)}{\Gamma(1+y)}\cdot\frac{\Gamma(N^L-n+\gamma-y)}{\Gamma(\gamma-y)}\mathrm{d}y,$$

with $\gamma = \frac{m(N^L-1)}{1-m}$. Let us remark that the above expression depends on the total number of individuals of the local community N^L, on the log-series θ parameter of the metacommunity SAD and on m.
The above integral can be numerically computed for a given the values of such parameters. In figure C.4, we insert the comparison between the SAD predicted by the neutral model for the BCI and the one obtained by fitting the empirical histogram (Preston plot) through another widely used distribution in theoretical ecology, the log-normal. As we can see, the neutral method, in addition of being biologically more informative, does also provide a better fit of the empirical data.

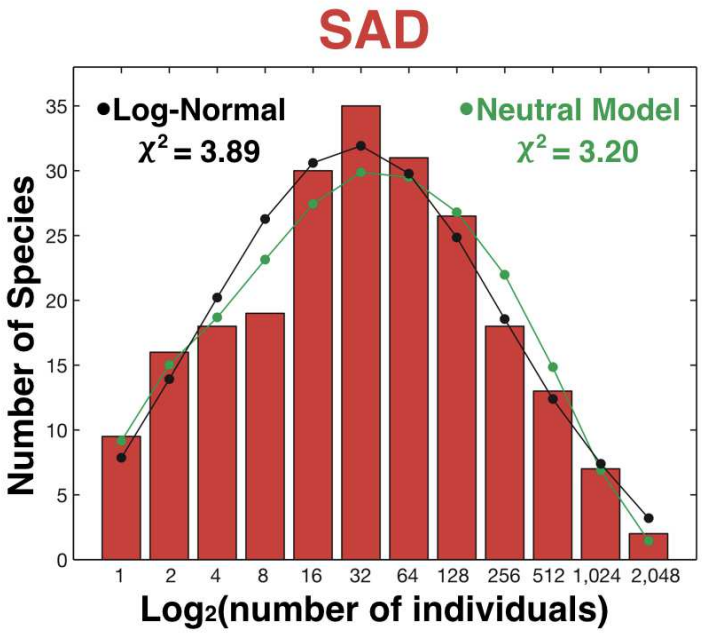


Figure C.4: Comparison between the neutral model's predictions (green) of the SAD and the fitting with a log-normal distribution (black). According to the standard χ^2 analysis, whose results are displayed in the picture, show the superiority of the first model compared to the second. The red histogram is the Preston plot computed for the empirical data of BCI. The figure is largely inspired by Figure 1 in Volkov et al., 2003.

Bibliography

Abe, Sumiyoshi and Attipat K. Rajagopal

2001 "Nonadditive conditional entropy and its significance for local realism", *Physica A: Statistical Mechanics and its Applications*, 289, 1, pp. 157-164. (Cited on p. 63.)

Abramowitz, Milton and Irene A. Stegun

1964 *Handbook of mathematical functions: with formulas, graphs, and mathematical tables*, Courier Corporation, vol. 55. (Cited on p. 31.)

Aczél, János and Zoltán Daróczy

1975 "On measures of information and their characterizations", *New York*. (Cited on p. 168.)

Adorisio, Matteo, Jacopo Grilli, Samir Suweis, Sandro Azaele, Jayanth R. Banavar and Amos Maritan

2009 *Spatial Maximum Entropy Modeling from Presence/Absence Tropical Forest Data*, arXiv: 1407.2425 [q-bio.PE]. (Cited on pp. vii, 51.)

Alonso, David, Annette Ostling and Rampal S. Etienne

2008 "The implicit assumption of symmetry and the species abundance distribution", *Ecology Letters*, 11, 2, pp. 93-105. (Cited on p. 106.)

Anderson, Marti J., Thomas O. Crist, Jonathan M. Chase, Mark Vellend, Brian D. Inouye, Amy L. Freestone, Nathan J. Sanders, Howard V. Cornell, Liza S. Comita, Kendi F. Davies et al.

2011 "Navigating the multiple meanings of β diversity: a roadmap for the practicing ecologist", *Ecology letters*, 14, 1, pp. 19-28. (Cited on p. 64.)

Augspurger, Carol K.

1984 "Seedling survival of tropical tree species: interactions of dispersal distance, light-gaps, and pathogens", *Ecology*, 65, 6, pp. 1705-1712. (Cited on p. 56.)

Azaele, Sandro, Stephen J. Cornell and William E. Kunin

2012 "Downscaling species occupancy from coarse spatial scales", *Ecological Applications*, 22, 3, pp. 1004-1014. (Cited on pp. vii, 22, 30, 51, 62, 126.)

Azaele, Sandro, Amos Maritan, Stephen J. Cornell, Samir Suweis, Jayanth R. Banavar, Doreen Gabriel and William E. Kunin

2015 "Towards a unified descriptive theory for spatial ecology: predicting biodiversity patterns across spatial scales", *Methods in Ecology and Evolution*, 6, 3, pp. 324-332. (Cited on pp. xi, 51, 54, 107, 120.)

Azaele, Sandro and Fabio Peruzzo

2016 "A phenomenological spatial model for macro-ecological patterns in species-rich ecosystems", *bioRxiv*, p. 074336. (Cited on p. 112.)

Azaele, Sandro, Simone Pigolotti, Jayanth R. Banavar and Amos Maritan

2006 "Dynamical evolution of ecosystems", *Nature*, 444, 7121, pp. 926-928, ISSN: 0028-0836, DOI: 10.1038/nature05320, http://dx.doi.org/10.1038/nature05320. (Cited on p. 114.)

Azaele, Sandro, Samir Suweis, Jacopo Grilli, Igor Volkov, Jayanth R. Banavar and Amos Maritan

2016 "Statistical mechanics of ecological systems: Neutral theory and beyond", *Rev. Mod. Phys.*, 88 (3 July 2016), p. 035003, DOI: 10.1103/RevModPhys.88.035003, http://link.aps.org/doi/10.1103/RevModPhys.88.035003. (Cited on pp. xi, xii, 106, 111, 119, 120, 123, 140.)

Baddeley, Adrian, Imre Bárány and Rolf Schneider

2007 "Spatial point processes and their applications", *Stochastic Geometry: Lectures given at the CIME Summer School held in Martina Franca, Italy, September 13–18, 2004*, pp. 1-75. (Cited on pp. vii, 5, 9, 11, 14-16, 27.)

Bak, Per

2013 *How nature works: the science of self-organized criticality*, Springer Science & Business Media. (Cited on p. 144.)

Baselga, Andrés

2010 "Multiplicative partition of true diversity yields independent alpha and beta components; additive partition does not", *Ecology*, 91, 7, pp. 1974-1981. (Cited on p. 168.)

Bertuzzo, Enrico, Francesco Carrara, Lorenzo Mari, Florian Altermatt, Ignacio Rodriguez - Iturbe and Andrea Rinaldo

2016 "Geomorphic controls on elevational gradients of species richness", *Proceedings of the National Academy of Sciences*, 113, 7, pp. 1737-1742. (Cited on p. 106.)

Bertuzzo, Enrico, Samir Suweis, Lorenzo Mari, Amos Maritan, Ignacio Rodriguez - Iturbe and Andrea Rinaldo

2011 "Spatial Effects on Species Persistence and Implications for Biodiversity", *Proceeding of the National Academy of Science of the United States of America*, 108, 11 (Mar. 2011), pp. 4346-4351. (Cited on p. 114.)

Besag, Julian Ernst

1977 "Comments on Ripley's paper", *Journal of the Royal Statistical Society: Series B*, 39, pp. 193-195. (Cited on p. 15.)

Birkinshaw, Mark

1994 "Radially-symmetric fourier transforms", in *Astronomical Data Analysis Software and Systems III*, vol. 61, p. 249. (Cited on p. 27.)

