UNIVERSITÀ DEGLI STUDI DI PADOVA

UNIVERSITÀ DEGLI STUDI DI PADOVA

Facoltà di Scienze Matematiche Fisiche e Naturali

Dipartimento di Fisica e Astronomia 'Galileo Galilei'

TESI DI LAUREA MAGISTRALE IN FISICA TEORICA

Spatial distribution of species in ecosystems using statistical mechanics techniques

Relatore: Prof. A. Maritan

Co-relatore: Dr. S. Suweis

Laureando: Adorisio Matteo

Dedicated to my Parents.

Contents

Introduction

Probably nobody took him literally when Boltzmann, almost a century ago, talking about statistical mechanics anticipated that

> *The wide perspectives opening up if we think of applying this science to the statistics of living beings, human society, sociology and so on, instead of only to mechanical bodies, can here only be hinted at in a few words.*

Since Boltzmann's statement statistical mechanics emerged as a natural theory that can investigate the properties of a system with an enormous amount of interacting entities . The scope of the formalism is almost as unlimited as the range of the natural phenomena and in principle it is applicable to matter in any state whatsoever. The success of statistical mechanics in revealing *universal properties* of inanimate matter is considered one of the most significant result. The term universal property is used in this context to emphasize the remarkable property of those systems, which seems physically unrelated, but share, somewhat unexpectedly, some non-trivial large scale properties. Somehow the extension of the statistical mechanics tools to investigate the living matter was a natural pathway driven by three main aspects.
One is related to the concept of complex system. Loosely speaking, a complex[1] system can be describe as a system where the combination "many entities + interactions" can drive it (e.g. ecosystems, the brain, financial markets) to develop macroscopical properties not directly deducible from the rules that "move" the single entities at some microscopical scale. For an outstanding example (not to mention the brain) we can think about ants colonies and their capacity to solve complex tasks [41].
The second is linked to a sort of *ubiquity* of certain "regularities" in a variety

[1] Although the term complexity as been mathematically quantified in the case of ecological community networks [1, 32, 51] its use is sometime abused.

of systems considered as complex. For example power law like distributions of meaningful system related quantities emerge in a variety of situations. For example the patch-size distribution of the vegetation in arid ecosystems [27, 45], the distribution of time intervals between a large earthquake and the next one [34] or the metabolic rates with the mass of an organism [7] follow a power law distribution. Reference [37] represents an elementary and nicely written review on power laws in various fields.

The last is linked to the quantity of data regarding lots of complex systems that in the last years started to grow at an incredible pace allowing the scientist to test their models. From the brief description above it seems that statistical mechanics is naturally built to explain collective behaviours, regardless the details of the entities and the kind of interactions, in the combination "many entities + interactions".

We will see that statistical mechanics and statistical inference are closely related [26]. This allow to take advantage of all the tools that have been developed to solve problems in statistical mechanics and use them in inference problems. Progressively, these methods have been applied to *a priori* unrelated disciplines and a prototypical example is the fact that the same class of models describing magnetic materials can be successfully applied to understand neural networks even if atoms and neurons have few things in common at the respective microscopic scale.

In general the three cornerstones of the natural sciences are: reproducible experiments, theory and simulations. In some sense, in the case of ecosystems and in particular the ones at large spatial scales that represents the central argument of this work, the role of experiments is missing because we have to face the fact that we have the system here and now. Furthremore, natural ecosystems are characterized by striking diversity of forms and functions. Their complexity derives by the tremendous number of mechanisms acting simultaneously onto the systems. After that observation the study of this kind of systems can seem hopeless and the methods of investigation can appear system dependent but the data collected in different part of our globe and at different scales revealed that ecosystems share common features that are system independent (an example can be the fact that there are signals that the micro- and macrorganisms spatial biodiversity are similarly organised [14]). In this direction, the great importance of microecology resides in the fact that the use of bacteria, algae and protozoa guarantees the reproducibility of the experiments. In fact, these species can be easily maintained in laboratory and have rapid generation times. On the contrary the macroecology suffers the lack of reproducibility of the experiments

due to the large spatiotemporal scales involved. The experiments can lead to the understanding and a possible unification of patterns that are observed in ecological systems at different spatial scales providing significant insights on their determinants thus linking the realm of micro and macro ecology.

With this big picture in mind we will focus on a general analysis of a maximum entropy model that at the end will be applied to the case study of a rainforest community to understand the spatial organisation of the community. Statistical mechanics will be present with two main roles: as a theory that can face the investigation of "many entities + interactions" and as an inference tool.

The chapters are organised as follows:

- **Chapter 1**
  We briefly explain the importance of considering explicit spatial models and we describe what kind of patterns is natural to investigate when facing the problem of characterising an ecosystem. We introduce the concept of β-diversity and briefly summarise some approaches that are used to investigate ecosystems.

- **Chapter 2**
  In this chapter the role of statistical mechanics emerges as an inference tool starting from the knowledge of partial information about the system. The central idea that will emerge is that if data are sufficient and the inference algorithm is good enough, some of the actual features of the system will eventually be detected.

- **Chapter 3**
  The tools introduced in the previous chapter will be applied to a case study to develop a spatial model maximum entropy with the aim to describe the structure of a rainforest ecosystem by means of the patterns introduced in Chapter 1. We will use the numerical algorithm explained in **Appendix A** and the theoretical tools exposed in detail in **Appendix B**.

- **Chapter 4**
  We summarise here the results and we compare our model to other approaches to understand the range of applicability and the future directions.

Chapter 1

Spatial structure and patterns in macroecology

Here we give a brief overview of the kind of patterns that can be studied in a quantitative approach to macroecology. The concepts of species area relationship, endemic area relationship and the analogue of a two point correlation function are introduced.

1.1 Spatial structure and patterns

The massive presence of detectable and sometimes repeated patterns in Nature stimulated the scientific community to develop theoretical models that try to explain this regularity. In particular, spatial ecology is a specific branch of ecology concerned with the identification of spatial patterns and their relationships to ecological phenomena. Organisms in nature are discrete entities that interact only within their immediate neighbourhood and therefore are neither distributed uniformly nor at random. They tend instead to form characteristic spatial patterns like patchy structures. This spatial variance in the environment creates diversity in communities of organisms, as well as it affects the community stability, dynamics and pattern generation. These spatial effects have been however ignored for a long time by most ecologists due to the difficulties related to their modelling.

From a theoretical point of view, in fact, the extension of non spatial models to spatial ones has brought to the increase of the number of observed phenomena [50].

Some early research was done on the spatial distribution of plants but this was predominantly from a statistical point of view. Little consideration was given to model the distributions observed, or to predict how spatial patterns could arise.

The idea that the presence of a spatial component in a model could lead to the formation of spatial patterns was first proposed by Alan Turing [52]. Turing showed that a system of two chemical species could exhibit spatially uniform steady states which were stable in the absence of diffusion processes but which were driven unstable by diffusion. If one think diffusion as a stabilizing mechanism, this result is quite counterintuitive and shows that complex phenomena can arise as the result of the interplay between fundamental units.
Turing's work has been then generalized to describe the generation of biological patterns using relatively simple set of interactions [35]

Once pattern are detected and described, we can seek to discover the determinants of pattern, and the mechanisms that generate and maintain those patterns. Anyway, by understanding the mechanism one can predict properties and test them. In fact, since "correlation doesn't mean causation" we must go beyond the analysis of statistical correlations between measured quantities. A comprehension of the underlining processes is therefore crucial to clarify the evolution of the species and their ecosystem.

At a glance an ecosystem may appear as very complex and a complete understanding of it is challenging. Indeed, an ecosystem is composed by various species of plants and animals with different biological characteristics and interacting in different ways (e.g. competition, symbiosis and predation). The existence of macro-ecological patterns suggests that there are some more general principles behind particular biological processes. Recent studies have, in fact, documented spatial patterns of microbial diversity that are qualitatively similar to those observed for plants and animals [14].

In the last years neutral models of biodiversity have been introduced [23].

Hubbell has given an extensive description of a unified neutral[1] theory of biodiversity relevant for the description of taxons where trophically similar[2] species such as trees in forests are competing for resources. In this kind of dynamic the system obeys a *zero-sum rule*, i.e. every death is rapidly followed by the birth of an individual belonging to the same or to a different species such that the total number of individuals is conserved.

This is at variance with the so-called "ecological niche theories" for which the fitness of the species to an environment is the relevant feature for the dynamic of the system. The neutral models provides a "null model" to which actual data can be compared and the influence of other mechanisms can be assessed. Neutral models achieved good results at reproducing empirically observed macroscopic patterns in communities such as tropical forests [53, 5]. Recently, McGill [33] proposed a minimal set of rules that different neutral models share and with which different communities pattern can be explained. Two of this assumptions are also useful to justify our model (see Chapter 3) and they are:

- Intraspecific clustering
- Individuals of different species are placed without regard to individuals of other species

The assertion that species are placed independently is also in apparent contradiction to the vast literature in ecology devoted to the study of species interactions. However, evidence for the importance of species interactions stems mostly from species-poor communities (see [55] and references therein).

1.2 Measuring biodiversity: spatially explicit indicators

One step forward in finding a rationale behind the overwhelming complexity of an ecosystem can be achieved by introducing quantities that can be measured.

[1] One of the major assumption of Hubbell is the biological *neutrality* of all the species in the ecosystem. In few words, all the species are equal competitors with the same per capita chances of dying and reproducing.

[2] living organisms exist within webs of interactions with other living creatures , the most important of which involve eating or being eaten (trophic interactions). We can say that two species are in the same trophic level when for example they compete for the same resources.

The ever increasing amount of data of real ecosystems is useful to test theories and no progress can be made until aspects of the community have been quantified. Without these data there is nothing to explain. Field biologists have defined lots of indices and quantities trying to capture some universal and recurrent property about biodiversity. Because of the difficulty of conducting controlled, replicated experiments at large scales, theory plays an important role in investigating these concepts.

Due to the complexity of systems under examination it is not easy to find a single and meaningful quantity describing the biodiversity of the system. At a first stage it can be defined as the number of coexisting species in the ecosystem in it but this definition lacks of spatial significance. In fact, as we mentioned above, spatial structure is generaly a fundamental feature characterizing an ecosystem. During the last years two different concepts appeared among others: the α-diversity and β-diversity. The first one is a measure of the biodiversity in a single place considered perfectly uniform (e.g. number of species). The latter instead is spatial explicit measure of the biodiversity.

To quantify the biodiversity of a system we will focus on the class of β-diversity indicators. In fact, we will consider two kind of quantities that have spatial meaning.
One is the relation between the number of species per sampled area, or the species area relation (SAR), and the other is the endemic area relationship (EAR) that express the mean number of species completely contained in an area (endemic in the area). Another function that will be quantitatively introduced in chapter 3 is the two-point correlation function of species occupancy. This correlation function can give important insights on the spatial organization and persistence of species.

1.2.1 Species Area Relationship

Using the words of MacArthur [30] *"To do science is to search for repeated patterns, not simply to accumulate facts"*. So the effort of the scientific community in developing theories about the way ecosystems or communities are organized, orbits around the attempt to discover robust patterns that can be quantified within systems and compared across them.

Hereafter we give a brief overview on the species area relation that will be investigated in this work with the techniques described in Chapter 2.

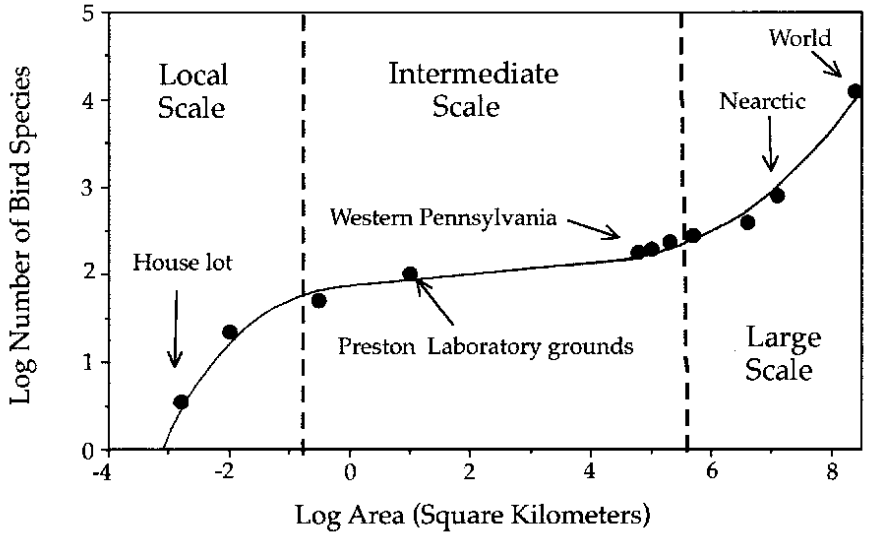

FIGURE 1.1: Example of the triphasic behaviour of the SAR curve. It is concave at local scales, approximately linear at regional scales, and finally convex at continental scales.

The SAR is defined as the average number of species S present in an area a of the ecosystem under study. This curve measures the species richness of the ecosystem and also the spatial variations of biodiversity. Since the early works of Arrehnius (e.g. [2]) this quantity has been investigated both theoretically and experimentally and the most commonly used form in literature is a power-law function like:

$$S = ca^z \tag{1.1}$$

Despite the seminal work of Gould [13], focused on the importance of the c coefficient, all the attention has been devoted to the exponent z. In fact, the behaviour of z has been experimentally investigated in relation to different variables [42]. The exponent z of the power law is in some sense related to biodiversity. If the number of individuals grows linearly with the sampled area [30], $z = 1$ describes the case of maximum biodiversity: the numbers of individuals and the number of species grows isometrically.

In real ecosystems $z < 1$ and the growth of the number of species with the area is sub-linear and the typical values of z range from 0.1 to 0.4. When observations are extended to very short and very large spatial scales, the SAR presents a triphasic behaviour as shown in Fig. 1.1.

The SAR under the hypothesis of random distribution of individuals

In this section we present one of the first models proposed for the SAR [9] based on the hypothesis of random distributed individuals.

Let's assume to have, in a given study area A, S species with abundances $\{N_i, i = 1, 2, ..., S\}$. If the N_i individuals of species i are randomly distributed in A, the probability of finding a particular individual of a given species in a sub area a is a/A. Therefore, the number of individuals, n_i, of a given species in the area a follows a binomial distribution:

$$p(n_i|a) = \binom{N_i}{n_i} \left(\frac{a}{A}\right)^{n_i} \left(1 - \frac{a}{A}\right)^{N_i - n_i} \tag{1.2}$$

so the probability of absence ($n_i = 0$) is $p(0|a) = (1 - a/A)^{N_i}$ and the one of presence of *at least* one individual of the species i is $1 - (1 - a/A)^{N_i}$. At this point we can give the expressione for the SAR in the case of random distributed species:

$$SAR(a) = S - \sum_{i=1}^{S} \left(1 - \frac{a}{A}\right)^{N_i} \tag{1.3}$$

which depends only on the abundaces N_i. It is thus possible study how the shape of the SAR changes as a function of the different species-abundace distribution [9, 21].
When we consider $a = A$ the probability $p(0|a) = (1 - a/A)^{N_i}$ exactly zero meaning that we have probability one to find a particular species. In this sense the random placement model is well defined for finite finite system and it can represents a good null model.
Anyway, in nature individuals of most species are rarely randomly distributed through space. Departures from randomness result in aggregated species with the effect that the probability of presence of the species in a sampling area a, should be less than that under random distribution. In Chapter 3 we will propose a random placement model based only on presence/absence data and not on species abundances but we will see this effect in our model where we consider an intraspecific interaction.

1.2.2 Endemic Area Relationship

The SAR curve gives information on the number of species present in a given area A. In this case a species, to be present, must have at least one individual inside the sampled area. A second quantity has been introduced in order to measure the number of species *completely* contained in the area A. We refer to it as the endemic area relation (EAR). Even if the EAR was less studied than SAR its use is fundamental to understand the behaviour of ecosystems under external disturbances such as habitat destruction. In principle one can think that the EAR and the SAR can be directly related but as we will show lately, only in a very special and biologically unrealistic case, when all species are randomly and independently distributed in space, is it possible to derive the EAR from the SAR [19]. Although species area relationship has traditionally been applied to estimate extinction rates following habitat loss, it is now established that applying EARs can yield very different and perhaps more accurate predictions

The EAR under random distribution of individuals

Under the hypothesis of randomly placed individuals it is easy to derive an expression for the EAR as well. Using the definition of the EAR given in section 1.2.2 we can thus consider the probability $p(N_i|a)$ of finding all the individuals N_i of the species i in the sub area a. All the species are independent and thus the EAR under the random placement of individuals is:

$$EAR(a) = \sum_{i}^{S} \left(\frac{a}{A}\right)^{N_i} \tag{1.4}$$

In Chapter 3 the EAR will be investigated and an expression for the random placement EAR in the case of presence absence framework will be derived and then compared to the same quantity for an interacting model.

1.2.3 Species turnover and two point correlation function

The spatial dispersion of individuals of species is central in ecological theory. Patchiness, or the degree to which individuals are aggregated or dispersed, is crucial to understand how a species uses resources, how it is used as a resource,

and to understand its reproductive biology. As the distance between sites increases, conditions of growth become more different, and it will become more likely that species found at one site do not occur at another. Thus, species composition will change as one moves across a region, a phenomenon called "turnover". It can be quantified with a 2-point correlation function for the species occurrence and it was proposed as a meaningful tool to acquire information about what controls diversity in ecological communities [10]. Since in this work we will work on a lattice we give an expression for this 2-point correlation function in Chapter 3.

1.3 Understanding spatial structure: different approaches

Different approaches have been proposed to build reliable ecological models all aimed to understand ecosystems structure, in particular to explain the SAR S-shaped curve. Before presenting our own approach (fully described in Chapter 3) and in order to understand how it differentiates from other commonly used methods, we give here a brief review of the state-of-the-art on ecosystem modelling.

