

STATISTICAL AND BIOLOGICAL PHYSICS SECTOR



Spatial Dynamics of Ecosystems

Thesis submitted for the degree of
Doctor Philosophiæ

Candidate:
Tommaso Zillio

Supervisor:
Prof. Amos Maritan

Thesis Advisor:
Prof. Cristian Micheletti

October, 21st 2005

*To Eva,
who understand plants
much better than me.*

Contents

Introduction	1
1 Spatial Patterns in Ecology	5
1.1 Community, Metacommunity and Trophical Similarity	5
1.2 The Law of Constant Density	7
1.3 α -Diversity vs β -Diversity	9
1.4 Relative Species Abundance (RSA)	10
1.5 Species Area Relation (SAR)	12
1.6 Measuring Spatial Turnover of Species	13
2 Modeling Ecosystems	25
2.1 Typologies of Models	25
2.1.1 The Lotka-Volterra Model of Predation.	25
2.1.2 Stochastic vs Deterministic Modeling	26
2.1.3 Discrete vs Continuous Modeling	27
2.2 Island Biogeography	27
2.3 Neutrality	30
2.4 The Unified Neutral Theory of Biodiversity - The Drift Model	31
3 The Voter Model With Speciation	35
3.1 Motivation	35
3.2 Definition	36
3.3 Master Equation Formulation and Solution	38
3.3.1 Lattice Master Equation	38
3.3.2 Lattice Solution and Total Probability	39
3.3.3 Continuum Master Equation and Solution	40
3.4 β -diversity	41
3.4.1 Master Equation and Solution on the Lattice	41

3.4.2	Solution in the Continuum	43
3.5	Scaling Relations	45
3.6	Simulations and Scaling Exponents	46
3.7	The Janzen-Connell effect	56
4	Spatial Patterns of Intertidal Vegetation Diversity	61
4.1	Motivation	61
4.2	Master Equation General Formulation and Solution	62
4.3	The Model Definition	64
4.3.1	One species plus the void	64
4.3.2	$N > 1$ species	66
4.4	Application to the Salt Marshes	67
5	Conclusions and Perspectives	73
A	Lattice Correlation Functions for Voter Model with Speciation.	77
A.1	1-Point Correlation Function (Generating Function).	77
A.2	1-Point Correlation Function (Fourier Transform)	79
A.3	2-Point Correlation Function	80
B	Relation Between $\sum_{\vec{x}} F_{\vec{x}}$ and Moments Ratio.	83
C	Asymptotic Behavior of $\sum_{\vec{x}} F_{\vec{x}}$	85
D	Scaling Relations	87
D.1	$1 + c - \beta > 0$	87
D.2	$-1 < 1 + c - \beta < 0$	88
D.3	$1 + c - \beta < -1$	89
E	Solution of Differential Equation in J.C. Voter Model	91
F	Correlation Functions for Intertidal Vegetation	93
F.1	1-Point Correlation Function	93
F.2	2-Point Correlation Function	94
F.3	Stationary Solutions	94
F.4	Analytic Solution for 2-point Correlation Function	95
F.4.1	tensorial part	97
F.4.2	spatial part	98

Introduction

*We're still in the Middle Ages in biodiversity research.
We're still cutting bodies open to see what organs are inside.*
(S.P.Hubbell)

Although the mathematical methodology was introduced in biology more than a couple of centuries ago [1], it is only recently that the tools developed in the realm of statistical physics have shown themselves as helpful in understanding complex problems arising in the diverse areas of biology. In some field this type of approach has become well established, as in the case of the study of biopolimers and proteins; in other cases this methodology is not so widespread but the interest is growing rapidly. This is the case of ecology.

There is now a growing availability of extensive datasets, thanks to organizations like the Center for Tropical Forest Science (CTFS) of the Smithsonian Institute, and his Forest Dynamics Plot project [2] which provide regular censuses of many forest plots all around the world; or the TIDE project (Tidal Inlet Dynamics and Environment) [3] monitoring tidal areas such as lagoons and estuaries. Moreover, novel techniques like spectral data analysis of satellite images (as we will see later) have made possible to have accurate data from vegetation patches that were before considered impossible to analyze quantitatively.

With more and more data available from many diverse fields of ecology it is of paramount importance to find models that are capable of explaining them. However it is not an easy task due to the sheer complexity of the ecosystems i.e. the inextricable web of different relations between the species present in the system. A model that will be able to explain quantitatively or even qualitatively a whole ecosystem is, and will be for a long time, completely beyond our possibilities. The ecological processes act simultaneously on very different scales of time and space [4]. Nevertheless

it is possible to focus on certain aspects of a certain subset of species, as we will precise later, and to construct models capable of explaining this simplified ecosystem.

In this work we are mainly interested to the study of the spatial distribution of the vegetation patches. We will utilize data that come from the tropical forests of Barro Colorado Island (Panama), and Yasuni (Peru), along with a different set of data about intertidal vegetation (the vegetation that is periodically submersed by the sea, due to the tide) in the venetian salt marshes. Due to the very recent availability of data¹, the field of spatial explicit ecology is not yet a well developed one, but it looks promising. Many models were proposed but we still lack a unifying perspective.

In the present work, for the first time we will be able to fit empirical data about spatial distribution of species in a meaningful way using spatial explicit ecological models. In particular we will use a neutral model to fit the data from tropical forests, where the number of species involved make unpractical to characterize each one of them. The result show that the neutral approximation is very well suited to fit the data, provided that an additional effect, called the Janzen-Connell one, is implemented. On the contrary, salt marshes vegetation distribution will be explained using a non-neutral model because the number of species involved is very low (3-4 species at most), and so it is possible to write and solve analytically a model able to distinguish between different species².

The core of this work consists of the analytical solution of two spatially explicit models. The solutions will be always verified with explicit computer simulations which are also used to calculate quantities not analytically accessible. In the case of the models developed for the tropical forests we will be also able to find a scaling form for it, linking together some ecologically relevant quantities that were considered independent. In both model, their analytical solution is utilized to fit the empirical data.

The organization of this research work is as follows:

¹ The first data about the spatial distribution of trees in the tropical forest are available since the early 80's, but the real study of clustering has begun much more recently, see for example [18]; the data about the intertidal vegetation were taken during the last years and the spatial distribution of species is analyzed for the first time in the present work.

² In fact the solution that we provide is valid for any number of species, but the application of this solutions looks unpractical if the number of species is much greater than here.

-
- In the First Chapter we will discuss some patterns identified in the last years by ecologists in this ecosystem, in their research of a meaningful measure of biodiversity, focusing in particular on the spatial patterns. We then will describe the two ecosystems studied in this work, the Tropical Forest and the Intertidal Vegetation in Salt Marshes.
 - In the Second Chapter we will give a brief survey of the past line of research relevant for this work. We will introduce the important concept of neutrality and outline some dynamical population model used so far.
 - In the Third Chapter we will study the voter model with speciation that we will use in a modified form to fit the data from the Tropical Forests. We will study also a scaling relationship for the original model.
 - In the Fourth Chapter we will examine a speciationless non-neutral model with the purpose to reproduce the spatial organization of the Salt Marshes, and discuss the fits that one can obtain.

And finally we will state some conclusion and perspectives, assessing the relevance of the work done and looking at possible improvement of our theories. We will discuss also the proper use of our models and how this can contribute to the research in ecology.

Chapter 1

Spatial Patterns in Ecology

*If a tree falls in the forest
with no one to hear him,
is the tree really fallen?*
(Koan Zen)

*If a man is talking in the forest,
and no woman is there to hear him,
is he still wrong?*
(Peter Kay)

In this chapter we will outline some of the patterns highlighted by ecologists during their study of the ecosystems.

1.1 Community, Metacommunity and Trophical Similarity

A complete and exhaustive study of an ecosystem is a formidable and probably impossible task with the tools currently available. Indeed, an ecosystem is a very complex object composed by many individuals of many different species with different sizes, diet, reproductive strategy, and so on. Most of the time there will be plants and animals (along with fungi, bacteria, etc.) in the same ecosystem. The relations among them can be very different: there can be competition, symbiosis, predator-prey or parasite-host relationship and so on. To further complicate the situation, ecosystems are not closed systems, and are subjected to immigration,

environmental and climate change, etc. In an effort to understand something from such a complex problem it is useful to fix some guidelines and restrict the observation to something simpler than the whole system.

A first and natural choice, that we will encounter later in the Lotka-Volterra model, is to consider only few species and to consider all the rest of the ecosystem as an external perturbation. The hope is to reconstruct the behavior of all the system starting from the simpler interaction of a subset of it [5].

Another approach, used in the main part of this work, is to focus on similar species and ignore the remaining environment (that is assumed to give the same conditions of life to all the considered species). To this end we will say that two species are *trophically similar* if:

- The species are similar (i.e. they have roughly the same lifespan, the same body size, etc.)
- The species compete for the same resources

An *ecological community* is the ensemble of all mutually trophically similar species living in a defined area (a single island, an isolated forest, etc.). A couple of synonyms for ecological community are *tropical level* and *local community*. If the same kind of species exist also on a larger spatial scale (like a continental scale), then we will speak of *metacommunity*, the main difference with the local community being that due to the large scale species belonging to the community may not directly compete or interact whatsoever.

The definition of trophic level implies some desirable characteristic: first of all it restrict all the interaction between considered species to the competition for the same resources, thus simplifying a lot the problem. Then it allows to formulate in the clearest way one of the biggest problem in ecology: why species have abundances so different among them? In the same trophic level the species are all similar, and so the answer should lie in their competitive dynamics. This is obviously a gross oversimplification that can be justified only *a posteriori*, by the results it produces.

Good datasets about single-trophic level ecosystem are today available from extensive studies on tropical forests [6]

In all the rest of this thesis, unless explicitly stated, we will assume that we are studying a single trophic level.

1.2 The Law of Constant Density

One of the simplest and most important characteristic of a generic ecosystem is the fact that, by definition, the ecosystem is *finite*. This means that there is an unavoidable physical constraint on the total number of individuals that can be packed into a given space. Space *per se* is not a resource, but it is often directly proportional to real resources (water, nutrients, light), and thus can act as a surrogate variable for these resources. Although the real resources are abundant, the very physical size of the organisms give an upper limit on the number of individuals that can live in a given area. Obviously, in a community of trophically dissimilar species, the number of organism per unit area will vary, depending on many factors such as their relative size, the energy demand and so on. However, if the community is composed solely of trophically similar species, one can guess that the density of organism should no vary too much from different local samples. MacArthur and Wilson [7] noticed that the number of individuals J in a local community depends linearly on the area A :

$$J = \rho A \quad (1.1)$$

ρ being the density of individuals. A striking example of this law is illustrated in Fig.(1.1). It is an enumeration of trees in a closed-canopy tropical forest, specifically in the 50-ha plot in Barro Colorado Island (BCI), Panama. The relationship holds very closely for five orders of magnitude (from $1m^2$ to $5 \cdot 10^5m^2$), despite the fact that the trees counted belong to many different species (> 300 species). While the linear shape of the average number of trees over the area is linear by definition¹, one should remark that in Fig.1.1 are also plotted the standard deviation, and that they are barely visible.

Now, there is only one way for this relationship to hold dynamically with such a precision: the landscape is saturated with individuals. This means that:

¹ Let's divide a given plot of area A in N small equal zones of area $a = A/N$. Let T be the total number of trees in A , and t_i the number of trees in the i -th area, $i = 1, \dots, N$. Then we have:

$$\langle t \rangle = \frac{\sum_i t_i}{N} = \frac{T}{A} a \quad (1.2)$$

that is linear in a , where we have used $\sum_i t_i = T$ and $N = A/a$. Remark that correctly $\rho = T/A$.

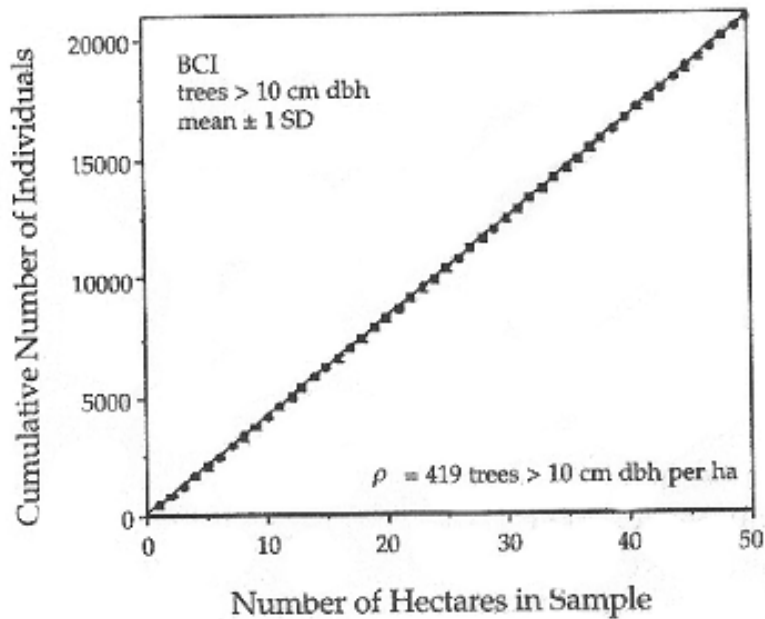


Figure 1.1: Number of individuals vs area in the 50-ha plot of tropical moist forest of Barro Colorado Island, Panama. The trees counted in this plot are the ones with a trunk diameter > 10 cm dbh (diameter at breast height) of all species. The points are the mean of one hundred random starting points for accumulation of area within the plot. One standard deviation about each mean is also plotted, but they are so small that they are barely visible. The mean density of individuals > 10 cm dbh is $\rho = 419$ individuals per hectare (see Eq.(1.1)). Picture taken from [9].

- Non-saturated landscapes are temporary: they are created by disturbances (treefalls, forest fires, etc.) and no unused resource goes unharvested for long time².
- A new individual can be born and grow only if another individual dies leaving some resource unused; or, using a game-theoretical approach: the dynamics of ecological communities are a zero-sum game.

It is thus perfectly natural to study models in which the number of individuals is constant; in particular for the study of spatial distribution it

² It is obviously possible that the disturbances are so severe and frequent that the landscape is in fact not saturated at all, showing open space and unused resources.

will make sense to study models defined on a lattice, with one individual per lattice site. We will allow empty sites depending of the magnitude of disturbances that we want to modelize . For the Tropical Closed-Canopy Forests we will not allow any empty space: a new birth will follow immediately any death. Such a model is developed in Chap.3. In the case of Salt Marshes and Intertidal Vegetation it is easy to see from the experimental data that the empty space plays a essential role in the dynamics of the system (i.e. the disturbances are more present). In this case the number of individuals is no more constant and we have to take into account explicitly the dynamics of the empty spaces. A model with these characteristics is presented in details in Chap.4.

1.3 α -Diversity vs β -Diversity

Presented with the overwhelming complexity of the ecosystems, and searching for meaningful ways to characterize them, field biologists have defined a great number of measurable quantities in their studies. Most of these quantities were defined to provide an estimation of the biodiversity of the sample ecosystem under examination. The aim of this operation is obvious: such a measure can be extremely useful for quantifying the environmental impact of man on a natural ecosystem. So far nobody has found a simple and meaningful quantity for the biodiversity of a system, and in fact the problem seems more difficult the more one learns about it. Obvious answers, such as “the biodiversity of a system is the number of species that lives in it” face at least three difficulties:

- An experimental one: It can be difficult, or impossible, to catalogue all the species in a given environment due to the rareness or the tendency of the individuals to avoid man.
- A metodological one: the classic definition of species as “the set of individuals that can have fertile offsprings” it is flawed because:
 - It deals only with species with sexual reproduction.
 - It is fuzzy because the transitive property doesn’t hold always.
 - It is difficult to apply on the field, because to verify directly it takes at least two generations of the species under examination.

- A theoretical one: it completely ignores the relative abundance of the species and the spatial variability of the ecosystem.

In the following paragraphs we will try to overcome the last difficulty stated. The former two are connaturated with the field methodology and are probably unavoidable.

We will divide the various quantities related with the biodiversity in two broad categories: α -diversity vs β -diversity. The former refers to the measures of the biodiversity in a single place considered perfectly uniform; the number of species being the most trivial example. The latter instead is concerned with a spatial explicit measure of the biodiversity, i.e. the spatial turnover of species. In the following sections we will examine the most important and useful quantities defined for each category. For the α -diversity we will examine the Relative Species Abundance, while for the β -diversity we will present the Species Area Relationship and the 2-point correlation function.

1.4 Relative Species Abundance (RSA)

The RSA, ϕ_N , is defined as the number of species having a given number N of individuals in the sample ecosystem. In a theoretical model or in the case of repeated measures in the same place it will be more convenient to use $\langle \phi_N \rangle$, the average number of species with N individuals. This quantity is used to determine how many rare species there are, respect to the abundant ones.

The first data for RSA were obtained by S. Corbet, an entomologist, in 1941. He collected the data for 620 species of butterflies in the Malay Peninsula [10]. After two years was published an article of Fisher, Corbet and Williams [11] with analogous results with a four-years light-trap study on the moths at Rothamsted Experimental Station in England. The resulting distribution (see Fig.1.2) was monotonic and long-tailed, with many rare species and just a few common. Fisher found that the data were reasonably fitted by a distribution of the form:

$$\langle \phi_N \rangle \propto \frac{\alpha^N}{N} \quad (1.3)$$

with $\alpha < 1$ as a parameter characterizing the system. Typical values for α are very close to 1. This distribution is known among ecologists as the Fisher Logseries³.

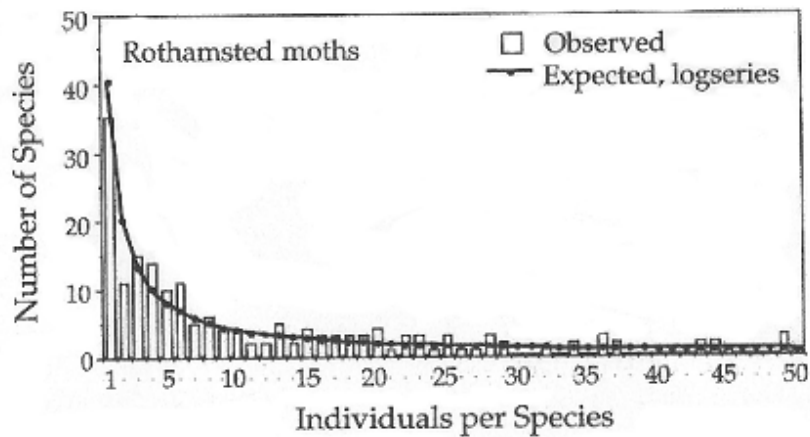


Figure 1.2: The data set of Fisher, Corbet and Williams, and the log series fit. Picture taken from [9].

Some years later, Preston [12] found that the logseries was unable to fit correctly the data he had assembled, primarily on bird species abundance. The main problem was that Preston's data were non-monotonical, showing a maximum for species with intermediate abundance. In order to obtain a clearer plot he made a log₂ binning of his data, with the additional prescription that species at the edge of the bin are equally divided between the two adjacent bins. This plot is known in the literature as the Preston plot. The obtained histogram appeared to be well fitted by a lognormal distribution (see Fig.1.3).

The main problem with these curves is that they both lack an underlying dynamical model. The Fisher Logseries was obtained by assuming that the true distribution of species abundances was a Gamma distribution (which is completely arbitrary) and the RSA was a Poisson sampling

³ This is because it resembles the Taylor expansion of a logarithm:

$$-\ln(1 - \alpha) = \alpha + \frac{\alpha^2}{2} + \frac{\alpha^3}{3} + \dots \quad (1.4)$$

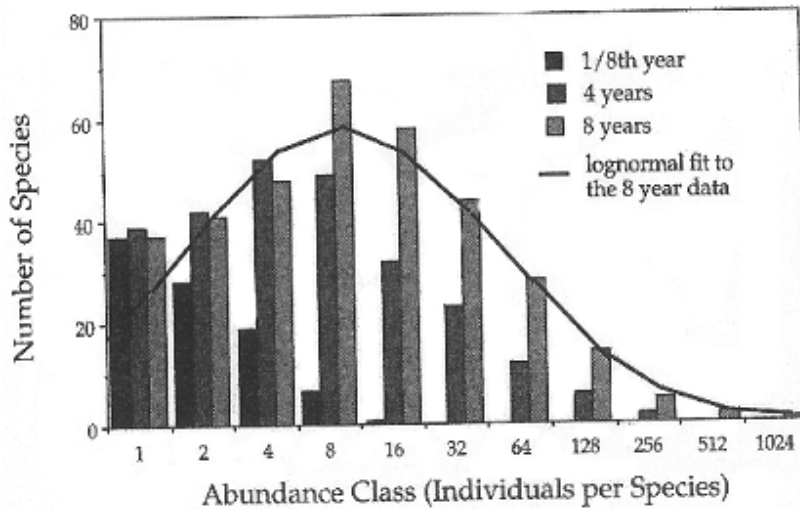


Figure 1.3: Lognormal fit to species abundance for the Rothamsted Field Station. Picture taken from [9].

of this distribution. The Preston Lognormal was justified on the basis that it is the result of a multiplicative random walk⁴. Apart from the lack of a convincing explanation of why the ecosystem should behave like a multiplicative random walk, one should remark that in such process the standard deviation of the Gaussian curve should grow with time, and this effect is completely absent in the data. Another problem is that the data show a “fat-tail” for rarest species, and this tail is not fitted by the lognormal. A tentative solution of these problems will be given in 2.4.

1.5 Species Area Relation (SAR)

The SAR is defined as the number of species S that lives, on average, in an area A of the ecosystem under scrutiny. The most used form for SAR [7] is a power law:

$$S \propto A^z \quad (1.5)$$

⁴An alternative explanation is given, among others, in [13]. It should also be noted that a multiplicative random walk with the constraint that the total number of individuals is fixed does not give a lognormal curve for the RSA, but a curve with a fat-tail for rarest species, as the experimental ones. See [14].

It should be noted, however, that such a form is unable to fit most of the data at once. Experimental SARs, indeed, show a behavior similar to a “piecewise power law”. The exponent z of the power law is in some sense a measure of biodiversity: the extreme example of a $z = 1$ would denote a maximum biodiversity, with each individual belonging to a different species⁵. Obviously in real ecosystems z is always less than one and the growth of the number of species with the area is sublinear. Depending on the specific ecosystem, the typical values of z range from 0.2 to 0.4. It has been advocated [16] [17] that this relation is caused by a self-similar distribution of the abundance of species, meaning that the value of the exponent z can be derived assuming that the probability that a species living in an area A is also living in an area $A/2$ is independent of A . But again this lack the explanation of the dynamical origin of the species area relation. Only recently there has been some attempt to try to explain the SAR with dynamical models. In fact, curves similar to the experimental ones have been numerically observed in simple specially-explicit lattice models, such as the voter model with speciation [15]. Despite the simplicity of those models, it is very hard to find an analytic expression for the exponent z . We will deal with this problems in Chap.3, where we will find a numerical value for z for the voter model with speciation, along with a scaling relation linking RSA and SAR.