Borda-de-Água, Luís, Paulo A. V. Borges, Stephen P. Hubbell and Henrique M. Pereira

2012 "Spatial scaling of species abundance distributions", *Ecography*, 35, 6, pp. 549-556. (Cited on p. 151.)

Box, George E. P. and George C. Tiao

2011 *Bayesian inference in statistical analysis*, John Wiley & Sons, vol. 40. (Cited on p. 41.)

Bracewell, Ronald Newbold

1986 *The Fourier transform and its applications*, McGraw-Hill New York, vol. 31999. (Cited on p. 27.)

Bray, J. Roger and John T. Curtis

1957 "An ordination of the upland forest communities of southern Wisconsin", *Ecological monographs*, 27, 4, pp. 325-349. (Cited on p. 64.)

Brose, Ulrich, Neo D. Martinez and Richard J. Williams

2003 "Estimating species richness: sensitivity to sample coverage and insensitivity to spatial patterns", *Ecology*, 84, 9, pp. 2364-2377. (Cited on p. 106.)

Brown, Emery N., Robert E. Kass and Partha P. Mitra

2004 "Multiple neural spike train data analysis: state-of-the-art and future challenges", *Nature neuroscience*, 7, 5, p. 456. (Cited on p. 5.)

Bunge, John and M. Fitzpatrick

1993 "Estimating the number of species: a review", *Journal of the American Statistical Association*, 88, 421, pp. 364-373. (Cited on pp. 105, 106.)

Bunge, John, Linda Woodard, Dankmar Böhning, James A. Foster, Sean Connolly and Heather K. Allen

2012 "Estimating population diversity with CatchAll", *Bioinformatics*, 28, 7, pp. 1045-1047. (Cited on p. 106.)

Buzas, Martin A. and Lee-Ann C. Hayek

1996 "Biodiversity resolution: an integrated approach", *Biodiversity Letters*, pp. 40-43. (Cited on p. 168.)

Carrara, Francesco, Florian Altermatt, Ignacio Rodriguez-Iturbe and Andrea Rinaldo

2012 "Dendritic connectivity controls biodiversity patterns in experimental metacommunities", *Proceedings of the National Academy of Sciences*, 109, 15, pp. 5761-5766. (Cited on p. 106.)

Cha, Maria and Qing Zhou

2014 "Detecting clustering and ordering binding patterns among transcription factors via point process models", *Bioinformatics*, 30, 16, pp. 2263-2271. (Cited on p. 5.)

Chao, Anne

2005 "Species estimation and applications", *Encyclopedia of statistical sciences*. (Cited on pp. xi, 131, 135.)

Chao, Anne and John Bunge

2002 "Estimating the number of species in a stochastic abundance model", *Biometrics*, 58, 3, pp. 531-539. (Cited on p. 105.)

Chao, Anne, Robin L. Chazdon, Robert K. Colwell and Tsung-Jen Shen

2005 "A new statistical approach for assessing similarity of species composition with incidence and abundance data", *Ecology letters*, 8, 2, pp. 148-159. (Cited on pp. 65, 77.)

2006 "Abundance-based similarity indices and their estimation when there are unseen species in samples", *Biometrics*, 62, 2, pp. 361-371. (Cited on p. 64.)

Chao, Anne and Chun-Huo Chiu

2016 "Species richness: estimation and comparison", *Wiley StatsRef: Statistics Reference Online*. (Cited on pp. xi, 105, 131, 135.)

Chao, Anne, Robert K. Colwell, Chih-Wei Lin and Nicholas J. Gotelli

2009 "Sufficient sampling for asymptotic minimum species richness estimators", *Ecology*, 90, 4, pp. 1125-1133. (Cited on pp. xi, 106.)

Chave, Jérôme

2004 "Neutral theory and community ecology", *Ecology Letters*, 7, 3 (Feb. 2004), pp. 241-253, ISSN: 1461023X, DOI: 10.1111/j.1461-0248.2003.00566.x, http://doi.wiley.com/10.1111/j.1461-0248.2003.00566.x. (Cited on pp. xi, 120.)

Chave, Jérôme, David Alonso and Rampal S. Etienne

2006 "Comparing models of species abundance", *Nature*, 441, 7089, E1, ISSN: 0028-0836. (Cited on pp. xi, 120.)

Chave, Jérôme and Egbert G. Leigh

2002 "A spatially explicit neutral model of β-diversity in tropical forests", *Theoretical population biology*, 62, 2, pp. 153-168. (Cited on pp. ix, 32, 77, 87.)

Cheetham, Alan H. and Joseph E. Hazel

1969 "Binary (presence-absence) similarity coefficients", *Journal of Paleontology*, pp. 1130-1136. (Cited on p. 64.)

Chisholm, Ryan A.

2007 "Sampling species abundance distributions: resolving the veil-line debate", *Journal of theoretical biology*, 247, 4, pp. 600-607. (Cited on pp. xi, 111, 121, 127.)

Chiu, Sung Nok, Dietrich Stoyan, Wilfrid S. Kendall and Joseph Mecke

2013 *Stochastic geometry and its applications*, John Wiley & Sons. (Cited on pp. 5, 6, 9, 11, 14, 25, 27, 29.)

Clarke, Andrew and Scott Lidgard

2000 "Spatial patterns of diversity in the sea: bryozoan species richness in the North Atlantic", *Journal of Animal Ecology*, 69, 5, pp. 799-814. (Cited on pp. 32, 87.)

Cliff, Andrew David and J. Keith Ord

1981 *Spatial processes: models & applications*, Taylor & Francis. (Cited on p. 5.)

Clifford, Harold Trevor and William Stephenson

1975 *An introduction to numerical classification*, Academic Press New York, vol. 229. (Cited on pp. 62, 64.)

Colwell, Robert K.

2009 "Biodiversity: concepts, patterns, and measurement", *The Princeton guide to ecology*, pp. 257-263. (Cited on p. 61.)

Colwell, Robert K. and Jonathan A. Coddington

1994 "Estimating terrestrial biodiversity through extrapolation", *Philosophical Transactions: Biological Sciences*, pp. 101-118. (Cited on p. 105.)

Condit, Richard, Peter S. Ashton, Patrick Baker, Sarayudh Bunyavejchewin, Savithri Gunatilleke, Nimal Gunatilleke, Stephen P. Hubbell, Robin B. Foster, Akira Itoh, James V. LaFrankie et al.

2000 "Spatial patterns in the distribution of tropical tree species", *Science*, 288, 5470, pp. 1414-1418. (Cited on pp. vii, 51, 57.)

Connell, Joseph H.

1971 "On the role of natural enemies in preventing competitive exclusion in some marine animals and in rain forest trees", *Dynamics of populations*. (Cited on p. 51.)

Corbet, A. Steven

1941 "The distribution of butterflies in the Malay Peninsula (Lepid.)" *Physiological Entomology*, 16, 10-12, pp. 101-116. (Cited on p. 105.)

Cox, David Roxbee and Valerie Isham

1980 *Point processes*, CRC Press, vol. 12. (Cited on p. 23.)

Cressie, Noel

2015 *Statistics for spatial data*, John Wiley & Sons. (Cited on pp. 9, 10, 15, 27.)

Crowther, Thomas W., Henry B. Glick, Kristofer R. Covey, Charlie Bettigole, Daniel S. Maynard, Stephen M. Thomas, Jeffrey R. Smith, G. Hintler, Marlyse C. Duguid, Giuseppe Amatulli et al.

2015 "Mapping tree density at a global scale", *Nature*, 525, 7568, pp. 201-205. (Cited on p. 105.)

Dale, Mark R.T.

2000 *Spatial pattern analysis in plant ecology*, Cambridge university press. (Cited on p. 14.)

Daley, Daryl J. and David Vere-Jones

2003 *An introduction to the theory of point processes: volume I: Elementary Theory and Methods*, Springer Science & Business Media. (Cited on pp. vii, 11, 21.)