1.3.1 Deterministic and stochastic dynamical models

Let's start considering one of the first spatially implicit model proposed by Levins to describe patch dynamics. It considers only the trade off between colonization and extinction. Levins made the simplifying assumption that all patches are of the same size and that migration is global, equally likely among any pair of populations and patches. The set of local populations inhabiting the network of patches is called the metapopulation[3], the size of which is given by the fraction of occupied patches, denoted by p. Anyway, the purpose of theoretical models is to isolate, for a theoretical study, some feature of real populations that happens to be of interest and not to account for as many real details as possible.

[3]In a nutshell the metapopulation approach consider a spatially structured community assembled into local patches linked by migration of organisms that has effect on local communities. This vision is different from the classical approach where in a population all individuals are equally likely to interact

In this purpose, Levins considering only the trade-off between colonization and extinction proposed a metapopulation dynamic described by the equation:

$$\frac{dp}{dt} = cp(1-p) - ep \tag{1.5}$$

where e is the rate of extinction and c rate of colonization[4]. This system presents a global stable equilibrium:

$$p^* = 1 - \frac{e}{c} \tag{1.6}$$

that for $e/c < 1$ imply that the metapopulation is predicted to persist in the habitat. Even if it is a logistic equation Levins' model was considered something new in population ecology and a first step towards further quantitative research in this field. Actually, in the literature regarding metapopulation models the Levins' model is often considered a mean field model since it is based on the assumptions that all patches are equally connected to other patches. The mean-field assumption is a good approximation when the physical environment is homogeneous. As conditions depart from those above, the mean-field approach becomes less and less appropriate. For example a lack of mixing can generate clumped distribution around individuals that deviate from the spatial averages. Heterogeneity in local environmental conditions becomes especially important if organisms only interact over short distances. Short range interactions have been identified by numerous theoretical models as a key mechanism able to maintain biodiversity [28, 8]. So the mean field approximation become weak when local populations have clumped distributions in space, either for environmental or dynamical reasons. In fact, in this case different populations are not likely to be equally connected.

In 1997, a work by Hanski et al. [16] in the field of metapopulation dynamics generalised this extinction-colonization dynamics in the case of an heterogeneous habitat to give some predictions for the species area relationship. In this case a pool of S species subdivided into a set of R patches is dynamically described by a generalization of equation (1.5) for the probability $p_{ij}(t)$ of species i being present on patch j at time t:

[4]This is the well known logistic equation $\dot{x} = \alpha x(1 - x/K)$ with $\alpha = c - e$ and "carrying capacity" $K = 1 - e/c$

$$\frac{dp_{ij}}{dt} = c_i(t)(1 - p_{ij}) - e_{ij}p_{ij} \tag{1.7}$$

where $c_i(t) \propto \sum_j p_{ij}$ is the colonization rate and e_{ij} the extinction rate for the species i in the island j.
Assuming particular expression for $c_i(t)$ and e_{ij} and studying the equilibrium probability $p^*_{ij} = \frac{c^*_i K_{ij}}{1+c^*_i K_{ij}}$ the model predicts z values ranging from 0.1 to 0.45[5].

Although the SAR is considered one of the most ubiquitous and robust pattern observed in different ecosystems at different scale and latitude, one should bear in mind that historically this power-law relation has been chosen to fit field data and its choice was not grounded on biological or dynamical considerations.
Beyond deterministic models, such as the one described by equation (1.7), we mentioned before the development of the so called "neutral models" [23] that gave a firm theoretical ground for observed biodiversity patterns.

The generalization of neutral models to describe spatial systems has been performed both analytically and numerically [58, 12, 43, 38]. Durrett and Levin [12] proposed the voter model with a speciation parameter equal for all the species [6] as a natural benchmark where to investigate the importance of spatial interactions in ecological systems.
Zillio et al. [58] using a spatially explicit master equation approach found an analytical expression for a two-point spatial correlation function thus giving this model an analytical predicting power for this spatial pattern. From the work of Zillio et al. one can also understand the importance of neutral models as null models. In fact, they also generalized the model to include the Janzen-Connell effect[7] to better reproduce the spatial patterns. As a result the voter model shows the triphasic behaviour for the SAR [38] and a dependence of z, the exponent of the power-law function, from the speciation rate.

[5] In the case of $c_i(t) \propto \sum_j p_{ij}(t)K_{ij}$ with K_{ij} the carrying capacity of the species i on patch j and choosing $e_{ij} \propto K_{ij}^{-1}$ (from empirical reasoning) one is able to find p^*_{ij}. Then a species area relantionship can be obtained as $\sum_i p^*_{ij}$.

[6] The voter model is a stochastic model. First defined by Liggett [29] has been thoroughly studied without speciation. The main results are that in a finite lattice, or in an infinite lattice with dimensionality $d \leq 2$ the system develop a stable state of monodominance (in our case we can think that a given species prevails on the other).

[7] Janzen and Connell postulated that there is an increased mortality rate of seeds and seedlings near adults that arises from the presence of pests that are host specific, i.e., specialized to that type of tree, and observational evidence supports this conclusion. This results in a negative density-dependent effect at short distance.

1.3.2 Maximum entropy approach

More recently, approaches based on information theory and in particular the method of maximum entropy[8] have also been proposed [4, 18] and in the work described here we will use this maximum entropy formalism endowing it with a spatial explicit component.
A key issue in a maximum entropy approach is to identify a group of mean quantities that represent a constraint for the system under examination. After that one can build a probability distribution that has these constraints as moments of the distribution. Even if the approach in [18] is not spatial explicit, the authors assume three quantities as fundamental to built a maximum entropy model to predict spatial pattern such as the SAR. All the ecology in this approach is in the selected constraints and these quantities are the total area of the ecosystem, total number of individuals and the total (sum over individuals) metabolic rate. Coupling this parameters to the maximum entropy method the authors are claim that they able to describe the SAR and EAR.

1.3.3 Spatial point processes

Other classical approaches are represented by the study of spatial point processes. This method focuses on understanding quantities like the SAR in terms of the underlying spatial abundance distribution [21, 15] rather than predicting the SAR from an explicit dynamical model. These processes are built to reproduce the behaviour of individuals (sessile organisms) structured in spatial clusters. Specifically they are built using simple rules: the centers of clusters are distributed in space with a constant density independent of each other. Each cluster is populated by a random number of individuals (drawn from a given distribution) and the distance of each individual from the center of the cluster is drawn from a given distribution. For example a central spatial point process is the homogeneous spatial Poisson point process. To built it imagine to subdivide a given area into sub areas and assign to them a set of n points with abundance drawn from a Poisson distribution with a certain mean. For each sub-region generate n couple of uniform number and use them as the coordinates of the n points. The Poisson spatial process represents a benchmark model and it serves for the construction of more complicated models. In Section 1.2.1 we presented the random placement model as a prototypical finite

[8] We will describe the method in the following Chapters

spatial point process. The Poisson process mentioned before can be considered as a version of random placement model for infinitely large areas and in fact the probability to find a particular species in the largest area becomes one only for an infinite area[9].

Using the Poisson spatial process Grilli et al. [15] can reproduce the triphasic behaviour of the SAR with the assumption of neutrality and independence of all the species.

Without the approximation of the infinite landscape assumption other models [57] has been considered to describe the clustering of species in finite landscapes. In [3] the authors use a wide class of spatial stochastic point processes to investigate the spatial structure and in particular the downscaling[10] of species in an ecosystem.

These point processes are formulated in continuous space and each point is placed regardless of any resolution used to subdivide the region. Superimposing a grid with a varying unit cell size on the study region, one can see how a fixed pattern of spatial locations of individuals looks under different resolutions.

Another "geometrical" approach [17], with the assumption that the spatial distribution of species is self-similar and fractal found a confirmation from field data on birds in Czech Republic [49] and claimed that a power-law SAR can be generated considering a community with a fractal spatial distribution.

Summary of the chapter

Without claiming to be complete we presented macroecological patterns with spatial significance. For the species area relationship (SAR), the most studied pattern, we sketched some method of investigation currently used. The endemic area relationship (EAR) and a two point correlation function to understand structure and spatial persistence of ecological communities were also mentioned. All these patterns will be presented again in Chapter 3 were we quantitatively analyse their behaviour using the framework of a maximum entropy model.

[9] The probability to find n out of N individuals in an area a for a given species is $P_{pois}(n) = \exp(-\mu)\frac{\mu^n}{n!}$ with $\mu = \frac{N}{A}a = \rho a$.Thus the probability to find at *least* one individual is $p = 1 - e^{-\rho a}$. Then the probability to find at least one individual for $a = A$ is $p = 1 - e^{-N}$ (less than one). We conclude that in general this model can be used for all species densities much greater than A^{-1}

[10] The word *downscaling* refers to the process of inferring the structure of the ecosystem at finer spatial scales starting from information at wider scale

In the next Chapter we describe the maximum entropy method as a general method of inference with which we justify our approach.

Chapter 2

Statistical mechanics and information theory

> "It is remarkable that a science which began with the consideration of games of chance should have become the most important object of human knowledge."
>
> Pierre-Simon Laplace

"The first reaction of nearly everybody on hearing of a mysterious principle called maximum entropy with a seemingly magical power of extracting more information from incomplete data than they contain, is disbelief.
The second reaction on sensing that there does seem to be something in it, is puzzlement. How is it possible that a quantity belonging to thermodynamics could escape from that setting and metamorphose itself into a principle of reasoning able to resolve logical ambiguities in situations that has nothing to do with thermodynamics?".

The words above are the ones of Jaynes about the maximum entropy method, an inference framework already present at the very beginning of statistical mechanics formalization.
In this Chapter we introduce the theoretical framework giving a brief description of the maximum entropy method and the rationale behind it. We can consider

this a key chapter because it presents the techniques that we use throughout our work.

2.1 Maximum entropy approach: a brief overview

An useful explanation of how to gain knowledge about a system characterised by a huge number of degrees of freedom starting from a minimal set of information, from empirical measurements, is given by the work of Jaynes [24, 26]. In Jaynes' works statistical physics and information theory walk side by side and the former is regarded as a way to draw inferences from partial information about the system under examination. Hereafter to explain the method of maximum entropy we will briefly describe a classical example and then generalize to give a more formal explanation.

To base our problem on a practical ground suppose that a dice, with the usual six faces, has been rolled a very large number of times; if we only know that the average result is $\bar{m} = 4.5$, we can ask what probability should we assign for the various outcomes 1 to 6. We call p_i the probability for the outcome i ($i = 1. \ldots, 6$).
At this point we only own the information about the average of the results and since:

$$\bar{m} = \sum_i i p_i = 4.5 \tag{2.1}$$

we can rule out the uniform distribution ($p_i = \frac{1}{6} \quad \text{for all} \quad i$) Jaynes has suggested that in this type of situation we should make the assignment by using the principle of maximum entropy that is to choose probability assignment for each configuration that maximize the function:

$$S[p] = S(p_1, \cdots, p_6) = -\sum_i p_i \ln p_i \tag{2.2}$$

subjected to the given constraints. In the case of $p_i = 0$ for some i the value of the corresponding summand $0 \ln 0$ is taken to be 0 consistent with the limit: $\lim_{p_i \to 0} p_i \ln p_i = 0$.
If we knew nothing but $\sum_i p_i = 1$ the result will be the uniform pdf $p_i = 1/6 \quad \forall i$. Apart from the mean value of the outcomes we have always the constrain on the normalization of the probability: $\sum_i p_i = 1$.

Maximizing $S[p]$ under this constraints we get[1]:

$$p_i = \frac{1}{Z(\lambda)} e^{\lambda i} \quad \text{with} \quad Z(\lambda) = \sum_i e^{\lambda i} \tag{2.3}$$

Due to the form of p_i it's easy to see that the *theoretical* mean value of the outcomes computed with p can be expressed as:

$$\langle m \rangle = \frac{\partial \ln Z(\lambda)}{\partial \lambda} \tag{2.4}$$

from which we can impose $\langle m \rangle = \bar{m}$:

$$\bar{m} = \frac{\partial \ln Z(\lambda)}{\partial \lambda} \tag{2.5}$$

Solving this equation (numerically) for λ we obtain a probability like the one shown in fig. **??**:

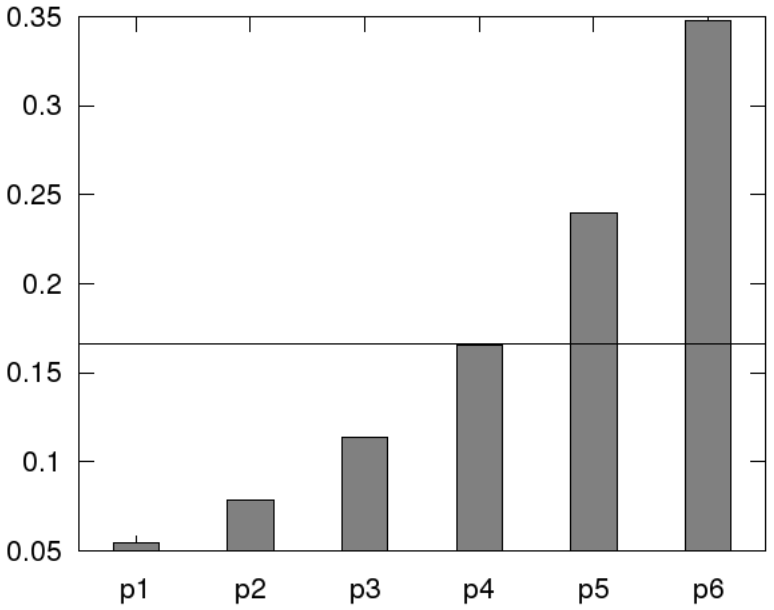

FIGURE 2.1: Probability for the unfair dice with $\sum_i i p_i = 4.5$. The straight line represents the $p_i = 1/6 \forall i$

This example is trivial but it shows the power of maximum entropy principle as a inference method and its data based starting point. It is reasonable to ask why the function (2.2) should be particular favoured as a selection criterion. Concluding this qualitative overview we can state that a criterion of maximum

[1] for the details see appendix B

entropy based on the functional (2.2) is highly consistent and any other choice of "information measure" will lead to inconsistencies (e.g. negative probabilities[2]). In fact Shore and Johnson (see [40] and references therein) views the maximization of entropy as a fundamental requirement for ensuring that inferences drawn from data satisfy basic self-consistency requirements of probabilities.
As Shannon wrote *"the real justification of these definitions,however, reside in their implication"* and using the words of Jaynes *"many years of use of the maximum entropy principle has not revealed any inconsistency; and of course we do not believe that one will ever be found."*

To introduce the maximum entropy principle in a more general form we can think to a typical situation where the state of our system can be specified by a variable $\vec{\sigma}$, the "micro state" of the system.
While statistical mechanics methods successfully describe macroscopic systems in terms of thermodynamic variables directly related to the microscopic behaviour, it is not immediately clear how to use its principles to build models of systems that are "not in thermal equilibrium". In other words it is not clear how to choose the right variables resembling the thermodynamic ones that can be manipulated using statistical mechanics techniques.

2.2 Maximum entropy probabilities

As mentioned before a typical information on a micro state of a system can be described by a vector $\vec{\sigma}$ of N (binary) components leading to $W = 2^N$ different states. To state the problem once and for all let us denote an *observed realization* of the system by $\hat{\sigma}$. Our goal, is to find a probability distribution $p_{\vec{\sigma}}$ which avoids bias and reproduces given empirical constraints. Here, the empirical constraint can be thought as a function $\phi(\hat{\sigma}) = \hat{\phi}$. The great advance provided by information theory lies in the discovery that there is a unique, unambiguous functional $S(\{p_{\vec{\sigma}}\})$ that quantifies the "amount of uncertainty" of a system:

$$S(\{p_{\vec{\sigma}}\}) \equiv -\sum_{\vec{\sigma}} p_{\vec{\sigma}} \ln p_{\vec{\sigma}} \tag{2.6}$$

[2] A mere fact is that the singularity of the logarithm in (2.6) ensure $p_i \geq 0$

The fact that Jaynes treated statistical mechanics as an "exercise" in inferential calculus somehow bridges statistical mechanics (in particular the works of Gibbs) and information theory in the particular case of *equal probabilities* ($p_{\vec{\sigma}} = 1/W, \forall i$) that substituted into equation (2.6) resembles the well known expression:

$$S = k_B \ln W \tag{2.7}$$

where k_B is the Boltzmann's constant and W the number of microscopical state of the system.

Actually, Gibbs imagined an *ensemble* of all the possible configurations of a system of N particles with each configuration being a potential microstate i of the entire system[3].
Therefore the macrostate of a system is described by a probability $p_{\vec{\sigma}}$ to be in the state $\vec{\sigma}$. Gibbs characterized the preferred macrostate as the one that maximizes the entropy given some constraints. If Clausius stated that the entropy of a system *tends* to increase Gibbs founded his work on the stronger assumption that entropy *must* increase up to the maximum value permitted by whatever constraints (conservation of energy, volume molecules numbers, etc.) are imposed.
The dice example we proposed above made use of a Lagrange multiplier λ to optimize the entropy subjected to a constrain. In fact the same method of Lagrange multipliers can be used to obtain the Gibbs distribution for the *canonical ensemble* in the case we fix the average energy of the system:

$$\bar{E} = \sum_{\vec{\sigma}} E_{\vec{\sigma}} p_{\vec{\sigma}}$$

Using β as a Lagrange multiplier we end up with the expression for p_i that characterizes the *canonical ensemble*[4]:

$$p_{\vec{\sigma}} = \frac{1}{Z} e^{-\beta E_{\vec{\sigma}}}$$

[3]introducing the ensemble Gibbs shifted a little the viewpoint. In fact he generalized the work of Boltzmann, that in fact considered only non interacting systems, to systems of interacting particles (for a crystal clear explanation see [25])

[4]β is directly linked to the temperature T of the system at equilibrium via the thermodynamic relation $\frac{1}{T} = \partial S / \partial \bar{E}$

In that case each of the microstate $\vec{\sigma}$ has a fixed number N of particles.
Obviously the analysis can be extended to include other constraints thus generating other distributions. For example the *grand canonical ensemble* is described by:

$$p_{\vec{\sigma}} = \frac{1}{Z} e^{-\beta E_{\vec{\sigma}} - \gamma N_{\vec{\sigma}}}$$

obtained with the additional constrain $\sum_{\vec{\sigma}} p_{\vec{\sigma}} N_{\vec{\sigma}} = \hat{N}$ (in this case γ is directly related to the chemical potential μ). Each microstate $\vec{\sigma}$ is characterized by $E_{\vec{\sigma}}$ and $N_{\vec{\sigma}}$.
In other words Gibbs' use of the second law of thermodynamic to predict equilibrium states was virtually identical in rationale with the maximum entropy inference method that we are going to describe. The experimental confirmation of Gibbs' thermodynamic predictions and the success of maximum entropy predictions outside thermodynamics are just two illustrations of the power of that theory.
Anyway, Gibbs classical statistical mechanics predictions also turned to be incorrect (e.g. specific heats) and this was as important as the right predictions. In fact in inductive inference and in particular in the maximum entropy theory if the predictions do not agree with the experimental data we have not "'failed' but instead discovered something new.
Some of the predictions could be wrong and those instances will open the door to new basic knowledge. Therefore, entropy and maximum entropy principle can be used as a tool to learn.