1.6 Measuring Spatial Turnover of Species

A meaningful suggestion on how to measure β -diversity in tropical forests comes from [18] (but see also [19][20][21]). The idea is to measure a function analogous to a 2-point correlation function. The exact definition is as follows: let’s choose two trees at a distance r apart each other; we are interested in the probability that this two trees are of the same species. We will denote this function by $F(r)$.

The first set of data with which we will calculate this function come from the tropical rainforests of Barro Colorado Island (BCI), Panama, and Yasuni, Peru, see Fig1.4. Both these location have a large plot censused completely with an area of 50 hectares for BCI (1Km x 500m) and 25 hectares

⁵ Assuming, as we have shown before, that the number of the individuals grows linearly with the area.

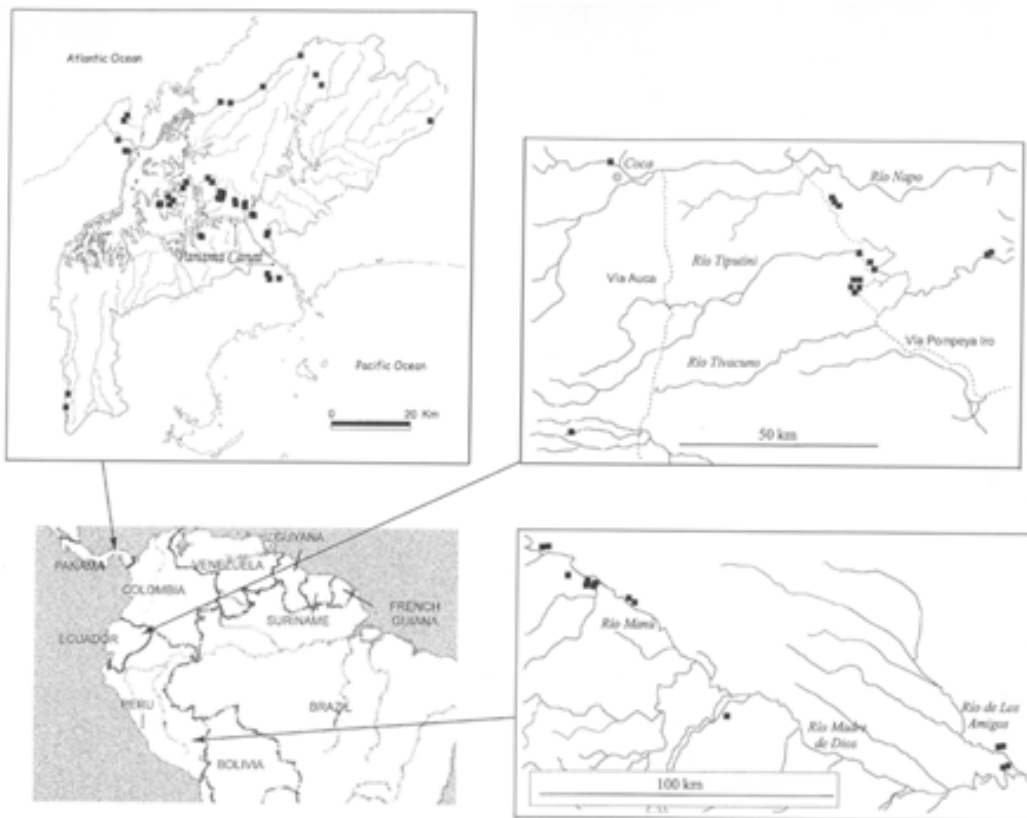


Figure 1.4: Map of the location of some of the tropical forest plots censused from the CTFS. In the present work we will use the data from The BCI plot, Panama, and Yasuni plot, Peru. Figure taken from the Supplementary material of [18]

(500m x 500m) for Yasuni. In particular the plot of 50 hectares of BCI was established in 1980 in the tropical moist forest of Barro Colorado Island in Gatun lake in central Panama. Five complete censuses (1981-1983, 1985, 1990, 1995, 2000, the sixth is underway) of the forest trees has been carried out. In each census, all free-standing woody stems of at least 10 mm diameter at breast height were identified, tagged, and mapped. About 350,000 individual trees have been censused over 20 years [22]. Then, both the locations have many plots of 1 hectare area at various distances, all completely censused.

One has to use two distinct procedures for the calculations of the 2-

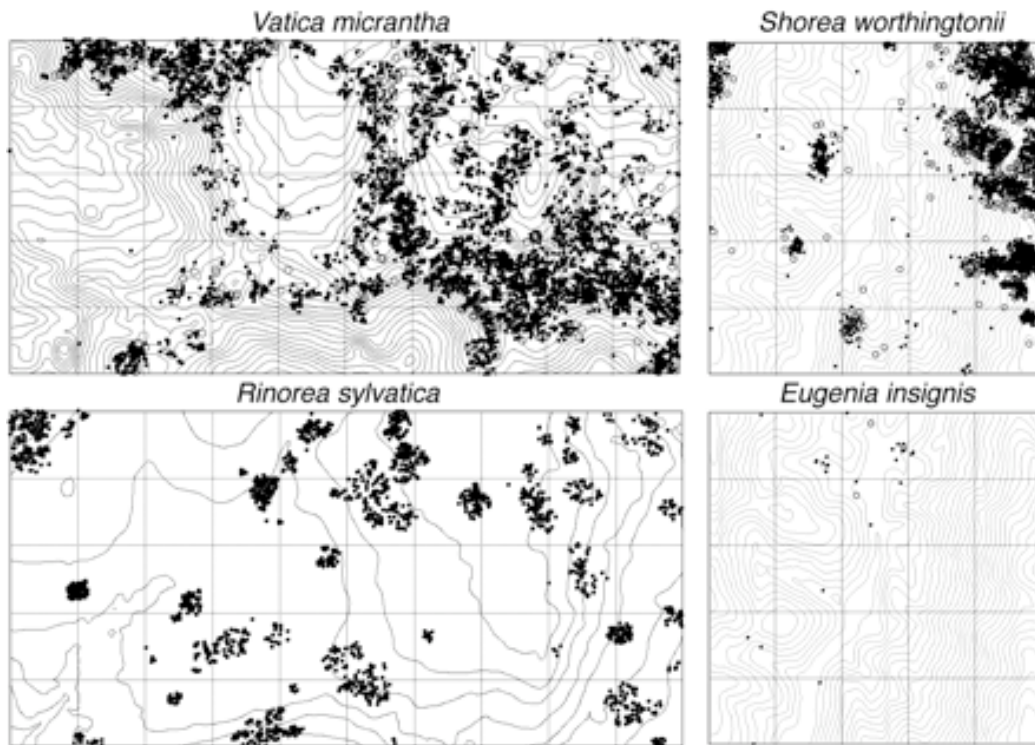


Figure 1.5: Some of the typical spatial patterns in the forest plot of BCI. Picture taken from [19]

point correlation function (from now on “the β -diversity function”). The first deals with distances smaller than the linear dimension of the biggest plot of the location: one counts all the pairs of trees at a distance r apart in said plot, and computes how many of them are conspecific. This accounts for the left part of Fig.1.6 and 1.7. For distances r greater than the linear extension of the biggest plots one takes two 1-ha plots at the chosen distance⁶ and for each plot (let’s call them A and B) and for each species present in at least one of the plots compute the density of the species in the considered plot⁷. Let’s denote this density by f_i^P where $P = A, B$ is the plot and $i = 1, \dots, N$ is the species. Then the β -diversity function $F(r)$ at the distance r (the probability that a tree randomly chosen from A and

⁶ Assumed much greater than 100m, the linear dimension of the smaller plots.

⁷ The density of a species in a plot is the number of the individuals of that species present in the plot divided by the total number of individuals in the plot.

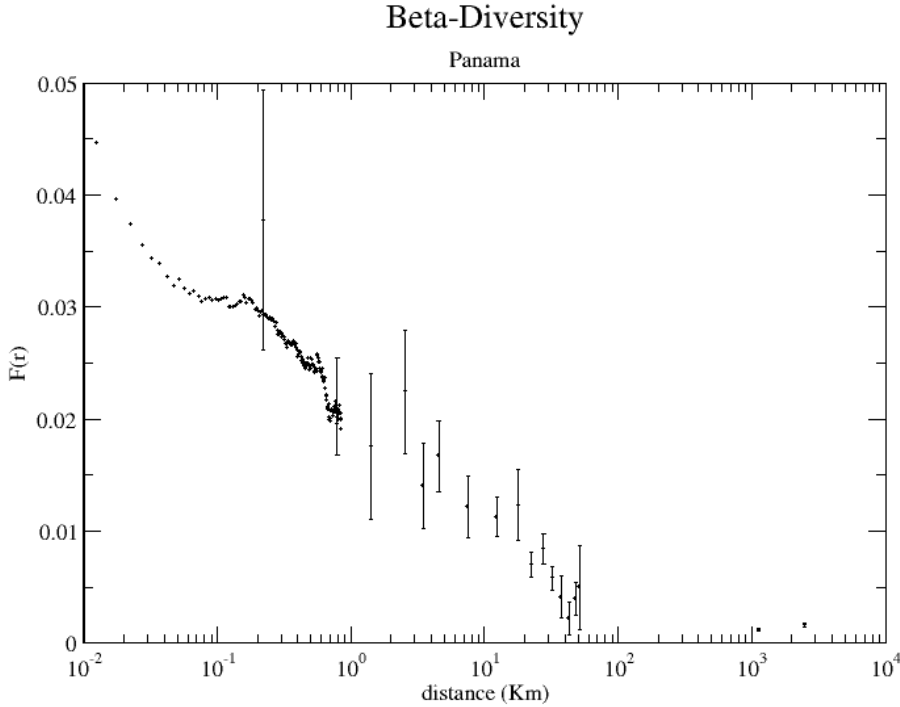


Figure 1.6: 2-point correlation function calculated for the tropical rainforest of Barro Colorado Island (BCI), Panama. The points at shorter distance are from the single 50-ha plot, while the ones at greater distance (with error bars) are calculated by the similarity of several 1-ha plots. Data from [18].

one randomly chosen from B are conspecific) is:

$$F(r) = \sum_{i=1}^N f_i^A f_i^B \quad (1.6)$$

This methods permit to measure the β -diversity function up to 2500Km⁸. The resulting data are shown in Fig.1.6 and 1.7. Remark that by definition we have $F(0) = 1$, due to the obvious fact that a tree is always conspecific with himself.

⁸ But one could wonder if this function has really some significance at such big distance.

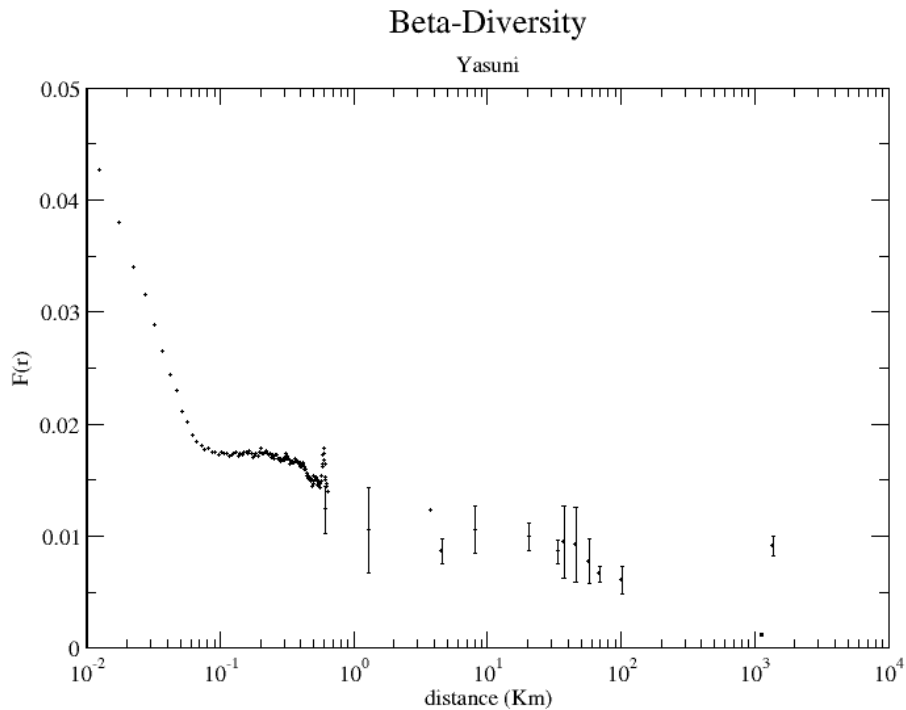


Figure 1.7: 2-point correlation function calculated for the tropical rainforest of Yasuni, Peru. The points at shorter distance are from the single 25-ha plot, while the ones at greater distance (with error bars) are calculated by the similarity of several 1-ha plots. Data from [18].

A first attempt to fit these curves has been carried out in the original article [18] and developed in [23], but the theories used were unable to fit completely the curve, lacking the features at very short distance, while the features at greater distance were correctly fitted by assuming a diffusion of the seeds with a kernel with finite mean dispersal distance. In Chap.3 we will derive a theory capable to fit correctly the data and explain the behavior of the curve.

In Chap.4 we will use a species-specific β -diversity, i.e. we will need the probability that two trees at a distance r are conspecific *and* belonging to a given species X . We will need also the probability that, of a couple of trees randomly chosen at distance r , one is of species X and the other is

of species Y . We then generalize the β -diversity function to $F_{i,j}(r)$, when the indices i and j indicate the chosen species for the two trees. It is also plainly obvious that:

$$F_{i,j}(0) = \delta_{i,j} f_i \quad (1.7)$$

where f_i is the density of species i in the considered plot and $\delta_{i,j}$ is the Kronecker delta function. With these definitions we will analyze the data from our second dataset, the intertidal vegetation [24] [25] [26] of venetian salt marshes [27] [28]. These data have been collected from a Compact Airborne Spectrographic Imager (CASI [30], characterized by 15 bands in the visible and near-infrared part of the spectrum, 1.3m resolution) and from a Multispectral Scanner (MS, 3 visible and 1 infrared bands, 0.5m geometrical resolution). Such data are part of a larger dataset acquired within the TIDE project [3]. The data were substantially interpreted using nearly-simultaneous ground truth observation of species presence within selected reference areas, whose location was accurately determined (with accuracy ± 1 cm) [29]. More details can be found also in [31] and [32].

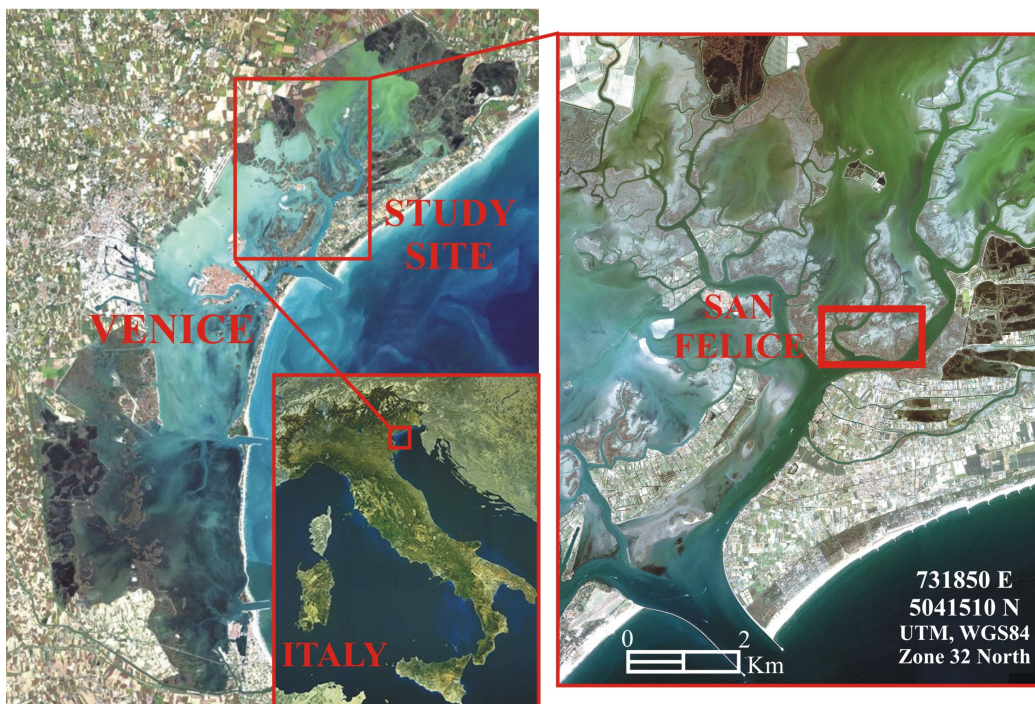


Figure 1.8: The site (barena di S.Felice) from we take the data about the salt marshes and intertidal vegetation. Courtesy of TIDE project.



Figure 1.9: Image of the S.Felice salt marsh taken from satellite. Courtesy of TIDE project.



Figure 1.10: Same image as before, but with the classification of the vegetation superimposed. Legend: maroon: soil, green: *spartina maritima*; yellow: *sarcocornia fruticosa*, pink: *limonium narbonense*. Courtesy of TIDE project.

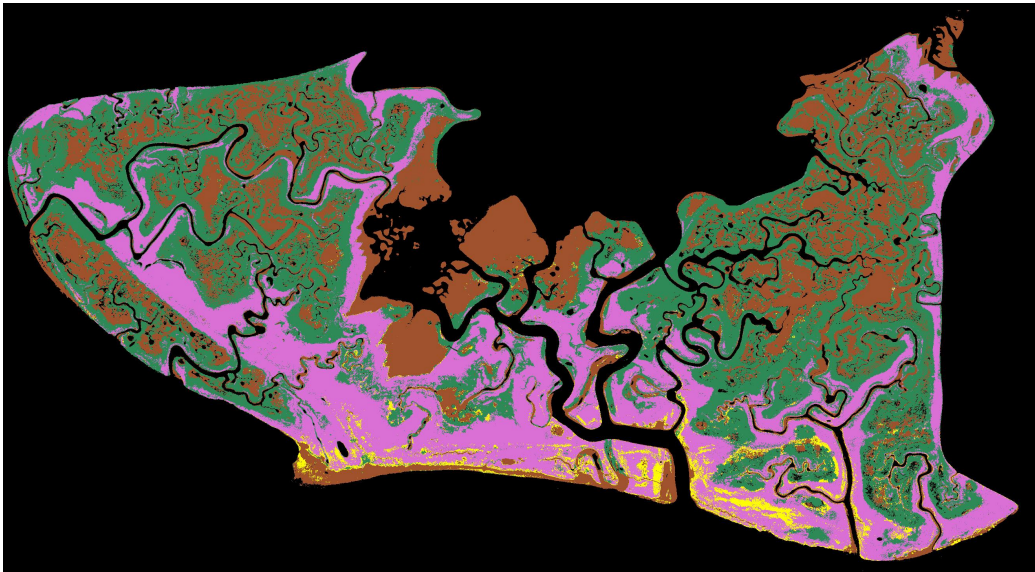


Figure 1.11: The actual part of the salt marsh used for the analysis. This is the maximum part of the scanned section outlined only by natural boundaries (i.e. not cropped by the scanner borders), with the exclusion of a section in which the classification was dubious for the presence of both water and plants. Legend: black: water; maroon: soil, green: *spartina maritima*; yellow: *sarcocornia fruticosa*, pink: *limonium narbonense*. Courtesy of TIDE project.

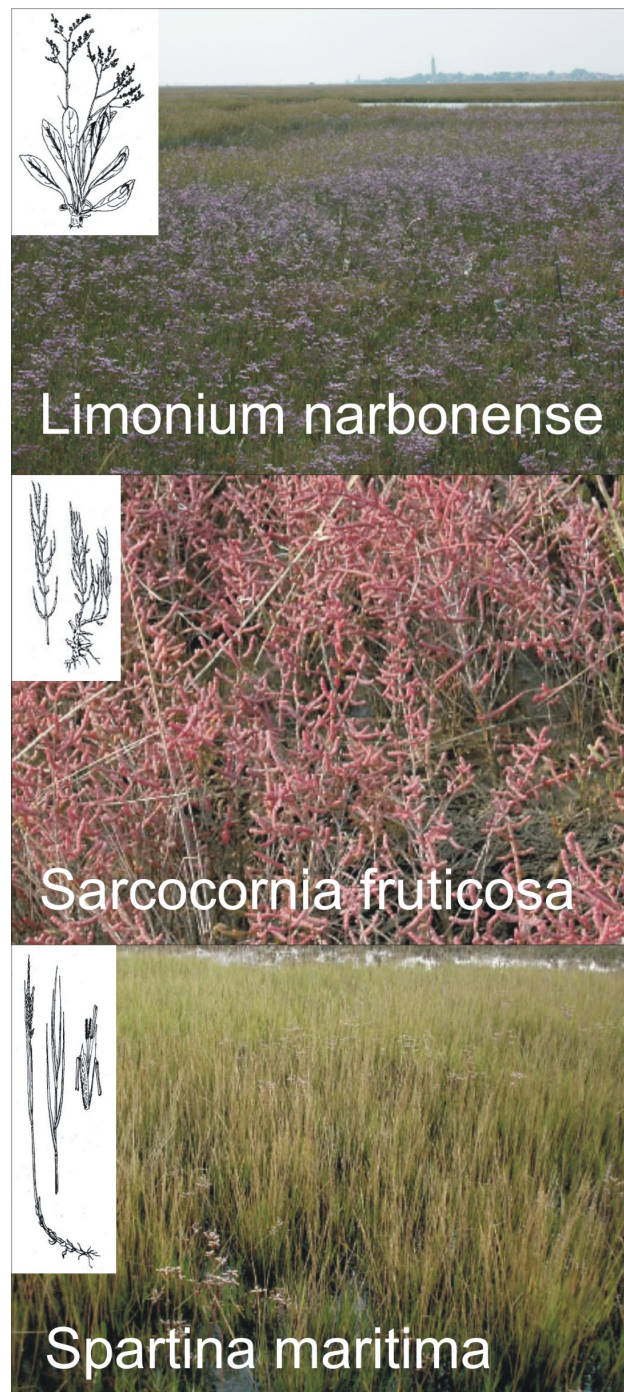


Figure 1.12: Samples of the vegetation present in the analyzed portion of salt marsh. Courtesy of TIDE project.