2007 *An introduction to the theory of point processes: volume II: General theory and Structure*, Springer Science & Business Media. (Cited on pp. 6, 7, 11.)

Darwin, Charles

1859 *On the origin of species by means of natural selection, or the preservation of favoured races in the struggle for life*, Murray. (Cited on pp. v, 172.)

Diggle, Peter J.

2003 *Statistical analysis of spatial point patterns*, ed. by Arnold, 2nd ed. (Cited on pp. 5, 23, 24, 43.)

2013 *Statistical analysis of spatial and spatio-temporal point patterns*, CRC Press. (Cited on pp. 9, 11, 36, 43, 51, 54, 79.)

Diggle, Peter J. and Richard J. Gratton

1984 "Monte Carlo methods of inference for implicit statistical models", *Journal of the Royal Statistical Society. Series B (Methodological)*, pp. 193-227. (Cited on p. 36.)

Doane, David P.

1976 "Aesthetic frequency classifications", *The American Statistician*, 30, 4, pp. 181-183. (Cited on p. viii.)

Engen, Steinar, Vidar Grøtan and Bernt-Erik Sæther

2011 "Estimating similarity of communities: a parametric approach to spatio-temporal analysis of species diversity", *Ecography*, 34, 2, pp. 220-231. (Cited on p. 64.)

Fisher, Ronald A., A. Steven Corbet and Carrington B. Williams

1943 "The relation between the number of species and the number of individuals in a random sample of an animal population", *The Journal of Animal Ecology*, pp. 42-58. (Cited on pp. x, xi, 117.)

Formentin, Marco, Alberto Lovison, Amos Maritan and Giovanni Zanzotto

2014 "Hidden scaling patterns and universality in written communication", *Physical Review E*, 90, 1, p. 012817. (Cited on p. 150.)

Freedman, David and Persi Diaconis

1981 "On the histogram as a density estimator: L_2 theory", *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete*, 57, 4, pp. 453-476. (Cited on pp. viii, 39, 40, 157.)

Gaston, Kevin J.

1996 "Species richness: measure and measurement", *Biodiversity: A Biology of Numbers and Difference*, pp. 77-113. (Cited on p. 62.)

Gibbs, Josiah Willard

1902 "Elementary principles in statistical physics", *The Collected Works of JW Gibbs (Yale University, New Haven, CT, 1957)*, 2. (Cited on p. 144.)

Gomes-Gonçalves, Erika, Henryk Gzyl and Silvia Mayoral

2014 "Density reconstructions with errors in the data", *Entropy*, 16, 6, pp. 3257-3272. (Cited on p. 161.)

Good, I. J. and G. H. Toulmin

1956 "The number of new species, and the increase in population coverage, when a sample is increased", *Biometrika*, 43, 1-2, pp. 45-63. (Cited on p. 105.)

Good, Irving J.

1953 "The population frequencies of species and the estimation of population parameters", *Biometrika*, 40, 3-4, pp. 237-264. (Cited on p. 62.)

Gower, John C.

1971 "A general coefficient of similarity and some of its properties", *Biometrics*, pp. 857-871. (Cited on p. 64.)

1985 "Measures of similarity, dissimilarity and distance", *Encyclopedia of statistical sciences*, 5, 397-405, p. 3. (Cited on p. 64.)

Gray, John S.

2000 "The measurement of marine species diversity, with an application to the benthic fauna of the Norwegian continental shelf", *Journal of Experimental Marine Biology and Ecology*, 250, 1, pp. 23-49. (Cited on p. 62.)

Grilli, Jacopo, Samir Suweis and Amos Maritan

2013 "Growth or reproduction: emergence of an evolutionary optimal strategy", *Journal of Statistical Mechanics: Theory and Experiment*, 2013, 10, P10020. (Cited on p. 146.)

Haegeman, Bart and Rampal S. Etienne

2010 "Entropy maximization and the spatial distribution of species", *The American Naturalist*, 175, 4, E74-E90. (Cited on pp. vii, viii.)

Hagmeier, Edwin M. and C. Dexter Stults

1964 "A numerical analysis of the distributional patterns of North American mammals", *Systematic Zoology*, 13, 3, pp. 125-155. (Cited on p. 64.)

Hardin, Garrett

1960 "The competitive exclusion principle", *science*, 131, 3409, pp. 1292-1297. (Cited on p. 172.)

Harte, John

2011 *Maximum entropy and ecology: a theory of abundance, distribution, and energetics*, Oxford University Press. (Cited on p. 111.)

Harte, John, Erin Conlisk, Annette Ostling, Jessica L. Green and Adam B. Smith

2005 "A theory of spatial structure in ecological communities at multiple spatial scales", *Ecological Monographs*, 75, 2, pp. 179-197. (Cited on p. 57.)

Harte, John, Adam B. Smith and David Storch

2009 "Biodiversity scales from plots to biomes with a universal species-area curve", *Ecology Letters*, 12, 8, pp. 789-797. (Cited on pp. x, 111, 115, 135-138, 151.)

Harte, John, Tommaso Zillio, Erin Conlisk and Adam B. Smith

2008 "Maximum entropy and the state-variable approach to macroecology", *Ecology*, 89, 10, pp. 2700-2711. (Cited on pp. 135, 136.)

He, Fangliang and Kevin J. Gaston

2003 "Occupancy, spatial variance, and the abundance of species", *The American Naturalist*, 162, 3, pp. 366-375. (Cited on p. 107.)

He, Fangliang and Stephen P. Hubbell

2003 "Percolation theory for the distribution and abundance of species", *Physical Review Letters*, 91, 19, p. 198103. (Cited on p. 107.)

He, Fangliang, Pierre Legendre and James V. LaFrankie

1997 "Distribution Patterns of Tree Species in a Malaysian Tropical Rain Forest", *Journal of Vegetation Science*, 8, 1, pp. 105-114. (Cited on pp. vii, 51, 57.)

Heip, Carlo H. R., Peter M. J. Herman and Karlin Soetaert

1998 "Indices of diversity and evenness", *Océanis (Paris)*, 4. (Cited on pp. 62, 63, 165.)

Hidalgo, Jorge, Jacopo Grilli, Samir Suweis, Amos Maritan and Miguel A. Muñoz

2016 "Cooperation, competition and the emergence of criticality in communities of adaptive systems", *Journal of Statistical Mechanics: Theory and Experiment*, 2016, 3, p. 033203. (Cited on p. 146.)

Hidalgo, Jorge, Jacopo Grilli, Samir Suweis, Miguel A. Muñoz, Jayanth R. Banavar and Amos Maritan

2014 "Information-based fitness and the emergence of criticality in living systems", *Proceedings of the National Academy of Sciences*, 111, 28, pp. 10095-10100. (Cited on p. 146.)

Hill, Mark O.

1973 "Diversity and evenness: a unifying notation and its consequences", *Ecology*, 54, 2, pp. 427-432. (Cited on pp. 62-64, 163-165, 167.)

Hodder, Ian and Clive Orton

1976 "Spatial analysis in archaeology". (Cited on p. 5.)

Horn, Henry S.

1966 "Measurement of "overlap" in comparative ecological studies", *The American Naturalist*, 100, 914, pp. 419-424. (Cited on p. 64.)

Hubalek, Zdenek

1982 "Coefficients of association and similarity, based on binary (presence-absence) data: an evaluation", *Biological Reviews*, 57, 4, pp. 669-689. (Cited on p. 64.)

Hubbell, Stephen P.

1997 "A unified theory of biogeography and relative species abundance and its application to tropical rain forests and coral reefs", *Coral reefs*, 16, 5, S9-S21. (Cited on p. ix.)

2001a *The unified neutral theory of biodiversity and biogeography*, Monographs in population biology, 32, Princeton University Press, Princeton, NJ. (Cited on pp. 174, 175.)