An information-based view of the rationale behind Jaynes' work is to recognize that we are concerned with the prediction of reproducible macroscopic behaviour from a description of the microscale trough probability distribution of the configurations.
To set up the mathematical framework we can think to start from a set of M measured quantities $\hat{\mathbf{\Phi}} = \{\hat{\phi}^{\mu}\}_{\mu=1}^{M}$ the experimental or empirical values of the set $\mathbf{\Phi} = \{\phi^{\mu}(\vec{\sigma})\}_{\mu=1}^{M}$.
We also introduce M parameters $\boldsymbol{g} = \{g_1, \cdots, g_M\}$ that will serve as Lagrange multipliers. We want to find the probability distribution $p(\vec{\sigma}|\boldsymbol{g})$ such that the mean quantities calculated with that $p(\vec{\sigma}|\boldsymbol{g})$ reproduce the measured ones:

$$\langle \phi^{\mu}(\vec{\sigma}) \rangle_{\boldsymbol{g}} \equiv \sum_{\{\vec{\sigma}\}} p(\vec{\sigma}|\boldsymbol{g}) \phi^{\mu}(\vec{\sigma}) = \hat{\phi^{\mu}} \tag{2.8}$$

At this point, in order to satisfy the fact that $p(\vec{\sigma}|\boldsymbol{g})$ must be normalized, we

can enlarge the set of multipliers and constraints adding a parameter g_0 linked to the constrain $\sum_{\{\vec{\sigma}\}} p(\vec{\sigma}|\boldsymbol{g}) = 1$ equivalent to take $\phi^0(\vec{\sigma}) = 1 = \hat{\phi}^0$.
Using p instead of $p(\vec{\sigma}|\boldsymbol{g})$ for readability, we want to find a maximum for the functional:

$$H[p] = S[p] - \sum_{\mu=0}^{M} g_\mu(\langle\phi^\mu(\vec{\sigma})\rangle_{\boldsymbol{g}} - \hat{\phi}^\mu) \tag{2.9}$$

notice that $\langle\phi^\mu(\vec{\sigma})\rangle_{\boldsymbol{g}}$ is a functional of p as well (see eq. (2.8)).

As we show in the appendix B this operation returns a probability distribution:

$$p(\vec{\sigma}|\boldsymbol{g}) = \exp\left(\sum_{\mu=0}^{M} g_\mu\phi^\mu(\vec{\sigma})\right) \tag{2.10}$$

Since the partition function Z in statistical mechanics ensures the normalization of the probability distribution we can rewrite equation (2.10) by eliminating g_0 as:

$$p(\vec{\sigma}|\boldsymbol{g}) = \frac{1}{Z(\boldsymbol{g})}\exp\left(\sum_{\mu>0} g_\mu\phi^\mu(\vec{\sigma})\right) \tag{2.11}$$

where $Z(\boldsymbol{g}) = \sum_{\{\vec{\sigma}\}}\exp\left(\sum_{\mu>0} g_\mu\phi^\mu(\vec{\sigma})\right)$. It is useful to remember that from $Z(\boldsymbol{g})$ we can obtain mean quantities as:

$$\langle\phi^\mu(\vec{\sigma})\rangle_{\boldsymbol{g}} = \frac{\partial \ln Z(\boldsymbol{g})}{\partial g_\mu} \tag{2.12}$$

Concluding this section we can identify $g_0 = -\ln Z(\boldsymbol{g})$ and $\beta\mathcal{H}(\vec{\sigma}) = -\sum_{\mu>0} g_\mu\phi^\mu(\vec{\sigma})$ as the Hamiltonian of the system.
The analogy with statistical mechanics is complete if we remember that the reduced free energy $F(\boldsymbol{g}) = -\ln Z(\boldsymbol{g})$.

The maximum entropy principle presented above is often invoked in order to justify the model (2.10) as the simplest one (i.e., with higher entropy) which is able to explain a given set of empirical averages. However, why choose the probability that maximizes the entropy? In Gibbs case, an equilibrium argument could be used and citing Boltzmann *"in most cases the initial state of a system will be a very unlikely state. From this state the system will steadily evolve towards more likely states until it has finally reached the most likely state, i.e. the state of thermal equilibrium"*.

Anyway, what about cases of general inference? As Shannon described, if entropy is thought of as a measure of information, then one should choose the probability that includes the least amount of information, just the information given. In other words, other probabilities could be chosen that satisfy the constraints, but the probability that maximizes the entropy is the least informative or as someone say, the most honest (based only on the data and not on other assumptions). In effect, on the space of probability distributions a "generalization" of the entropy, the Kullback-Liebler divergence can be thought as a "metric" (it lacks the property of symmetry, see Appendix B). Thus with this tool we can measure distances between probability distribution.

Even though we ended up with a probability distribution that is but the Gibbs' distribution valid at thermodynamical equilibrium, Jaynes saw the Gibbs' formalism of equilibrium statistical mechanics as a general form of statistical inference that could be extended to non-equilibrium systems, as well as to other problems requiring prediction from insufficient data. Indeed, a general non-equilibrium "state" is not a single stationary state as in the equilibrium case but rather a "path" through state space.

2.3 The inverse problem: from $\hat{\mathbf{\Phi}}$ to $\boldsymbol{g}$

Typically the problem considered by statistical mechanics is to find the observables associated with a given statistical model described by a probability distribution. Usually this is called the *direct problem*.
In our work we faced the opposite task, the *inverse problem*. As described in chapter 3, given a vector $\hat{\mathbf{\Phi}}$ of empirical averages we had to find a coupling vector $\boldsymbol{g}^*$ such that the maximum entropy probability (2.10) was able to reproduce $\hat{\mathbf{\Phi}}$ when $\boldsymbol{g} = \boldsymbol{g}^*$.
The method used in this work is presented in Appendix B Section B.2 where we show that to find $\boldsymbol{g}^*$ we can optimize the function

$$H(\boldsymbol{g}) = \ln Z(\boldsymbol{g}) - \sum_{\mu>0} \hat{\phi}^\mu g_\mu \tag{2.13}$$

The choice of the function above can be justified using information theory where the Kullback-Liebler divergence (see Appendix B) can be considered a

"distance" on the space of probability distributions. Thinking in this terms we will use the Kullback-Liebler divergence to find the coupling vector $\boldsymbol{g}^*$ that minimise the "distance" between the *true* probability distribution and the inferred one.

Summary of the chapter

We presented maximum entropy method as a general inference technique. Its characteristics places it at a crossroads between several disciplines such as statistical physics, information theory, quantitative biology [44, 48] and theoretical ecology [4, 18]. Based on some quantitative knowledge about the system we showed how it is possible to choose a probability distribution for the configurations of the system subjected to macroscopic constraints.
In the next Chapter we will describe how we developed and tested our model starting from this inference method.

Chapter 3

Inferring ecosystem organization with a spatial maximum entropy model

In this chapter we explain how using the techniques briefly presented in Chapter 2 we are able to investigate the emergent organization of a tropical forest. We try to describe observed pattern counting only on mean occurrence and pairwise correlations extracted from presence/absence data. We will analyse two cases. Firstly we focus on the whole ecosystem (considering all the species). Then, since it is difficult to include rare species in census counts, having a theory or a method that reliably predicts useful patterns is of enormous value. Thus we will analyse the case of rare species as well.

3.1 MaxEnt at works: the case of a tropical forest

Characterizing an ecosystem assembly can be done using different information. The number of individuals, or the abundance, of a species in an area is a fundamental ecological parameter (for example it is crucial when making management and conservation decisions).
However, unless the scale is very fine or localized (e.g., in an accessible habitat or permanent forest plot devoted to scientific investigation), abundance information is not available. At coarse or regional scales for many species, information

on commonness and rarity is, at best, limited to a map of their presence or absence.

Although not as rich in information as the species abundances data where the number of individuals for each species is known, presence/absence data are crucial to the understanding of ecosystem functioning. Presence absence data store a "low level" information that anyway is sufficient to investigate biodiversity indicators such as the species-area relation, the endemic area relation and the correlation function which can reveal important information of the underlying community structure.

From abundance data to presence absence data

We focus our analysis on spatial data coming from the tropical rainforest of Barro Colorado Island (BCI), Panama. The censused data represent position of all the individuals in an area of 50 hectares (500×1000 m^2) containing about 350000 individuals subdivided in about $S = 300$ species. For the BCI ecosytem we own information on the abundaces of species but in order to test our approach we transform abundaces in presence/absence dataPresence/absence data has been obtained dividing the surveyed area in $N = 256$ cells and assigned at each of the cell a variable σ_i^α for $i \in \{1, \cdots, N\}$ and $\alpha \in \{1, \cdots, S\}$ such that $\sigma_i^\alpha = 1$ when species α is present at the cell i and $\sigma_i^\alpha = 0$ if it is absent. Applying this procedure we end up with lattice like configurations $\vec{\sigma}^\alpha$, one for each species (see Figure 3.1).

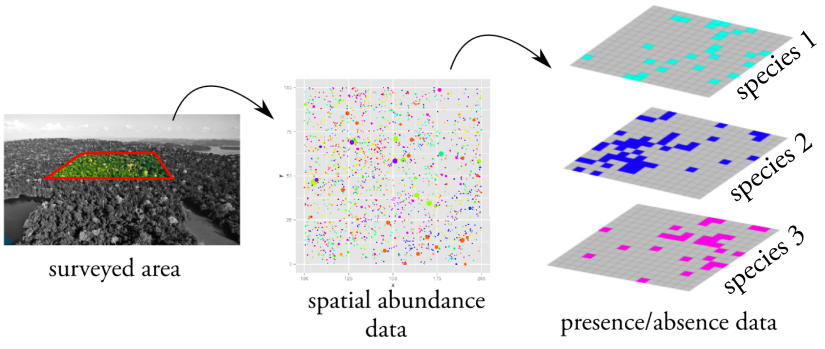

FIGURE 3.1: From spatial data to presence absence data

3.1.1 The spatial correlation function

We subdivided the area in N plots and we defined the occupancy two point correlation function (hereafter correlation function) on the lattice. Given two sites i and j at distance r_{ij} (measured in units of nearest neighbours sites) we define the correlation function for the species α as:

$$C^{\alpha}(r) = \frac{\sum\limits_{i \neq j} \sigma_i^{\alpha} \sigma_j^{\alpha} \delta(r - r_{ij})}{\sum\limits_{i \neq j} \delta(r - r_{ij})} \tag{3.1}$$

where δ is the Kronecker delta. Averaging over the species we define the correlation function for the entire ecosystem made of S species:

$$C(r) = \frac{1}{S} \sum_{\alpha} C^{\alpha}(r) \tag{3.2}$$

The results about the correlation function will be described in Section 3.3.2 where we will compare our model to the one of randomly distributed species.

3.1.2 The spatially explicit maximum entropy model

We define the configuration of the system (in our case the BCI) as $\{\vec{\sigma}^{\alpha}\}_{\alpha=1}^{S}$ with $\sigma_i^{\alpha} \in \{0, 1\}$. Following the analysis presented in Chapter 2 we can build our maximum entropy model.
First of all we identify the constraints $\hat{\mathbf{\Phi}}_{\boldsymbol{\alpha}}$. We assume that all the fundamental information of the spatial organization of a species α are enclosed in two constraints, its occurrence[1] $\hat{M}^{\alpha}$ defined as:

$$\hat{\phi}_{\alpha}^{1} = \hat{M}_{\alpha} = \sum_{i} \hat{\sigma}_i^{\alpha} \tag{3.3}$$

and its nearest neighbours correlation function[2] $\hat{E}^{\alpha}$ defined as:

$$\hat{\phi}_{\alpha}^{2} = \hat{E}^{\alpha} = \sum_{\langle i,j \rangle} \hat{\sigma}_i^{\alpha} \hat{\sigma}_j^{\alpha} \tag{3.4}$$

[1] Given that we have defined the problem on a lattice the occurrence of species must not be confused with its abundance. Actually, increasing the resolution the occurrence tends to the abundance

[2] The normalised nearest neighbour correlation function ($r = 1$) is defined as $C_1^{\alpha} = \frac{1}{N_b} \sum_{\langle i,j \rangle} \sigma_i^{\alpha} \sigma_j^{\alpha}$ where N_b is the number of nearest neighbours pairs

where the notation $\sum_{\langle i,j \rangle}$ means summation over nearest neighbours sites. This quantity is directly linked to equation (3.1) with $r = 1$.
We consider only pairwise interactions neglecting all higher order interactions. Obviously the dominance of pairwise interactions on the higher order ones is not automatically satisfied and this aspect as been discussed [46, 47] (for maximum entropy models applied to study the properties of neural activity). To recall the notation of Chapter 2 we identify:

$$\hat{\mathbf{\Phi}}_{\boldsymbol{\alpha}} = \{\hat{\phi}^1_\alpha, \hat{\phi}^2_\alpha\} = \{\hat{M}_\alpha, \hat{E}_\alpha\} \tag{3.5}$$

In addition we introduce a simplification concerning the interaction between different species: we consider the species as non interacting deciding to describe only the intraspecific interaction.
At a first glance this can be regarded as a really strong approximation but as shown in Volkov et al. [54] for the BCI forest the effects of interspecific pairwise interactions are relatively weak compared with intraspecific ones. Volkov et al. using a maximum entropy approach and a dynamical approach described by stochastic birth/death equations, also suggests that higher order interactions involving three species are negligible. Furthermore, the few studies that extensively estimate interaction strength suggest that distributions of interaction strength tend to be skewed toward few strong and many weak interactions [56].
At this point we have all the ingredients that are necessary to build our maximum entropy model. Thus assuming that the community patterns are only linked to the constraints imposed by $\hat{\mathbf{\Phi}}_{\boldsymbol{\alpha}}$ we have to find the maximum entropy distribution that is least biased by the information not taken into account.

Let us define:

$$\mathbf{\Phi}_{\boldsymbol{\alpha}} = \{\phi^1_\alpha, \phi^2_\alpha\} = \Big\{ \sum_i \sigma^\alpha_i, \sum_{\langle i,j \rangle} \sigma^\alpha_i \sigma^\alpha_j \Big\} \tag{3.6}$$

and the conjugated couplings (the Lagrange multipliers of Chapter 2):

$$\boldsymbol{g}_{\boldsymbol{\alpha}} = \{h_\alpha, J_\alpha\} \tag{3.7}$$

Following the prescription of the principle of maximum entropy we end up with:

$$p(\vec{\sigma}^\alpha | \boldsymbol{g}_{\boldsymbol{\alpha}}) = \frac{1}{Z(\boldsymbol{g}_{\boldsymbol{\alpha}})} \exp\left(J_\alpha \sum_{\langle i,j \rangle} \sigma^\alpha_i \sigma^\alpha_j + h_\alpha \sum_i \sigma^\alpha_i \right) \tag{3.8}$$

Fixing the nearest neighbour two-point correlation and the mean species occurrence we obtain the standard Ising model[3].
Due to the fact that we do not have interspecific interactions (individuals interact only with conspecifics), our probability distribution $\mathbb{P}(\{\vec{\sigma}^\alpha\})$ for the system of S species factorizes out such as:

$$\mathbb{P}(\{\vec{\sigma}^\alpha\}) = \prod_{\alpha=1}^{S} p(\vec{\sigma}^\alpha|\boldsymbol{g_\alpha}) \tag{3.9}$$

Using a physical expression we can define a sort of "ecosystem" reduced Hamiltonian as:

$$\beta H(\{\vec{\sigma}^\alpha\}) = -\sum_\alpha \left(J_\alpha \sum_{\langle i,j\rangle} \sigma_i^\alpha \sigma_j^\alpha + h_\alpha \sum_i \sigma_i^\alpha \right) \tag{3.10}$$

For ecosystems, h_α may be interpreted as a uniform parameter capturing environmental effects that may favour the presence of the species α ($h_\alpha > 0$) or its absence ($h_\alpha < 0$).
Instead, J_α is an intraspecific coupling and $J_\alpha > 0$ favours clustering of the species whereas $J_\alpha < 0$ suppresses it (see fig. 3.3 for an example). This two behaviours have a biological explanation and in particular the case of $J_\alpha < 0$ can be related to the Janzen-Connell effect. In fact according to this effect in tropical rainforests there exist pathogens that specifically target a tree species making the areas directly surrounding the parent tree (the seed producing tree) inhospitable for the survival of seedlings. Since the pathogens are found commonly around the parent tree those seedlings that are farthest from their parents have a competitive advantage and the resulting effect can be that the given species is over-dispersed on the region.