Chapter 2

Modeling Ecosystems

An ideal world is left as an exercise to the reader.
(Paul Graham, “On Lisp”)

In this chapter we will give a brief outline of the more recent concepts of population ecology from the mathematical side. It is beyond of the purpose of this work to give a complete survey of all the work done in this ever-increasing topic: we will sketch only the most relevant aspect relating to our research.

2.1 Typologies of Models

In the attempt of write a meaningful ecological model, many ways have been explored. The various proposals of models can be roughly divided in few categories. Here below we will examine briefly two of the main dichotomies in the formulation of the models: the stochastic-deterministic one and the discrete-continuous one. As an example of a deterministic continuous model we will present the Lotka-Volterra model of predation. These examples are intended to show how different will be our approach of modeling respect to the ‘classical’ one.

2.1.1 The Lotka-Volterra Model of Predation.

One of the first ecological mathematical models, and by far the most famous, is the Lotka-Volterra prey-predator model [33][34]. This is the first

known model that address the issue of the existence of a stable equilibrium situation in an ecosystem. The model consists in a system of two differential equations:

$$\begin{aligned}\dot{x} &= x(a - by) \\ \dot{y} &= y(-c + dx)\end{aligned}\tag{2.1}$$

with $a, b, c, d > 0$, x represent the number of preys, y the number of predators. This system has been widely studied, and display many non-trivial feature, such as periodic oscillation around the fixed point $x = c/d$, $y = a/b$, which is marginally stable. Similar oscillation has also been observed in real systems (see for example [35])¹. the Lotka-Volterra model can be extended to more species and more complicate interactions, but this is beyond the scope of this work.

2.1.2 Stochastic vs Deterministic Modeling

The first characteristic of the Lotka-Volterra model that one remarks is that it's perfectly deterministic: it does not allow for stochastic environmental variation nor implements any noisy behavior. In fact the role of noise in ecological systems has just begun to be recognized. One of the consequence of the strict determinism of the Lotka-Volterra model is that any extinction is impossible in it (see below). Indeed, experimental data on prey-predator systems show an oscillatory behavior with nearly constant period, as predicted, but the data themselves are very noisy. In the following section we will rather adopt the opposite point of view: the noise will be the central characteristic of our models. The very dynamic of them will be a stochastical one, and the solutions that we will find will be solution in the statistical sense.

¹ It's interesting also to remark that the purpose of Volterra was not to find a model able to reproduce the oscillations, which he has not observed. His aim was, instead, to explain the variation of the population of fishes in the Trieste gulf after WWII: it happened that the war reduced greatly fishery activities and so it was natural to expect that the population of fishes was grown in the meantime; instead it diminished seriously. Now, fishery in the Volterra model act like the parameter c : a suppressor of predators (it is assumed negligible in the prey species when compared to predation). At equilibrium the parameter c is irrelevant for what concerns the predators, but affect greatly the preys: if we diminish fishery (i.e. we have a smaller c) the prey species diminish accordingly. So the model was successful to explain a paradoxical situation.

2.1.3 Discrete vs Continuous Modeling

The second major problem of the Lotka-Volterra model is the fact that the population is expressed as a real number and not an integer one. This causes all sort of problems when dealing with the analytical solutions of the system, especially when generalized to more than two species. In particular it should be highlighted the fact that, being the axes ($x = 0$ and $y = 0$) an invariant set under the dynamic Eq.(2.1), then no solution starting with both x and y greater than zero can ever cross one of the axes. In physical terms this means that in the Lotka-Volterra model the extinction is impossible. This is linked to the problem of how to interpret exactly x and y , or, more precisely, how the numerical values of x and y translates in actual population values. Due to the fact that in Nature all populations are composed of individuals, one should recognize that it is correct to represent a population with a real number only if the population is very large and so the corrections associated to the discreteness are negligible. But when the population become very low, then one should use a discrete variable, so to allow for the extinction of species. A correction in this direction for the Lotka-Volterra model will not be attempted in this work. We simply remark that all the models that we will develop later will use discrete values for the population variables.

2.2 Island Biogeography

One of the most important work in the modern study of ecology is the book of MacArthur and Wilson on the island biogeography [7]. In this work they provide a simple theory for the fact that the number of different species coexisting in a given island has a functional dependence from various parameters like the distance from the mainland, the area of the island, and so on. To be more specific, an island of greater area and nearer to the coast host generally more species than a smaller one, or a more distant one from the mainland. To explain this dependence, MacArthur and Wilson postulated that the number of species in the island varies as an effect of immigration and local extinction, and that the rates of these two effects in a given island depends only on the number of the species present at that moment on the island itself. In particular, one assumes that the rate of im-

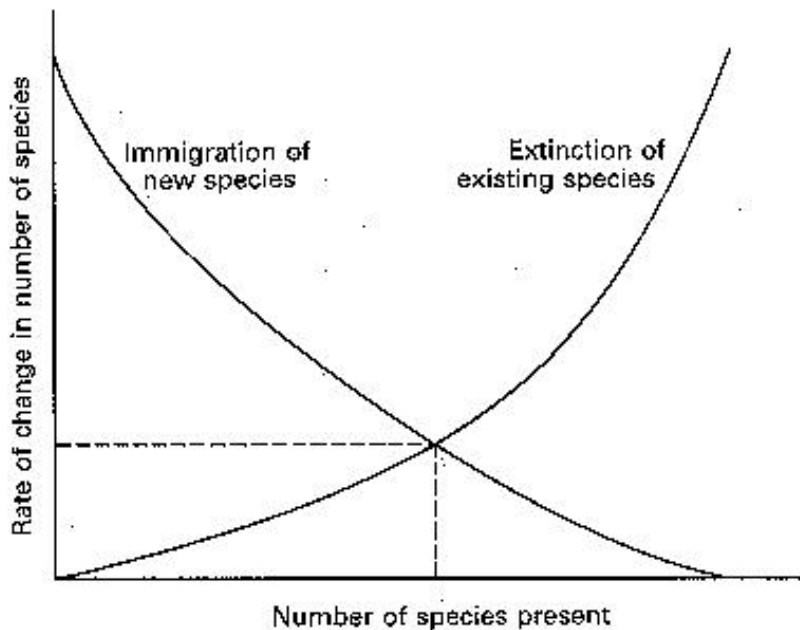


Figure 2.1: The basic graphical model of equilibrium in the MacArthur-Wilson model. Picture taken from [8].

migration is a decreasing function of the number of species, because of the difficulties to settle in an overcrowded environment. On the other hand, the extinction rate will be an increasing function of the number of species, because the more crowded is the island, the more likely is that a species become extinct². Then, obviously, there exist a stable equilibrium when the immigration rate and the extinction rate are of the same magnitude, see Fig 2.1. This theory is fairly general, not requiring a specific shape of the curves, but only the fact that they are monotonic. The exact shape of the curves can be influenced by the environmental conditions. For instance, a larger island will have a lower extinction rate, and an island nearer to the mainland will have a greater immigration rate (see Fig.2.2). With these assumptions, the position of the equilibrium point varies qualitatively as expected.

² Equivalently, if the island can support only a fixed number of individual, then more species means less individuals for each species on average, and it's very well known that rarer species are in greater danger of extinction.

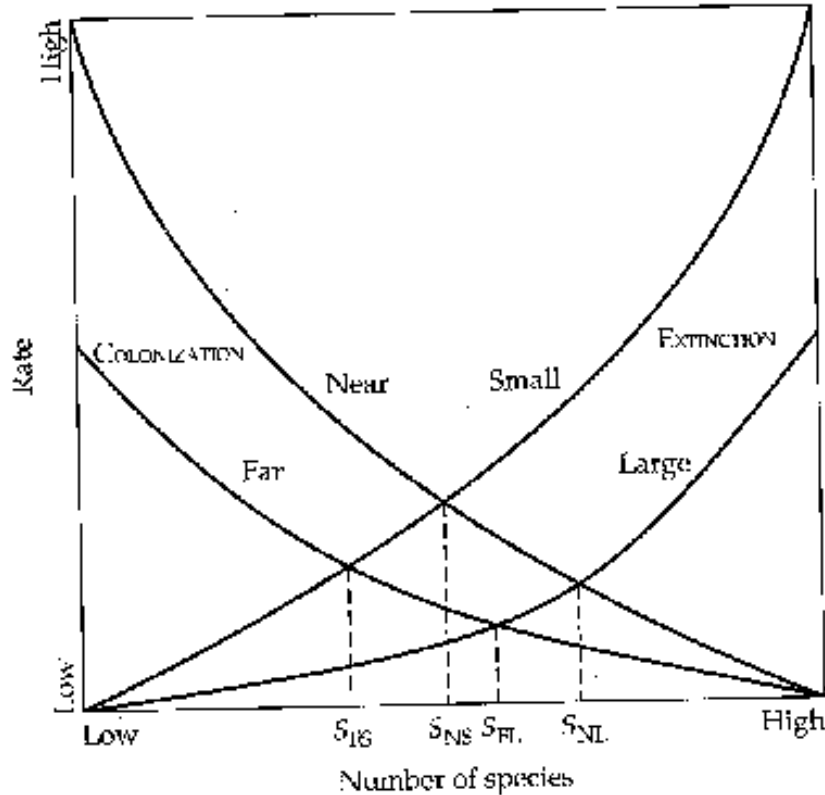


Figure 2.2: Multiple immigration and extinction curves indicating effects of differences in size and isolation on equilibria. Picture taken from [8].

What it's really important about this model (and, quite surprisingly, the authors themselves didn't stress in the original publication) is the fact that the final number of species is a *dynamical* equilibrium. At the time of the publication (1967) this was a revolutionary approach to modeling ecosystems. Let's also remark that the immigration and extinction rate depends only on the number of species present in the island, not their abundance or what kind of species they are. This means that in the theory of island biogeography one can substitute a species with another one with completely different habits and characteristics without changing the behavior of the system. This is an important point, and we devote to it the next section.

2.3 Neutrality

Hypothesis is a toll which can cause trouble if not used properly. We must be ready to abandon our hypothesis as soon as it is shown to be inconsistent with the facts.

(W.I.B.Beveridge)

The concept of neutrality was developed in the field of molecular evolution by Kimura [36] to provide an explanation to mutations that are evolutionarily neutral (i.e. that provide no particular advantage or disadvantage in the genetic pool). According to this theory, many patterns of molecular evolution can be explained postulating that the overwhelming majority of possible mutations at molecular level don't bring a relevant change in the fitness of the molecule (or the fitness of the organism to which she belongs). Or, to put it in other words, the original molecule and the mutated one performs the same work with the same efficiency. This approach is not antievolutionistic, in fact quite the opposite: in neutral theory the evolutionarily relevant mutations aren't ruled out, only they are recognized to be extremely rare. The key idea is that most of the time the genes perform a sort of random walk in the space of sequence. It may also be that neutral mutations become relevant ones due to the changing environment.

The neutral assumptions was extended, in some sense, to the field of ecology by Hubbell [9]. While in the theory of Kimura, most of the mutations were functionally equivalent, Hubbell now assumes that *every individual in the same trophic level is functionally equivalent (i.e. undergoes the same dynamical rules) regardless of the species it belongs to*. It is important to stress that the neutral theory applies at the individuals level and not at the species level: the functional equivalence is enforced on a per capita basis. This means that a rare species may undergo a different dynamic than a more abundant one, provided that the rules that specifies the dynamics (that can depend on the abundance of the species) are the same for all the species.³

³ It is worth to report here the original definition of Hubbell: "The essential defining characteristic of a neutral theory in ecology is [...] the complete identity of the ecological interaction rules affecting all organism on a per capita basis". This means that in neutral models the different species have no differential advantage i.e. species cannot be selected

The assumption of individual equivalence means that we are neglecting fitness differences of the various species against the stochasticity of the demographic process, by which we mean the randomness associated with the birth and death of a species. This is obviously only a first approximation and not a complete theory of demographic processes, but can be useful as a null-model i.e. as an aid to interpret data [37]. In fact, we should note that the assumption of ecological equivalence gives no hint about what the dynamical rules precisely are, but only that the rules are the same for all species. Some authors [38][39] claim that there are no direct evidence of neutrality in real ecosystems. This comment in fact misses the point: the neutrality assumption is just a zero-order hypothesis which has the relevant advantage that leads often to analytically tractable models. Furthermore, models obeying the neutral hypothesis often compare surprisingly well with experimental data, as we will see in the following chapters. Last, but not least, the law of constant density show that, though neutrality can be not strictly true, there are ecosystem that behave as if individuals of different species would be equivalent. The debate about the usefulness of Neutral Theory is still alive [40] [41].

2.4 The Unified Neutral Theory of Biodiversity - The Drift Model

In his book about the Neutral Theory [9], after having stated the neutral hypothesis, Hubbell presents an implementation of it. While this model is very simple to state and define precisely, its solution is a bit involved, and was given later [42] [43]. This model is known as “unified neutral theory of biodiversity”, or, more simple, as “Hubbell’s drift model”⁴. the model is defined as follows: let us consider a community with a fixed number J of individuals, ‘surrounded’ by a larger community, that we will call a metacommunity, as defined in the previous chapter, in which the relative abundance of the species are kept fixed. At each discrete time step one randomly chosen individual dies, leaving an empty slot. With a probability $1 - m$, m being a parameter of the model, the empty slot is filled

in an evolutionary manner, not existing any relevant difference between them.

⁴ Obviously, other models have been proposed along these lines. See [44]

with the offspring of an random individual belonging to the community in that moment. With a probability m there is an immigration event: the individual is replaced by another one belonging to a species present in the metacommunity, with probability proportional to the abundance of that species in the metacommunity itself (that we assume being distributed according to a Fisher Logseries, Eq.(1.3); we will justify this assumption a few lines below).

The natural mathematical framework for such a model is the theory of stochastic processes. It can be easily shown that the probability that the k -th species has n individuals evolves with time according to the following birth and death master equation:

$$\dot{P}_{n,k}(t) = d_{n+1,k}P_{n+1,k}(t) + b_{n+1,k}P_{n+1,k}(t) - (b_{n,k} + d_{n,k})P_{n,k}(t) \quad (2.2)$$

where $b_{n,k}$ and $d_{n,k}$ are respectively the birth and the death rates. It is interesting to remark that, if we pose:

$$\begin{aligned} b_{n,k} &= bn \\ d_{n,k} &= dn \end{aligned} \quad (2.3)$$

with b and d constant, then the equilibrium solution of the master equation is the Fisher Logseries, Eq.(1.3), with $\alpha = b/d$, from which we can infer that $b < d$. This is the dynamic that we expect at continental level (i.e. metacommunity level), and so the metacommunity of our model will have abundances distributed accordingly.

The drift model is then defined by the master equation Eq.(2.2) and the following birth and death rates:

$$\begin{aligned} b_{n,k} &= (1-m)\frac{n}{J}\frac{J-n}{J-1} + m\frac{\mu_k}{J_M}\left(1 - \frac{n}{J}\right) \\ d_{n,k} &= (1-m)\frac{n}{J}\frac{J-n}{J-1} + m\left(1 - \frac{\mu_k}{J_M}\right)\frac{n}{J} \end{aligned} \quad (2.4)$$

where μ_k is the abundance of the k -th species in the metacommunity and J_M is the total number of individuals in the metacommunity. Remark that, when $m = 0$ and $j \ll J$ then we have again Eq.(2.3), with $b = d = 1/(1-J)$, and in fact this is the dynamic that we expect for the metacommunity. It is possible to find a stationary solution for this model by satisfying detailed

balance; and after averaging over μ one obtains a curve of abundance that can be compared with experimental data:

$$\langle \phi_n \rangle = \theta \frac{J!}{n!(J-n)!} \frac{\Gamma(\gamma)}{\Gamma(J+\gamma)} \int_0^\gamma \frac{\Gamma(n+\gamma)}{\Gamma(1+\gamma)} \frac{\Gamma(J-n+\gamma-y)}{\Gamma(\gamma-y)} e^{-y\theta/\gamma} dy \quad (2.5)$$

where $\langle \phi_n \rangle = \sum_k P_{n,k}^{(stat)}$ is the average number of species having n individuals, $\gamma = \frac{m(J-1)}{1-m}$, and θ is the biodiversity parameter, equivalent to Fisher's α (see [9] and [42] for details).

When we plot a fit of this solution on the experimental data from BCI and we compare it with a fit done with the Preston Lognormal (see Fig.2.3) one finds that it's hard to discard either one due to the experimental errors. But on the theoretical side, the neutral solution is on better ground: it is the equilibrium solution of a dynamical model; while the lognormal is hard to justify in these terms: as we said before the lognormal is the solution of a multiplicative random walk, but his variance should grow linearly with time, unlike the data.

Some final observation about the drift model:

- It is clear that this model is a generalized version of the theory of island biogeography but with the focus on the dynamic of the single individual instead of the dynamics of the species.
- This model does not implement spatial differences, apart from the distinction between community and metacommunity (which is well defined only for island or isolated places).

In the next chapter we will generalize a similar model with adding explicitly the space, we will solve it for the 2-point correlation function and we will find a scaling relation linking this function with SAR and RSA.

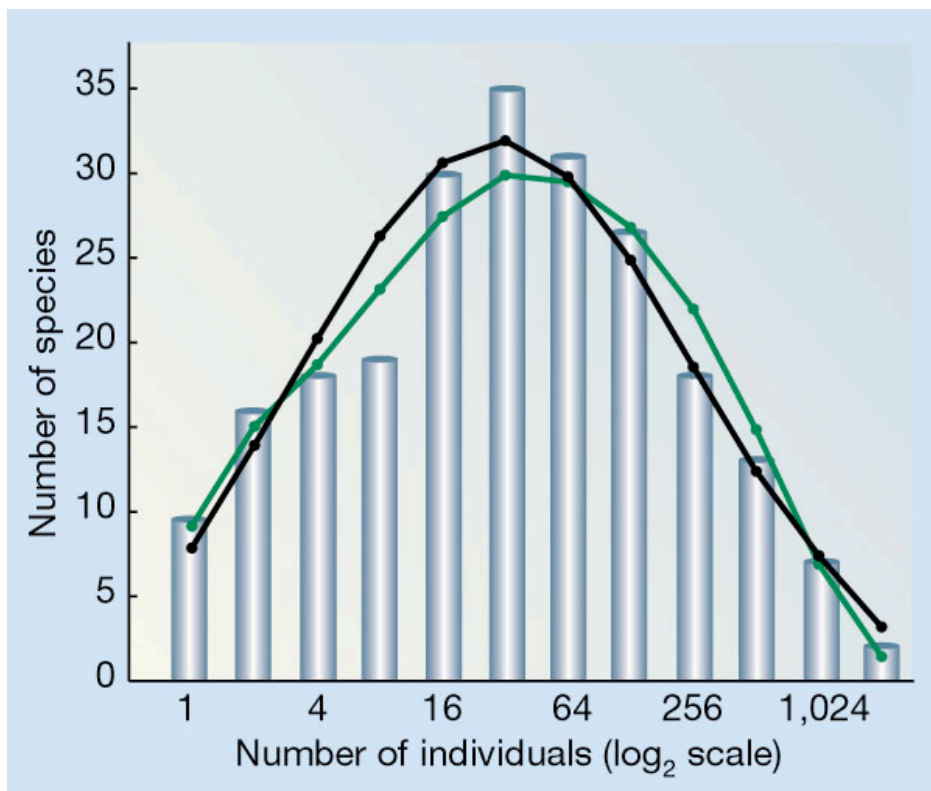


Figure 2.3: Data for the RSA from the BCI plot, along with a fit with the Preston Lognormal (black) and the analytic solution to the Hubbell's Drift Model (green). Notice that the difference between the two curves is minimal if compared with the noise of the data. Figure taken from [42]

Chapter 3

The Voter Model With Speciation

Justifying space exploration because we get non-stick frying pans is like justifying music because it is good exercise for the violinists right arm.

(Richard Dawkins)

3.1 Motivation

Spatial explicit models are much rarer in the literature, due both to the additional difficulty of taking into account a community that is different from point to point and to the past lack of data about the spatial turnover of species. With the purpose of providing a context, here below we list, without any claim of completeness, a couple of important breakthrough on the theoretical line of research that leads to the present work.

In 1996 Durrett e Levin published an article on the voter model model with speciation [15] which contained a numerical study about the behavior of the Species-Area Relationship. We will describe in detail this model below, and we will study it in depth, both with analytical tools and numerical simulations. We now remark only that in their article Durrett and Levin identify for the voter model with speciation a typical length of the order $1/\sqrt{\nu}$ where ν is the speciation rate. (see also [45], [46] [47])

More recently Leigh e Chave [23] have found an analytical solution of the voter model with speciation for an analogous of the 2-point correlation function. They found that this simple model, that implements only diffusion and speciation, can not explain the β -diversity data of the trop-

ical forest. Their solution is analogous to the one that we will give later, but the formalism of our solution is such to permit also the introduction of density-dependent processes, that we will found necessary to explain the real data from tropical forests.

The voter model was first defined by Liggett [48] without the speciation process, and in this form it has been thoroughly studied (see [49], [50], [51], [52], [53], [54], [55]), the main result being that in a finite lattice, or in an infinite lattice with dimensionality $d \leq 2$, after an initial transient the system attain monodominance, i.e. all the sites have the same label (which can represent the species of an individual, or an opinion in a voting system, etc.); and it is clear that such a system in monodominance is perfectly stable. Instead, in an infinite lattice with dimensionality $d > 2$ there is the coexistence of many labels without a single one being dominant over the others¹. Obviously, the presence of the speciation process change completely the model, and now the monodominance is possible only when the lattice is much smaller compared with the typical length imposed by speciation, and even then it is not a stable state but can be invaded by another species introduced by the speciation process. All these statement will be made precise in the following sections. For further details about the voter model with speciation see [56] [57] [58]. The work described in this chapter has been published in [59]

3.2 Definition

Let we have a d -dimensional cubic lattice, with side L . At each side there is a variable holding a label, that specifies the species (or the opinion, if the model is meant to represent a voting system) that is present in the site (one and only one species/opinion is availed at each site at the same time). We can start with one of the many possible configuration at will (one can show that the initial configuration is unimportant for our goals), then at each time step we update the system as it follows:

1. We choose a site at random (this simulate a tree's death or a change of opinion in the site).

¹ As can be easily shown, (see section 3.3.2) the relative abundances of the various labels in these conditions are constant over time.

2. With probability ν we put on the site a label which is not already present on the system (this simulate a speciation/immigration event, or a new opinion/candidate arising).
3. With probability $1 - \nu$ we replace the label with the label of a random chosen neighbor (this simulate a reproduction event, or an opinion spreading).