2001b "The unified neutral theory of species abundance and diversity", *Princeton University Press*, 79, pp. 96-97. (Cited on pp. ix, 91, 97.)

2015 "Estimating the global number of tropical tree species, and Fisher's paradox", *Proceedings of the National Academy of Sciences*, 112, 24, pp. 7343-7344. (Cited on pp. 143, 148.)

Hughes, Jennifer B., Jessica J. Hellmann, Taylor H. Ricketts and Brendan J. M. Bohannan

2001 "Counting the uncountable: statistical approaches to estimating microbial diversity", *Applied and environmental microbiology*, 67, 10, pp. 4399-4406. (Cited on p. 105.)

Hui, Cang, Gordon A. Fox and Jessica Gurevitch

2017 "Scale-dependent portfolio effects explain growth inflation and volatility reduction in landscape demography", *Proceedings of the National Academy of Sciences*, 114, 47, pp. 12507-12511. (Cited on p. 151.)

Hui, Cang and Melodie A. McGeoch

2007 "Modeling species distributions by breaking the assumption of self-similarity", *Oikos*, 116, 12, pp. 2097-2107. (Cited on p. 66.)

Hui, Cang, Melodie A. McGeoch and Marié Warren

2006 "A spatially explicit approach to estimating species occupancy and spatial correlation", *Journal of Animal Ecology*, 75, 1, pp. 140-147. (Cited on p. vii.)

Hull, Pincelli M., Simon A. F. Darroch and Douglas H. Erwin

2015 "Rarity in mass extinctions and the future of ecosystems", *Nature*, 528, 7582, pp. 345-351. (Cited on p. 146.)

Hurlbert, Stuart H.

1971 "The nonconcept of species diversity: a critique and alternative parameters", *Ecology*, 52, 4, pp. 577-586. (Cited on p. 164.)

Illian, Janine, Antti Penttinen, Helga Stoyan and Dietrich Stoyan

2008 *Statistical analysis and modelling of spatial point patterns*, John Wiley & Sons, vol. 70. (Cited on pp. 5, 7, 9-11, 15, 25, 27, 29, 43, 51.)

Ionita-Laza, Iuliana, Christoph Lange and Nan M. Laird

2009 "Estimating the number of unseen variants in the human genome", *Proceedings of the National Academy of Sciences*, 106, 13, pp. 5008-5013. (Cited on p. 105.)

Jaccard, Paul

1900 *Contribution au problème de l'immigration post-glacière de la flore alpine: etude comparative de la flore alpine du massif du Wildhorn, du haut bassin du Trient et de la haute vallée de Bagnes*. (Cited on p. 64.)

1908 "Nouvelles recherches sur la distribution florale", *Bulletin de la Société Vandoise des Sciences Naturelles*, 44, 163, pp. 223-270. (Cited on p. 64.)

Janzen, Daniel H.

1970 "Herbivores and the number of tree species in tropical forests", *The American Naturalist*, 104, 940, pp. 501-528. (Cited on p. 51.)

Johnson, Norman L., Adrienne W. Kemp and Samuel Kotz

2005 *Univariate discrete distributions*, John Wiley & Sons, vol. 444. (Cited on p. 67.)

Jost, Lou

2006 "Entropy and diversity", *Oikos*, 113, 2, pp. 363-375. (Cited on pp. 63, 64, 163-165, 167, 168.)

2007 "Partitioning diversity into independent alpha and beta components", *Ecology*, 88, 10, pp. 2427-2439. (Cited on pp. 163, 164, 167, 168.)

2010a "Independence of alpha and beta diversities", *Ecology*, 91, 7, pp. 1969-1974. (Cited on p. 168.)

2010b "The relation between evenness and diversity", *Diversity*, 2, 2, pp. 207-232. (Cited on pp. 62, 63.)

Keylock, Christopher J.

2005 "Simpson diversity and the Shannon–Wiener index as special cases of a generalized entropy", *Oikos*, 109, 1, pp. 203-207. (Cited on p. 167.)

Kitzes, Justin and John Harte

2015 "Predicting extinction debt from community patterns", *Ecology*, 96, 8, pp. 2127-2136. (Cited on pp. 111, 135, 137.)

Knuth, Kevin H.

2006 "Optimal data-based binning for histograms", *arXiv preprint physics/0605197*. (Cited on pp. viii, 39-42.)

Krebs, Charles J. et al.

1989 *Ecological methodology*, tech. rep., Harper & Row New York. (Cited on pp. 64, 65.)

Kulczyński, Stanisław

1928 *Die Pflanzenassoziationen der Pieninen*, Imprimerie de l'Université. (Cited on p. 64.)

Kunin, William E., John Harte, Fangliang He, Cang Hui, R. Todd Jobe, Annette Ostling et al.

2017 "Upscaling biodiversity: estimating the Species-Area Relationship from small samples", *Ecological Monographs*. (Cited on p. 105.)

Lance, Godfrey N. and William T. Williams

1967 "Mixed-Data Classificatory Programs I - Agglomerative Systems", *Australian Computer Journal*, 1, 1, pp. 15-20. (Cited on p. 64.)

Lande, Russell

1996 "Statistics and partitioning of species diversity, and similarity among multiple communities", *Oikos*, pp. 5-13. (Cited on pp. 63, 64, 168.)

Lee, Tsung-Dao and Chen-Ning Yang

1952 "Statistical theory of equations of state and phase transitions. II. Lattice gas and Ising model", *Physical Review*, 87, 3, p. 410. (Cited on p. 145.)

Legendre, Pierre and Loic F. J. Legendre

2012 *Numerical ecology*, Elsevier, vol. 24. (Cited on p. 63.)

Leigh, Egbert Giles, S. Joseph Wright, Edward Allen Herre and Francis E. Putz

1993 "The decline of tree diversity on newly isolated tropical islands: a test of a null hypothesis and some implications", *Evolutionary Ecology*, 7, 1, pp. 76-102. (Cited on p. ix.)

Lennon, Jack J., Patricia Koleff, J.J.D. Greenwood and Kevin J. Gaston

2001 "The geographical structure of British bird distributions: diversity, spatial turnover and scale", *Journal of Animal Ecology*, 70, 6, pp. 966-979. (Cited on p. 64.)

Lewontin, Richard C.

1972 "The apportionment of human diversity", in *Evolutionary biology*, Springer, pp. 381-398. (Cited on p. 168.)

Locey, Kenneth J. and Jay T. Lennon

2016 "Scaling laws predict global microbial diversity", *Proceedings of the National Academy of Sciences*, 113, 21, pp. 5970-5975. (Cited on p. 105.)

Lotka, Alfred James

1956 *Elements of Physical Biology*. (Cited on p. 172.)

Lunde, Asger and Robert F. Engle

1998 "Trades and Quotes: A Bivariate Point Process". (Cited on p. 5.)

MacArthur, Robert H.

1960 "On the relative abundance of species", *The American Naturalist*, 94, 874, pp. 25-36. (Cited on p. 107.)

1965 "Patterns of species diversity", *Biological reviews*, 40, 4, pp. 510-533. (Cited on p. 165.)

MacArthur, Robert H. and Edward O. Wilson

2016 *The Theory of Island Biogeography*, Princeton University Press. (Cited on p. 173.)

Magurran, Anne E.

1988 "Why diversity?", in *Ecological diversity and its measurement*, Springer, pp. 1-5. (Cited on p. 165.)

2005 "Species abundance distributions: pattern or process?", *Functional Ecology*, 19, 1, pp. 177-181. (Cited on pp. xi, 120.)

2013 *Measuring biological diversity*, John Wiley & Sons. (Cited on pp. xi, 61-63, 107, 120.)

Magurran, Anne E. and Peter A. Henderson

2003 "Explaining the excess of rare species in natural species abundance distributions", *Nature*, 422, 6933, pp. 714-716. (Cited on pp. 62, 123.)