3.2 The inferred couplings

With the maximum entropy probability (3.8) we evaluate the mean values $\langle M^\alpha \rangle_{\boldsymbol{g}}$ and $\langle E^\alpha \rangle_{\boldsymbol{g}}$ (evaluation of mean values is explained in Appendix A).
Imposing:

$$\langle M^\alpha \rangle_{\boldsymbol{g}} = \hat{M}^\alpha \quad \text{and} \quad \langle E^\alpha \rangle_{\boldsymbol{g}} = \hat{E}^\alpha \tag{3.11}$$

[3]actually the Ising model with couplings $\bar{J}$ and $\bar{h}$ is usually expressed with binary variables $s_i = \{-1, 1\}$. The change of variable $s_i = 2\sigma_i - 1$ links the two formalisms and one obtain the relation between the couplings (h, J) and $(\bar{h}, \bar{J})$: $\bar{J} = J/4$ and $\bar{h} = h/2 + J$.

we are able to find the couglings $\boldsymbol{g} = (h, J)$ (see Appendix B Section B.2). The results are presented and commented in Figures 3.4 - 3.2.
The analysis comprehends three sets of inferred couplings:

- **Random Placement Model**
  It is equivalent to a non interacting case ($J^\alpha = 0$). The value of h^α_{rpm} is calculated knowing only M_α as explained in Section 3.3.1
- **Interacting model**
  Both h^α and J^α are inferred knowing $\hat{M}_\alpha$ and $\hat{E}_\alpha$ calculated from the empirical configurations $\hat{\sigma}^\alpha_i$
- **Scrambled data**
  In this case h^α_{rnd} and J^α_{rnd} are inferred knowing $\hat{M}_\alpha$ (the same as before) and $\hat{E}^\alpha_{rnd}$ the nearest neighbours correlation function of random a configuration obtained by reshuffling the σ^α_i among the various sites

Using this analysis we compare the results for the three different situations listed above with the aim to understand the "degree of randomness" for the BCI ecosystem. Then with the inferred coupling of the interacting model we will investigate the correlation function the SAR and the EAR considering all the species and only the rarest ones[4].

3.2.1 Highlights

The aim of the previous analysis was to reveal the "degree of randomness" of the BCI system analysed at scale imposed by setting $N = 256$. Despite the few cases with $J_\alpha < 0$ it is evident from Figure 3.2 that the J's of the interacting model are positive and significantly different from the J_{rnd}'s of the random spatial distribution. This suggests that the species in the ecosystem are clustered. Figure 3.3 shows an example of two species with quite the same $\hat{M}_\alpha$ but opposite J_α values.

The additional analysis for the h_{rnd}'s and J_{rnd}'s was conducted to show that the case of the random placement model ($J_\alpha = 0 \quad \forall\alpha$) is equivalent to spatially uncorrelated species. In fact, the h_{rnd}'s are equals to the h_{rpm}'s of the random placement model for all the values of $\hat{M}_\alpha$ (Figure 3.4). On the other

[4]Here we consider rare a species that is present in a fraction from to of the sites

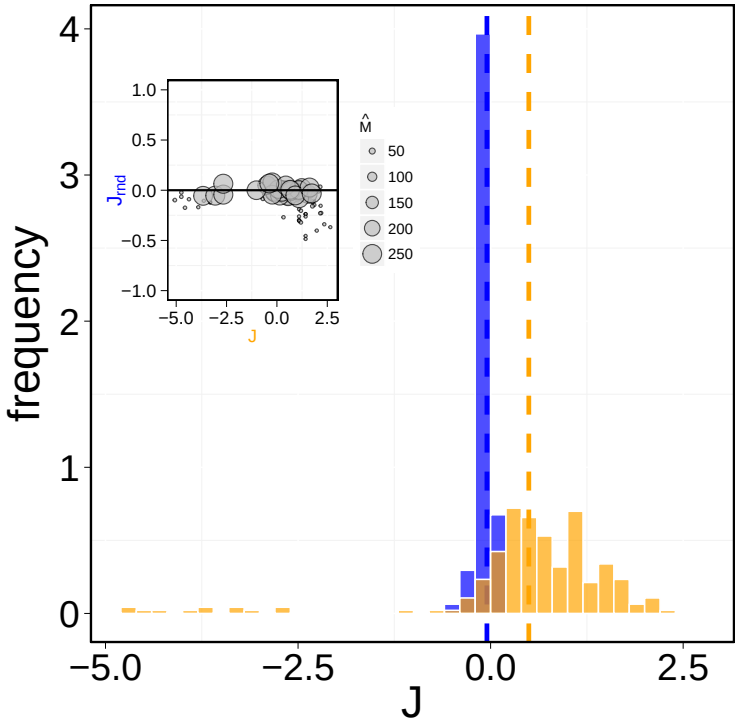

FIGURE 3.2: Histograms of the couplings J's obtained for the interacting model (orange) and for scrambled data (blue). The inset shows the scatter plot between the two set of J's. Each circle has a dimension proportional to $\hat{M}^{\alpha}$. As expected the J_{rnd}'s inferred for scrambled data have a pronounced pick around zero. The positive mean of the J's relative to BCI configuration reveals that species tend to form clustered structures. The two dashed line represent the mean of the two histograms.

hand the J_{rnd}'s (Figure 3.2) have almost zero mean and a small variance (i.e. $\sigma^2 \approx 0.007$). Instead, the comparison between the h's of the interacting model and the h_{rpm} of the random placement model shows significant differences between the two sets (Figure 3.5). For completeness Figure 3.6 shows the comparison between the h's and the h_{rpm}'s.

Another evidence of the non randomness of the ecosystem organisation is shown in Figure 3.7 where we compare the nearest neighbours correlation for a random configuration to the one of the BCI system at fixed $\hat{M}_{\alpha}$. The fact that $\hat{E}$ is greater than $\hat{E}_{rnd}$ signals the tendency of species to form clusters. This characteristic is also shown in Figure 3.8 where the total "energy" of the interacting case is greater then the one of the random placement. Anyway, the top right part of this plot shows equality between the two models. These are the cases of both negative J_{α} and h_{α} (J_{rnd} and h_{rnd}) and small values of $\hat{E}$ and $\hat{M}$ ($\hat{E}_{rnd}$ and $\hat{M}$).

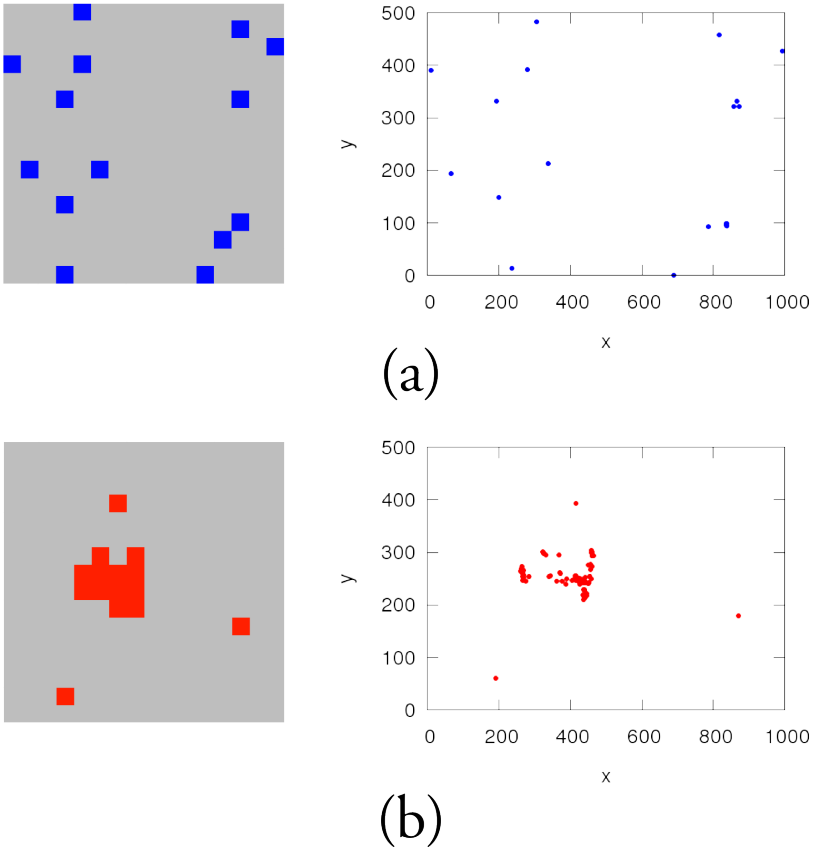

FIGURE 3.3: An example of two species with opposite J values. Case a) the lattice configuration (left) and the real spatial distribution (right) of a species with $J_\alpha < 0$ b) the same as in case a) for a species with $J_\alpha > 0$. In the case a) $\hat{M}_\alpha = 14$ and in the case b) $\hat{M}_\alpha = 15$.

3.3 Patterns from our model: results and discussion

In the previous section we have presented the maximum entropy model used in this work. In this section we use it to explain how we can investigate biodiversity indicators introduced in Chapter 1.

One of the main goal of this work is to investigated the structure of the ecosystem in terms of β-diversity (i.e. correlation function, SAR and EAR), using only the knowledge of $\hat{M}_\alpha$ and $\hat{E}_\alpha$ for each of the species.

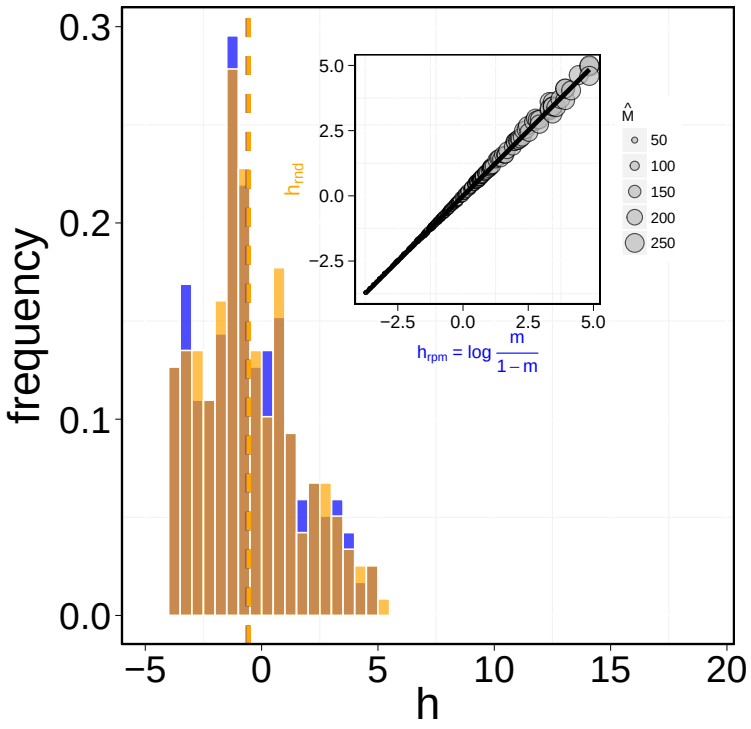

FIGURE 3.4: Histograms of the h's obtained (see section 3.3.1) with the random placement model (blue) and with scrambled data (orange). The inset shows the scatter plot between the two set of h's with the straight line representing equality. Each circle as a dimension representing the value of $\hat{M}^{\alpha}$. The inset shows that as expected h's values for the random placement model and for scrambled data are highly correlated (the darker area of the histogram is the overlapping region). The dashed lines represent the mean of the histograms and in this case are practically equal

3.3.1 A preliminary result: the relation between M_{α} and h_{α} in the case $J_{\alpha} = 0$

To build our model we made the approximation of non interacting species. In this brief section we explain the simplest model in which intraspecific interaction is suppressed as well. We call it the random placement model and we use it as a basis of comparison for our results. The random placement model was introduced for the first time by Coleman [9] to study the SAR under different hypothesis on the species abundances distribution. Our framework does not require species abundances and the random placement model we use is based on the species "abundances" measured by the number of occupied sites.

When intraspecific interactions are absent, $J = 0$ (the case of infinite temperature for the Ising model), presence or absence in a particular site is indipendent

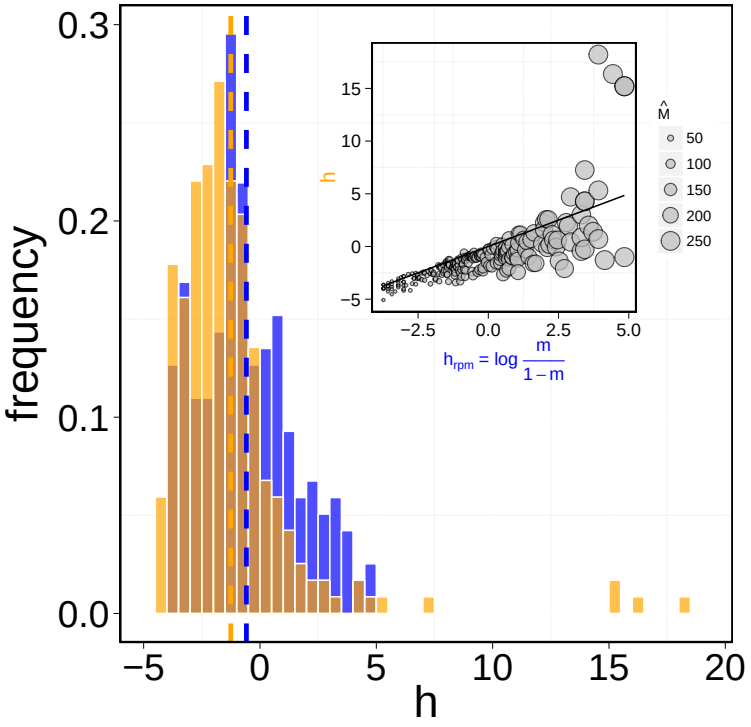

FIGURE 3.5: Histograms of the h's inferred in the case of interacting model (orange) and random placement model (blue). The inset shows the scatter plot between the two sets of h's with the solid line representing equality. The size of the circles in the inset is proportional to $\hat{M}^\alpha$. The dashed line in the histograms represent the mean values of the h's. The outliers for the interacting model in the range $15 < h < 20$ correspond to cases of high $\hat{M}^\alpha$ and $\hat{E}^\alpha$.

from the other sites.

In this case the probability $p(\vec{\sigma}|\boldsymbol{g})$ introduced in section 3.1 takes the form:

$$p(\vec{\sigma}^\alpha, h_\alpha) = \frac{1}{Z(h_\alpha)} \exp\left(h_\alpha \sum_i \sigma_i^\alpha\right) \tag{3.12}$$

and the partition function Z:

$$Z(h_\alpha) = \sum_{\{\vec{\sigma}\}} \exp\left(h_\alpha \sum_i \sigma_i^\alpha\right) = \left(1 + e^{h_\alpha}\right)^N \tag{3.13}$$

thus,

$$\langle M(\vec{\sigma}^\alpha)\rangle_{h_\alpha} = \frac{\partial \ln Z(h_\alpha)}{\partial h_\alpha} \tag{3.14}$$

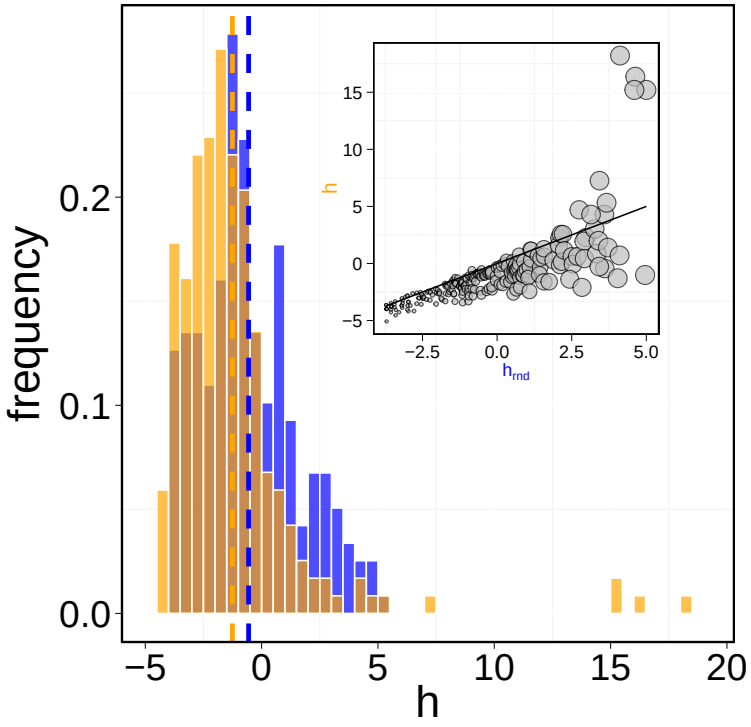

FIGURE 3.6: The information summarised by this figure is practically the same of the figure 3.5 due to the fact that h_{rnd}'s and h_{rpm}'s are practically equal (see fig. 3.4).

Imposing the condition that averages must reproduce observed ones means:

$$\langle M(\vec{\sigma}^{\alpha})\rangle_{h_\alpha} = N\frac{1}{1+e^{-h_\alpha}} = \hat{M}_\alpha \tag{3.15}$$

Defining $\hat{m}_\alpha = \frac{\hat{M}_\alpha}{N}$ the previous equation fixes h_α to:

$$h_\alpha = \ln\left(\frac{\hat{m}_\alpha}{1-\hat{m}_\alpha}\right). \tag{3.16}$$

Thus imposing only the constrain to reproduce the mean occurrence $\hat{M}^\alpha$ is equivalent to fix the coupling h_α. This results we'll be useful in the following when we will analyse the β-diversity In fact for each one of the patterns we will analyse the one predicted by the random placement model and by the interacting one.