Some worthwhile notes:

- we define “time step” as the update of a single site; “whole update” as the update of L^d sites, which correspond to a “mean” update of the whole lattice.
- This model is neutral in the sense of Hubbell [9], as we have described it in the previous chapter. Then the only way that we have to distinguish the different species/opinion is the label that we will give them.
- A consequence of neutrality is that if we can solve the dynamics of a single species/opinion, then we have information about the whole system, being all the species/opinions equivalent. This is a very useful feature from a theoretical point of view, and we’ll make extensive use of it.
- Each site of our model is always occupied i.e. at no time there can be an empty site. This modelize the law of constant density that we have stated in the first chapter.
- Here we have defined the model on a d -dimensional cubic lattice for simplicity, but the model can be defined on an arbitrary lattice. Indeed, the simulations in $d = 2$ has been carried on on the triangular lattice to eliminate anisotropies. The quantities on which we are interested are rather insensitive to the precise underlying lattice, as we have demonstrated numerically.
- The neighbors of a site can be defined in several ways, but the most common choice is to define as neighbours all the sites within a given distance R_1 . In the majority of our calculations and simulations we

will use $R_1 = 1$. In this case in the square lattice we are left with the freedom to use one of the following two definitions: the neighbours are the four sites that are adjacent to a given site in the horizontal or vertical direction, or the eight sites adjacent to a given site in horizontal, vertical and diagonal direction (Von Neumann neighbors). Both choices lead to identical results, but the former is easier manageable in theoretical development, while the latter leads to a faster equilibrium in simulations. In the triangular lattice the nearest neighbours are always the six sites adjacent to a given site.

3.3 Master Equation Formulation and Solution

3.3.1 Lattice Master Equation

We now focalize upon a single chosen species/opinion's dynamics. To do it, we rewrite the labels in the following way:

- We put the label 1 if on the site there is the chosen species/opinion.
- We put the label 0 if on the site there is another species.

We will call this model the "Single Species Voter Model With Speciation".

Now, let $S_{\vec{x}}^t$ be the value of the label at site \vec{x} and at time t in a given realization of the model. Let us fix the initial conditions i.e. $S_{\vec{x}}^0$. Then, define:

$$P_{\vec{x}}^t \equiv \langle S_{\vec{x}}^t \rangle \quad (3.1)$$

where brackets denote the average over different realization of the system with the same initial conditions (remark $P_{\vec{x}}^0 = S_{\vec{x}}^0$).

At this point we can write the master equation² of this process, in discrete space-time:

$$P_{\vec{x}}^{t+1}(S_{\vec{x}} = 1) = P_{\vec{x}}^t(S_{\vec{x}} = 1) + \frac{1}{N} \left(- P_{\vec{x}}^t(S_{\vec{x}} = 1) \right) \quad (3.2)$$

² For the theory of stochastic processes, and master equations in particular, the reader is referred to [61] [62] [63] [64] [65] [66]

$$+ \frac{1-\nu}{2d} \sum_{\vec{y}(\vec{x})} P^t(S_{\vec{x}} = 1, S_{\vec{y}} = 1) \quad (3.3)$$

$$+ \frac{1-\nu}{2d} \sum_{\vec{y}(\vec{x})} P^t(S_{\vec{x}} = 0, S_{\vec{y}} = 1) \quad (3.4)$$

where $\vec{y}(\vec{x})$ are the nearest neighbors of \vec{x} .

This simplifies in:

$$P_{\vec{x}}^{t+1} = P_{\vec{x}}^t \left(1 - \frac{1}{N}\right) + \frac{1-\nu}{2dN} \sum_{\mu=1}^d (P_{\vec{x}+\mu}^t + P_{\vec{x}-\mu}^t) \quad (3.5)$$

where $N = L^d$, the sum is over the canonical orthonormal base and the time unit is a single time step. We can easily transform this equation in a continuous-time one:

$$\frac{\partial P_{\vec{x}}^t}{\partial t} = -P_{\vec{x}}^t + \frac{1-\nu}{2d} \sum_{\mu=1}^d (P_{\vec{x}+\mu}^t + P_{\vec{x}-\mu}^t) \quad (3.6)$$

where the time unit is a whole update. This equation can be rewritten in the following form:

$$\frac{\partial P_{\vec{x}}^t}{\partial t} = \frac{1-\nu}{2d} \Delta_{\vec{x}} P_{\vec{x}}^t - \nu P_{\vec{x}}^t \quad (3.7)$$

where $\Delta_{\vec{x}}$ is the discrete laplacian.

3.3.2 Lattice Solution and Total Probability

This equation can be solved for $P_{\vec{x}}^t$ in two ways: by the use of a generating function or by Fourier transform. Both procedures are performed in details in Appendix A. The solution that one obtains is

$$P_{\vec{x}}^t = e^{-t} \sum_{\vec{y}} S_{\vec{y}}^0 I_{\vec{x}-\vec{y}} \left(\frac{1-\nu}{d} t \right) \quad (3.8)$$

where we have defined:

$$I_{\vec{x}}(t) \equiv \prod_{i=1}^d I_{x_i}(t) \quad (3.9)$$

and $I_a(x)$ is the modified Bessel³ function I of order a .

³ For the theory of Bessel functions the reader is referred to [67] and [68].

It's interesting to calculate the behaviour of the quantity $\langle n \rangle_t = \sum_{\vec{x}} P_{\vec{x}}^t$ (mean number of individual belonging to a single species) respect to time.

To calculate it, we recall that [60]:

$$\sum_{j=-\infty}^{+\infty} a^j I_j(t) = e^{\frac{1}{2}(a+a^{-1})t} \quad (3.10)$$

then, putting $a = 1$ we have:

$$\sum_{j=-\infty}^{+\infty} I_j(t) = e^t \quad (3.11)$$

Then:

$$\langle n \rangle_t = \langle n \rangle_{t=0} e^{-\nu t} \quad (3.12)$$

Remark that for $\nu = 0$ the number of individual is conserved (in mean). In practice, the fluctuations around that average value are so prominent that in all finite lattices, and in all infinite lattices with dimensionality $d \leq 2$, the dynamics end in a state of monodominance (consensus), this meaning that all the sites of the lattice share the same species/opinion.

When $\nu \neq 0$, this last equation assures that on average all species become extinct, meaning that even in a monodominance situation there is a finite probability that a speciation event is able to invade the system.

3.3.3 Continuum Master Equation and Solution

Now we want to write a continuous-space equation for our model. With this purpose we define the probability density function:

$$P(\vec{x}, t) \equiv P_{\vec{x}}^t a^{-d} \quad (3.13)$$

where a is the lattice spacing and $a \rightarrow 0$. In this limit we have also:

$$\Delta_{\vec{x}} = a^2 \nabla^2 \quad (3.14)$$

where ∇^2 is the Laplacian operator in the continuum.

In the continuum limit we have:

$$\frac{\partial P}{\partial t}(\vec{x}, t) = \nabla^2 P(\vec{x}, t) - \gamma^2 P(\vec{x}, t) \quad (3.15)$$

where $\gamma^2 = 2d\nu/a^2$ and the time has been rescaled $t \rightarrow 2dt/a^2$ (in other words, 1 time unit is now $2d/a^2$ complete updates).

To solve this equation we use again Fourier transforms.

$$\frac{\partial \hat{P}}{\partial t}(\vec{p}, t) = - (p^2 + \gamma^2) \hat{P}(\vec{p}, t) \quad (3.16)$$

Then:

$$\hat{P}(\vec{p}, t) = \hat{P}(\vec{p}, 0) e^{-(p^2 + \gamma^2)t} \quad (3.17)$$

Antitransforming we obtain the solution:

$$P(\vec{x}, t) = e^{-\gamma^2 t} \int d^d y P(\vec{x} - \vec{y}, 0) \frac{e^{-\frac{y^2}{4t}}}{(4\pi t)^{d/2}} \quad (3.18)$$

It's interesting to remark that:

$$\int d^d x P(\vec{x}, t) = e^{-\gamma^2 t} \int d^d x P(\vec{x}, 0) \quad (3.19)$$

that is a result analogous to the one stated before in the discrete case.

3.4 β -diversity

3.4.1 Master Equation and Solution on the Lattice

As explained in Chapter 1, in ecology the term β -diversity refers to the spatial turnover of species in the landscape. A proposed measure of this quantity [18] is defined in this way: first we calculate the probability $F(r)$ that 2 randomly chosen trees at distance r are conspecific, then the quantity $-\frac{\partial F}{\partial r}(r)$ is a measure of β -diversity.

Our aim now is to write a master equation for $F(r)$ and to solve it. To fulfill this goal we begin calculating the equation for the probability $P_{\vec{x}, \vec{y}}^n$ that at the sites \vec{x} and \vec{y} there is the same species. We can fix *a priori* an important boundary condition: $P_{\vec{x}, \vec{x}}^n = 1$. So, the master equation for this quantity is, when $\vec{x} \neq \vec{y}$:

$$P_{\vec{x}, \vec{y}}^{n+1} = P_{\vec{x}, \vec{y}}^n \left(1 - \frac{2}{N}\right) + \frac{1 - \nu}{2dN} \sum_{\mu=1}^d \left(P_{\vec{x}+\mu, \vec{y}}^n + P_{\vec{x}-\mu, \vec{y}}^n + P_{\vec{x}, \vec{y}+\mu}^n + P_{\vec{x}, \vec{y}-\mu}^n \right) \quad (3.20)$$

Then we impose traslational invariance (i.e. $P_{\vec{x},\vec{y}}^n = F_{\vec{x}-\vec{y}}^n$) so we can rewrite, for $\vec{x} \neq 0$:

$$F_{\vec{x}}^{n+1} = F_{\vec{x}}^n \left(1 - \frac{2}{N}\right) + \frac{1-\nu}{dN} \sum_{\mu=1}^d (F_{\vec{x}+\vec{\mu}}^n + F_{\vec{x}-\vec{\mu}}^n) \quad (3.21)$$

with the boundary condition $F_0^n = 1$.

This equation can be solved on the lattice with the aid of Fourier Transforms, as can be seen in details in Appendix A. The solution is:

$$F_{\vec{x}} = \left(\int_{-\pi < q_i < \pi} \frac{d^d \vec{q}}{(2\pi)^d} \frac{1}{l(\vec{q})} \right)^{-1} \int_{-\pi < p_i < \pi} \frac{d^d \vec{p}}{(2\pi)^d} e^{i\vec{p} \cdot \vec{x}} \frac{1}{l(\vec{p})} \quad (3.22)$$

where:

$$l(\vec{p}) = -\frac{2}{N} + 2\frac{1-\nu}{dN} \sum_{\mu=1}^d \cos(p_{\mu}) \quad (3.23)$$

It appears that this approach to the solution is suitable to give a closed form in 1 dimension. Again the detailed calculation are in Appendix A, and the final solution is:

$$F_r = e^{-r/\xi} \quad (3.24)$$

where:

$$\xi = \left(\ln \frac{1-\nu}{1-\sqrt{\nu(2-\nu)}} \right)^{-1} \quad (3.25)$$

is the typical lenght of the solution. We can remark the following facts:

1. $\xi > 0$ when $0 < \nu < 1$.
2. When $\nu \rightarrow 0$ we have $\xi \rightarrow \infty$ i.e. at equilibrium with vanishing speciation there is only one species that dominate a finite landscape.
3. When $\nu \rightarrow 1$ we have $\xi \rightarrow 0$ i.e. when the speciation is unity then reproduction is forbidden and $F_r = 0$ if $r \neq 0$.
4. When $0 < \nu < 1$ we have $\frac{d\xi}{d\nu} < 0$ i.e. when speciation lowers each single species tend to occupy more space (equivalently: to have more individuals). In a finite landscape this also mean that there are less species.

For $d > 2$ this approach does not give any closed form solution. With the goal of find an approach able to yield a closed form, and with the future perspective of implementing other effects in the model,

3.4.2 Solution in the Continuum

An easier way to calculate $F(r)$ is to transform the discrete model equation in a continuum one, and treat with the available tools for solving differential equations.

To simplify our job, we can write the lattice model Eq.(3.21) as:

$$F_{\vec{x}}^{n+1} - F_{\vec{x}}^n = (LF^n)_{\vec{x}} \quad \forall \vec{x} \quad (3.26)$$

where:

$$L_{\vec{x},\vec{y}} = \mathcal{L}_{\vec{x},\vec{y}} - \delta_{\vec{x},0} \mathcal{L}_{0,\vec{y}} \quad (3.27)$$

$$\mathcal{L}_{\vec{x},\vec{y}} = \begin{cases} -\frac{2}{N} & \text{if } \vec{x} = \vec{y} \\ \frac{1-\nu}{dN} & \text{if } |\vec{x} - \vec{y}| = a \\ 0 & \text{elsewhere} \end{cases} \quad (3.28)$$

Remark that in this formulation the boundary condition at the origin is now implemented in the model equation, and does not need to be imposed afterwards.

The probability distribution function (PDF) in the continuum is $F(\vec{x}, t) = F_{\vec{x}}^n a^{-d}$ with $t = n\tau$, τ being the microscopic time unit. Then we can state the following correspondences:

$$\begin{aligned} F_{\vec{x}}^{n+1} - F_{\vec{x}}^n &\rightarrow a^d \dot{F}(\vec{x}, t) \\ (LF^n)_{\vec{x}} &\rightarrow 2a^d \left(-\alpha F(\vec{x}, t) + D\nabla^2 F(\vec{x}, t) + \dots \right) - \delta_{\vec{x},0} \sum_{\vec{y}} \mathcal{L}_{0,\vec{y}} F_{\vec{y}}^n \\ \sum_{\vec{y}} \mathcal{L}_{0,\vec{y}} F_{\vec{y}}^n &= -\frac{2}{N} + \frac{1-\nu}{dN} \sum_{\vec{y}:|\vec{y}|=a} F_{\vec{y}}^n \\ &\xrightarrow{a \rightarrow 0} -\frac{2}{N} + 2\frac{1-\nu}{N} F_1^n \\ &\equiv -2\tau b \end{aligned} \quad (3.29)$$

where the quantity F_1 is the limit of $F_{\vec{x};|\vec{x}|=a}$ as $a \rightarrow 0$. We don't know this limit, so the resulting multiplicative parameter b is another free parameter of the model and can't be fixed like we have done for the solution on the lattice.

Performing all the substitutions and reorganizing the result we obtain:

$$\dot{F}(\vec{x}, t) = 2 \left(D\nabla^2 F(\vec{x}, t) - \alpha F(\vec{x}, t) \right) + 2b\delta^d(\vec{x}) \quad (3.30)$$

Imposing the stationarity ($\dot{F}(\vec{x}, t) = 0$), we can solve this equation as follows. Using Fourier transforms:

$$D\vec{p}^2 F_S(\vec{p}) + \alpha F_S(\vec{p}) = b \quad (3.31)$$

Solving in $P_0(\vec{p})$ and anti-transforming the equation yields (with $\gamma^2 = \frac{\alpha}{D} = \frac{2dv}{a^2}$, $s = b/D$, $r = |\vec{x}|$):

$$F_S(\vec{x}) = s \int \frac{d^d \vec{p}}{(2\pi)^d} \frac{e^{i\vec{p}\cdot\vec{x}}}{\vec{p}^2 + \gamma^2} \quad (3.32)$$

$$= \frac{s}{(2\pi)^d} \int_0^\infty d\lambda \int d^d \vec{p} e^{i\vec{p}\cdot\vec{x} - \lambda(\vec{p}^2 + \gamma^2)} \quad (3.33)$$

$$= \frac{s}{(2\pi)^d} \int_0^\infty d\lambda \left(\frac{\pi}{\lambda}\right)^{\frac{d}{2}} e^{-\frac{r^2}{4\lambda} - \lambda\gamma^2} \quad (3.34)$$

$$= \frac{s}{(4\pi)^{\frac{d}{2}}} \left(\frac{r}{2\gamma}\right)^{1-\frac{d}{2}} \int_0^\infty dz z^{-\frac{d}{2}} e^{-\frac{\gamma r}{2}\left(z+\frac{1}{z}\right)} \quad (3.35)$$

where, in the last equality, we use the scaled variable $z = 2\gamma r$ while $r \equiv |x|$.

Then we let $z = e^l$ and using formula 9.6.24 from [60] we conclude:

$$F_S(r) = \frac{s\gamma^{d-2}}{(2\pi)^{\frac{d}{2}}} (\gamma r)^{\frac{2-d}{2}} K_{\frac{2-d}{2}}(\gamma r) \quad (3.36)$$

where $K_\mu(x)$ is the modified Bessel K function of order μ .

It could be interesting to see the limiting behaviour for big and small distances:

- For $\gamma r \ll 1$ we have:

$$F_S(r) \simeq \begin{cases} \frac{s\gamma^{d-2}}{4\pi^{d/2}} \Gamma\left(\frac{2-d}{2}\right) & d < 2 \\ -\frac{s}{2\pi} \ln(\gamma r) & d = 2 \\ \frac{s}{(4\pi)^{d/2}} \Gamma\left(\frac{d-2}{2}\right) r^{2-d} & d > 2 \end{cases} \quad (3.37)$$

- For $\gamma r \gg 1$ we have:

$$F_S(r) \simeq \frac{s\gamma^{d-2}}{(2\pi)^{\frac{d}{2}}} \sqrt{\frac{\pi}{2}} (\gamma r)^{\frac{1-d}{2}} e^{-\gamma r} \quad (3.38)$$

3.5 Scaling Relations

A picture is worth a thousand words.
(ancient chinese proverb)

And from a formula you can generate a thousand pictures.
(T.Zillio)

To measure the α -diversity of our model we define the species abundance relation as the function that give the number of species with n individuals. Let's assume that this relation scales⁴ like:

$$f(n, \nu, A) = n^{-\beta} \hat{f}(n\nu^a, A\nu^{d/2}) \quad (3.39)$$

where ν is the speciation parameter and A the area (or volume) of the sample (remember that $\nu^{-1/2}$ has the dimension of a length). This relation defines the exponents a and β .

Let's also assume that the number of species scales like⁵:

$$S(\nu, A) = A^z \hat{S}(A\nu^{d/2}) \quad (3.40)$$

(defining the exponent z), and that the fraction of species $\phi(n, \nu, A) \equiv f(n, \nu, A)/S(\nu, A)$ scales like:

$$\phi(n, \nu, A) = n^{-b} \hat{\phi}(n\nu^a, A\nu^{d/2}) \quad (3.41)$$

(defining the exponent b).

Now, it's easy to demonstrate that:

$$\sum_{\vec{x}} F_{\vec{x}} = \frac{\langle n^2 \rangle}{\langle n \rangle} \quad (3.42)$$

(see Appendix B), and we can also show that, for $\nu \rightarrow 0$:

$$\sum_{\vec{x}} F_{\vec{x}} \sim \begin{cases} \nu^{-d/2} & d < 2 \\ \nu^{-1} & d > 2 \end{cases} \quad (3.43)$$

⁴ Finite-size scaling techniques such the ones used in the present work are become increasingly popular in the field of ecology. See [69] [70].

⁵ Recent research has established that there is a link between species spatial turnover data and the Species-Area Relationship. See [17] and [71] for details.

(see Appendix C).

Guided by this relation, we assume that:

$$a = \min(1, d/2) \quad (3.44)$$

Now we need the further hypothesis that $\hat{f}(x, y) \sim x^c g(y)$ for $x \rightarrow 0$.

Since:

$$S = \sum_n f(n, \nu, A) \sim \nu^{a(\beta-1)} \int_{\nu^a}^{\infty} dx x^{-\beta} \hat{f}(x, A\nu^{d/2}) \quad (3.45)$$

$$A = \sum_n n f(n, \nu, A) \sim \nu^{a(\beta-2)} \int_{\nu^a}^{\infty} dx x^{1-\beta} \hat{f}(x, A\nu^{d/2}) \quad (3.46)$$

we have three cases:

1. $1 + c - \beta > 0$: this leads to $b = 1$ and

$$z = \begin{cases} 0 & d < 2 \\ \frac{d-2}{d} & d > 2 \end{cases} \quad (3.47)$$

2. $-1 < 1 + c - \beta < 0$: this leads to $1 < b < 2$ and the scaling relation

$$a(2 - b) = \frac{d}{2}(1 - z) \quad (3.48)$$

3. $1 + c - \beta < -1$: this leads to $b = \beta - c > 2$, $c = -\max(1, d/2)$ and $z = 1$.

Detailed calculations for the three cases are performed in Appendix D.

3.6 Simulations and Scaling Exponents

*Simulations are like miniskirts,
they show a lot and hide the essentials.*
(Hubert Kirrman)

The model has then been explicitly simulated on a computer. The simulation were carried out in hypercubic lattices of various sizes and of dimensionality $d = 1, 2, 3$. In $d = 2$ we also used the exagonal lattice, as a test to verify the insensitiveness to the microscopic details. The goals of these simulations was:

1. Verify the solution for the 2-point correlation function, both in the discrete (in $d = 1$, Eq.(3.24)) and the continuous version (Eq.(3.36)).
2. Provide various instances of RSA and SAR curves with the aim of collapsing them using the theory presented in the previous section, thus having a numerical estimate of the critical exponents a , b and z .

Regarding point 1 above, Fig.3.1 and Tab.3.1 shows the beta-diversity from some simulations of the voter model with speciation in $d = 2$, on a 200×200 lattice, varying the speciation. The graph displays an excellent fit between the analytical solution and the explicit simulation of the model.

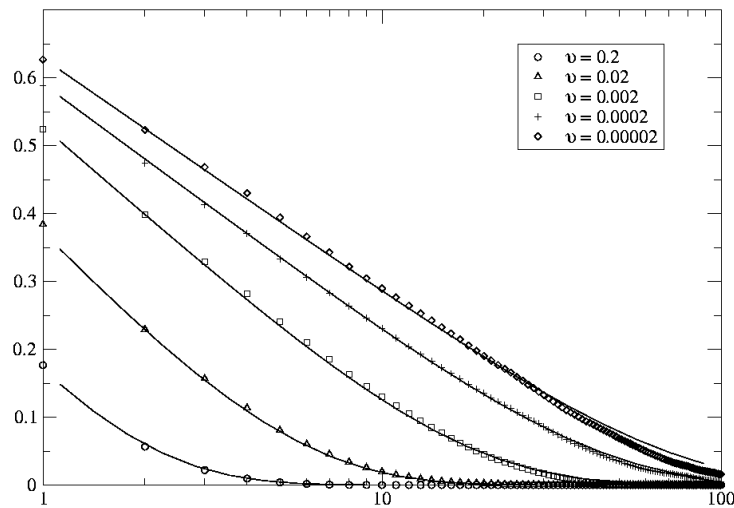


Figure 3.1: Points: simulations. Lines: theoretical fits. This graph shows some simulations with varying speciation in $d = 2$ and the relative theoretical fits (which parameters are listed in Tab.3.1).