Magurran, Anne E. and Brian J. McGill

2011 *Biological diversity: frontiers in measurement and assessment*, Oxford University Press. (Cited on p. ix.)

Mao, Chang Xuan and Robert K. Colwell

2005 "Estimation of species richness: mixture models, the role of rare species, and inferential challenges", *Ecology*, 86, 5, pp. 1143-1153. (Cited on p. 106.)

Matthews, Thomas J. and Robert J. Whittaker

2014 "Neutral theory and the species abundance distribution: recent developments and prospects for unifying niche and neutral perspectives", *Ecology and evolution*, 4, 11, pp. 2263-2277. (Cited on pp. xi, 120.)

McGill, Brian J., Rampal S. Etienne, John S. Gray, David Alonso, Marti J. Anderson, Habtamu Kassa Benecha, Maria Dornelas, Brian J. Enquist, Jessica L. Green, Fangliang He et al.

2007 "Species abundance distributions: moving beyond single prediction theories to integration within an ecological framework", *Ecology letters*, 10, 10, pp. 995-1015. (Cited on p. 106.)

McIntosh, Robert P.

1967 "An index of diversity and the relation of certain concepts to diversity", *Ecology*, 48, 3, pp. 392-404. (Cited on p. 62.)

Menzel, Peter, Kim Lee Ng and Anders Krogh

2016 "Fast and sensitive taxonomic classification for metagenomics with Kaiju", *Nature communications*, 7. (Cited on p. 152.)

Møller, Jesper, Rasmus Plenge Waagepetersen et al.

2004 *Statistical inference and simulation for spatial point processes*, Chapman & Hall/CRC, (cited on p. 5.)

Morisita, Masaaki

1959 "Measuring of interspecific association and similarity between communities", *Mem. Fac. Sci. Kyushu Univ. Series E*, 3, pp. 65-80. (Cited on p. 64.)

Morlon, Hélene, George Chuyong, Richard Condit, Stephen Hubbell, David Kenfack, Duncan Thomas, Renato Valencia and Jessica L. Green

2008 "A general framework for the distance–decay of similarity in ecological communities", *Ecology letters*, 11, 9, pp. 904-917. (Cited on pp. ix, 30, 51, 54, 57, 60, 64, 77-79, 94, 99.)

Muneepeerakul, Rachata, Enrico Bertuzzo, Heather J. Lynch, William F. Fagan, Andrea Rinaldo and Ignacio Rodriguez-Iturbe

2008 "Neutral metacommunity models predict fish diversity patterns in Mississippi-Missouri basin", *Nature*, 453, 7192, p. 220. (Cited on p. 119.)

Myers, Norman, Russell A. Mittermeier, Cristina G. Mittermeier, Gustavo A. B. Da Fonseca and Jennifer Kent

2000 "Biodiversity hotspots for conservation priorities", *Nature*, 403, 6772, pp. 853-858. (Cited on p. 143.)

Nekola, Jeffrey C. and Peter S. White

1999 "The distance decay of similarity in biogeography and ecology", *Journal of Biogeography*, 26, 4, pp. 867-878. (Cited on pp. ix, 97.)

Newman, Mark E. J. and Gerard T. Barkema

1999 *Monte Carlo Methods in Statistical Physics*, Oxford University Press: New York, USA. (Cited on p. 144.)

Neyman, Jerzy and Elizabeth L. Scott

1958 "Statistical approach to problems of cosmology", *Journal of the Royal Statistical Society. Series B (Methodological)*, pp. 1-43. (Cited on p. 22.)

Nishimori, Hidetoshi

2001 *Statistical physics of spin glasses and information processing: an introduction*, Clarendon Press, vol. 111. (Cited on p. 144.)

Ohser, Joachim

1983 "On estimators for the reduced second moment measure of point processes", *Statistics: A Journal of Theoretical and Applied Statistics*, 14, 1, pp. 63-71. (Cited on p. 36.)

Orlitsky, Alon, Ananda Theertha Suresh and Yihong Wu

2016 "Optimal prediction of the number of unseen species", *Proceedings of the National Academy of Sciences*, p. 201607774. (Cited on p. 105.)

Ornstein, Leonard S. and Frits Zernike

1914 "Accidental deviations of density and opalescence at the critical point of a single substance", *Proc. Akad. Sci.*, 17, pp. 793-806. (Cited on p. 15.)

Ostling, Annette, John Harte and Jessica Green

2000 "Self-similarity and clustering in the spatial distribution of species", *Science*, 290, 5492, pp. 671-671. (Cited on p. 57.)

Palm, Conrad

1943 *Intensitätsschwankungen im Fernsprechverkehr: Untersuchungen über die Darstellung auf Fernsprechverkehrsprobleme anwendbarer stochastischer Prozesse*, PhD thesis, C. Palm. (Cited on p. 14.)

Palmer, Michael W. and Peter S. White

1994 "Scale dependence and the species-area relationship", *The American Naturalist*, 144, 5, pp. 717-740. (Cited on p. 99.)

Peebles, Phillip James Edwin

1974 "The nature of the distribution of galaxies", *Astronomy and Astrophysics*, 32, p. 197. (Cited on p. 5.)

Peters, James A.

1968 "A computer program for calculating degree of biogeographical resemblance between areas", *Systematic Biology*, 17, 1, pp. 64-69. (Cited on pp. 64, 65.)

Pielou, Evelyn Chris

1969 "An introduction to mathematical ecology." *An introduction to mathematical ecology.*. (Cited on p. 62.)

1975 "Ecology diversity", *J. Wiley and Sons, New York*. (Cited on p. 62.)

Plotkin, Joshua B., Jérôme Chave, Peter S. Ashton and Joseph Travis

2002 "Cluster analysis of spatial patterns in Malaysian tree species", *The American Naturalist*, 160, 5, pp. 629-644. (Cited on pp. ix, 78, 79, 99, 126.)

Plotkin, Joshua B., Matthew D. Potts, Nandi Leslie, N. Manokaran, James LaFrankie and Peter S. Ashton

2000 "Species-area curves, spatial aggregation, and habitat specialization in tropical forests", *Journal of theoretical biology*, 207, 1, pp. 81-99. (Cited on pp. vii, 22, 30, 33, 35, 36, 51, 54, 56, 57, 94, 106, 126.)

Preston, Frank W.

1948 "The commonness, and rarity, of species", *Ecology*, 29, 3, pp. 254-283. (Cited on pp. vi, 62, 121, 123.)

1962a "The canonical distribution of commonness and rarity: Part I", *Ecology*, 43, 2, pp. 185-215. (Cited on p. 75.)

1962b "The canonical distribution of commonness and rarity: Part II", *Ecology*, 43, 3, pp. 410-432. (Cited on p. 75.)

Quesada, Jose Antonio, Inmaculada Melchor and Andreu Nolasco

2017 "Point process methods in epidemiology: application to the analysis of human immunodeficiency virus/acquired immunodeficiency syndrome mortality in urban areas", *Geospatial Health*, 12, 1. (Cited on p. 5.)

Réjou-Méchain, Maxime and Olivier J. Hardy

2011 "Properties of similarity indices under niche-based and dispersal-based processes in communities", *The American Naturalist*, 177, 5, pp. 589-604. (Cited on p. 64.)

Renkonen, Olavi

1938 *Statistisch-ökologische Untersuchungen über die terrestrische Käferwelt der finnischen Bruchmoore*, PhD thesis, Societas zoologica-botanica Fennica Vanamo. (Cited on p. 64.)

Rényi, Alfred

1961 "On measures of entropy and information", in *Fourth Berkeley symposium on mathematical statistics and probability*, vol. 1, pp. 547-561. (Cited on pp. 64, 168.)

Ricotta, Carlo

2005 "On hierarchical diversity decomposition", *Journal of Vegetation Science*, 16, 2, pp. 223-226. (Cited on p. 168.)