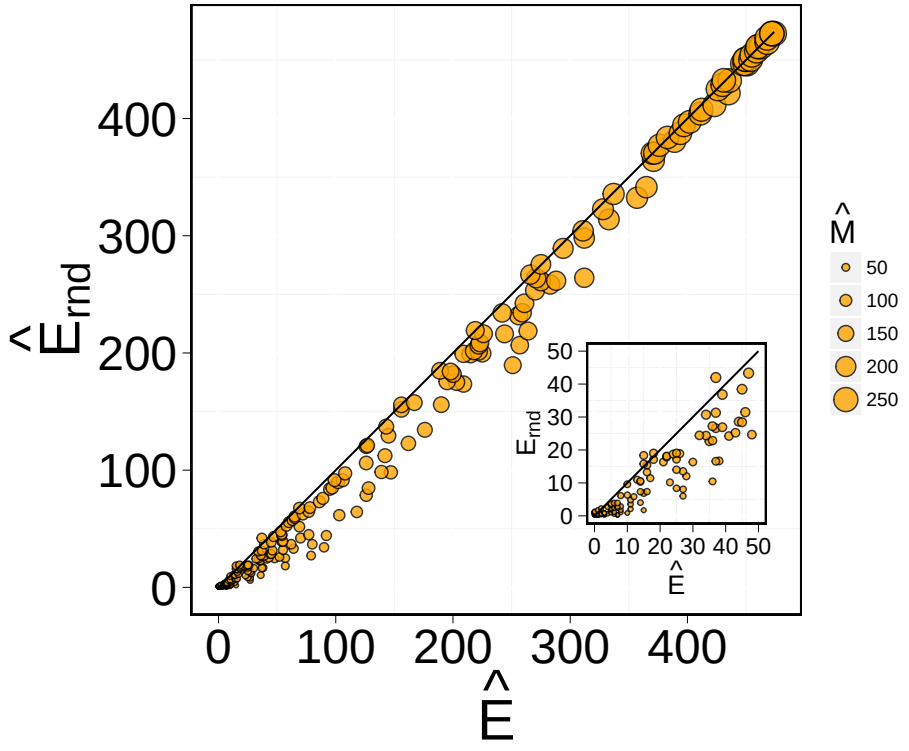

FIGURE 3.7: The nearest neighbours correlation of species configurations $\hat{E}$ is compared to the ones obtained for configurations with the same $\hat{M}^\alpha$ and a random spatial distribution of species' presence sites. The size of the circles represents the value $\hat{M}^\alpha$ and the solid line equality. For big values of $\hat{M}^\alpha$ (upper right) $\hat{E}$ is practically equal to $\hat{E}_{rnd}$. Instead it is greater for intermediate (central part) and lower (bottom left) values of $\hat{M}^\alpha$. The inset shows a zoom for the configurations of rare species (lower $\hat{M}^\alpha$)

3.3.2 Correlation function C(r)

As we said before, one of the first analysis involved the correlation function defined in equation (3.2).
Under the hypothesis of $J_\alpha = 0$ it's trivial to see that $\langle\sigma_i\sigma_j\rangle_{h_\alpha} = \langle\sigma_i\rangle_{h_\alpha}\langle\sigma_j\rangle_{h_\alpha}$ does not depend on the distance r_{ij}:

$$C^\alpha_{rpm} = \frac{\sum_{\{\vec{\sigma}\}} e^{h_\alpha \sum_k \sigma_k}\sigma_i\sigma_j}{Z(h_\alpha)} = \left(\frac{1}{1+e^{-h_\alpha}}\right)^2 = \left(\frac{\hat{M}_\alpha}{N}\right)^2 \tag{3.17}$$

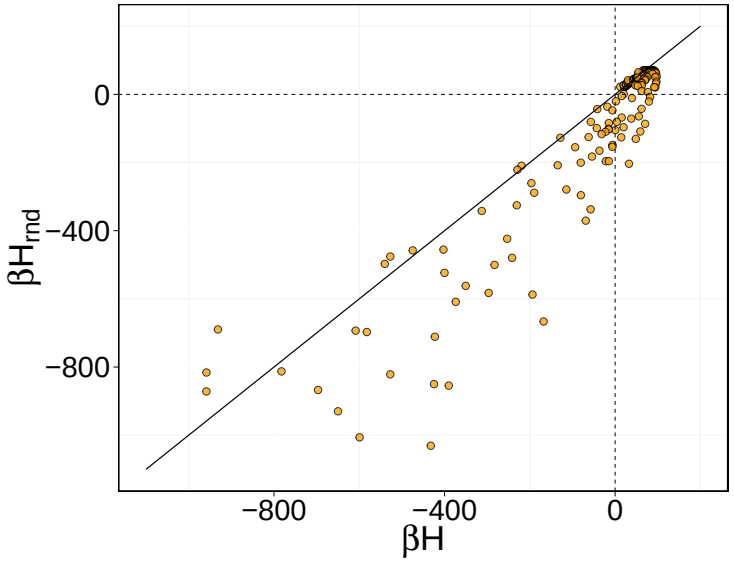

FIGURE 3.8: **Total "energy".** This figure shows the relation between βH^{α}_{rnd} and βH^{α}. The upper right corner corresponds to the region where both βH^{α}_{rnd} and βH^{α} are positive related to the cases of negative couplings and small $\hat{E}$ and $\hat{M}$

Since h_α is directly determined by $\hat{M}_\alpha$ the correlation function become:

$$C_{rpm}(r) = \frac{1}{S}\sum_{\alpha} \hat{m}^2_\alpha \tag{3.18}$$

In general we expect that the correlation function of a system decays for long distances r. On the contrary, looking at the data (fig. 3.9) this ecosystem presented a peculiar behaviour. In fact the two point correlation function computed from the data decays until a certain distance and then begin to increase. This behaviour can be explained looking at how species are distributed over the region. In fact, they accumulate on the border and this effect coupled with the fact that large distances have poor statistics imply the feature observed. The fact that species are densely present on the edges is due to topographical and soil nutrient variability effect. In effect, it was found that between 30% and 40% of the species are non randomly distributed with respect to soil nutrient variation [6]. For this reason we studied the two point correlation function truncated at shorter distances and we analysed the system up to a distance $r < r_{max} \approx 12$ lattice units.

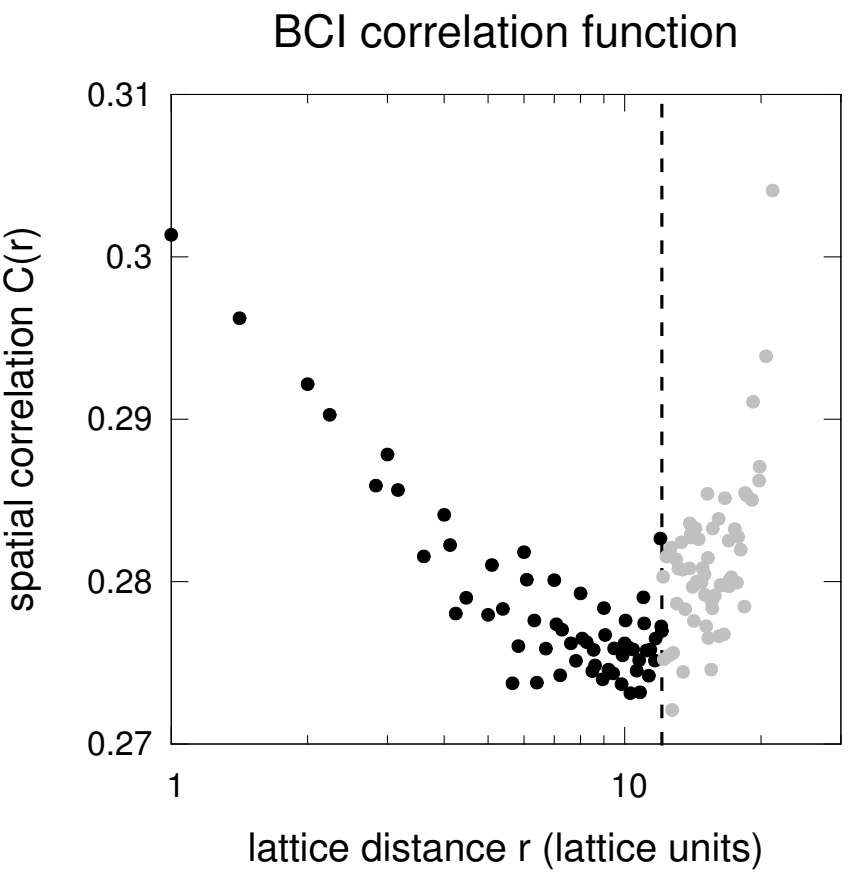

FIGURE 3.9: The correlation function extracted from species configurations. Poor statistics and some peculiarities in species spatial distribution are responsible for greater values of the correlation function at larger distances. The dashed vertical line shows the maximum distance considered in our study, $r_{max} \approx 12$ lattice units.

Correlation function results

The behaviour of the correlation function in the two analysed cases is pretty well described by the interacting model only at short distances (Figures 3.10 and 3.11). It is evident that a random placement model is insufficient to describe the correlation function and that the system is far from being randomly organised (in agreement with the analysis of the couplings J_α's summarised by the fig. 3.2). Introducing an interaction between nearest neighbours sites improve the results in particular in the case of rare species. Imposing $\hat{E}_\alpha = \langle E_\alpha \rangle_{\boldsymbol{g}}$ ensures that the correlation function at $r = 1$ is the same for the interacting model and the data. We think that the aforementioned peculiarities in the topography and soil nutrients distribution over the field are responsible for the differences between the correlation function of the interacting model and the one of the data. In

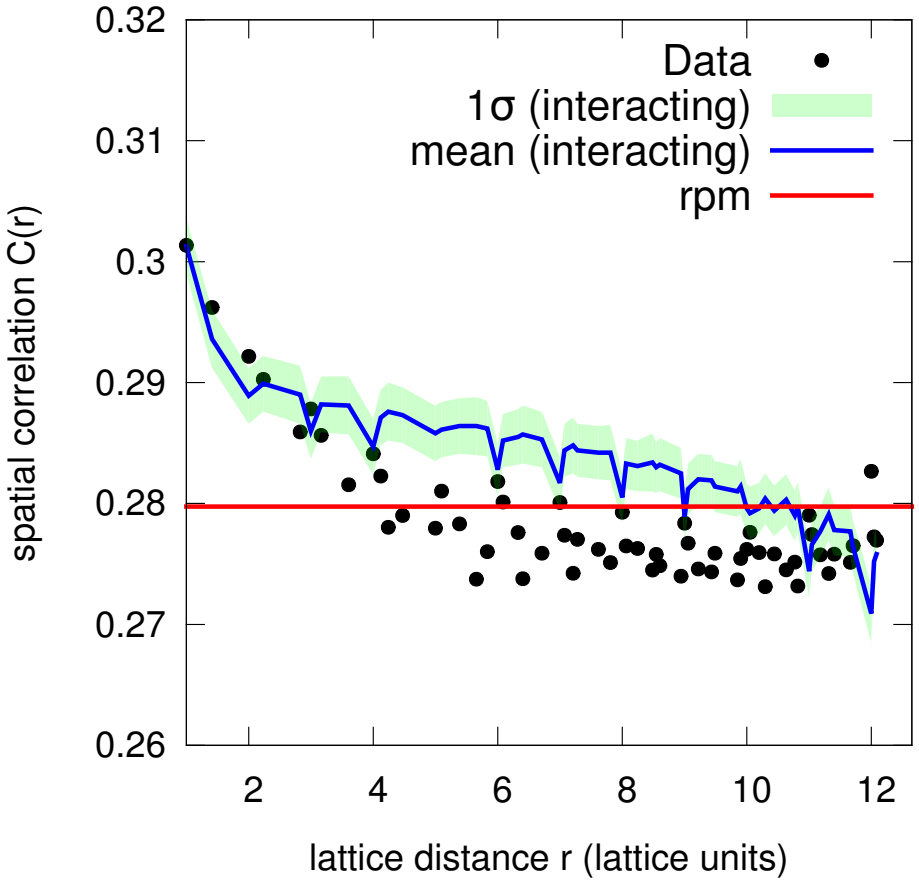

FIGURE 3.10: **The correlation function for all the species**. The analysis was developed with $N = 256$ sites. The red line represents the correlation function for the random placement model (eq. 3.18). The blue line is the correlation for the interacting model expressed by (3.2) (numerical).

fact, variability on the conditions of the environment are more likely to occur between distant sites and that can influence the behaviour of the correlation function.

3.3.3 The species area relationship (SAR)

We introduced the SAR in chapter 1 as a biodiversity indicator capable to synthesize spatial properties of species assembled community and we defined it as the average number of species S present in an area A.
This means that in the area A we count all the species that have *at least* one

individual. We can express it as:

$$SAR(A) = \sum_{\alpha=1}^{S} \Big\langle \chi_A(\{\sigma\}) \Big\rangle_{\boldsymbol{g}_\alpha} \tag{3.19}$$

where

$$\chi_A(\{\sigma\}) = \begin{cases} 1, & \text{if } \exists \text{ i} \in \text{A} : \sigma_i = 1 \\ 0, & \text{otherwise} \end{cases}$$

This mathematical expression simply states that we count a species every time we find a cell were it is present.

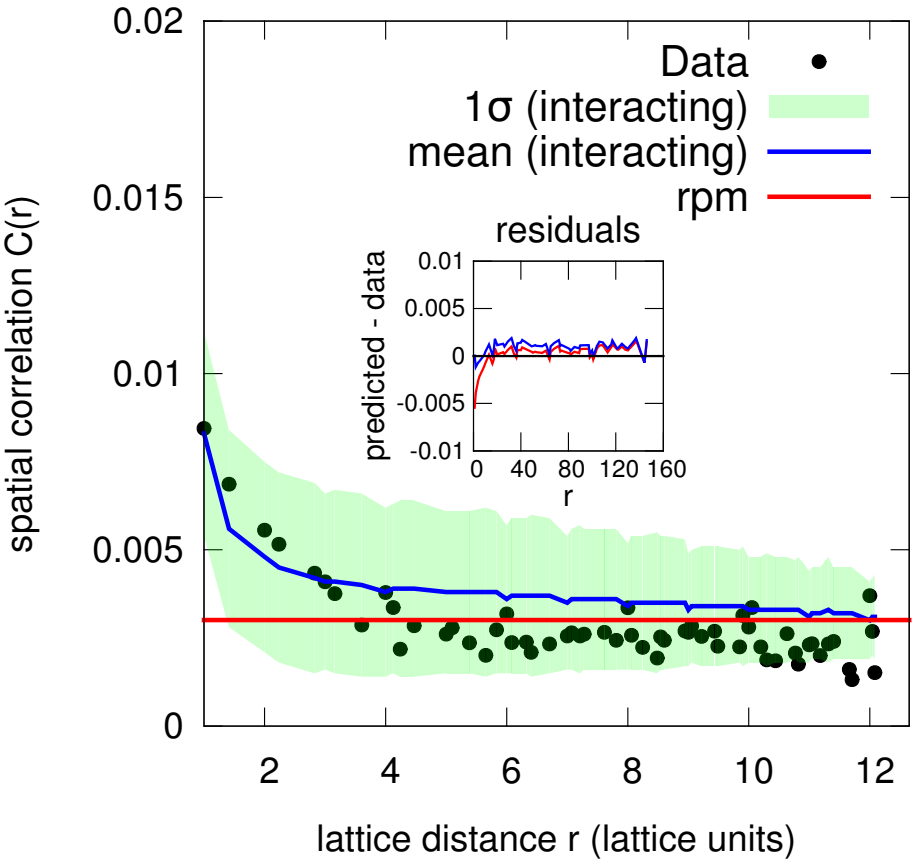

FIGURE 3.11: **The correlation function for rare species** The analysis was developed with $N = 256$ sites. The black points representing data are completely contained in the 1σ confidence interval of the interacting model that is able to reproduce the decay of correlation function at short distances.

The SAR in the random placement model

Our theoretical framework does not require an hypothesis on the form of species abundances to derive an expression for the SAR. Using only presence/absence data and more precisely the number of occupied sites we can deduce an analytical expression for the SAR in a random placement model using equation (3.19) in the non interacting case ($J_\alpha = 0$).

We define the probability of having an individual of the species α in the site i as

$$p_{i,\alpha} \equiv \langle \delta_{\sigma_i^\alpha,1} \rangle_{h_\alpha} = \frac{1}{1+e^{-h_\alpha}} \tag{3.20}$$

Given the expression for h_α (eq. 3.16), $p_{i,\alpha}$ becomes

$$p_{i,\alpha} = \hat{m}_\alpha = \frac{\hat{M}_\alpha}{N} \tag{3.21}$$

independent of the particular site.
If we denote $|a|$ the size of the sampled area[5] the contribution to the SAR at area a by the species α is:

$$SAR_\alpha(a) = \sum_{k=1}^{|a|} \binom{|a|}{k} p_{i,\alpha}^k (1-p_{i,\alpha})^{|a|-k} = 1-(1-p_{i,\alpha})^{|a|} = 1-(1-\hat{m}_\alpha)^{|a|} \tag{3.22}$$

Concluding, the SAR for a random placement model can be expressed as

$$SAR_{rpm}(a) = \sum_{\alpha=1}^{S} SAR_\alpha(a) = S - \sum_{\alpha=1}^{S} (1-\hat{m}_\alpha)^{|a|} \tag{3.23}$$

The case for $J_\alpha \neq 0$ cannot be analytically solved so we computed the SAR numerically.

Species Area Relationship results

Figures 3.12 and 3.13 show respectively the SAR for all the species in the ecosystem and for the rare ones. In both the models the SAR obviously converge to the total species richness due to the fact that the probability to find a species on the largest surveyed area is one.
The fact that the correlation function overestimate the one of the data for

[5] In our case the area is measured in number of plot so $|a|$ range from 1 to N

distances greater then 20 lattice units it is reflected on the prediction of the SAR. In fact if we think the correlation function as the probability that two sites have in common the same species a lower value of the correlation function means a lower probability to find an individual of the same species. To quantify the reliability of the predicted SAR with the two models we evaluated the difference between the predicted species richness with the one extracted from the data (inset Figures 3.12 and 3.13). Although the differences between prediction and data in both models present a peak at an intermediate area the differences for the interacting model are always smaller (approximately by a factor of 2).
Furthermore, the overall positive mean of the J_α's (see Figure 3.2) produce a clustering of the individuals that has the effect of decreasing the mean number of species respect to the one of the random placement model. Although our approach does not consider the species abundances the results for the random placement model agree with the conclusions of previous works [39] where the authors report inadequacy of the random placement model for three rainforest ecosystems.