Regarding point 2 above, a series of simulation was runned for the determination of the RSA curve, with fixed lattice size ($L = 200$ for $d = 2$ and $L = 100$ for $d = 1, 3$, L being the side of the hypercube used). Another series of simulations was carried out to determine the SAR curve, in which we vary both the speciation rate the size of the lattice. We defined $S(\nu, A)$ as the mean number of species in a simulation with speciation rate ν lattice size $A = L^d$. In particular we were not using the “nested area” method to obtaining the number of species, but we were only taking the total number of species in the lattice.

Table 3.1: Value of the parameters for the fitting curves in Fig.3.1.

ν	Species	$1/\gamma$	s
0.2	16883	1.33	0.28
0.02	3752	4.35	0.23
0.002	682	14.3	0.19
0.0002	122	35.7	0.16
0.00002	40	58.8	0.15

Once obtained the curves by averaging over multiple runs (up to 1000) of the simulations, the curves themselves were collapsed using our previously described scaling forms. The actual collapse was done using the algorithm described in [72] with some slight modification on the form of the main functional, to suit best our needs. In brief, this algorithm allows, given a scaling form and a number of curves to be collapsed, to find the exponent values that yields the optimal collapse. This way we can have a numerical estimation of the value of the collapsing exponent present in our theory.

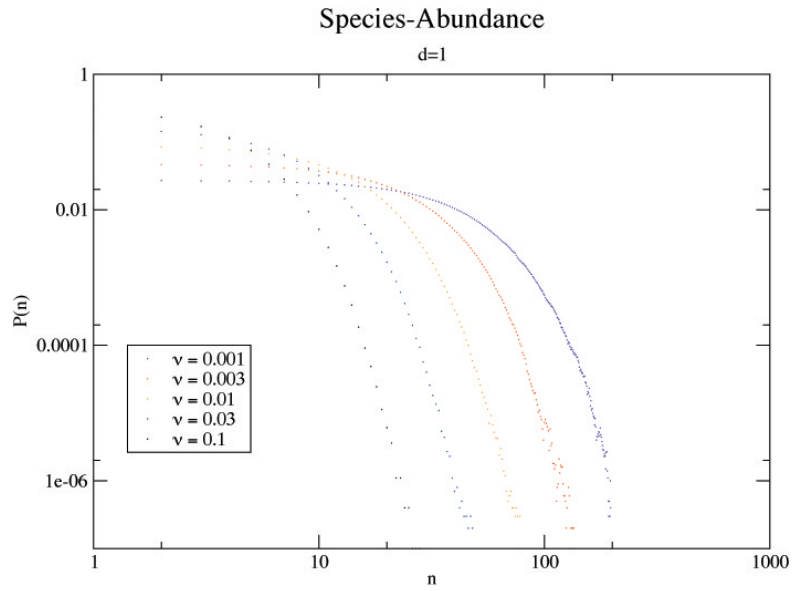
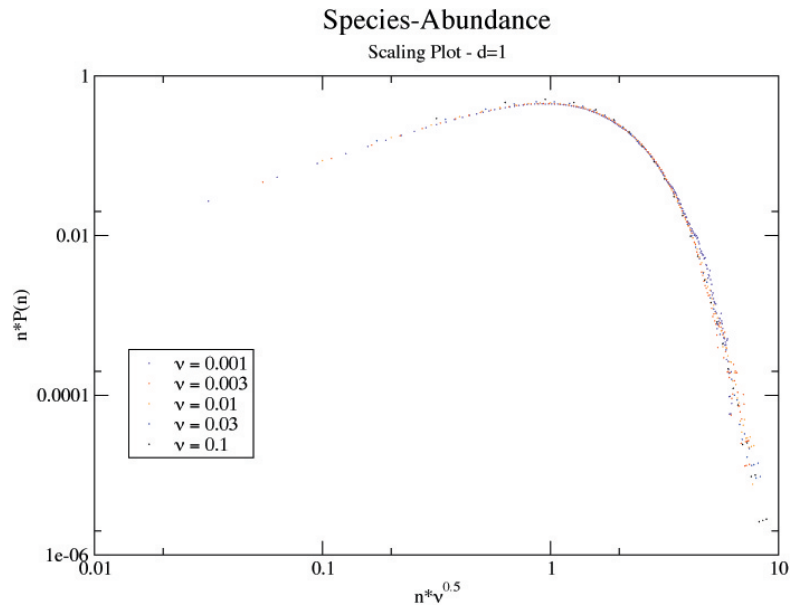
In the following figures (Fig.3.2-3.11) we will present the original RSA and SAR data for $d = 1, 2, 3$ and their scaling collapse.

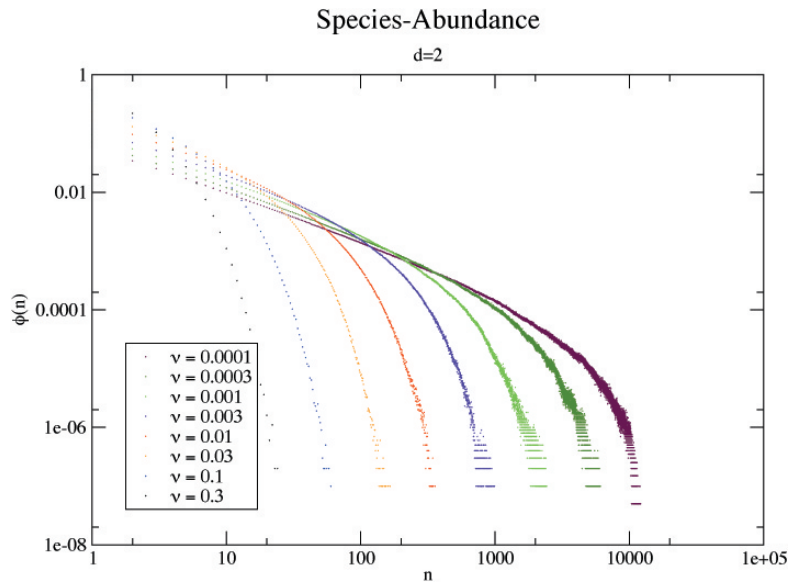
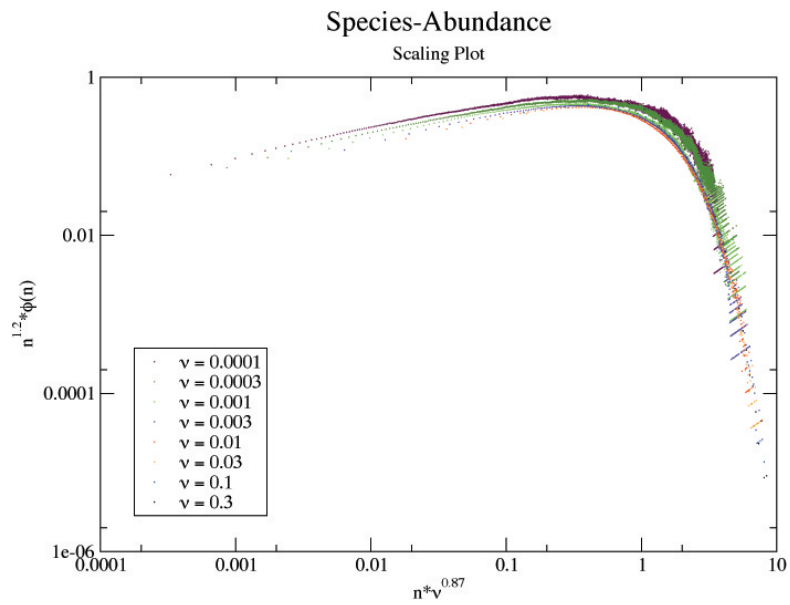
The values of the scaling exponents obtained by our optimization of the collapse are collected in Tab.3.2

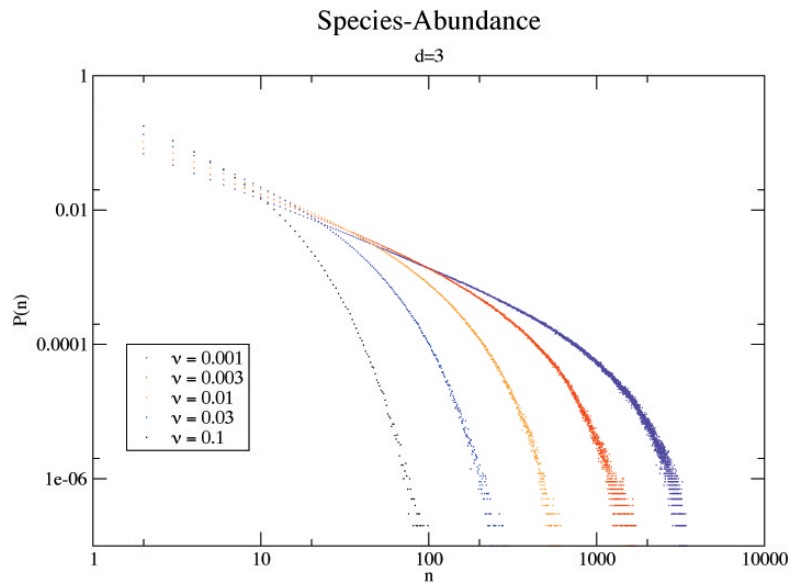
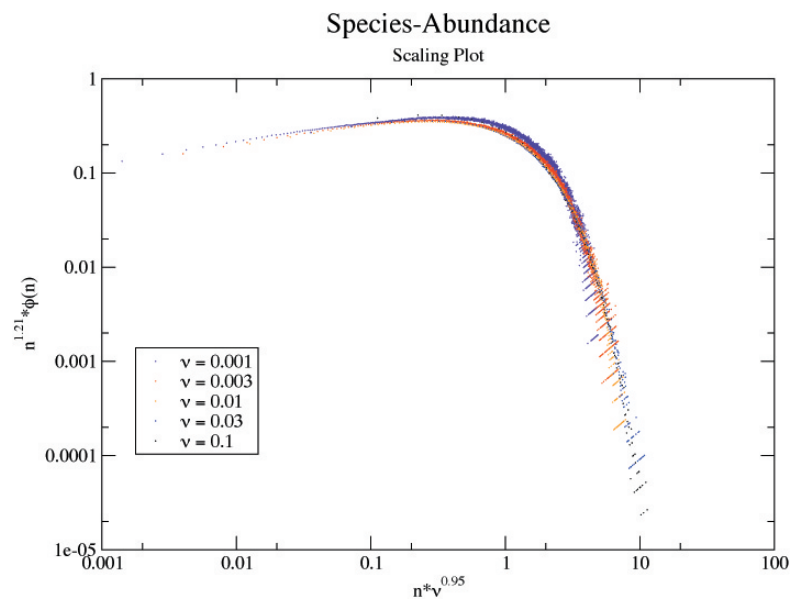
Table 3.2: Scaling exponents for $d = 1, 2, 3$ determined from the scaling collapse of the normalized RSA and SAR plots.

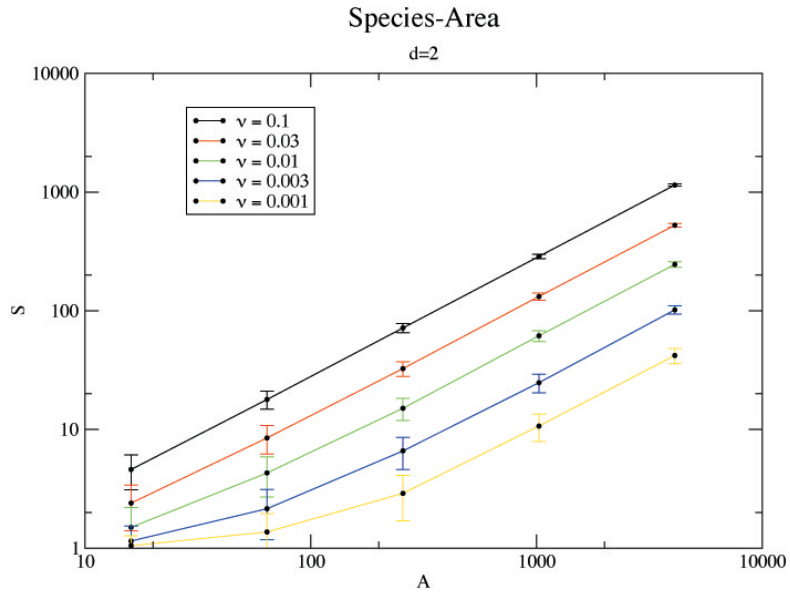
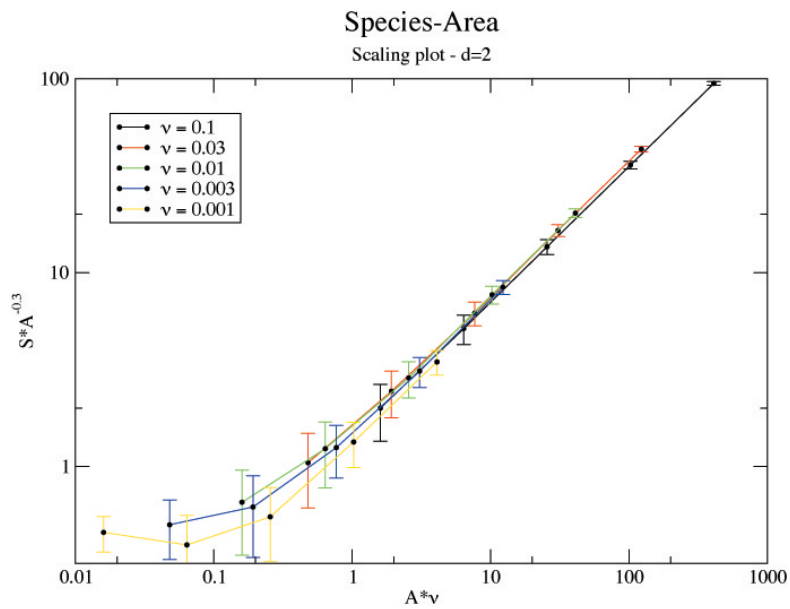
d	a	b	z
1	0.5	1.05	0.03
2	0.87	1.2	0.3
3	0.95	1.21	0.5

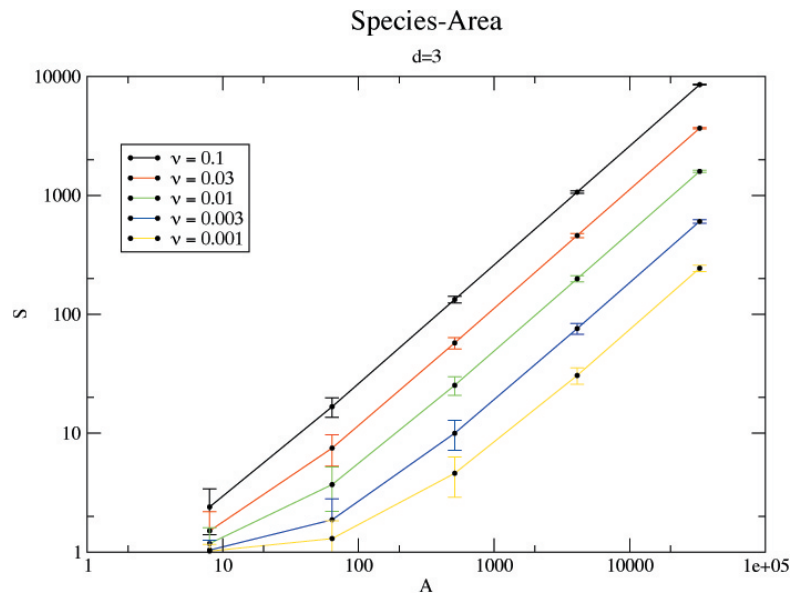
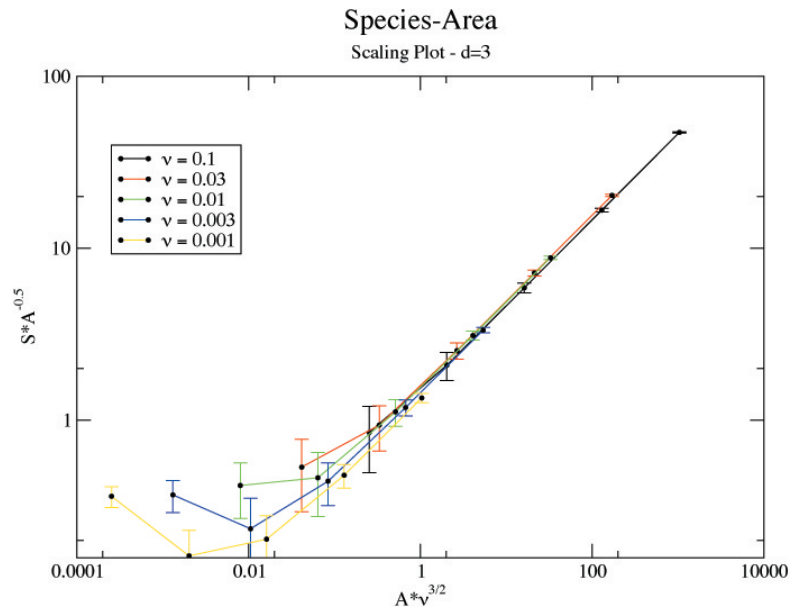
As one can easily verify, in all the three cases we have $1 < b < 2$, and the corresponding scaling relation Eq.(3.48) is satisfied with very good approximation. The main deviation of the data from our theory is in $d = 2$, the upper critical dimension for diffusive processes, where $a = 0.87$,

Figure 3.2: Species-Abundance relation in $d = 1$.Figure 3.3: Species-Abundance scaling collapse in $d = 1$.

Figure 3.4: Species-Abundance relation in $d = 2$.Figure 3.5: Species-Abundance scaling collapse in $d = 2$.

Figure 3.6: Species-Abundance relation in $d = 3$.Figure 3.7: Species-Abundance scaling collapse in $d = 3$.

Figure 3.8: Species-Area relation in $d = 2$.Figure 3.9: Species-Area scaling collapse in $d = 2$.

Figure 3.10: Species-Area relation in $d = 3$.Figure 3.11: Species-Area scaling collapse in $d = 3$.

while our theory predicts a value of 1. Our data are suggestive of logarithmic corrections, that are nevertheless very hard to verify numerically being masked by the global power-law shape of the curve.

An alternative and independent method useful for finding directly the value of a is to plot the value of $\langle n^2 \rangle / \langle n \rangle$ vs. ν for different simulations, because, as we know from the previous section,

$$\frac{\langle n^2 \rangle}{\langle n \rangle} \sim \nu^{-\alpha} \quad (3.49)$$

This quantity is easily calculated using the RSA curve.

As can be seen on the following figures (Fig.3.12-3.14), the value of α found with this method is always compatible with the one obtained from the scaling collapse. This supports our claim that the strange value for a in $d = 2$ is due to some pathological property of the model at the critical dimension and it's not due to the particular method we use to measure it.

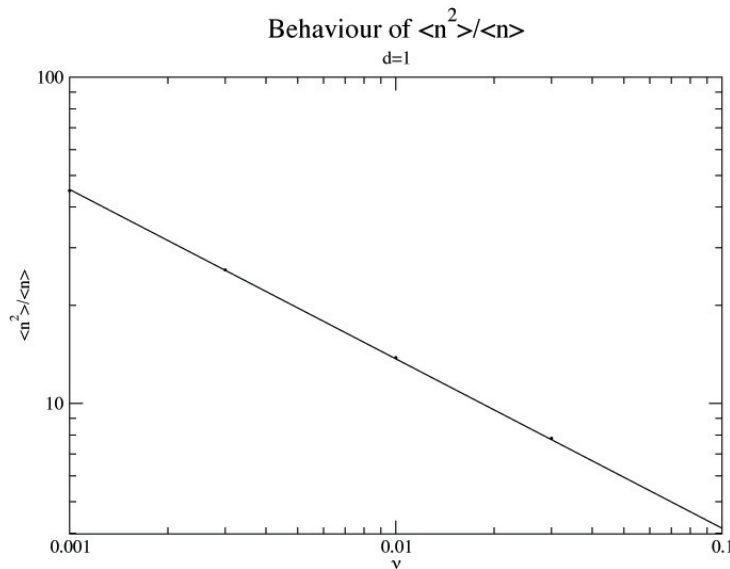


Figure 3.12: Behavior of $\langle n^2 \rangle / \langle n \rangle$ vs. ν for $d = 1$. The line is the linear regression on the log-log plot that yields $\langle n^2 \rangle / \langle n \rangle = 1.25\nu^{-0.52}$.

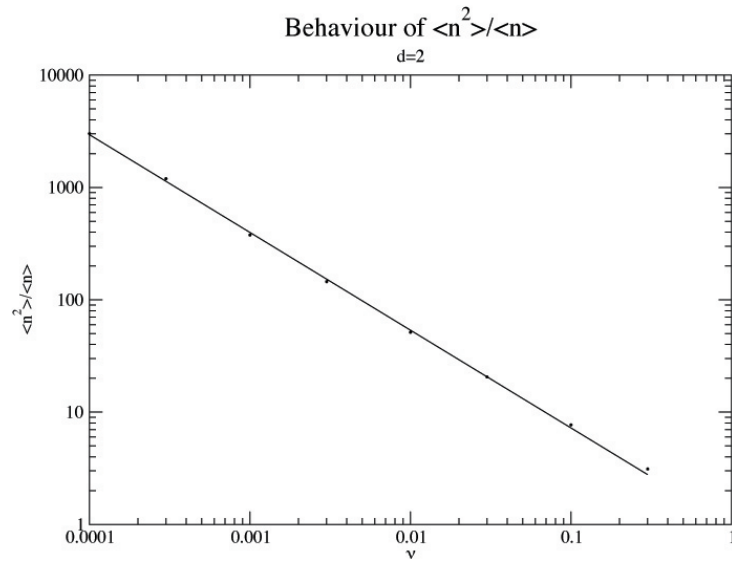


Figure 3.13: Behavior of $\langle n^2 \rangle / \langle n \rangle$ vs. ν for $d = 2$. The line is the linear regression on the log-log plot that yields $\langle n^2 \rangle / \langle n \rangle = 0.975\nu^{-0.87}$.