2010 "On beta diversity decomposition: trouble shared is not trouble halved", *Ecology*, 91, 7, pp. 1981-1983. (Cited on p. 168.)

Ripley, Brian D.

1976 "The second-order analysis of stationary point processes", *Journal of applied probability*, pp. 255-266. (Cited on p. 15.)

1977 "Modelling spatial patterns", *Journal of the Royal Statistical Society. Series B (Methodological)*, pp. 172-212. (Cited on p. 15.)

2005 *Spatial statistics*, John Wiley & Sons, vol. 575. (Cited on p. 15.)

Rogers, David J. and Taffee T. Tanimoto

1960 "A computer program for classifying plants." *Science (New York, NY)*, 132, 3434, pp. 1115-1118. (Cited on p. 64.)

Romesburg, C.H.

1984 "Cluster Analysis for Researches Lifetime Learning", *Belmont, CA*. (Cited on p. 64.)

Rosindell, James, Stephen P. Hubbell and Rampal S. Etienne

2011 "The unified neutral theory of biodiversity and biogeography at age ten", *Trends in ecology & evolution*, 26, 7, pp. 340-348. (Cited on p. 178.)

Sanlı, Ceyda and Renaud Lambiotte

2015 "Local variation of hashtag spike trains and popularity in twitter", *PloS one*, 10, 7, e0131704. (Cited on p. 150.)

Schiffers, Katja, Frank M. Schurr, Katja Tielbörger, Carsten Urbach, Kirk Moloney and Florian Jeltsch

2008 "Dealing with virtual aggregation–a new index for analysing heterogeneous point patterns", *Ecography*, 31, 5, pp. 545-555. (Cited on pp. viii, 15, 43, 51, 53.)

Scott, David W.

1979 "On optimal and data-based histograms", *Biometrika*, 66, 3, pp. 605-610. (Cited on p. 39.)

2015 *Multivariate density estimation: theory, practice, and visualization*, John Wiley & Sons. (Cited on pp. viii, 39.)

Shannon, Claude E.

1948 "A mathematical theory of communication", *The Bell System Technical Journal*, 27, 3–4, pp. 379-423, 623-656. (Cited on p. 64.)

Shannon, Claude E. and Warren Weaver

1949 "The mathematical theory of communication. Urbana, Ill", *Univ. Illinois Press*, 1, p. 17. (Cited on pp. 62, 64.)

Shimatani, Kenichiro

2001 "Multivariate point processes and spatial variation of species diversity", *Forest Ecology and Management*, 142, 1, pp. 215-229. (Cited on pp. ix, 35, 63, 66, 70, 72, 77.)

2002 "Point processes for fine-scale spatial genetics and molecular ecology", *Biometrical Journal*, 44, 3, pp. 325-352. (Cited on pp. 5, 24.)

2010 "Spatially explicit neutral models for population genetics and community ecology: Extensions of the Neyman–Scott clustering process", *Theoretical population biology*, 77, 1, pp. 32-41. (Cited on p. 24.)

Shimatani, Kenichiro and Yasuhiro Kubota

2004 "Quantitative assessment of multispecies spatial pattern with high species diversity", *Ecological Research*, 19, 2, pp. 149-163. (Cited on pp. ix, 63, 70.)

Simini, Filippo, Marta C. González, Amos Maritan and Albert-László Barabási

2011 "A universal model for mobility and migration patterns", *arXiv preprint arXiv:1111.0586*. (Cited on p. 150.)

Simpson, Edward H.

1949 "Measurement of diversity." *Nature*. (Cited on pp. 62-64.)

Slik, J. W. Ferry, Víctor Arroyo-Rodríguez, Shin-Ichiro Aiba, Patricia Alvarez-Loayza, Luciana F. Alves, Peter Ashton, Patricia Balvanera, Meredith L. Bastian, Peter J. Bellingham, Eduardo Van Den Berg et al.

2015 "An estimate of the number of tropical tree species", *Proceedings of the National Academy of Sciences*, 112, 24, pp. 7472-7477. (Cited on pp. x, xi, 62, 111, 115, 119, 123, 131, 140, 143, 148.)

Smith, Benjamin and J. Bastow Wilson

1996 "A consumer's guide to evenness indices", *Oikos*, pp. 70-82. (Cited on p. 62.)

Soininen, Janne, Robert McDonald and Helmut Hillebrand

2007 "The distance decay of similarity in ecological communities", *Ecography*, 30, 1, pp. 3-12. (Cited on p. 75.)

Sokal, Robert R. and Peter H. A. Sneath

1963 "Principles of numerical taxonomy". (Cited on p. 65.)

Sørensen, Thorvald

1948 "A method of establishing groups of equal amplitude in plant sociology based on similarity of species and its application to analyses of the vegetation on Danish commons", *Biol. Skr.*, 5, pp. 1-34. (Cited on pp. ix, 64, 65.)

Stanley, H. Eugene

1999 "Scaling, universality, and renormalization: Three pillars of modern critical phenomena", *Reviews of modern physics*, 71, 2, S358. (Cited on pp. xii, 143, 145.)

Steinbauer, Manuel Jonas, Klara Dolos, Björn Reineking and Carl Beierkuhnlein

2012 "Current measures for distance decay in similarity of species composition are influenced by study extent and grain size", *Global Ecology and Biogeography*, 21, 12, pp. 1203-1212. (Cited on p. 99.)

Stone, Charles J.

1984 *An Asymptotically Optimal Histogram Selection Rule*, tech. rep. 34, Department of Statistics, California University, Berkeley, California. (Cited on pp. viii, 39, 40, 157, 158, 160, 161.)

Stoyan, Dietrich and Helga Stoyan

1994 *Fractals, random shapes and point fields: methods of geometrical statistics*. (Cited on pp. x, 5, 6, 9, 11, 14, 26, 35, 44, 81.)

1996 "Estimating pair correlation functions of planar cluster processes", *Biometrical Journal*, 38, 3, pp. 259-271. (Cited on p. 24.)

Sturges, Herbert A.

1926 "The choice of a class interval", *Journal of the american statistical association*, 21, 153, pp. 65-66. (Cited on pp. viii, 39, 40.)

Suweis, Samir, Enrico Bertuzzo, Lorenzo Mari, Ignacio Rodriguez-Iturbe, Amos Maritan and Andrea Rinaldo

2012 "On species persistence-time distributions", *Journal of theoretical biology*, 303, pp. 15-24. (Cited on p. 106.)

Tanaka, Ushio, Yosihiko Ogata and Dietrich Stoyan

2008 "Parameter Estimation and Model Selection for Neyman-Scott Point Processes", *Biometrical Journal*, 50, 1, pp. 43-57. (Cited on pp. 24, 87.)

Taylor, L. Roy

1978 "Bates, Williams, Hutchinson-a variety of diversities", *Diversity of insects faunas*, pp. 1-18. (Cited on p. 62.)

Ter Steege, Hans, Nigel C. A. Pitman, Daniel Sabatier, Christopher Baraloto, Rafael P. Salomão, Juan Ernesto Guevara, Oliver L. Phillips, Carolina V. Castilho, William E. Magnusson, Jean-François Molino et al.

2013 "Hyperdominance in the Amazonian tree flora", *Science*, 342, 6156, p. 1243092. (Cited on pp. x, xi, 131, 140, 143, 148.)

Ter Steege, Hans, Daniel Sabatier, Sylvia Mota de Oliveira, William E. Magnusson, Jean-François Molino, Vitor F. Gomes, Edwin T. Pos and Rafael P. Salomão

2017 "Estimating species richness in hyper-diverse large tree communities", *Ecology*, 98, 5, pp. 1444-1454. (Cited on pp. 62, 111, 115, 128, 135, 143.)

Thomas, Marjorie

1949 "A generalization of Poisson's binomial limit for use in ecology", *Biometrika*, 36, 1/2, pp. 18-25. (Cited on p. 30.)