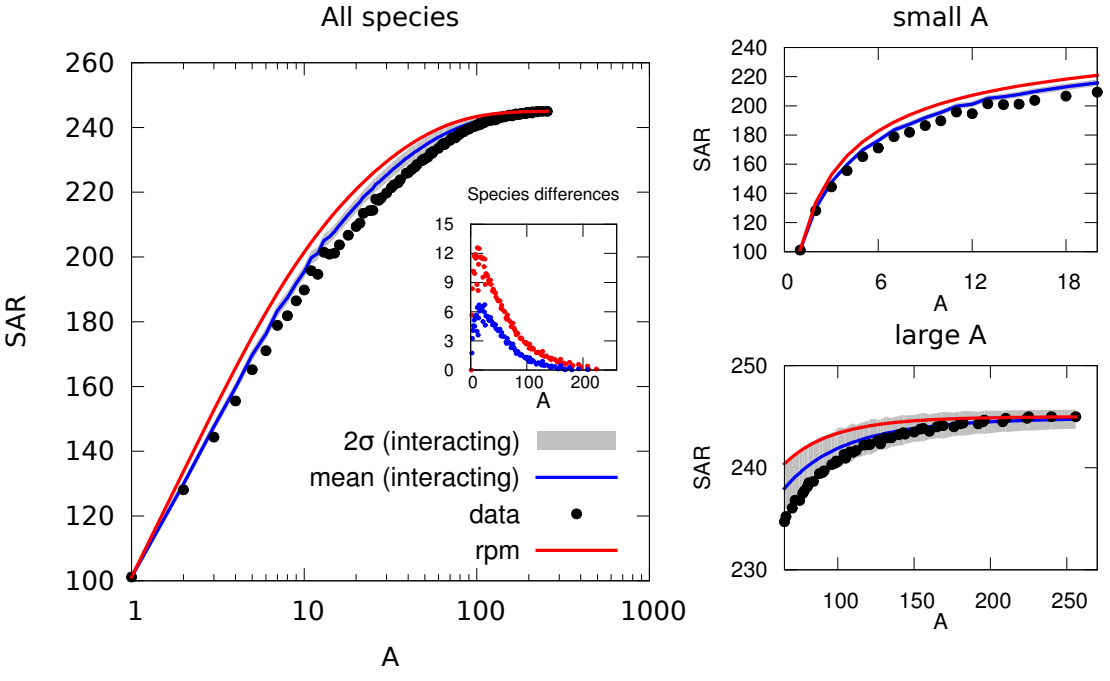

FIGURE 3.12: **SAR for all the species** The SAR over all the sampled areas (left) with the inset that shows the differences between data and the two models (red for random placement and blue for interacting model); the right column shows a magnification for small (top) and large regions (bottom). The grey area represents the 2σ confidence interval for the interacting model

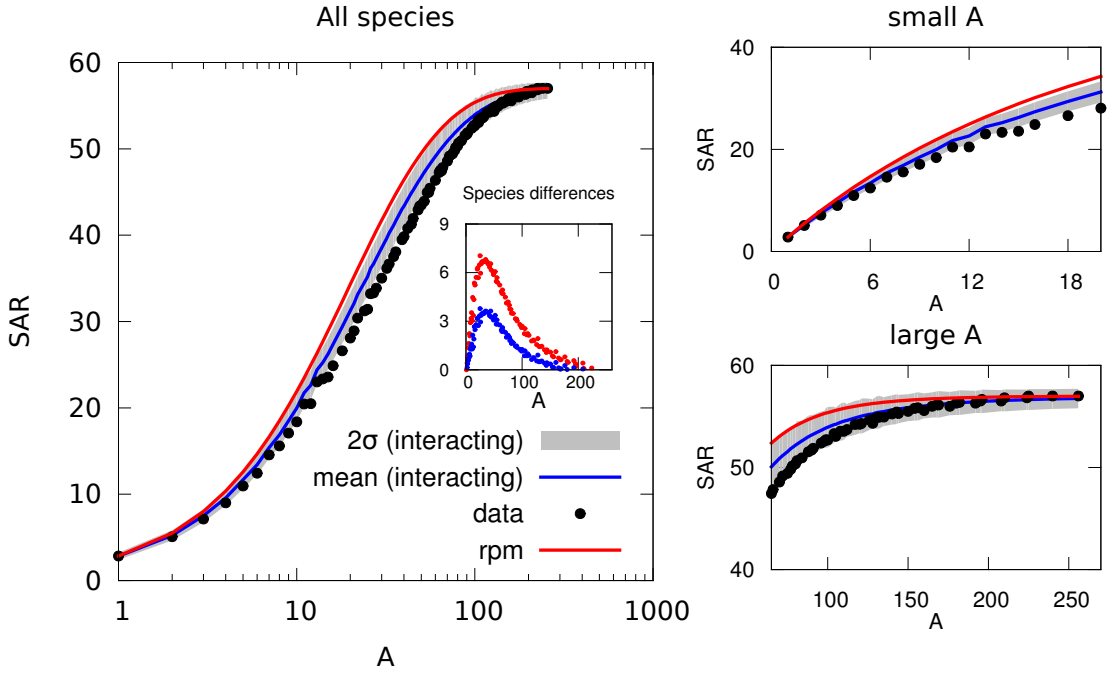

FIGURE 3.13: **SAR for the rare species** The SAR over all the sampled areas (left) with the inset that shows the differences between data and the two models (red for random placement and blue for interacting model); the right column shows a magnification for small (top) and large regions (bottom). The grey area represents the 2σ confidence interval for the interacting model

3.3.4 The endemic area relationship

In the same way we computed the SAR here we give an expression for the endemic area relationship (EAR). Due to the fact that a species is called endemic in the area a if it is completely contained in it we must find the area a that completely contains the species [20].
It is equivalent to say that the species is absent in a^c the complement of a. Taking advantage of this fact and of the independence of the species we write the EAR for area a as:

$$EAR(a) = \sum_{\alpha=1}^{S} EAR_\alpha(a) \tag{3.24}$$

where $EAR_\alpha(a)$ reads:

$$EAR_\alpha(a) = \left\langle \delta_{\sum_{i\in a^c} \sigma_{i,0}} \right\rangle_{\boldsymbol{g}_\alpha} \tag{3.25}$$

The Kronecker delta imposes the condition of endemicity of α over the area a. Using the definition of $\langle\cdot\rangle_{\boldsymbol{g}_\alpha}$ we explicit the expression 3.25 as:

$$EAR_\alpha(a) = \frac{Z_a(\boldsymbol{g}_\alpha)}{Z_A(\boldsymbol{g}_\alpha)} \tag{3.26}$$

Here $Z_a(\boldsymbol{g}_\alpha)$ and $Z_A(\boldsymbol{g}_\alpha)$ are the partition function evaluated respectively in a subconfiguration of area a and on the whole lattice of area A.

EAR in the random placement model

As before, the random placement model is a specific case of 3.26, the one with $J_\alpha = 0$. Thus the contribution of species α to the EAR is:

$$EAR_\alpha^{rpm}(a) = \frac{Z_a(h_\alpha)}{Z_A(h_\alpha)} = \left(\frac{1}{1+e^{h_\alpha}}\right)^{A-|a|} \tag{3.27}$$

The fact that h_α is fixed by the mean occupation $\hat{m}_\alpha$ allows us to rewrite 3.27 as:

$$EAR_\alpha^{rpm}(a) = (1-\hat{m}_\alpha)^{A-|a|} \tag{3.28}$$

To conclude we want to show a property that relates the SAR and EAR *only* in the case of a random placement model. In fact they can be obtained one from the other:

$$EAR_{rpm}(a) = S - SAR_{rpm}(A-a) \tag{3.29}$$

Thus only in a very special and biologically unrealistic case, when all species are randomly and independently distributed in space, it is possible to derive the EAR from the SAR. Since the EAR can be used to estimate the species extinction rate by habitat loss, the one obtained from the SAR has been used in the past but as shown in [20] this produces overestimate extinction rates.

Endemic Area Relationship results

Figures 3.15 and 3.14 show respectively the EAR for all the species in the ecosystem and for the rare ones. In both the cases the EAR is practically zero for the initial range of small areas. In fact a sampling frame with an area of a size sufficient to contact the species for the first time is always less than the sample area needed to encompass the entire range of the species. The SAR

is constructed from sample areas of first contact, and the EAR is constructed from areas of last contact and for this reason the EAR grows very slowly at the beginning. Then, as shown in the two figures, it is "forced" to reach the total number of species in the largest plot (loosely speaking all the species are endemic in the largest area). For the EAR the interacting model seems to

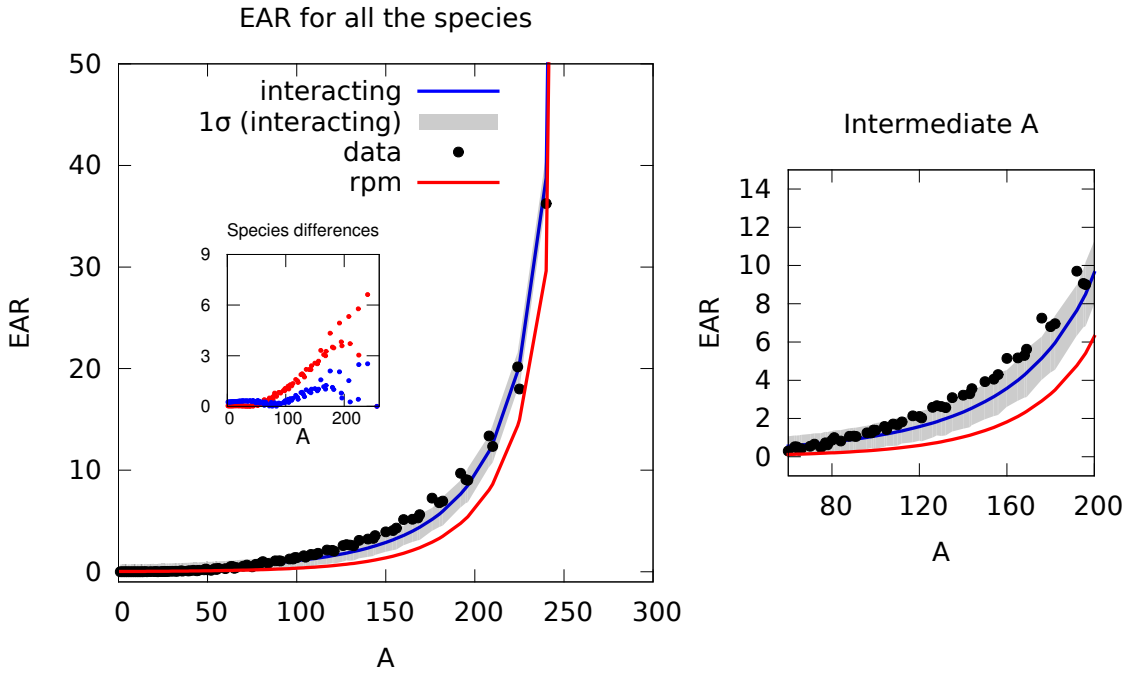

FIGURE 3.14: **EAR for all the species.** The EAR over all the sampled areas (left) with a zoom of the intermediate area region (right). The two figures show the curve of the interacting model (blue) with the 1σ confidence interval (grey area) and the one for the random placement (red). The data are represented by black dots. The inset shows the differences between the two models. As in the case of the SAR the differences between the two models are sensible (approximately a factor of 2).

better reproduce the data within the 1σ confidence interval. However, the data are completely contained in the 2σ confidence interval. The inset of the main plot in Figures 3.15 and 3.14 shows that for $A \approx 100$ (plot units) the random placement model for presence/absence data drastically deviates from the data. In this case the introduction of the interaction improves the prediction because with a $J_\alpha > 0$ a species tends to cluster and there is a higher probability to sample an area that completely contains the given species. A clear example is shown in Figure 3.3. Somehow, the same considerations made for the SAR apply here. In fact in the case of the SAR the random placement model overestimates the data while here it underestimates the EAR extracted from the data. Even if in (3.29) we state that the EAR is the "mirror image" of the SAR we stress

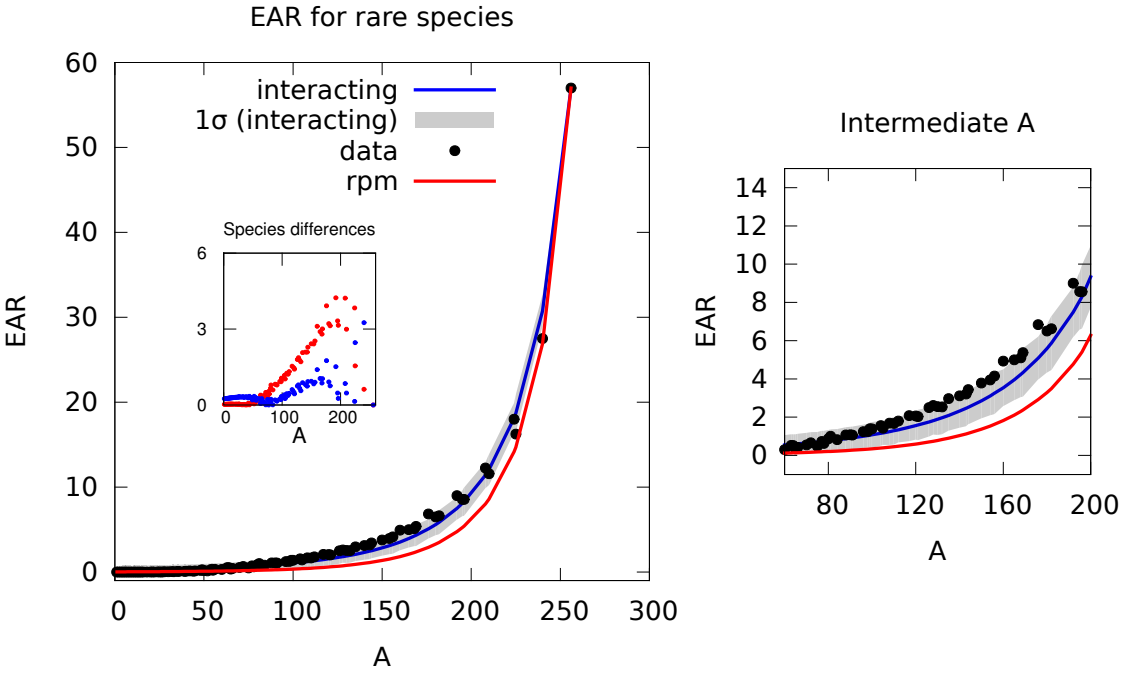

FIGURE 3.15: **EAR for the rare species.** The EAR over all the sampled areas in the case of rare species. The interacting model (blue) and the random placement model (red) reach the same value for the largest A because the every species is completely contained in the surveyed area. As in the case of the analysis for all the species the EAR for the random placement model underestimate the data by approximately a factor of two.

that this is not a rule and so an overestimating SAR does not automatically translate to an EAR that underestimates the data [20].

Summary of the chapter

Using the techniques introduced in Chapter 2 we suggested the use of a spatially explicit maximum entropy model to understand the spatial structure of an ecosystem. Starting from species abundances data we shifted to a "low level" information framework were only presence or absence of the species are known. Although the knowledge of the average presence of a given species in an area and the probability that two neighbouring sites have in common the given species constitutes the minimal information about spatial organisation of species they are sufficient to deduce the global spatial organisation of species on the entire

area. We tested the model comparing the results for predicted biodiversity indicators to the data of a rainforest in Barro Colorado Island, Panama.

Chapter 4

Conclusion and perspectives

This project was thought and developed with the spirit of a feasibility study. We studied a maximum entropy model with the purpose of investigating the spatial organisation of a rainforest ecosystem. The theoretical framework we propose is based on a minimal quantity of information regarding the system: i) the average presence of a given species in an area and ii) the probability that two neighbouring sites have in common the given species. For this reason we can build the model using only presence/absence data instead of species abundances.

We think that this kind of approach is appealing from both the practical and theoretical point of view. The practical reason is that detailed data of species abundances are present only for few cases limited to restricted geographical regions and a world wide characterisation of species abundances sounds like a titanic task for obvious reasons. On the other hand we think that the theoretical reason is closely linked to the concept of *universality*, an idea that has a deep meaning in Physics and in particular in statistical mechanics. Loosely speaking universality emerge macroscopically when the processes are independent of the microscopic details and analysing a system using presence/absence data can be regarded as a coarse grained analysis. Thus, developing a framework based on a "low level" kind of information can help to identify essential and universal features characterising living systems. Indeed, the interest regarding the β-diversity of an ecosystem and in particular the Species Area Relationship (introduced in Chapter **??**) goes beyond the analysis tailored to a particular ecological system. The universal shape of this emergent pattern naturally attracts contributions from multiple fields of investigation.

The results presented in Chapter 3 are a first step in this direction. In fact, the spatial correlation function predicted by the random placement model and the interacting model are very different, in particular the interacting model better reproduces the data at small distances (Figures 3.10 and 3.11), but this aspect does not translate to completely different shapes for the SAR and EAR. In other words, the correlation function can be sensible to environmental factors [39] (e.g. topography, soil differentiation, water resources) but this has minor effects on the SAR and EAR that can be fairly well reproduced by the interacting model improving the prediction of the random placement model.

To highlight the novelty of our approach among the other modelling frameworks, we briefly report the essential ingredients of two works that use the maximum entropy approach to investigate the ecosystems structure 1.3.2. Harte et al. [18] use the maximum entropy model based on the knowledge of some "high level" prior information. Apart from the ecosystem total area the authors consider other three ecosystem related quantities: the total number of species in that area, the total number of individuals across those species, and the metabolic energy rate summed over all those individuals. Using these information they can investigate the SAR and other ecosystem meaningful patterns. Nevertheless, in their predictions space is never explicitly considered.
The other work we cite is the one of Azaele et al. [4] on the characterisation of ecosystem structure from the point of view of interspecific interaction. In our case we build the maximum entropy model considering only intraspecific interaction and a "microstate" of the system is described by σ_i^α for the species α in the plot i. Furthermore, we infer the couplings from a single realization.
Instead in [4] they consider a "microstate" $\vec{\sigma} = (\sigma_1, \dots, \sigma_S)$ containing only the species presence or absence in the whole area and they infer the couplings from the first and the second empirical moments $\langle \sigma_\alpha \rangle_{emp}$ and $\langle \sigma_\alpha \sigma_\beta \rangle_{emp}$ calculated from an high number of "microstates" collected over the area of interest. Thus in this case the authors neglect spatial correlations.
In this way they end up with the maximum entropy probability

$$p(\vec{\sigma}) = \frac{1}{Z} \exp\Big(\frac{1}{2} \sum_{\alpha \neq \beta} J_{\alpha\beta} \sigma_\alpha \sigma_\beta + \sum_\alpha h_\alpha \sigma_\alpha\Big) \tag{4.1}$$

with $J_{\alpha\beta}$ and h_α determined by the empirical averages. The goal of their analysis is to reveal the network of interactions between the species and we briefly described it to understand how our model can be considered a complementary

investigation.