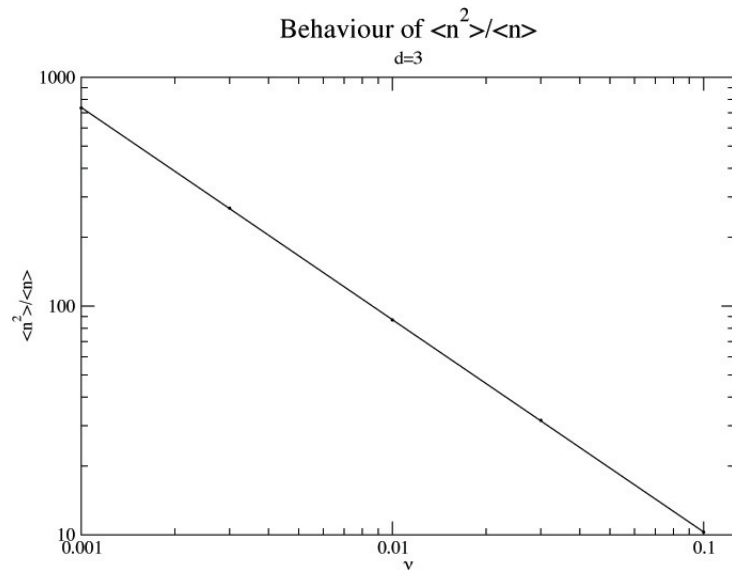


Figure 3.14: Behavior of $\langle n^2 \rangle / \langle n \rangle$ vs. ν for $d = 3$. The line is the linear regression on the log-log plot that yields $\langle n^2 \rangle / \langle n \rangle = 1.22\nu^{-0.93}$.

3.7 The Janzen-Connell effect

*In the fight between you and the world,
back the world.*

(Frank Zappa)

As it is apparent by the shape of our solution Eq.(3.36), the simple voter model with speciation is unable to fit the experimental data about β -diversity (Fig.1.6 and 1.7). In other words, speciation and diffusion alone can't explain the spatial patterns that we find in Tropical Forests. It appears that we could be able to reproduce the overall shape of the data if there were a negative density-dependent [73][74] effect for shorter distances. In practice we are searching for an ecological effect that disfavor the birth (or favor the death) of a tree if there are too many of his conspecific in their neighborhood.

Such an effect was discovered experimentally in the early seventies independently by Janzen [75] and Connell [76], who remarked that a seed is more likely to germinate, and a seedling more likely to survive, if they are far from conspecific adult trees. This effect alter substantially the dynamics of the system we have so far described. Janzen and Connell postulated that this increased mortality rate of seed and seedlings near conspecific adults arises from the presence of pests that are host specific i.e., specialized to that type of tree, and experimental evidence supports this conclusion [77].

To implement the Janzen-Connell effect in the voter model with speciation, we will add the following modification: each time a new tree is born there is a probability that it will immediately dies directly proportional to the number of his conspecifics within a given radius R (called the J.C. radius). This obviously does not apply if the tree was placed there by a speciation event, since there aren't any conspecific of it in the lattice by definition.

Following step-by-step the algorithm described above, one can show that there is an effective speciation rate equal to:

$$\begin{aligned} \nu_{\text{eff}} &= \nu + \nu(1 - \nu)p_d(\vec{x}) + \nu(1 - \nu)^2 p_d^2(\vec{x}) + \dots \\ &= \nu \sum_{n=0}^{\infty} ((1 - \nu)p_d(\vec{x}))^n \end{aligned}$$

$$= \frac{\nu}{1 - (1 - \nu)p_d(\vec{x})} \quad (3.50)$$

where $p_d(\vec{x})$ is the probability that an offspring will die due to the Janzen-Connell effect when there is a conspecific adult a vector \vec{x} away. In other words, ν_{eff} is the probability that at the end of the process in the chosen site there will be a tree from a new species, while $1 - \nu_{\text{eff}}$ is the probability that the new tree will be chosen from among its neighbours. Using this effective speciation rate in Eq.(3.21), one obtains (for $\vec{x} \neq 0$):

$$F_{\vec{x}}^{n+1} = F_{\vec{x}}^n \left(1 - \frac{2}{N}\right) + \frac{1}{dN} \left(1 - \frac{\nu}{1 - (1 - \nu)p_d(\vec{x})}\right) \sum_{\mu=1}^d (F_{\vec{x}+\mu}^n + F_{\vec{x}-\mu}^n) \quad (3.51)$$

along with the boundary condition $F_0^n = 1$.

Using the same procedure as before, we obtain:

$$\dot{F}(\vec{x}, t) = 2 \left(D \nabla^2 F(\vec{x}, t) - \frac{\alpha}{1 - p_d(\vec{x})} F(\vec{x}, t) \right) + 2 \frac{b}{1 - p_d(\vec{x})} \delta^d(\vec{x}) \quad (3.52)$$

with the same α and D as before. We will define $\alpha_{\text{eff}}(x) \equiv \frac{\alpha}{1 - p_d(x)}$. If $p_d(x)$ is piecewise constant, a general solution of this equation in $d = 2$ for a single zone in which $p_d(x)$ is constant is (see Appendix E):

$$\frac{s\gamma_0^2}{2\pi} K_0(\gamma_0 r) + c I_0(\gamma_0 r) \quad (3.53)$$

In the standard Voter Model, $c = 0$ because $I_0(r) \rightarrow \infty$ as $r \rightarrow \infty$. Instead, in the Voter Model Modified with Janzen-Connell we assume:

$$\alpha_{\text{eff}} = \begin{cases} \alpha_0 & \text{if } |\vec{x}| < R \\ \alpha_1 & \text{if } |\vec{x}| > R. \end{cases} \quad (3.54)$$

And thus:

$$F(r) = \begin{cases} c_0 K_0(\gamma_0 r) + c_1 I_0(\gamma_0 r) & r < R \\ c_2 K_0(\gamma_1 r) & r > R \end{cases} \quad (3.55)$$

with $c_0 = \frac{s\gamma_0^2}{2\pi}$. Here we have followed the standard approach of solving the equations separately in the two regions and then imposing the continuity of the function and its derivative⁶ at $r = R$. This condition has been

⁶ It is necessary to impose the continuity also of the derivative because it is a solution of a differential equation of second order in space, and a non-continuous derivative will yields some spurious delta functions when derived two time in space.

enforced along with the boundary condition as $r \rightarrow \infty$ (that forbids the presence of the I_0 term in the most far segment from the origin).

The continuity of the function and his derivative at $r = R$ is assured solving:

$$\begin{cases} c_0 K_0(\gamma_0 R) + c_1 I_0(\gamma_0 R) & = c_2 K_0(\gamma_1 R) \\ c_0 K_0'(\gamma_0 R) + c_1 \gamma_0 I_0'(\gamma_0 R) & = c_2 \gamma_1 K_0'(\gamma_1 R) \end{cases} \quad (3.56)$$

which provides the values of c_1 and c_2 .

To summarize, our final expression Eq.(3.55) depends on 4 parameter: c_0 , γ_0 , γ_1 and R . While it may seems that 4 parameter are really too many, we want to remember that the theory explained in [18] has 3 parameters, and was unable to fit the data at shorter distance, lacking the Janzen-Connell effect.

The fits (Fig. 3.15 and 3.16) were performed by minimizing the square deviation from the data and taking into account the square deviations of each point. The resulting values for each parameter of the fit are collected in Tab.3.3. The inverse of the γ 's are typical length of the system; we expect that $\gamma_0^{-1} \sim R$, and in both fit they are of the same order.

Table 3.3: Values of the parameter in Eq.(3.55) obtained by the fits.

Plot	R (m)	γ_0^{-1} (m)	γ_1^{-1} (Km)	c_0
Panama	46	68	210	0.012
Yasuni	86	69	23500	19

Independent measures show that the order of magnitude of R is correct [78]. In particular it appears that the Janzen-Connell radius that we obtain from the fit is overestimated by a factor of two, more or less. One suggested explanation is that the experimental procedure used to measure the J.C. effect depends on the time of the observation of the forest, and that in the last 20 years of observation the measure of this radius was growing. However, this point deserve some further examination, so we can't say the final word on it.

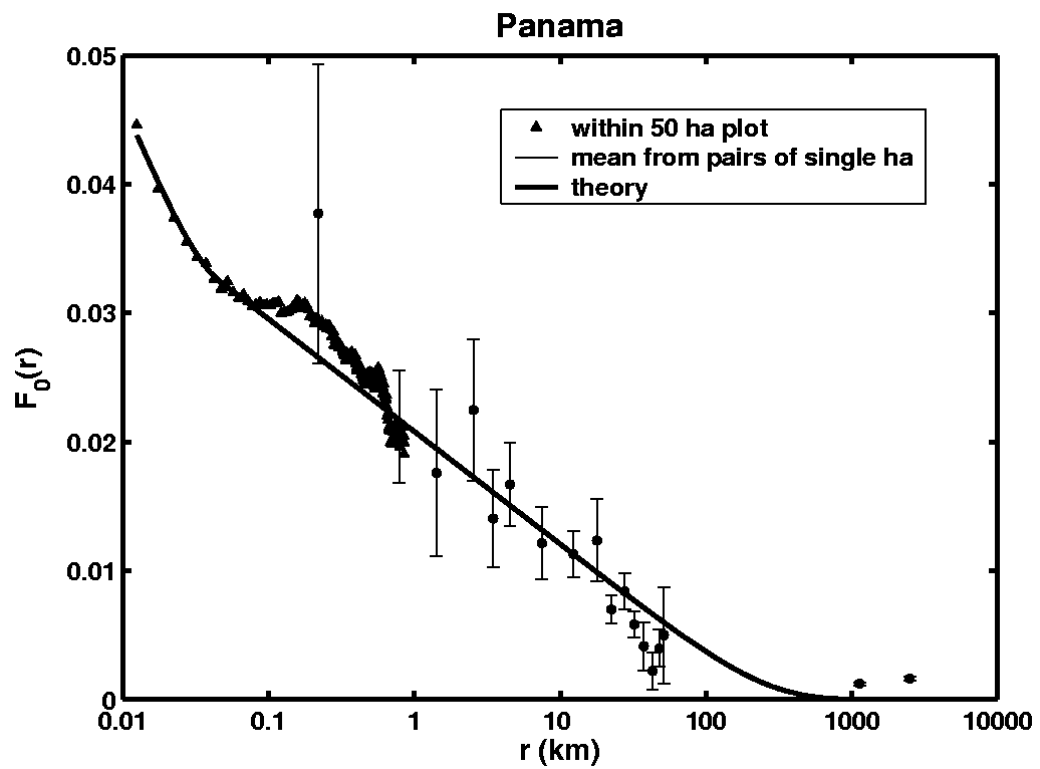


Figure 3.15: Fit of the voter model with speciation and Janzen-Connell effect to the data from BCI, Panama.

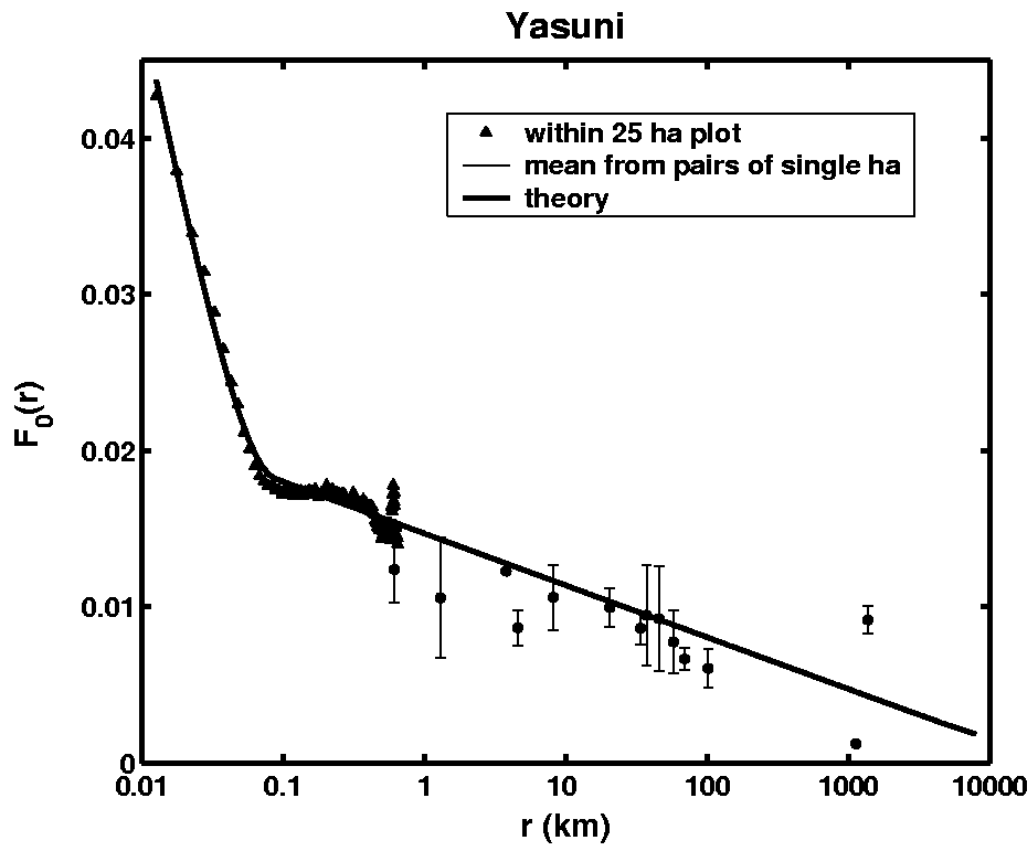


Figure 3.16: Fit of the voter model with speciation and Janzen-Connell effect to the data from Yasuni, Peru.

Chapter 4

Spatial Patterns of Intertidal Vegetation Diversity

People don't do hard things gratuitously; no one will work on a harder problem unless it is proportionately (or at least $\log(n)$) more rewarding.
(Paul Graham)

4.1 Motivation

Now we will describe a model capable to describe the β -diversity of the venetian salt marshes. Before the definition of the actual model we will describe a general master equation formalism for a linear spatial-explicit model on a lattice without speciation. Then we will specialize the formalism for the intertidal vegetation that we intend to model. In this model we will take into account also empty territory¹, and we will not follow the prescription of neutrality. The simplicity of the model and the little number of parameters will permit us to have a transparent picture of the non-neutrality in action.

The dynamic of the model will be as follows: each site can propagate its content (being it an individual of a species or the void) to the nearest neighboring sites (mimicking in this way the reproduction by propagules or, in the case of empty sites, a 'desertification' process). Also, each indi-

¹ Thus breaking the law of constant density at the level of the single site

vidual present on the lattice has a certain (species-dependent) probability of dying out on his own, and each empty site can give born of an individual of any species (with a species-dependent rate, this reproduce the presence of dormant seeds scattered around the territory).

4.2 Master Equation General Formulation and Solution

Let's have a lattice with N sites $x = 1, \dots, N$, occupied with S interacting species. We denote the specie at site x with $\sigma_x = 1, \dots, S$. Each site can host one individual, or being empty ($\sigma_x = 0$). Let's also formally denote the generic state of the system as $\vec{\sigma} = (\sigma_1, \dots, \sigma_N)$. Then we are interested in an equation for the probability of being in a given state $\vec{\sigma}$ at the n -th time step, namely $P^n(\vec{\sigma})$. We can write the master equation:

$$P^{n+1}(\vec{\sigma}) = P^n(\vec{\sigma}) + \sum_{\vec{\sigma}'} [W_{\vec{\sigma},\vec{\sigma}'} P^n(\vec{\sigma}') - W_{\vec{\sigma}',\vec{\sigma}} P^n(\vec{\sigma})] \quad (4.1)$$

We will assume that the time step is such that a single event can occur, and that the W 's do not depend on n . Moreover, we assume the following form for W that guarantees that the transition probabilities for each site will depend only on the state of the site considered and his neighbors:

$$W_{\vec{\sigma},\vec{\sigma}'} = \sum_x W_{\vec{\sigma},\vec{\sigma}'}(x) \quad (4.2)$$

when:

$$W_{\vec{\sigma},\vec{\sigma}'}(x) = C \prod_{x \neq y} \delta_{\sigma_y, \sigma'_y} \cdot \left(I_{\sigma_x, \sigma'_x} + \sum_{z(x)} N_{\sigma_x, \sigma'_z} \right) \quad (4.3)$$

where $z(x)$ are the nearest neighbors of x , and the constant C will be useful later for tuning the time scale. In this equation, operator I is responsible of the action 'on site' while operator N will take care of the diffusion effects.

In order to be able to write a master equation for the 1-point and 2-point correlation function and ensure the conservation of probability, we have to require (as seen a posteriori) that:

$$\sum_{\sigma} I_{\sigma, \sigma'} = k \quad (4.4)$$

$$\sum_{\sigma} N_{\sigma, \sigma'} = \frac{1-k}{q(x)} \quad (4.5)$$

where $q(x)$ is the number of nearest neighbors of x , and $0 \leq k \leq 1$ is a parameter balancing the relative importance of 'on site' processes and diffusion processes (the greater k is, the more 'on site' processes are important). To ensure this last condition, we require:

$$I_{\sigma,\sigma'} = kL_{\sigma,\sigma'}^{(1)} \quad (4.6)$$

$$N_{\sigma,\sigma'} = \frac{1-k}{q(x)}L_{\sigma,\sigma'}^{(2)} \quad (4.7)$$

with:

$$\sum_{\sigma} L_{\sigma,\sigma'}^{(i)} = 1 \quad (4.8)$$

then we can look at the elements of $L_{\sigma,\sigma'}^{(i)}$ as transition probabilities.

We define the 1-point and 2-point correlation function respectively as:

$$P_x^n(\tau) \equiv \langle \delta_{\tau,\sigma_x} \rangle^{(n)} = \sum_{\vec{\sigma}} P^n(\vec{\sigma}) \delta_{\tau,\sigma_x} \quad (4.9)$$

(remark that $\sum_{\tau} P_x^n(\tau) = 1$), and:

$$P_{x,y}^n(\varepsilon, \tau) \equiv \langle \delta_{\varepsilon,\sigma(x)} \delta_{\tau,\sigma(y)} \rangle \quad (4.10)$$

We can obtain a close equation for both of them (see Appendix F), and imposing stationarity and translational invariance (and symmetrizing the 2-point correlation function) we have:

$$P(\tau) = \sum_{\varepsilon} \left(kL_{\tau,\varepsilon}^{(1)} + (1-k)L_{\tau,\varepsilon}^{(2)} \right) P(\varepsilon) \quad (4.11)$$

and

$$\begin{aligned} F_x(\varepsilon, \tau) = & \frac{1}{2} \sum_a \left[kL_{\varepsilon,a}^{(1)} F_x(a, \tau) + kL_{\tau,a}^{(1)} F_x(\varepsilon, a) + \right. \\ & + \frac{1-k}{2d} \sum_{i=1}^d \left(L_{\varepsilon,a}^{(2)} \left(F_{x+e_i}(a, \tau) + F_{x-e_i}(a, \tau) \right) + \right. \\ & \left. \left. + L_{\tau,a}^{(2)} \left(F_{x+e_i}(\varepsilon, a) + F_{x-e_i}(\varepsilon, a) \right) \right) \right] \quad (4.12) \end{aligned}$$

for $x \neq 0$, where $q(x) = 2d$, d is the space dimensionality of the system and e_i are the lattice versors and

$$F_x(\varepsilon, \tau) = \frac{1}{2} \left(\hat{F}_x(\varepsilon, \tau) + \hat{F}_x(\tau, \varepsilon) \right) \quad (4.13)$$

is the symmetrized 2-point correlation function. When $x = 0$ instead, we have:

$$F_0(\varepsilon, \tau) = \delta_{\varepsilon, \tau} P(\tau) \quad (4.14)$$

We assume now that:

$$L^{(2)} = \mathbf{1} \quad (4.15)$$

that is, each individual can give birth only to conspecific offsprings. Under this hypothesis both equations can be solved (see Appendix F). The solution for the 2-point correlation function is:

$$F_x = \sum_{a,b} v_{a,b}^R c_{a,b} f(x, \mu_{a,b}) \quad (4.16)$$

where the formalism utilized is explained fully in Appendix F. This equation will give us all the correlation functions $F_x(\varepsilon, \tau)$ (the indexes ε and τ are contained implicitly in the term $v_{a,b}^R$) in a numerically computable form (even analytical in $d = 1$). The great advantage of this system is that it permits to obtain the correlation functions at the desired precision without the explicit simulation of the model, thus avoiding all the problems correlated with simulations (sufficient statistics, sampling errors, long simulation times, sloppy numerical definition of equilibrium state, etc.).

We are now free with the choice of one among the many models that can fit inside this formalism. In the next section we will outline the explicit embodiment of our theory, our guides will be simplicity and the need to keep the numbers of parameter at a minimum.

4.3 The Model Definition

4.3.1 One species plus the void

Different models can be fitted in this framework specifying the elements of $L_{\sigma, \sigma'}^{(i)}$, i.e. the transition probabilities. Our first assumption is that each species can diffuse only conspecific seeds, and that means that the matrix $L^{(2)}$ is diagonal, as was assumed before: $L_{\sigma, \sigma'}^{(2)} = \delta_{\sigma, \sigma'}$ (remark that for consistency, even the empty sites can 'diffuse', a process that we will call 'desertification').

For now, in our model we put $S = 1$, i.e. each site can be occupied by species 1 or be empty (in the next section we will deal also with the case

with multiple species). Now the last thing to specify is the matrix:

$$L^{(1)} = \begin{pmatrix} L_{0,0}^{(1)} & L_{0,1}^{(1)} \\ L_{1,0}^{(1)} & L_{1,1}^{(1)} \end{pmatrix} \quad (4.17)$$

For our model we choose the following matrix form:

$$L^{(1)} = \begin{pmatrix} 1 - v & d \\ v & 1 - d \end{pmatrix} \quad (4.18)$$

with the following interpretation: kd it's the rate of mortality 'on site' of the trees, $it's the rate of birth 'on site' (i.e. not caused by diffusion, this meaning that can be caused by a dormant seed that suddenly germinate), and $1 - k$ is the diffusion rate.$

Now it should be noted that if we can change the three coefficients k, v, d , preserving the ratios of the rates $kv, kd, 1 - k$ we obtain the same model as before except for a change in the time scale (thus not affecting equilibrium properties). To see how to change them, one writes the following system of equations:

$$\begin{cases} 1 - k = c(1 - k') \\ kv = ck'v' \\ kd = ck'd' \end{cases} \quad (4.19)$$

where c is the change of scale of time. Solving these equations we arrive at the following symmetry:

$$\begin{cases} v \rightarrow av \\ d \rightarrow ad \\ k \rightarrow \frac{k}{a+k(1-a)} \end{cases} \quad (4.20)$$

where $0 \leq a \leq 1$ is an auxiliary variable and the change of time scale is given by:

$$c = \frac{a}{a+k(1-a)} \quad (4.21)$$

Please remark that regardless of the value of a , if k has the value of 0 or 1 it will not change its value (i.e. $k = 0$ and $k = 1$ are fixed point under this transformation), as it should be: the case of pure diffusion and the case of pure 'on site' process must be invariant.