Tovo, Anna

2014 *Spatial Aggregation in Ecology. Models and Data Analysis*, Master Thesis, Università degli Studi di Padova. (Cited on p. viii.)

Tovo, Anna and Marco Favretti

2017 "The distance decay of similarity in tropical rainforests. A spatial point processes analytical formulation", *Theoretical Population Biology*, in press. (Cited on pp. x, 75-77, 80.)

Tovo, Anna, Marco Formentin, Marco Favretti and Amos Maritan

2016 "Application of optimal data-based binning method to spatial analysis of ecological datasets", *Spatial Statistics*, 16, pp. 137-151. (Cited on pp. viii, 11, 14, 15, 22, 30, 39, 119, 126.)

Tovo, Anna, Samir Suweis, Marco Formentin, Marco Favretti, Igor Volkov, Jayanth R. Banavar, Sandro Azaele and Amos Maritan

2017 "Upscaling Species Richness and Abundances in Tropical Forests", *Science Advances*, 3, 10, DOI: https://doi.org/10.1126/sciadv.1701438. (Cited on pp. xii, 30, 106, 118, 139.)

Tsallis, Constantino

2001 "I. Nonextensive statistical mechanics and thermodynamics: Historical background and present status", in *Nonextensive statistical mechanics and its applications*, Springer, pp. 3-98. (Cited on p. 167.)

Tsallis, Constantino, Renio S. Mendes and Anel R. Plastino

1998 "The role of constraints within generalized nonextensive statistics", *Physica A: Statistical Mechanics and its Applications*, 261, 3, pp. 534-554. (Cited on p. 63.)

Tuomisto, Hanna

2010 "A consistent terminology for quantifying species diversity? Yes, it does exist", *Oecologia*, 164, 4, pp. 853-860. (Cited on p. 168.)

Veech, Joseph A. and Thomas O. Crist

2010a "Diversity partitioning without statistical independence of alpha and beta", *Ecology*, 91, 7, pp. 1964-1969. (Cited on p. 168.)

2010b "Toward a unified view of diversity partitioning", *Ecology*, 91, 7, pp. 1988-1992. (Cited on p. 168.)

Vere-Jones, David

1970 "Stochastic models for earthquake occurrence", *Journal of the Royal Statistical Society. Series B (Methodological)*, pp. 1-62. (Cited on p. 5.)

Volkov, Igor, Jayanth R. Banavar, Fangliang He, Stephen P. Hubbell and Amos Maritan

2005 "Density dependence explains tree species abundance and diversity in tropical forests." *Nature*, 438, 7068 (Dec. 2005), pp. 658-61, ISSN: 1476-4687, DOI: 10.1038/nature04030, http://www.ncbi.nlm.nih.gov/pubmed/16319890. (Cited on pp. xi, 105, 111.)

Volkov, Igor, Jayanth R. Banavar, Stephen P. Hubbell and Amos Maritan

2003 "Neutral theory and relative species abundance in ecology." *Nature*, 424, 6952 (Aug. 2003), pp. 1035-7, ISSN: 1476-4687, DOI: 10.1038/nature01883, http://www.ncbi.nlm.nih.gov/pubmed/12944964. (Cited on pp. 178, 179.)

2007 "Patterns of relative species abundance in rainforests and coral reefs", *Nature*, 450, 7166, pp. 45-49, http://dx.doi.org/10.1038/nature06197. (Cited on pp. xi, xii, 112, 120, 139.)

Volterra, Vito

1927 *Variazioni e fluttuazioni del numero d'individui in specie animali conviventi*, vol. 2, pp. 31-113. (Cited on p. 172.)

Wang, Ji-Ping Z. and Bruce G. Lindsay

2005 "A penalized nonparametric maximum likelihood approach to species richness estimation", *Journal of the American Statistical Association*, 100, 471, pp. 942-959. (Cited on p. 106.)

Watson, David M., David A. Roshier and Thorsten Wiegand

2007 "Spatial ecology of a parasitic shrub: patterns and predictions", *Austral Ecology*, 32, pp. 359-369. (Cited on p. 24.)

White, Ethan P., Katherine M. Thibault and Xiao Xiao

2012 "Characterizing species abundance distributions across taxa and ecosystems using a simple maximum entropy model", *Ecology*, 93, 8, pp. 1772-1778. (Cited on pp. 62, 111.)

Whittaker, Robert H.

1960 "Vegetation of the Siskiyou mountains, Oregon and California", *Ecological monographs*, 30, 3, pp. 279-338. (Cited on p. 61.)

1972 "Evolution and measurement of species diversity", *Taxon*, pp. 213-251. (Cited on pp. 61, 75, 167, 168.)

Wiegand, Thorsten, Savitri Gunatilleke, Nimal Gunatilleke and Toshinori Okuda

2007 "Analyzing the spatial structure of a Sri Lankan tree species with multiple scales of clustering", *Ecology*, 88, 12, pp. 3088-3102. (Cited on p. 24.)

Wiegand, Thorsten and Kirk A. Moloney

2013 *Handbook of spatial point-pattern analysis in ecology*, CRC Press. (Cited on pp. 9-11, 14, 15, 23, 27, 43-45, 50, 51, 63.)

Wolda, Henk

1981 "Similarity indices, sample size and diversity", *Oecologia*, 50, 3, pp. 296-302. (Cited on p. 64.)

1983 "Diversity, diversity indices and tropical cockroaches", *Oecologia*, 58, 3, pp. 290-298. (Cited on pp. 63, 64.)

Yule, George U. and Maurice G. Kendall

1950 "An Introduction to the Theory of Statistics", *Charles Griffin, London*. (Cited on p. viii.)

Zillio, Tommaso, Jayanth R. Banavar, Jessica L. Green, John Harte and Amos Maritan

2008 "Incipient criticality in ecological communities." *Proceedings of the National Academy of Sciences of the United States of America*, 105, 48 (Dec. 2008), pp. 18714-7, ISSN: 1091-6490, DOI: 10.1073/pnas.0807380105. (Cited on p. 143.)

Zillio, Tommaso and Fangliang He

2010 "Inferring species abundance distribution across spatial scales", *Oikos*, 119, 1, pp. 71-80. (Cited on p. 151.)

Zipkin, Joseph R., Frederic P. Schoenberg, Kathryn Coronges and Andrea L. Bertozzi

2016 "Point-process models of social network interactions: parameter estimation and missing data recovery", *European Journal of Applied Mathematics*, 27, 3, pp. 502-529. (Cited on p. 5.)

List of Symbols

$(\Omega, \mathcal{A}, \mathbb{P})$	probability space
$(\rho_{\boldsymbol{X}}, \mu_{\boldsymbol{X}}, \gamma_{\boldsymbol{X}})$	cluster parameters of a Neyman-Scott proces $\boldsymbol{X}$
$1_{\boldsymbol{X}}^{B}$	presence indicator of $\boldsymbol{X}$ associated with B
$\alpha_{\boldsymbol{X}}$	spatial-dependent alpha-diversity index
χ_B	characteristic function of B
$\beta_{\boldsymbol{X}}$	cluster parameter of an exponential cluster process $\boldsymbol{X}$ or spatial-dependent beta-diversity index
$\chi(A, B)$	spatial density of Sørensens's similarity index between regions A and B
$\chi_{\boldsymbol{X},\infty}$	asymptotic value of the similarity decay function
$\chi_{\boldsymbol{X}}$	similarity decay function of the superposed spatial point process $\boldsymbol{X}$
$\hat{\gamma}_{\boldsymbol{X}}$	empirical scaling faction for the similarity decay function
$\hat{\chi}_{\boldsymbol{X}}(\cdot; a)$	estimator for the similarity decay function computed for cells of area a
$\underline{\hat{M}}$	optimal bin number according to Knuth's method
$\hat{\xi}_{p^*}$	ξ parameter of the negative binomial SAD at the sample scale p^*
$\hat{s}_{\boldsymbol{X}}$	estimation of the statistic $s_{\boldsymbol{X}}$ of a point process $\boldsymbol{X}$
$\lambda_{\boldsymbol{X}}$	intensity function of the point process $\boldsymbol{X}$
$\mathbb{I}$	indicator function
$\mathbb{P}^x$	Palm probability measure of the point process $\boldsymbol{X}$ at the location x
$\mathbf{P}_{\boldsymbol{X}}^{x}$	Palm distribution of the point process $\boldsymbol{X}$ at the location x
$\mathcal{B}_r$	ball centred in the origin with radius r in $\mathbb{R}^2$
$\mathcal{N}$	random measure
$\mathcal{N}_{\boldsymbol{X}}$	random measure associated to the point process $\boldsymbol{X}$