Concluding, the Ising model in our work emerged in a natural way as a second order spatially explicit maximum entropy model for binary data. For this reason the role of Statistical Mechanics is twofold. On the one hand it is present as a natural framework to study systems with an high number of entities and on the other one it has all the characteristics to be considered a powerful inference method.

4.1 Perspectives

We conclude this thesis with a few clues for further work. Some of them can be considered as straightforward extensions, the others request much more effort and an extension of the theory presented in this thesis.

- **Application to other databases**
  A natural extension is the application of this method to other species abundance databases to test the validity and robustness of our results obtained for the BCI forest.

- **Upscaling and downscaling of biodiversity**
  If we know the number of species at a specific regional scale what can we say for smaller (downscaling) and larger (upscaling) areas of the system ? A first step towards the answer could be the analysis of the BCI restricted to a sub-areas (not necesseraly connected) to understand what kind of information we can extract starting from this partial knowledge.

- **Criticality** Empirical evidence has proliferated that living systems might operate at the vicinity of critical points with examples ranging from spontaneous brain activity to flock dynamics (see [22, 36] and references therein). In a recent work [31] the authors focused on the statistical properties of inferred models and argued that inference procedures are likely to yield models which are close to a phase transition. In other words, following their results we should end up with the couplings $\boldsymbol{g}$ in a neighborhood of the critical point[1] $\boldsymbol{g}_c$ where $\det \hat{\boldsymbol{\chi}}$, the determinant of the Fisher information matrix, diverges in the case of infinite systems (see fig. 4.1). We

[1] For the Ising model with variable $\sigma_i = \{0, 1\}$ the critical point is $\boldsymbol{g}_c = (h_c, \beta J_c) \approx (-3.52, 1.76)$

will investigate if these results can be used to develop a method usefull to characterise a living system as *critical*.

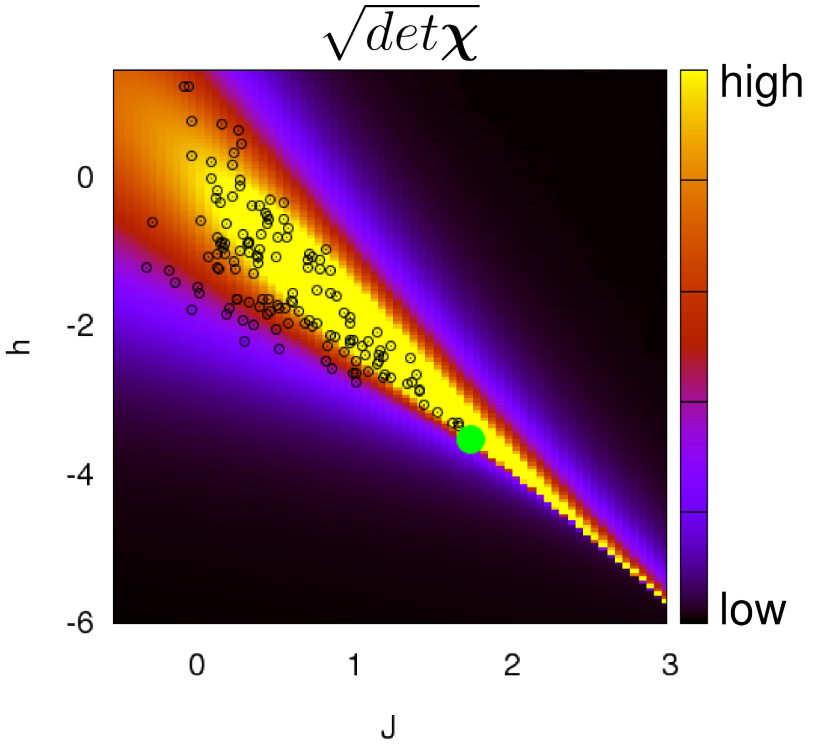

FIGURE 4.1: This figure shows the determinant of the Fisher information matrix (coloured background) on the (h, J) plane with superimposed inferred couplings (black open circles) for the BCI ecosystem (see section 3.2). The big green point represents the critical point (h_c, J_c). A big fraction of the inferred couplings occupy a region where $det\hat{\chi}$ assumes high values.

- **Detection of Janzen Connell effect**

  In Chapter 3 we mention the Janzen-Connell effect and the fact that this biological effect can be related to the cases with $J_\alpha < 0$. Comparing the list of species having $J_\alpha < 0$ with the species for which the Janzen-Connell effect has been detected on the field we found 5 common cases out of 11. Thus, another direction of this work should be to understand the validity of our model to give some hints on the presence of this effect.

Appendix A

Simulation technique: the Wang Landau algorithm

A.1 The Wang Landau algorithm

Even if the 2d Ising model was exactly solved in the case of zero external field, generalizations of it to comprehend the behaviour of the system in an external field or to analyse important quantities of finite size systems have no exact solution and one has to resort to expansion techniques or to simulation approach. Computer simulation now plays a major role in statistical physics, particularly for the study of phase transitions and critical phenomena. The standard Markov Chain Monte Carlo (MCMC) methods, like the Metropolis algorithm, allow to obtain a random sample from a certain probability distribution for which direct sampling is difficult due to the large number of possible states the system can have. Refinements to the Metropolis algorithm to solve the critical slowing down[1] have been proposed (e.g. the Swedsen and Wang algorithm and the Wolff's one).
These algorithms can be classified as cluster flip algorithms because, unlike the Metropolis one that use single spin flip techniques, they implement flip of entire clusters of equal spins. More recently new and efficient algorithms have begun to play a role in allowing simulation to achieve the resolution which is needed to accurately characterize the investigated systems. One of this is the Wang-Landau algorithm of which we explain the crucial aspects here after.

[1] At the critical point observables' fluctuations increase and the relaxation time tend to diverge. Consequences are that data obtained around the critical parameter are not so reliable.

A.1.1 How to calculate the density of state

One of the most important quantities in statistical physics is the density of states (DOS) $g(E)$, i.e. the number of all possible states (or configurations) for an energy level E of the system, but direct estimation of this quantity has not been the goal of simulations. Instead, most conventional Monte Carlo algorithms such as Metropolis importance sampling, Swendsen-Wang cluster flipping, etc. generate a canonical distribution $P(E) \sim g(E)e^{-\beta E}$ at a given temperature ($\beta = \frac{1}{k_B T}$). Such methods do not allow to make prediction for a wider range of temperatures and multiple runs are required if we want to know thermodynamic quantities over a significant range of temperatures. If we can estimate the density of states $g(E)$ with high accuracy for all energies, we can then construct canonical distributions at essentially any temperature and this allows us to slove the inverse problem introduced at the end of Chapter 2. Given that the density of states does not depend on temperature the model is essentially "solved" when one knows the $g(E)$ of the system described by the Hamiltonian H. The main object of statistical mechanics approach is the partition function Z . If $\boldsymbol{\sigma}$ is a certain configuration of the system with energy E, then the partition function can be rewritten using the $g(E)$:

$$Z(\beta) = \sum_{\{\boldsymbol{\sigma}\}} e^{-\beta H(\boldsymbol{\sigma})} = \sum_{E} g(E) e^{-\beta E} \tag{A.1}$$

Then for example the statistical average of an observables f directly related to the energy E is:

$$\langle f(E) \rangle_\beta \equiv \frac{\sum_E f(E) g(E) e^{-\beta E}}{Z(\beta)} \tag{A.2}$$

The Wang-Landau algorithm is an iterative procedure where an histogram of the energy distribution is generated during a random walk in the energy space. At the beginning the method starting from no knowledge of $g(E)$, gradually approach the true profile of the density of states by "guiding" the random walker to visits more frequently those energy regions where the density of states is smaller.

There could also be the case where one wants to determine quantities ρ (e.g. the order parameter or the correlation function) that are non directly linked to the energy. Therefore the random walking must be performed on a two dimensional

space (E, ρ) and the density of states $g(E, \rho)$ will be 2-dimensional too. In this case the statistical average of observables $f(\rho)$ must be calculated as:

$$\langle f(\rho)\rangle_\beta \equiv \frac{\sum_E \sum_\rho f(\rho) g(E,\rho) e^{-\beta H(E,\rho)}}{Z(\beta)} \tag{A.3}$$

where now $Z(\beta) = \sum_E \sum_\rho e^{-\beta H(E,\rho)}$.

In the following section we will present the structure of the Wang-Landau algorithm both in the case of one and two dimensional random walking. In the first case the system under examination is the two dimensional Ising model with nearest neighbours interaction in zero external field described by the reduced Hamiltonian:

$$\beta H = -J \sum_i \sum_{j=nn(i)} \sigma_i \sigma_j \equiv -JE \tag{A.4}$$

When we consider also the presence of an external perturbation then our Hamiltonian is:

$$\beta H = -\frac{J}{2} \sum_i \sum_{j=nn(i)} \sigma_i \sigma_j + h \sum_i \sigma_i \equiv -JE - hM \tag{A.5}$$

In both cases J represents the coupling interaction and h is an external influence. Regardless of the values the variables σ_i can assume ($\{-1, +1\}$ or $\{0, 1\}$) we call E the interaction energy and M the magnetization of the system.

A.2 The Wang-Landau scheme

With the aim to show how the Wang-Landau works we will refer to the Hamiltonian (A.4). In this method random walk in discrete energy space of a spin system is performed by flipping spins in a random manner. A random walker without any bias tends to visit regions of energy where $g(E)$ is greater. The random walk is developed by flipping spin at random but instead of implementing the Metropolis recipe to choose the probability of acceptance of the random move, the idea of the WL sampling is to visit the space in a manner that allows us to obtain a "flat" energy histogram. In fact if we visit the energy state E

with probability $P(E) \sim \frac{1}{g(E)}$ then a flat histogram is generated for the energy distribution. Since the density of states $g(E)$ is not known *a priori* all $g(E)$ are set equal to a common constant value, say 1 at the beginning of a simulation. At every step of the random walk $g(E)$ is modified by a multiplicative factor $f > 1$ and the updated $g(E)$ is used for the next step of random walk. The modification factor f is controlled carefully in the following iterations and finally when $f \approx 1$ the density of states $g(E)$ converges very close to its true value. The accuracy of the estimated density of states depends on many factors such as the final value of the modification factor, flatness criterion or system size.

What follows is a kind of flow chart that explains how the Wang-Landau algorithm develops.

1. ***Initialization.*** At the very beginning of the simulation, we can not know *a priori* the $g(E)$ so we choose $g(E) = 1$, $h(E) = 0$ (the histogram of energy that will become "flat") and $f = e > 1$ (the Euler number)

2. ***Random Move.*** Spin a flip at random. We jump from state i to j

3. ***Accept or reject ?*** The move is accepted or rejected depending on the following probability:

$$p_{acc}(i \longrightarrow j) = min\left(1, \frac{g(E_i)}{g(E_j)}\right) \tag{A.6}$$

   This means that we accept the move if $g(E_j) < g(E_i)$ or if r (uniform random number in $[0, 1]$) is such that $r < \frac{g(E_i)}{g(E_j)}$. In the contrary we retain the configuration i.

4. If j is accepted then $g(E_j) = fg(E_j)$ and $h(E_j) = h(E_j) + 1$. In the case i is accepted $g(E_i) = fg(E_i)$ and $h(E_i) = h(E_i) + 1$

5. Operations 2, 3 and 4 must be repeated till a flatness criterion for $h(E)$ is satisfied. Generally one check for flatness after 10^4 Monte Carlo sweeps (one sweep correspond to a number of spin flip equal to the total number of spins). Given that a perfect flatness cannot be reached one criterion could be $\frac{min_E h(E)}{\langle h(E)\rangle} > p$. The parameter p can be choosen regarding the precision one wants to achieve and the complexity of the system. $\langle h(E)\rangle$ is the mean value of the histogram over the energy bins.

6. ***Reset and update*** When flatness is reached the histogram $h(E)$ must be resetted ($h(E) = 0$) and $f \rightarrow f^{\frac{1}{n}}$ for $n \geq 2$

7. Steps from 2 to 5 must be repeated until f fall below a predefined value close to unity (e.g. $1 + 10^{-6}$)

A.2.1 Ergodicity and detailed balance

Every MC simulation algorithm must satisfy two properties: *ergodicity* and *detailed balance*.
Briefly by ergodicity we mean that starting from any configuration we can reach by repeated random moves any other possibile configuration of the system. Any MC algorithm based on single spin flip satisfies the request of ergodicity because in a sequence of successive steps we can find a path that connects the starting configuration to any other.

The request to satisfy the detailed balance equation impose that the probabilities of moving into a state or leaving it is the same. At the very beginning the density of states $g(E)$ in the WL algorithm change rapidly. This means that the acceptance rule is modified during the simulation by the factor f. For this reason at this stage the detailed balance condition can not be satisfied. When $f \to 1$ the $g(E)$ undergoes finer and finer adjustmenents and at this point the detailed balance condition is very close to be obeyed.
Using (A.6) the ratio between ingoing and outgoing probabilty can be expressed like:

$$\frac{P(E_i \to E_j)}{P(E_j \to E_i)} = \frac{g(E_i)}{g(E_j)} \tag{A.7}$$

In fact when f is close to unity $g(E)$ is weakly modified and the above equation becomes the detailed balance condition:

$$\frac{1}{g(E_i)} P(E_i \to E_j) = \frac{1}{g(E_j)} P(E_j \to E_i) \tag{A.8}$$

with $P(E_i) \sim \frac{1}{g(E_i)}$.

Appendix B

Details of various calculations

B.1 The maximum entropy solution

Suppose that the discrete variable σ describing the system can take values $(\sigma_1, ..., \sigma_W)$, $W = 2^N$ in the case of N binary variables, and that we have M different functions of σ,

$$\phi^\mu(\sigma), \quad \mu = 1, \cdots, M \tag{B.1}$$

We want them to have expectations values:

$$\hat{\phi}^\mu(\sigma) = \langle \phi^\mu(\sigma) \rangle = \sum_\sigma p_\sigma \phi^\mu(\sigma) \tag{B.2}$$

togheter with:

$$\sum_\sigma p_\sigma = 1 \tag{B.3}$$

To find the p_σ's that have maximum entropy subjected to all the constraints simultaneously we add $\boldsymbol{g} = (g_0, \cdots, g_M)$ Lagrange multipliers and then we apply a variation:

$$\delta\Big(S + (g_0 - 1)\sum_\sigma p_\sigma + \sum_{\mu=1}^{M} g_\mu \sum_\sigma p_\sigma \phi^\mu(\sigma)\Big) =$$

$$\sum_\sigma \Big(\frac{\partial S}{\partial p_\sigma} + (g_0 - 1) + \sum_{\mu=1}^{M} g_\mu \phi^\mu(\sigma)\Big)\delta p_\sigma = 0$$

using $\frac{\partial S}{\partial p_\sigma} = -1 - \ln p_\sigma$ the result is:

$$p_\sigma = \exp\Big[g_0 + \sum_{\mu=1}^{M} g_\mu \phi^\mu(\sigma)\Big] \tag{B.4}$$

The normalization (B.3) implies:

$$\exp(g_0) \sum_\sigma \exp\Big[\sum_{\mu=1}^{M} g_\mu \phi^\mu(\sigma)\Big] = 1 \tag{B.5}$$

Defining the partition function as:

$$Z(g_1, \cdots, g_M) = \sum_\sigma \exp\Big[\sum_{\mu=1}^{M} g_\mu \phi^\mu(\sigma)\Big] \tag{B.6}$$

we obtain $g_0 = -\ln Z(g_1, \cdots, g_M)$.
It is now clear how to write down explicitly the expectation value $\hat{\phi}^\mu$ using B.4:

$$\hat{\phi}^\mu = e^{g_0} \sum_\sigma \phi^\mu(\sigma) e^{\sum_{\mu=1}^{M} g_\mu \phi^\mu(\sigma)} \tag{B.7}$$

equivalent to

$$\hat{\phi}^\mu = \frac{\partial \ln Z}{\partial g_\mu} \tag{B.8}$$

Substituting the expression (B.3) inside the entropy functional we can calculate his extreme value that is:

$$S^* = -g_0 - \sum_{\mu=1}^{M} g_\mu \hat{\phi}^\mu \tag{B.9}$$

As written by Jaynes *"our Lagrange multiplier argument has the nice feature that it gives us the answer instantaneously. It has the bad feature that after we done it, were not quite sure it is the answer"*.
In effect $\delta S = 0$ shows only that the entropy is stationary and to complete the argument we must show that (B.9) is the global property rather than just a local extremum or a stationary point.

To prove that (B.4) has this global property let us suppose that we have in addition another pdf q_σ that also satisfy the same constraints (B.2) of p_σ.
We define:

$$D_{KL}[q, p] = \sum_\sigma q_\sigma \ln \frac{q_\sigma}{p_\sigma} = -S[q] - \sum_\sigma q_\sigma \ln p_\sigma \tag{B.10}$$

it is known as the Kullback-Liebler divergence and it represents a sort of "metric" (it is not symmetric, $D_{KL}[p,q] \neq D_{KL}[q,p]$) on the space of probability distributions. We show that $D_{KL}[q,p] \geq 0$. Using $\ln x \leq x - 1$ we can write:

$$\sum_{\sigma} q_\sigma \ln \frac{p_\sigma}{q_\sigma} \leq \sum_{\sigma} q_\sigma \Big(\frac{p_\sigma}{q_\sigma} - 1 \Big) = 0 \tag{B.11}$$

that means:

$$S[q] \leq -\sum_{\sigma} q_\sigma \ln p_\sigma \tag{B.12}$$

At this point we can substitute (B.4) for p_σ obtaining:

$$S[q] \leq -\sum_{\sigma} q_\sigma \sum_{\mu=0}^{M} g_\mu \phi^\mu(\sigma) = -\sum_{\mu=0}^{M} g_\mu \hat{\phi}^\mu = S^* = S[p] \tag{B.13}$$

In words, within the family of all distributions q that satisfy the constraints B.2 the distribution that achieves the maximum entropy is the canonical distribution p given in eq. B.4.