It is possible to construct the invariants of this symmetry, and it is easily seen that, for $0 < k < 1$, we have only two independent invariants:

$$\zeta = \frac{kd}{1-k}; \quad \gamma = \frac{kv}{1-k} \quad (4.22)$$

If now we substitute $L^{(i)}$ into Eq.(4.11), we find that at equilibrium we have²:

$$\begin{aligned} P(0) &= \frac{d}{v+d} = \frac{\zeta}{\zeta+\gamma} \\ P(1) &= \frac{v}{v+d} = \frac{\gamma}{\zeta+\gamma} \end{aligned} \quad (4.23)$$

These are the equilibrium density for the species and the empty sites, and are invariant for the transformation above because they can be expressed in term of invariants only.

4.3.2 $N > 1$ species

In this case the matrix $L^{(1)}$ is of the form:

$$L^{(1)} = \begin{pmatrix} 1 - \sum_{i=1}^N v_i & d_1 & \dots & d_N \\ v_1 & 1 - d_1 & 0 & 0 \\ \vdots & 0 & \ddots & 0 \\ v_N & 0 & 0 & 1 - d_N \end{pmatrix} \quad (4.24)$$

meaning that an empty site can become a site occupied with each of the species, and an occupied site can become empty, but an occupied site can't become occupied by another species without being empty in the meantime. The matrix $L^{(2)}$ is still the identity $L_{\sigma,\sigma'}^{(2)} = \delta_{\sigma,\sigma'}$.

Again we can construct a transformation that leaves unchanged the actual probabilities, and we can find the invariants of this transformation:

$$\zeta_i = \frac{kd_i}{1-ki}; \quad \gamma_i = \frac{kv_i}{1-k} \quad (4.25)$$

²When substituting $L^{(i)}$ into Eq.(4.11) one finds that the values of $P(\tau)$ are found by calculating $(L^{(1)} - \mathbf{1})P = 0$

Substituting $L^{(i)}$ into Eq.(4.11) we can find the stationary solution for the densities of species:

$$\begin{aligned} P(0) &= \frac{1}{1 + \sum_{j=1}^N v_j/d_j} = \frac{1}{1 + \sum_{j=1}^N \gamma_j/\zeta_j} \\ P(i \neq 0) &= \frac{v_i/d_i}{1 + \sum_{j=1}^N v_j/d_j} = \frac{\gamma_i/\zeta_i}{1 + \sum_{j=1}^N \gamma_j/\zeta_j} \end{aligned} \quad (4.26)$$

4.4 Application to the Salt Marshes

The use of this model for simulating the intertidal vegetation and fit the experimental data faces many approximations and inaccuracies, namely:

- The experimental data can classify the vegetation with a resolution of 0.5m. This means that we can not see the single individuals, and each pixel of the data is assigned by a majority rule. So our first approximation is to neglect the minority part of the vegetation in each pixel.
- The classification is done by analyzing the spectral photography of the zone of interest. The passage from the spectral photograph to the actual classification of the vegetation is not exempt from errors, though it can be done with remarkable precision [30] [31] .
- It should also be remarked that the various species that compose the intertidal vegetation are very different in size and density over the territory. We will consider our lattice as coincident with the grid given by the data, in which each site has a side of 0.5m, and so contains more than one individual. Thus our coefficients are not the ones of birth, death and diffusion of a single individual, but they are the effective ones for a $0.5 \times 0.5 \text{m}^2$ zone of vegetation with the additional hypothesis that the pixel can behave like a single individual in a certain sense.

The fits are performed on all the mixed β -diversities for every barena considered, minimizing the sum of all the squared deviation from the data.

If N is the number of species present in a given barena, excluding the void (soil), then the number of effective parameters n_P required by the

model is:

$$n_P = 2N \quad (4.27)$$

while the number of β -diversity curves generated n_C is:

$$n_C = \frac{(N + 1)(N + 2)}{2} \quad (4.28)$$

Then the number of parameter against the number of curves goes as in Tab.4.1; the strikingly feature here is that we are thus able to fit the β -diversity data with less than one parameter per curve!

Table 4.1: Number of parameters required (n_P) against the number of curves generated (n_C) for our model.

N	n_P	n_C
1	2	3
2	4	6
3	6	10
4	8	15

Here we present the graphs of the pure β -diversities (i.e. $F_x(i, i)$) and the relative theoretical fits from our model for the salt marsh of S.Felice and two of its subsets, Zone A and Zone C. In Zone A the presence of *Sarcocornia Fructicosa* was so small (79 pixels over a total of nearly 40.000) that we eliminated it completely to have clearer data. The parameters of the fits are presented in the tables below. Numerically it appears that the curves are fairly insensible to small variation in the parameters.

The coefficient γ (ζ) should represent the ratio between the probability that an individual born from a dormant seed (that an individual die) and the probability of diffusion by propagules. However, due to the impossibility of an empirical measure of such quantities, we can't confront our theory with the field data apart from the fit presented here. A deeper analysis of this theory is now under development.

The model has, nevertheless, also an interest on his own, providing a solution to a fairly general class of equations, and his applications look promisingly.

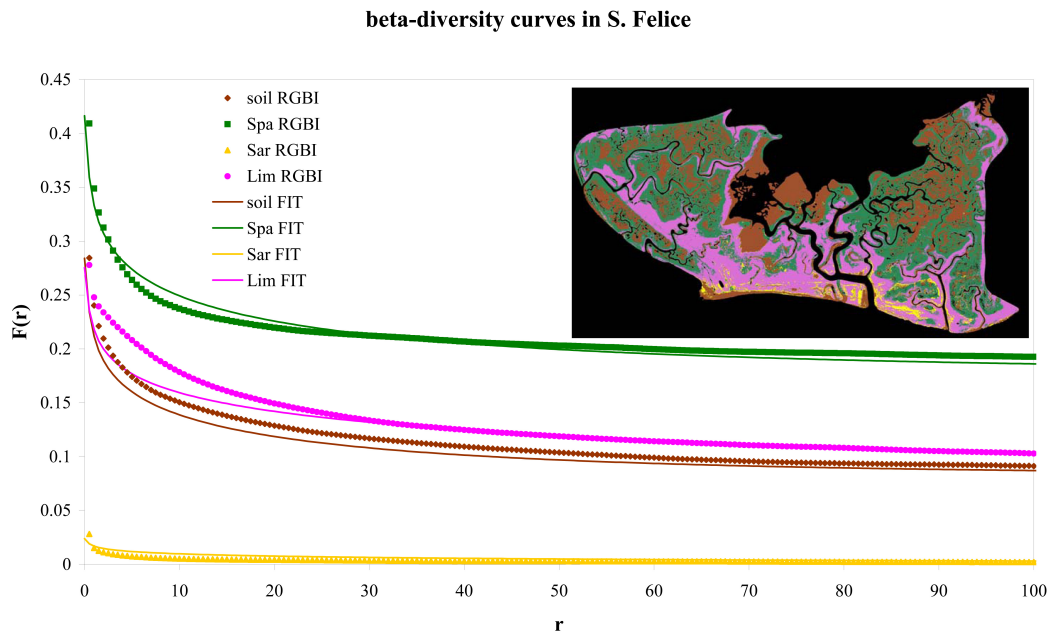


Figure 4.1: β -diversity curves for S.Felice salt marsh (labeled with RGBi) and their theoretical fit.

Table 4.2: Parameters from the fit of β -diversity data in S.Felice.

Species	γ	ζ
<i>spartina maritima</i>	3.77e-5	2.58e-5
<i>sarcocornia fruticosa</i>	3.22e-7	3.82e-6
<i>limonium narbonense</i>	1.48e-6	1.52e-6

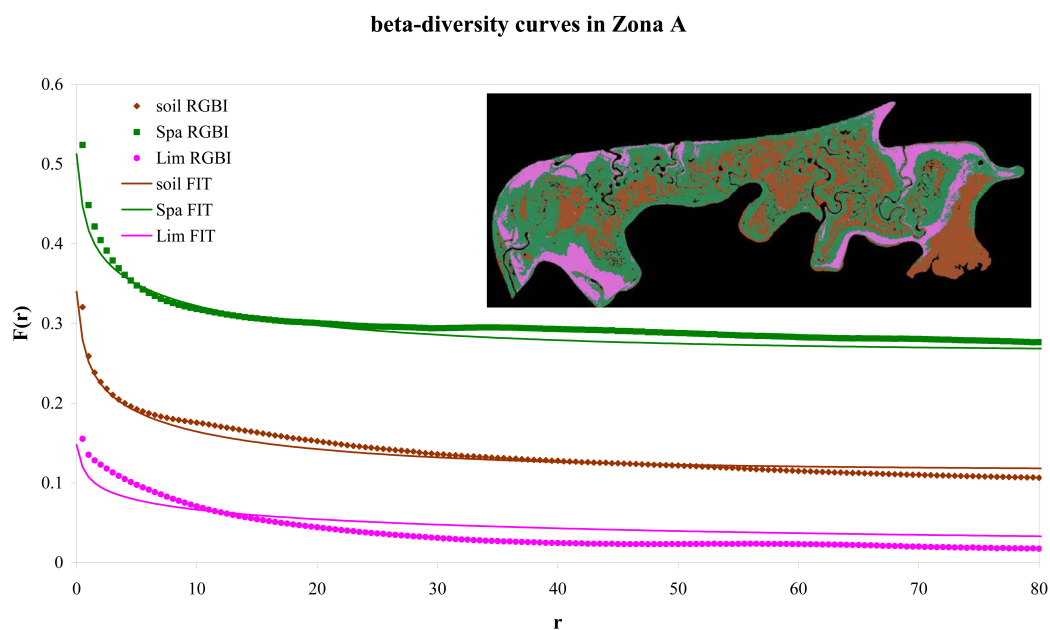


Figure 4.2: β -diversity curves for a subset (zone A) of S.Felice salt marsh (labeled with RGBi) and their theoretical fit.

Table 4.3: Parameters from the fit of β -diversity data in the subset of S.Felice named Zona A.

Species	γ	ζ
<i>spartina maritima</i>	1.00e-4	6.63e-5
<i>limonium narbonense</i>	3.93e-6	9.04e-6

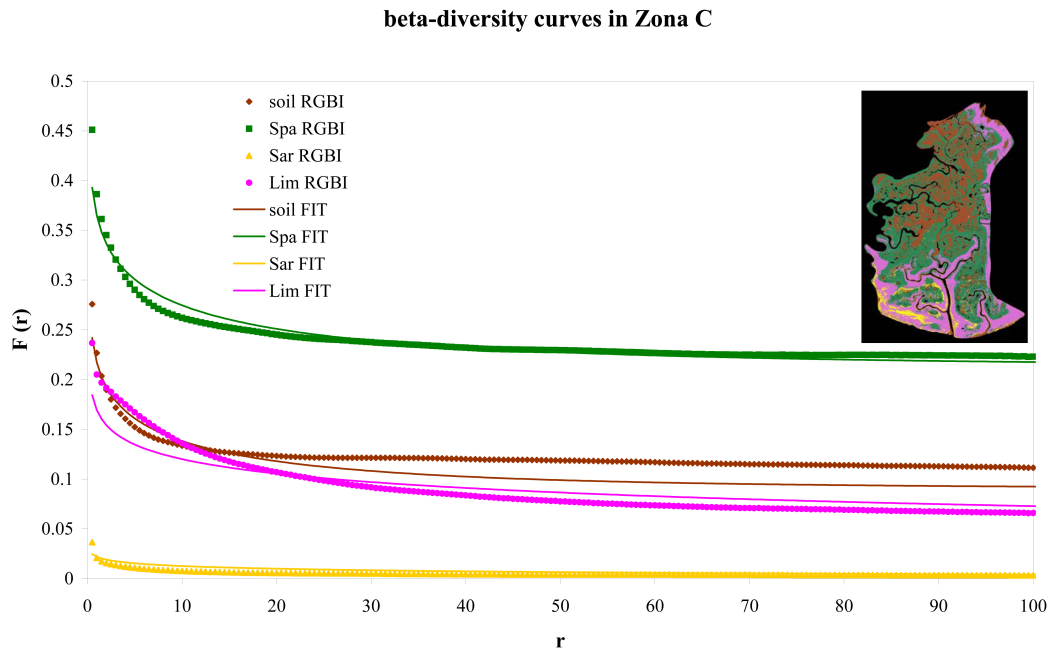


Figure 4.3: β -diversity curves for a subset (zone C) of S.Felice salt marsh (labeled with RGBi) and their theoretical fit.

Table 4.4: Parameters from the fit of β -diversity data in the subset of S.Felice named Zone C.

Species	γ	ζ
<i>spartina maritima</i>	9.07e-5	5.91e-5
<i>sarcocornia fruticosa</i>	4.31e-7	4.15e-6
<i>limonium narbonense</i>	9.32e-7	1.27e-6

Chapter 5

Conclusions and Perspectives

A scientist is somebody who constantly questions, generates falsifiable hypotheses, and collects data from well designed experiments – the kind of people who brush their teeth on only one side of their mouth to see whether brushing your teeth has any benefit.

(Richard Smith, in "Doctors Are Not Scientists"
(BMJ: 19 June 2004))

As we said in the Introduction of the present work, the field of spatial ecology is a young and still developing one. Much work is yet to be done, and many problems are still waiting for a solution.

What we have shown in the present work is that relatively simple models, with the important feature of being analytically solvable, are apt to fit experimental data with very good approximation. Some precisations are now due.

First and foremost, the model presented here are *simulations* of the reality, but not *emulations*. That is to say, we are able to reproduce some of the observed patterns, but we certainly do not claim that the dynamics of the system is exactly the one displayed by our models, nor that effects different from the ones modeled are absent in real ecosystems. What we claim is that others effect are, to a certain extent shown by the accuracy of the fits, ignorable *from a statistical point of view*. An example of this concept is the notion of neutrality in the tropical forests: it is clearly a hypothesis neglecting many details of the real dynamics. Anybody looking at a map of the distribution patterns of vegetation, such as Fig.1.5, will recognize instantly

that the various species have strong preferences regarding the terrain in which they live, and that these preferences vary among species. Also in the salt marshes the various species show the same behavior, due to the different degree of terrain salinity in which the plants prefer to grow. In our models we have ignored both the species preferences and whatsoever effect due to the terrain (all our models are solved in an uniform terrain).

Thus, it appears that many important characteristics of the dynamic are in fact negligible if we limit ourselves to the examination of β -diversity. A correct explanation of the causes of this observation should wait for some more accurate research, because the elements we have now are not sufficient to draw a definite conclusion. Nevertheless it seems useful to advance some hypothesis.

The β -diversity can be a “robust” function of the raw data, meaning that there are many configurations that give the same shape of the curves. This can also be rephrased as: the β -diversity does not characterize strongly the data set. Such an effect will allow many models, producing different configurations, to have the same β -diversity curve. An alternative, but closely related explanation is that other effects, absent in our models, are important at the level of the single plant/tree, but are statistical irrelevant when considering the β -diversity. Whether this fact is a coincidence or the ecosystem is tuned for some reason to this outcome, we do not know. What we can claim is that at the level of spatial similarity the system behaves “as is” the only relevant effects are the ones modeled (neutrality, diffusion and Janzen-Connell for tropical forests; non-neutrality, diffusion and birth from dormant seeds in salt marshes).

Secondly, in this work we have taken into consideration only the spatial turnover of species. One can easily verify that the voter model with speciation and Janzen-Connell effect is able to fit also the Relative Species Abundance (and in fact the resulting curve is rather similar to the solution of the Hubbell’s drift model, showing that the introduction of spatial explicitness has not a strong influence on some important quantity), or the Species-Area Relationship, but it is not possible to fit all three curves *at the same time*. In other words one has to change the parameter of the model depending on which quantity one wants to reproduce. This is a strong indicator that our model can’t be a an explanation of what really happens in a tropical forest, but only a “statistical approximation” of the reality.

The consideration above can be really puzzling: the models are good and at the same time they are not good. The explanation comes from the theory of critical phenomena: many models that have different 'microscopical' rules display the same behavior at large scale. This effect is called "universality". Universality is very useful because it permits the study of a simple (and often analytically solvable) model for a complex system, on the grounds that when we do the thermodynamic limit (i.e. when the system is really big) the simple model behaves alike the complex system. Now, it appears that our models shows universality properties for what concerns spatial turnover, but not for other quantities. It is just because of universality that we can't affirm that the dynamic in our models it's the real one: there will probably be some similar model with a slight different dynamic that yields results similar to ours. Our merit is to have found for the first time simple models capable to reproduce real data.

Finally, one can rightly ask what's the use of such models if they do not implement the real dynamic. As the Hubbell's Drift Model, our aim has been to obtain the greater amount of information possible from the simplest possible model. This way we are able to use our models as a null models. For instance, the original Voter Model with Speciation was a null model, and it was not able to reproduce the real data. Then, having such a null model, one is able to ask "what are we missing now?". It appeared evident that we lacked some negative density-dependent effect at short distance, and it was natural to introduce the Janzen-Connell effect. This allow us now to claim that the Janzen-Connell effect plays an important role in the spatial turnover of species, while other features, such as the terrain dishomogeneity seem to play a secondary role. In other words, the use of a null model lead us to an important insight regarding a complex system. So this is the use intended for the models presented in this work: to show the results of simple processes and to be a comparison against which we have to weight the data. Because it is impossible to understand the real significance of the data if we do not have some expectation or theory about them.

Appendix A

Lattice Correlation Functions for Voter Model with Speciation.

Math is like love – a simple idea but it can get complicated
(R.Drabek)

Here are contained the calculation for the solution of 1-point and 2-point correlation function of the Voter Model With Speciation on the lattice. There are (at least) two equivalent ways of obtaining the 1-point correlation function from Eq.(3.7): solving it with the method of generating function and by the use of Fourier Transform. Both are presented here in details. The 2-point correlation function will be obtained only by Fourier Transform.

A.1 1-Point Correlation Function (Generating Function).

This is a generalization of a calculation done by Krapivsky to solve the discrete voter model without speciation (i.e. with $\nu = 0$). See [49] for further details.

Let:

$$G(\vec{a}, t) \equiv \sum_{\vec{x}} \left(P_{\vec{x}}^t \prod_{i=1}^d a_i^{x_i} \right) \quad (\text{A.1})$$

Then Eq.(3.7) can be rewritten:

$$\frac{\partial G}{\partial t}(\vec{a}, t) = \left(\frac{1-\nu}{2d} \left(\sum_{i=1}^d (a_i + a_i^{-1}) - 2d \right) - \nu \right) G(\vec{a}, t) \quad (\text{A.2})$$

that can be easily solved with respect to time:

$$G(\vec{a}, t) = G(\vec{a}, 0) \exp \left[\left(\frac{1-\nu}{2d} \sum_{i=1}^d (a_i + a_i^{-1}) - 1 \right) t \right] \quad (\text{A.3})$$

Making use of

$$\sum_{j=-\infty}^{+\infty} a^j I_j(t) = e^{\frac{1}{2}(a+a^{-1})t} \quad (\text{A.4})$$

(see [60], 9.6.33), where $I_j(t)$ is the modified Bessel I function of order j , we can write:

$$G(\vec{a}, t) = G(\vec{a}, 0) e^{-t} \prod_{i=1}^d \sum_{j_i=-\infty}^{+\infty} a_i^{j_i} I_{j_i} \left(\frac{1-\nu}{d} t \right) \quad (\text{A.5})$$

Remembering that:

$$G(\vec{a}, 0) = \sum_{\vec{x}} \left(S_{\vec{x}}^0 \prod_{i=1}^d a_i^{x_i} \right) \quad (\text{A.6})$$

we have:

$$G(\vec{a}, t) = e^{-t} \sum_{\vec{y}} S_{\vec{y}}^0 \prod_{i=1}^d \sum_{j_i=-\infty}^{+\infty} a_i^{j_i+y_i} I_{j_i} \left(\frac{1-\nu}{d} t \right) \quad (\text{A.7})$$

$$= e^{-t} \sum_{\vec{y}} S_{\vec{y}}^0 \sum_{\vec{j}} a_i^{j_i} I_{j_i-y_i} \left(\frac{1-\nu}{d} t \right) \quad (\text{A.8})$$

Then, by confronting this last equation with the definition of $G(\vec{a}, t)$, we have the solution presented in the main text:

$$P_{\vec{x}}^t = e^{-t} \sum_{\vec{y}} S_{\vec{y}}^0 I_{\vec{x}-\vec{y}} \left(\frac{1-\nu}{d} t \right) \quad (\text{A.9})$$

where we have defined:

$$I_{\vec{x}}(t) \equiv \prod_{i=1}^d I_{x_i}(t) \quad (\text{A.10})$$

A.2 1-Point Correlation Function (Fourier Transform)

Let $\hat{P}^t(\vec{p})$ be the Fourier transform of $P_{\vec{x}}^t$, i.e.:

$$\hat{P}^t(\vec{p}) = \sum_{\vec{x}} e^{-i\vec{p}\cdot\vec{x}} P_{\vec{x}}^t \quad (\text{A.11})$$

$$P_{\vec{x}}^t = \frac{1}{L^d} \sum_{\vec{p}} e^{i\vec{p}\cdot\vec{x}} \hat{P}^t(\vec{p}) \quad (\text{A.12})$$

where:

$$\vec{p} = \frac{2\pi}{L} \vec{n} \quad (\text{A.13})$$

and \vec{n} is a vector which components $n_i = 0, \dots, L-1$.

If we let $L \rightarrow \infty$ then Eq.(A.12) becomes:

$$P_{\vec{x}}^t = \int_{-\pi < p_i < \pi} \frac{d^d \vec{p}}{(2\pi)^d} e^{i\vec{p}\cdot\vec{x}} \hat{P}^t(\vec{p}) \quad (\text{A.14})$$

Applying the Fourier transform to Eq.(3.6) we obtain:

$$\frac{\partial \hat{P}^t}{\partial t}(\vec{p}) = \left(-1 + \frac{1-\nu}{d} \sum_{i=1}^d \cos p_i \right) \hat{P}^t(\vec{p}) \quad (\text{A.15})$$

Solving as before with respect to time:

$$\hat{P}^t(\vec{p}) = \hat{P}^0(\vec{p}) \exp \left(-t + \frac{1-\nu}{d} t \sum_{i=1}^d \cos p_i \right) \quad (\text{A.16})$$

Anti-transforming:

$$P_{\vec{x}}^t = e^{-t} \int_{-\pi < p_i < \pi} \frac{d^d \vec{p}}{(2\pi)^d} e^{i\vec{p}\cdot\vec{x}} \hat{P}^0(\vec{p}) \exp \left(\frac{1-\nu}{d} t \sum_{i=1}^d \cos p_i \right) \quad (\text{A.17})$$

$$= e^{-t} \sum_{\vec{y}} P_{\vec{x}-\vec{y}}^0 \prod_{i=1}^d \int_{-\pi}^{\pi} \frac{dp_i}{(2\pi)} e^{ip_i y_i} \exp \left(\frac{1-\nu}{d} t \cos p_i \right) \quad (\text{A.18})$$

Then, using the integral representation:

$$I_n(z) = \frac{1}{\pi} \int_0^{\pi} d\vartheta e^{z \cos \vartheta} \cos(n\vartheta) \quad (\text{A.19})$$

(see [60], 9.6.19), we find again Eq.(3.8).