$\mathcal{P}(k \mid n, p)$	sampling probability, i.e. conditional probability that a species has abundance k at the scale p, given that it has n individuals in the whole area A
$\mathcal{P}(n \mid r, \xi)$	negative binomial functional form of the SAD at a whole forest scale
$\mathcal{P}_{binom}(k \mid n, p)$	binomial distribution of parameters n and p
$\mathcal{P}_{sub}(k \mid p)$	functional form of the SAD at a local scale p
$\mathcal{S}_{\boldsymbol{X}}$	support of $\mathcal{N}_{\boldsymbol{X}}$
$\mathcal{W}$	obervation window
$\mathscr{N}$	set of counting measures
μ	Lebesgue measure on $\mathbb{R}^2$
$\nu_{2,\boldsymbol{X}}$	second moment measure of $\boldsymbol{X}$
$\nu_{[2],\boldsymbol{X}}$	second factorial moment measure of the point process $\boldsymbol{X}$
$\nu_{\boldsymbol{X}}$	intensity measure of the point process $\boldsymbol{X}$
Ω_{0-10}	relative neighbourhood density
$\overline{x_{\boldsymbol{X}}}$	value of the random variable $x_{\boldsymbol{X}}$ for a realisation of the process $\boldsymbol{X}$
ϕ_n	number of species in a community having n individuals
Π_k	boxcar function of the k^{th} bin
π_k	probability mass of the k^{th} bin
$\rho_{\boldsymbol{X}}$	second moment density of the point process $\boldsymbol{X}$
$\sigma_{\boldsymbol{X}}$	standard deviation of a bi-variate Gaussian distribution in a modified Thomas process
$\mathbf{P}_{\boldsymbol{X}}$	number distribution of the point process $\boldsymbol{X}$
$\boldsymbol{X}$	spatial point process
$\boldsymbol{X} \cup \boldsymbol{Y}$	superposition of the point processes $\boldsymbol{X}$ and $\boldsymbol{Y}$
$\boldsymbol{X}_c$	representative cluster's process in a Neyman-Scott process
$\boldsymbol{X}_p$	parents' process in a Neyman-Scott process
$\mathrm{JAC}(A, B)$	Jaccard index of similarity between disjoint regions A and B
$\mathrm{SØR}(A, B)$	Sørensen index of similarity between disjoint regions A and B
${}^{q}D_{\alpha}$	alpha-diversity or Hill's number of order q

$b_{\boldsymbol{X}}$	cluster parameter of a Cauchy cluster process $\boldsymbol{X}$
D	Simpson's diversity index
$d(\boldsymbol{X}, u)$	contact distance between the point process $\boldsymbol{X}$ and a point $u \in \mathbb{R}^2$
D''	Gini-Simpson's index
D'	Simpson's dominance index
$d_{\gamma_{\boldsymbol{X}}}$	dispersal kernel of a Neyman-Scott process $\boldsymbol{X}$, depending on the cluster's set of parameters $\gamma_{\boldsymbol{X}}$
f^{pol}	f function in polar coordinates
$F_{\boldsymbol{X}}$	contact distribution function of the point process $\boldsymbol{X}$
$G_{\boldsymbol{X}}$	nearest-neighbour distance distribution function of the point process $\boldsymbol{X}$
$g_{\boldsymbol{X}}$	pair correlation function of the point process $\boldsymbol{X}$
$H_{\gamma_{\boldsymbol{X}}}$	distribution function of $h_{\gamma_{\boldsymbol{X}}}$
$h_{\gamma_{\boldsymbol{X}}}$	density function of the distance parent-daughter of a Neyman-Scott representative cluster
I_q	diversity index of order q
I_{an}	Knuth's anisotropy index
J_n	Bessel function of the first kind of order n
$J_{\boldsymbol{X}}$	J-function of the point process $\boldsymbol{X}$
K_n	modified Bessel function of the second kind of order n
k_R	kernel function of bandwith R
$K_{2,\boldsymbol{X}}$	Shiffers's K_2-function of the point process $\boldsymbol{X}$
$K_{\boldsymbol{X}}$	reduced second moment function of the point process $\boldsymbol{X}$
$L_{\boldsymbol{X}}$	L-function of the point process $\boldsymbol{X}$
N^*	total number of individuals at the sample scale p^*
N_p	total number of individuals at a local scale p
$P(n \mid p)$	RSA at the spatial scale p
p^*	sampling scale
p_s	relative abundance of species s

p_{pred}	critical sampling scale to have an upscaling estimate precision around 5%
$R_{\boldsymbol{X}}$	fixed dispersal radius of a Matérn cluster process $\boldsymbol{X}$
$r_{\boldsymbol{X}}$	mean cluster radius of a Neyman-Scott process $\boldsymbol{X}$
$r_{\max}$	maximum considered distance
S^*	total number of species at the sample scale p^*
S_p	total number of species at a local scale p
S_{pred}	predicted number of species via an upscaling method
$v_{\boldsymbol{X}}^B$	vacancy indicator of $\boldsymbol{X}$ associated with B
$\mathcal{N}$	σ-field of $\mathscr{N}$
Chao_{wor}	Chao's upscaling method based on sampling without replacement
CSR	complete spatial randomness hypothesis or process
LS	log-series upscaling method
NB	negative binomial upscaling method
RSA	relative species abundance
SAD	species-abundance distribution
SAR	species-area relationship

Index

L

M

N

O

P

Q

R

S

T

U

V

W

Z

Acknowledgments

I would like to thank the Mathematics Department "Tullio Levi-Civita" of the University of Padova for giving me the opportunity, through these three years of Doctoral School, of an important personal and intellectual growth. In particular, I wish to thank my PhD colleagues, who have accompanied me during this wonderful journey, always sharing thoughts and ideas. I also wish to thank all the Department's staff for the constant help and disposability.

A special thanks goes to my thesis' advisor, Prof. Marco Favretti, for his constant dedication and his ability to instil security and passion in every aspect of the research work.

Another important thanks goes to Marco Formentin and to my collaborators of the LIPh Lab, the Laboratory of Interdisciplinary Physics of the Physics and Astronomy Department "Galileo Galilei", in particular to Amos Maritan and Samir Suweis. The insightful discussions with them have always constituted a fundamental source of inspiration in my work.

I am indebted to all my abroad colleagues, especially Sandro Azaele of the Department of Applied Mathematics of the Leeds University, who has welcomed me in my research trips and involved me in interesting and important research projects.

I am extremely grateful to Prof. Kenichiro Shimatani and Prof. Cang Hui who have accepted the demanding task of reviewing this PhD thesis. Their insightful comments and suggestions have inspired new ideas and have addressed me to new possible developments.

Last, but not least, I would like to thank my parents, who have always supported and encouraged me in every step I took. They have always pointed me in the right direction and they never made me forget what are the important things which are worth pursuing in life.
A special thanks goes to my brother Giovanni and my sister Maria, who have represented a constant example to follow.

My last thanks goes to Gianluca, who plays a fundamental role in my life. I owe him more than I can tell.