Concluding, we want to notice that the inequalities $S[p] \geq 0$ and $K[q,p] \geq 0$ (the last evaluated in the special case $q_\sigma = 1/W$) give us the range in which $S[p]$ is contained:

$$0 \leq S[p] \leq \ln W \tag{B.14}$$

The two interval extremes correspond respectively to complete certainty $p_\sigma = \delta_{\sigma\sigma'}$ and complete uncertainty where having no more information than $\sum_\sigma p_\sigma = 1$ we choose $p_\sigma = 1/W$.

B.2 Convex optimization: the gradient descent algorithm

What follows briefly explain how to infere the vector $\boldsymbol{g}$ of couplings capable of reproducing mean observed values.

Consider a convex, differentiable function $H(\boldsymbol{g}) : \mathbb{R}^M \to \mathbb{R}$. Then for each point $\boldsymbol{g}$ it exists a gradient $\nabla H(\boldsymbol{g}) = (\partial_{g_1}, \cdots, \partial_{g_M})\, H(\boldsymbol{g})$ and a positive semidefinite *Hessian* matrix $\boldsymbol{\chi}(\boldsymbol{g})$ with elements $\chi_{\mu\nu} = \partial_{g_\mu}\partial_{g_\nu} H(\boldsymbol{g})$.
Using the gradient we can define a descent direction $\boldsymbol{v} = -\nabla H(\boldsymbol{g})$.

This means that for all $\boldsymbol{g}$ it exists an ϵ such that:

$$H(\boldsymbol{g} - \epsilon \nabla H(\boldsymbol{g}) \leq H(\boldsymbol{g}) \tag{B.15}$$

To find the minimum $\boldsymbol{g}^*$ of $H(\boldsymbol{g})$ means to solve $\nabla H(\boldsymbol{g}) = 0$. If it can't be analytically solved, starting from a point $\boldsymbol{g}^{(0)}$ we can built an iterative scheme:

$$\boldsymbol{g}^{(k+1)} = \boldsymbol{g}^{(k)} - \epsilon \nabla H(\boldsymbol{g}^{(k)}) \tag{B.16}$$

that ensures we can reach a *global minimum* $\boldsymbol{g}^*$ of $H(\boldsymbol{g})$.

Applying this ideas to the function (for a justification see below):

$$H(\boldsymbol{g}) = \ln Z(\boldsymbol{g}) - \sum_{\mu>0} \hat{\phi}^\mu g_\mu \tag{B.17}$$

allows us to find the couplings $\boldsymbol{g}^*$ such that $\langle \phi^\mu \rangle_{\boldsymbol{g}^*} = \hat{\phi}^\mu$.
In fact, since $\partial_{g_\mu} H(\boldsymbol{g}) = \langle \phi^\mu \rangle_{\boldsymbol{g}} - \hat{\phi}^\mu$, finding a global minimum is equivalent to state:

$$\langle \phi^\mu \rangle_{\boldsymbol{g}} = \hat{\phi}^\mu \tag{B.18}$$

We only have to prove that this minimum exists and it is unique and so that $H(\boldsymbol{g})$ is convex (i.e. $\partial^2_{g_\mu g_\nu} H \geq 0$).

For this purpose we define the *susceptibility* matrix (the Fisher information metric in information theory):

$$\chi_{\mu\nu} = -\frac{\partial^2 F}{\partial g_\mu \partial g_\nu} \tag{B.19}$$

where $F(\boldsymbol{g}) = -\ln Z(\boldsymbol{g})$.
It represents the Hessian of $H(\boldsymbol{g})$ and its properties ensure that the minimum, if exists, is global.
In fact we briefly demonstrate that $\chi_{\mu\nu}$ is a positive semidefinite matrix. Just to make things more fluid to read we indicate $\langle \cdot \rangle_{\boldsymbol{g}} \equiv \langle \cdot \rangle$, $p(\vec{\sigma}|\boldsymbol{g}) \equiv p_\sigma$ and $\phi^\mu(\vec{\sigma}) \equiv \phi^\mu_\sigma$.
From the definition (B.19) it follows that:

$$\chi_{\mu\nu} = \langle \phi_\mu \phi_\nu \rangle - \langle \phi_\mu \rangle \langle \phi_\nu \rangle = \sum_\sigma p_\sigma \left(\phi^\mu_\sigma - \langle \phi^\mu \rangle \right) \left(\phi^\nu_\sigma - \langle \phi^\nu \rangle \right) \tag{B.20}$$

Using this rearrangement it is straightforward to prove that for any vector $\boldsymbol{x}$ the quadratic form $\sum_{\mu,\nu>0} x_\mu \chi_{\mu,\nu} x_\nu$ is greater or equal to zero. Indeed,

$$\begin{aligned}\sum_{\mu,\nu>0} x_\mu \chi_{\mu\nu} x_\nu &= \sum_\sigma p_\sigma \left[\sum_\mu x_\mu \left(\phi_\sigma^\mu - \langle\phi^\mu\rangle\right)\right]\left[\sum_\nu x_\nu \left(\phi_\sigma^\nu - \langle\phi^\nu\rangle\right)\right] \\ &= \left\langle \left[\sum_\nu x_\nu \left(\phi_\sigma^\nu - \langle\phi^\nu\rangle\right)\right]^2 \right\rangle \geq 0 \end{aligned} \tag{B.21}$$

in fact this result holds for every p_σ.

The use of the function (B.17) can be justified using the the Kullback-Liebler divergence. It was defined in the previous section We recall here that it is defined as:

$$D_{KL}[q,p] = \sum_\sigma q_\sigma \ln \frac{q_\sigma}{p_\sigma} = -S[q] - \sum_\sigma q_\sigma \ln p_\sigma \tag{B.22}$$

Although it doesn't satisfy the symmetry condition nor the triangular inequality it has some of the properties of a metric. It is always non negative (see (B.11)) and is zero if and only if $p = q$ moreover it is a convex function in p and q (see [11]).
If q is the true probability distribution that generates the observed averages and p is the maximum entropy probability (B.4) the Kullback-Liebler reads:

$$D_{KL}[q,p] = \sum_\sigma q_\sigma \ln \frac{q_\sigma}{p_\sigma} = -S[q] + \ln Z(\boldsymbol{g}) - \sum_\sigma q_\sigma \sum_{\mu>0} g_\mu \phi_\sigma^\mu \tag{B.23}$$

Since that:

$$\sum_\sigma q_\sigma \phi_\sigma^\mu = \hat{\phi}^\mu \tag{B.24}$$

we rewrite (B.23) as:

$$D_{KL}[q,p] = -S[q] + \ln Z(\boldsymbol{g}) - \sum_{\mu>0} g_\mu \hat{\phi}^\mu \tag{B.25}$$

The expression above is the function (B.17) up to a constant. Regarding it as a function of $\boldsymbol{g}$ the minimum of the "distance" between p and q is found when $\langle\phi^\mu(\sigma)\rangle_{\boldsymbol{g}} = \hat{\phi}^\mu$.

Bibliography

[1] Allesina, S., and Tang, S. Stability criteria for complex ecosystems. *Nature 483*, 7388 (Mar. 2012), 205–8.

[2] Arrhenius, O. Species and area. *Journal of Ecology 9*, 1 (1921), 95–99.

[3] Azaele, S., Cornell, S., and Kunin, W. Downscaling species occupancy from coarse spatial scales. *Ecological Applications 22*, 3 (2012), 1004–1014.

[4] Azaele, S., Muneepeerakul, R., Rinaldo, a., and Rodriguez-Iturbe, I. Inferring plant ecosystem organization from species occurrences. *Journal of theoretical biology 262*, 2 (Jan. 2010), 323–9.

[5] Azaele, S., Pigolotti, S., Banavar, J. R., and Maritan, A. Dynamical evolution of ecosystems. *Nature 444*, 7121 (Dec. 2006), 926–8.

[6] Baldeck, C. a., Harms, K. E., Yavitt, J. B., John, R., Turner, B. L., Valencia, R., Navarrete, H., Davies, S. J., Chuyong, G. B., Kenfack, D., Thomas, D. W., Madawala, S., Gunatilleke, N., Gunatilleke, S., Bunyavejchewin, S., Kiratiprayoon, S., Yaacob, A., Supardi, M. N. N., and Dalling, J. W. Soil resources and topography shape local tree community structure in tropical forests. *Proceedings. Biological sciences / The Royal Society 280*, 1753 (Feb. 2013), 20122532.

[7] Banavar, J. R., Cooke, T. J., Rinaldo, A., and Maritan, A. Form, function, and evolution of living organisms. *Proceedings of the National Academy of Sciences of the United States of America 111*, 9 (Mar. 2014), 3332–7.

[8] Chesson, P. Mechanisms of maintenance of species diversity. *Annual review of Ecology and Systematics* (2000).

[9] Coleman, B. On random placement and species-area relations. *Mathematical Biosciences 215* (1981), 191–215.

[10] Condit, R., Pitman, N., Leigh, E. G., Chave, J., Terborgh, J., Foster, R. B., Núñez, P., Aguilar, S., Valencia, R., Villa, G., Muller-Landau, H. C., Losos, E., and Hubbell, S. P. Beta-diversity in tropical forest trees. *Science (New York, N.Y.) 295*, 5555 (Jan. 2002), 666–9.

[11] Cover, T., and Thomas, J. *Elements of information theory*. Wiley-interscience, 2012.

[12] Durrett, R., and Levin, S. Spatial Models for Species-Area Curves. *Journal of Theoretical Biology 179*, 2 (Mar. 1996), 119–127.

[13] Gould, S. An allometric interpretation of species-area curves: the meaning of the coefficient. *American Naturalist 114*, 3 (1979), 335–343.

[14] Green, J., and Bohannan, B. J. M. Spatial scaling of microbial biodiversity. *Trends in ecology & evolution 21*, 9 (Sept. 2006), 501–7.

[15] Grilli, J., Azaele, S., Banavar, J., and Maritan, A. Spatial aggregation and the speciesarea relationship across scales. *Journal of theoretical biology* (2012), 1–16.

[16] Hanski, I., and Gyllenberg, M. Uniting Two General Patterns in the Distribution of Species. *Science 275*, 5298 (Jan. 1997), 397–400.

[17] Harte, J., Kinzig, A., and Green, J. Self-similarity in the distribution and abundance of species. *Science 334*, 1999 (1999).

[18] Harte, J., Zillio, T., Conlisk, E., and Smith, A. Maximum entropy and the state-variable approach to macroecology. *Ecology 89*, May 1975 (2008), 2700–2711.

[19] He, F., and Hubbell, S. P. Species-area relationships always overestimate extinction rates from habitat loss. *Nature 473*, 7347 (May 2011), 368–71.

[20] He, F., and Hubbell, S. P. Species-area relationships always overestimate extinction rates from habitat loss. *Nature 473*, 7347 (May 2011), 368–71.

[21] He, F., and Legendre, P. Species diversity patterns derived from species-area models. *Ecology* (2002).

[22] Hidalgo, J., Grilli, J., and Suweis, S. Emergence of criticality in living systems through adaptation and evolution: Practice Makes Critical. *arXiv preprint arXiv: ...* (2013).

[23] Hubbell, S. P. *The Unified Neutral Theory of Biodiversity and Biogeography*. 2011.

[24] Jaynes, E. Information theory and statistical mechanics. *Physical review* (1957).

[25] Jaynes, E. Gibbs vs Boltzmann entropies. *Am. J. Phys* (1965).

[26] Jaynes, E. *Probability theory: the logic of science*. 2003.

[27] Kéfi, S., Rietkerk, M., Alados, C. L., Pueyo, Y., Papanastasis, V. P., Elaich, A., and de Ruiter, P. C. Spatial vegetation patterns and imminent desertification in Mediterranean arid ecosystems. *Nature 449*, 7159 (Sept. 2007), 213–7.

[28] Kerr, B., Riley, M. a., Feldman, M. W., and Bohannan, B. J. M. Local dispersal promotes biodiversity in a real-life game of rock-paper-scissors. *Nature 418*, 6894 (July 2002), 171–4.

[29] Liggett, T. M. *Stochastic Interacting Systems: Contact, Voter and Exclusion Processes*. 1999.

[30] MacArthur, R. H., and Wilson, E. O. *The theory of island biogeography*, vol. 1. Wiley-interscience, 1967.

[31] Mastromatteo, I., and Marsili, M. On the criticality of inferred models. *arXiv s* (2011), 1–6.

[32] May, R. Qualitative stability in model ecosystems. *Ecology 54*, 3 (1973), 638–641.

[33] McGill, B. J. Towards a unification of unified theories of biodiversity. *Ecology letters 13*, 5 (May 2010), 627–42.

[34] Mega, M., Allegrini, P., Grigolini, P., Latora, V., Palatella, L., Rapisarda, A., and Vinciguerra, S. Power-Law Time Distribution of Large Earthquakes. *Physical Review Letters 90*, 18 (May 2003), 188501.

[35] Meinhardt, H. *Models of biological pattern formation.* 1982.

[36] Mora, T., and Bialek, W. Are biological systems poised at criticality? *Journal of Statistical Physics* (2011), 1–21.

[37] Newman, M. Power laws, Pareto distributions and Zipf's law. *Contemporary Physics 46*, 5 (Sept. 2005), 323–351.

[38] Pigolotti, S., and Cencini, M. Speciation-rate dependence in species-area relationships. *Journal of theoretical biology 260*, 1 (Sept. 2009), 83–9.

[39] Plotkin, J. B., Potts, M. D., Leslie, N., Manokaran, N., Lafrankie, J., and Ashton, P. S. Species-area curves, spatial aggregation, and habitat specialization in tropical forests. *Journal of theoretical biology 207*, 1 (Nov. 2000), 81–99.

[40] Pressé, S., Ghosh, K., Lee, J., and Dill, K. a. Principles of maximum entropy and maximum caliber in statistical physics. *Reviews of Modern Physics 85*, 3 (July 2013), 1115–1141.

[41] Reid, C. R., Sumpter, D. J. T., and Beekman, M. Optimisation in a natural system: Argentine ants solve the Towers of Hanoi. *The Journal of experimental biology 214*, Pt 1 (Jan. 2011), 50–8.

[42] Rosenzweig, M. L. *Species diversity in space and time*, vol. 19. Wiley-interscience, 1995.

[43] Rosindell, J., and Cornell, S. J. Species-area relationships from a spatially explicit neutral model in an infinite landscape. *Ecology letters 10*, 7 (July 2007), 586–95.

[44] Roudi, Y., Nirenberg, S., and Latham, P. E. Pairwise maximum entropy models for studying large biological systems: when they can work and when they can't. *PLoS computational biology 5*, 5 (May 2009), e1000380.

[45] Scanlon, T. M., Caylor, K. K., Levin, S. a., and Rodriguez-Iturbe, I. Positive feedbacks promote power-law clustering of Kalahari vegetation. *Nature 449*, 7159 (Sept. 2007), 209–12.

[46] Schneidman, E., Berry, M. J., Segev, R., and Bialek, W. Weak pairwise correlations imply strongly correlated network states in a neural population. *Nature 440*, 7087 (Apr. 2006), 1007–12.

[47] Schneidman, E., Still, S., Berry, M. J., and Bialek, W. Network Information and Connected Correlations. *Physical Review Letters 91*, 23 (Dec. 2003), 238701.

[48] Seno, F., Trovato, A., Banavar, J., and Maritan, A. Maximum Entropy Approach for Deducing Amino Acid Interactions in Proteins. *Physical Review Letters 100*, 7 (Feb. 2008), 078102.

[49] Sizling, A. L., and Storch, D. Power-law species-area relationships and self-similar species distributions within finite areas. *Ecology Letters 7*, 1 (Jan. 2004), 60–68.

[50] Solé, R., Valls, J., and Bascompte, J. Spiral waves, chaos and multiple attractors in lattice models of interacting populations. *Physics Letters A 166* (1992), 123–128.

[51] Suweis, S., Grilli, J., and Maritan, A. Disentangling the effect of hybrid interactions and of the constant effort hypothesis on ecological community stability. *Oikos*, October (Nov. 2013), no–no.

[52] Turing, A. The chemical basis of morphogenesis. *Bulletin of mathematical biology 237*, 641 (1956), 37–72.

[53] Volkov, I., Banavar, J. R., Hubbell, S. P., and Maritan, A. Neutral theory and relative species abundance in ecology. *Nature 424*, 6952 (Aug. 2003), 1035–7.

[54] Volkov, I., Banavar, J. R., Hubbell, S. P., and Maritan, A. Inferring species interactions in tropical forests. *Proceedings of the National Academy of Sciences of the United States of America 106*, 33 (Aug. 2009), 13854–9.

[55] Wiegand, T., Huth, A., Getzin, S., Wang, X., Hao, Z., Gunatilleke, C. V. S., and Gunatilleke, I. a. U. N. Testing the independent species' arrangement assertion made by theories of stochastic geometry of biodiversity. *Proceedings. Biological sciences / The Royal Society 279*, 1741 (Aug. 2012), 3312–20.

[56] Wootton, J. T., and Emmerson, M. Measurement of Interaction Strength in Nature. *Annual Review of Ecology, Evolution, and Systematics 36*, 1 (Dec. 2005), 419–444.

[57] Zillio, T., and He, F. Modeling spatial aggregation of finite populations. *Ecology 93*, 11 (Nov. 2010), 2497–8.

[58] Zillio, T., Volkov, I., Banavar, J. R., Hubbell, S. P., and Maritan, A. Spatial scaling in model plant communities. *Physical review letters 95*, 9 (2005), 098101.