A.3 2-Point Correlation Function

We can solve the model directly on the lattice with periodic boundary conditions, going in Fourier transform (following the conventions like Eq.(A.12)). This way, the eq.(3.26) becomes:

$$\hat{F}^{n+1}(\vec{p}) = \hat{F}^n(\vec{p}) (1 + l(\vec{p})) - \int_{-\pi < q_i < \pi} \frac{d^d \vec{q}}{(2\pi)^d} \hat{F}^n(\vec{q}) l(\vec{q}) \quad (\text{A.20})$$

where:

$$l(\vec{p}) = -\frac{2}{N} + 2 \frac{1-\nu}{dN} \sum_{\mu=1}^d \cos(p_\mu) \quad (\text{A.21})$$

It's easy to see that $l(\vec{p}) < 0 \quad \forall \vec{p}$ if $0 < \nu \leq 1$.

We can now write down the solution at stationarity:

$$F_{\vec{x}} = g \int_{-\pi < p_i < \pi} \frac{d^d \vec{p}}{(2\pi)^d} e^{i\vec{p} \cdot \vec{x}} \frac{1}{l(\vec{p})} \quad (\text{A.22})$$

where:

$$g \equiv \int_{-\pi < q_i < \pi} \frac{d^d \vec{q}}{(2\pi)^d} \hat{F}(\vec{q}) l(\vec{q}) \quad (\text{A.23})$$

but imposing the boundary condition $F_0 = 1$ we can also write:

$$g = \left(\int_{-\pi < q_i < \pi} \frac{d^d \vec{q}}{(2\pi)^d} \frac{1}{l(\vec{q})} \right)^{-1} \quad (\text{A.24})$$

So we see that the multiplicative constant can be fixed in the lattice model.

This approach to the solution is suitable to give a closed form in 1 dimension. Let $d = 1$, $z \equiv e^{ip}$ and $r = |x|$:

$$F_r = g \int_{|z|=1} \frac{dz}{2iz\pi} \frac{z^r}{\frac{(1-\nu)}{2} \left(z + \frac{1}{z} \right) - 1} \quad (\text{A.25})$$

$$= \frac{g}{i\pi} \int_{|z|=1} dz \frac{z^r}{(1-\nu)z^2 - 2z + (1-\nu)} \quad (\text{A.26})$$

Where the path of integration is the unitary circle.

This can be computed using the theory of complex variables i.e the value of the integral is equal to the sum of the residues at the poles of the integrand inside the path of integration times $2\pi i$. The poles of the integrand are located at:

$$Z_{\pm} = \frac{1 \pm \sqrt{1 - (1-\nu)^2}}{1-\nu} \quad (\text{A.27})$$

Remembering that $0 \leq \nu < 1$ we see that both the solutions are reals and positive.

We can also see that in the case $\nu = 0$ our integral has no poles in the complex plane (because there is only one singularity in $z = 1$ and is not a pole, being of multiplicity 2) and so the integral is vanishing. This means that $g = \infty$ and so the problem can't be normalized.

Instead, when we have $0 < \nu < 1$ we have 2 simple poles. To see if this poles are inside the unitary circle we simply have to check if $Z_{\pm} < 1$. We have that Z_- is always inside the circle and Z_+ always outside. So we have:

$$F_r = \frac{g}{i\pi} 2\pi i \frac{1}{1-\nu} \frac{Z_-^r}{Z_- - Z_+} \quad (\text{A.28})$$

$$= \frac{2g}{1-\nu} \frac{Z_-^r}{Z_- - Z_+} \quad (\text{A.29})$$

Imposing g such that $F_0 = 1$ at the end we have:

$$F_r = Z_-^r \quad (\text{A.30})$$

$$= \left(\frac{1 - \sqrt{\nu(2-\nu)}}{1-\nu} \right)^r \quad (\text{A.31})$$

$$= e^{-r/\xi} \quad (\text{A.32})$$

where:

$$\xi = \left(\ln \frac{1-\nu}{1 - \sqrt{\nu(2-\nu)}} \right)^{-1} \quad (\text{A.33})$$

is the typical length of the solution.

Appendix B

Relation Between $\sum_{\vec{x}} F_{\vec{x}}$ and Moments Ratio.

We want to demonstrate that on the lattice:

$$\sum_{\vec{x}} F_{\vec{x}} = \frac{\langle n^2 \rangle}{\langle n \rangle} \quad (\text{B.1})$$

First we remark that (by traslational invariance of F_x):

$$\frac{1}{L^d} \sum_{x,y} P_{x,y} = \sum_x F_x \quad (\text{B.2})$$

where $P_{x,y}$ is the probability that the trees at x and at y share the same species. Then, denoting with σ_x the species of the trees in the site x , we have:

$$\begin{aligned} \frac{1}{L^d} \sum_{x,y} P_{x,y} &= \frac{1}{L^d} \sum_{x,y} \delta_{\sigma_x, \sigma_y} \\ &= \frac{1}{L^d} \left(L^d + \sum_{x \neq y} \delta_{\sigma_x, \sigma_y} \right) \end{aligned} \quad (\text{B.3})$$

The last summation can be evaluated in the following way: if the species σ_x has k individuals, the sum over y yield $k - 1$. If now we sum over all sites x in which there is an individual of the same species, we obtain $k(k - 1)$. So we can write:

$$\frac{1}{L^d} \sum_{x,y} P_{x,y} = 1 + \frac{1}{L^d} \sum_{\text{species}} k(k - 1)$$

$$\begin{aligned}
&= 1 + \frac{1}{L^d} \sum_k f_k k(k-1) \\
&= 1 + \frac{S}{L^d} \sum_k \phi_k k(k-1) \tag{B.4}
\end{aligned}$$

where f_k is the number of species with k individuals; while ϕ_k is the fraction of species with k individual and S the number of species in the sample. This three quantities are related by $f_k = S\phi_k$.

We can simplify the last result noting that:

$$\sum_k f_k k = L^d \tag{B.5}$$

and:

$$\sum_k \phi_k k^i = \langle n^i \rangle \tag{B.6}$$

Substituting this two relation, along with $L^d/S = \langle n \rangle$, in Eq.(B.4) we obtain our thesis.

Appendix C

Asymptotic Behavior of $\sum_{\vec{x}} F_{\vec{x}}$

We already know the exact solution for the β -diversity function on the lattice, Eq.(3.22). His Fourier transform is:

$$\hat{F}(\vec{p}) = \frac{1}{l(\vec{p})} \left(\int_{-\pi}^{\pi} \frac{d^d q}{(2\pi)^d} \frac{1}{l(\vec{q})} \right)^{-1} \quad (\text{C.1})$$

with $l(\vec{p})$ given by Eq.(3.23). Through some manipulation, we can rewrite:

$$\hat{F}(\vec{p}) = \frac{1}{K(\vec{p})} \left(\int_{-\pi}^{\pi} \frac{d^d q}{(2\pi)^d} \frac{1}{K(\vec{q})} \right)^{-1} \quad (\text{C.2})$$

with:

$$K(\vec{p}) = \nu + \frac{1-\nu}{d} \sum_{\mu=1}^d (1 - \cos p_{\mu}) \quad (\text{C.3})$$

Now:

$$\begin{aligned} \sum_{\vec{x}} F_{\vec{x}} &= \hat{F}(\vec{p}=0) \\ &= \frac{1}{K(0)} \left(\int_{-\pi}^{\pi} \frac{d^d q}{(2\pi)^d} \frac{1}{K(\vec{q})} \right)^{-1} \\ &\sim \frac{d}{\nu(2\pi)^d} \left(\int_{-\pi}^{\pi} d^d q \frac{1}{d\nu + q^2} \right)^{-1} \end{aligned} \quad (\text{C.4})$$

where we have used the following approximation: for $\nu \rightarrow 0$ and $p \rightarrow 0$ we have $K(\vec{p}) \sim \nu + p^2/d + O(p^4)$. Now we distinguish two cases:

Case 1 $d < 2$: We change variable in the integral: $s = \sqrt{d\nu} q$ so we have:

$$\begin{aligned} \int_{-\pi}^{\pi} d^d q \frac{1}{d\nu + q^2} &= (d\nu)^{\frac{d-2}{2}} \int_{-\pi/\sqrt{d\nu}}^{\pi/\sqrt{d\nu}} d^d s \frac{1}{1+s^2} \\ &\simeq (d\nu)^{\frac{d-2}{2}} \int_{-\infty}^{\infty} d^d s \frac{1}{1+s^2} \\ &\sim \nu^{\frac{d-2}{2}} \end{aligned} \quad (\text{C.5})$$

Then:

$$\sum_{\vec{x}} F_{\vec{x}} \sim \nu^{-d/2} \quad (\text{C.6})$$

Case 2 $d > 2$: We carry out the limit $\nu \rightarrow 0$ in the integral:

$$\begin{aligned} \int_{-\pi}^{\pi} d^d q \frac{1}{d\nu + q^2} &\rightarrow \int_{-\pi}^{\pi} d^d q \frac{1}{q^2} \\ &\sim \int_0^{\pi} dq q^{d-3} \end{aligned} \quad (\text{C.7})$$

and we can easily see that the singularity in $\vec{q} = 0$ is harmless and the integral behaves like a constant for $\nu \rightarrow 0$. Then we have:

$$\sum_{\vec{x}} F_{\vec{x}} \sim \nu^{-1} \quad (\text{C.8})$$

Appendix D

Scaling Relations

We need to study the equations:

$$S = \sum_n f(n, \nu, A) \sim \nu^{a(\beta-1)} \int_{\nu^a}^{\infty} dx x^{-\beta} \hat{f}(x, A\nu^{d/2}) \quad (\text{D.1})$$

$$A = \sum_n n f(n, \nu, A) \sim \nu^{a(\beta-2)} \int_{\nu^a}^{\infty} dx x^{1-\beta} \hat{f}(x, A\nu^{d/2}) \quad (\text{D.2})$$

So we distinguish three cases:

D.1 $1 + c - \beta > 0$

In this case both the integrals in Eq.(D.1) and Eq.(D.2) converge to a constant as $\nu \rightarrow 0$. Then we have:

$$S \sim \nu^{a(\beta-1)} F_S(A\nu^{d/2}) \quad (\text{D.3})$$

$$A \sim \nu^{a(\beta-2)} F_A(A\nu^{d/2}) \quad (\text{D.4})$$

and from the second of this relations we can easily see that $F_A(y) \sim y$ and $a(\beta - 2) = d/2$. Using our assumption for a , i.e. Eq.(3.44), we have:

$$\beta = 2 - \frac{d}{2a} = \begin{cases} 1 & d < 2 \\ \frac{4-d}{2} & d > 2 \end{cases} \quad (\text{D.5})$$

Moreover, confronting Eq.(D.3) with Eq.(3.40) we have:

$$z = -\frac{2}{d}a(\beta - 1) = \begin{cases} 0 & d < 2 \\ \frac{d-2}{d} & d > 2 \end{cases} \quad (\text{D.6})$$

Now, dividing member-to-member Eq.(3.39) by Eq.(D.3) we obtain:

$$\phi(n, \nu, A) = n^{-1} \hat{\phi}(n\nu^a, A\nu^{d/2}) \quad (\text{D.7})$$

and, by confrontation with Eq.(3.41), we find $b = 1$.

D.2 $-1 < 1 + c - \beta < 0$

In this case the integrals in Eq.(D.2) converges but the one in Eq.(D.2) does not converge to a constant as $\nu \rightarrow 0$. Then we have:

$$S \sim \nu^{ac} F_S(A\nu^{d/2}) \quad (\text{D.8})$$

$$A \sim \nu^{a(\beta-2)} F_A(A\nu^{d/2}) \quad (\text{D.9})$$

and from the second of this relations we can easily see that $F_A(y) \sim y$ and $a(\beta - 2) = d/2$ as above. Using our assumption for a , i.e. Eq.(3.44), we have again:

$$\beta = 2 - \frac{d}{2a} = \begin{cases} 1 & d < 2 \\ \frac{4-d}{2} & d > 2 \end{cases} \quad (\text{D.10})$$

But now, confronting Eq.(D.8) with Eq.(3.40) we have:

$$z = -\frac{2ac}{d} = \begin{cases} -c & d < 2 \\ \frac{-2c}{d} & d > 2 \end{cases} \quad (\text{D.11})$$

Again, dividing member-to-member Eq.(3.39) by Eq.(D.8) we obtain:

$$\phi(n, \nu, A) = n^{c-\beta} \hat{\phi}(n\nu^a, A\nu^{d/2}) \quad (\text{D.12})$$

and, by confrontation with Eq.(3.41), we find:

$$b = \beta - c = \begin{cases} 1 - c = 1 + z & d < 2 \\ 2 - c - \frac{d}{2} = 2 - \frac{d}{2}(1 - z) & d > 2 \end{cases} \quad (\text{D.13})$$

With a bit of algebra, this last equation can be rewritten as a scaling relation:

$$a(2 - b) = \frac{d}{2}(1 - z) \quad (\text{D.14})$$

D.3 $1 + c - \beta < -1$

In this case both the integrals in Eq.(D.1) and Eq.(D.2) don't converge to a constant as $\nu \rightarrow 0$. Then we have:

$$S \sim \nu^{ac} F_S(A\nu^{d/2}) \quad (\text{D.15})$$

$$A \sim \nu^{ac} F_A(A\nu^{d/2}) \quad (\text{D.16})$$

By the second of these equations we have $F_A(y) \sim y$ and:

$$c = -\frac{ad}{2} = \begin{cases} -1 & d < 2 \\ -\frac{d}{2} & d > 2 \end{cases} \quad (\text{D.17})$$

Then, confronting Eq.(D.15) with Eq.(3.40) we have:

$$z = -\frac{2ac}{d} = 1 \quad (\text{D.18})$$

Moreover we have, proceeding as before, $b = \beta - c > 2$.

Appendix E

Solution of Differential Equation in J.C. Voter Model

In this section we will restrict to the case $d = 2$, but the procedure is fairly general and can be repeated for different values of d with only a few minor changes. As seen in the main text, a stationary solution of Eq.(3.21) is the modified Bessel function K_0 (Eq.(3.36)). Of course, this is not the only solution; indeed all function that satisfy:

$$- D\nabla^2 F(\vec{x}) + \alpha F(\vec{x}) = b\delta^d(\vec{x}) \quad (\text{E.1})$$

are valid solutions. Now, consider the homogeneous equation:

$$- D\nabla^2 H(\vec{x}) + \alpha H(\vec{x}) = 0. \quad (\text{E.2})$$

All the functions of the form

$$F(\vec{x}) + cH(\vec{x}), \quad (\text{E.3})$$

where F is a particular solution of (E.1), H is a solution of (E.2) and c is a constant, are solutions of (E.1). We have already found a particular solution F of (E.1) solving the standard voter model, so now we need only to solve (E.2). We will assume $H(\vec{x}) = H(r)$, $r \equiv |\vec{x}|$. Then one can write

$$\nabla^2 H(r) = r^{1-d} \frac{\partial}{\partial r} \left(r^{d-1} \frac{\partial}{\partial r} H(r) \right) \quad (\text{E.4})$$

$$= \frac{d-1}{r} H'(r) + H''(r). \quad (\text{E.5})$$

We define $G(r)$ such that $H(r) \equiv r^\zeta G(r)$, and from (E.2) we get an equation for $G(r)$:

$$r^2 G''(r) + r(2\zeta + d - 1)G'(r) + (\zeta(d - 2 + \zeta) - r^2 \gamma^2)G(r) = 0 \quad (\text{E.6})$$

where $\gamma^2 = \alpha/D$. Our goal is to find a solution that is well behaved at the origin. Recall that ζ is a free variable – we choose $2\zeta + d - 1 = 1$, i.e. $\zeta = \frac{2-d}{2}$. This leads to

$$r^2 G'''(r) + rG''(r) - (\gamma^2 r^2 + \zeta^2)G(r) = 0 \quad (\text{E.7})$$

and setting $\gamma r \equiv t$ and $G(r) \equiv y(t)$ we find:

$$t^2 y''(r) + t y'(t) - (t^2 + \zeta^2)y(t) = 0 \quad (\text{E.8})$$

Setting $d = 2$ yields $\zeta = 0$ and the general solution of the equation is

$$y(t) = C_1 I_0(t) + C_2 K_0(t), \quad (\text{E.9})$$

where K_0 and I_0 are the modified Bessel functions. $C_2 = 0$ because $K_0(t)$ is not regular in the origin. Thus the general solution of (E.1) is

$$\frac{s\gamma_0^2}{2\pi} K_0(\gamma_0 r) + c I_0(\gamma_0 r) \quad (\text{E.10})$$

Appendix F

Correlation Functions for Intertidal Vegetation

F.1 1-Point Correlation Function

We want to obtain a close equation for the 1-point correlation function:

$$P_x^n(\tau) \equiv \langle \delta_{\tau, \sigma_x} \rangle^{(n)} = \sum_{\vec{\sigma}} P^n(\vec{\sigma}) \delta_{\tau, \sigma_x} \quad (\text{F.1})$$

(remark that $\sum_{\tau} P_x^n(\tau) = 1$). Thus:

$$P_v^{n+1}(\tau) = P_v^n(\tau) + \sum_x \sum_{\vec{\sigma}, \vec{\sigma}'} [W_{\vec{\sigma}, \vec{\sigma}'}(x) P^n(\vec{\sigma}') - W_{\vec{\sigma}', \vec{\sigma}}(x) P^n(\vec{\sigma})] \delta_{\tau, \sigma_v} \quad (\text{F.2})$$

The contribution from $x \neq v$ is zero since Eq.(4.3) implies that $\sigma_v = \sigma'_v$. Indeed, interchanging $\vec{\sigma}$ with $\vec{\sigma}'$:

$$\sum_x \sum_{\vec{\sigma}, \vec{\sigma}'} W_{\vec{\sigma}, \vec{\sigma}'}(x) P^n(\vec{\sigma}') \delta_{\tau, \sigma_v} = \sum_x \sum_{\vec{\sigma}, \vec{\sigma}'} W_{\vec{\sigma}', \vec{\sigma}}(x) P^n(\vec{\sigma}) \delta_{\tau, \sigma'_v} \quad (\text{F.3})$$

$$= \sum_x \sum_{\vec{\sigma}, \vec{\sigma}'} W_{\vec{\sigma}', \vec{\sigma}}(x) P^n(\vec{\sigma}) \delta_{\tau, \sigma_v} \quad (\text{F.4})$$

Thus it can be easily seen (using Eq.(4.4) and Eq.(4.5)) that:

$$P_x^{n+1}(\tau) = (1 - C) P_x^n(\tau) + C \sum_a \left[I_{\tau, a} P_x^n(a) + \sum_{z(x)} N_{\tau, a} P_z^n(a) \right] \quad (\text{F.5})$$

It should be noted that changing the constant $C > 0$ we can obtain different equations for the same process by changing only the time scale, i.e. the properties at equilibrium will not be affected.

F.2 2-Point Correlation Function

Proceeding in the same way one can obtain a close form for the 2-point correlation function:

$$P_{x,y}^n(\varepsilon, \tau) \equiv \langle \delta_{\varepsilon, \sigma(x)} \delta_{\tau, \sigma(y)} \rangle \quad (\text{F.6})$$

i.e. the probability that at the n -th time step x and y are occupied by ε and τ respectively. Remark that $P_{x,x}^n(\varepsilon, \tau) = \delta_{\varepsilon, \tau} P_x^n(\tau)$ and that

$$\sum_{\varepsilon, \tau} P_{x,y}^n(\varepsilon, \tau) = 1 \quad (\text{F.7})$$

By the same procedure described before it is obtained (with $x \neq y$):

$$\begin{aligned} P_{x,y}^{n+1}(\varepsilon, \tau) &= (1 - 2C)P_{x,y}^n(\varepsilon, \tau) + \\ &+ C \sum_a \left[I_{\varepsilon,a} P_{x,y}^n(a, \tau) + I_{\tau,a} P_{x,y}^n(\varepsilon, a) + \right. \\ &\left. + \sum_{z(x)} N_{\varepsilon,a}(x) P_{z,y}^n(a, \tau) + \sum_{z(y)} N_{\tau,a}(y) P_{z,y}^n(\varepsilon, a) \right] \quad (\text{F.8}) \end{aligned}$$

F.3 Stationary Solutions

At stationarity, in a square d -dimensional lattice with periodic boundary conditions, we look for a solution which is translationally invariant: $P_x^n(\tau) \equiv P(\tau)$. Eq.(F.5) then becomes:

$$P(\tau) = \sum_{\varepsilon} \left(kL_{\tau,\varepsilon}^{(1)} + (1 - k)L_{\tau,\varepsilon}^{(2)} \right) P(\varepsilon) \quad (\text{F.9})$$

Remark that this equation has always at least one solution. This can be seen with the following argument: in order to be a solution, $P(\tau)$ has to be a (right) eigenvector of the matrix

$$A_{\tau,\varepsilon} = \delta_{\tau,\varepsilon} - \left(kL_{\tau,\varepsilon}^{(1)} + (1 - k)L_{\tau,\varepsilon}^{(2)} \right) \quad (\text{F.10})$$

and his eigenvalue has to be zero. Now remember that the sum of each column of the matrices $L^{(i)}$ has to be 1, by definition (Eq.(4.8)). Then each column of A must sum to zero. This in turn implies that there is a left eigenvector with zero as eigenvalue, and so that $\det(A) = 0$, then there exist at least one right eigenvector with eigenvalue zero.

Analogously, we define $P_{x,y}^n(\varepsilon, \tau) = \hat{F}_{x-y}^n(\varepsilon, \tau)$. Then it is convenient to use a symmetrized version of this quantity as:

$$F_x(\varepsilon, \tau) = \frac{1}{2} \left(\hat{F}_x(\varepsilon, \tau) + \hat{F}_x(\tau, \varepsilon) \right) \quad (\text{F.11})$$

(so, obviously, $F_x(\varepsilon, \tau) = F_x(\tau, \varepsilon)$). With the help of this last definition Eq.(F.8) becomes (when $x \neq 0$):

$$\begin{aligned} F_x^{n+1}(\varepsilon, \tau) &= (1 - 2C)F_x^n(\varepsilon, \tau) + \\ &+ C \sum_a \left[kL_{\varepsilon,a}^{(1)}F_x(a, \tau) + kL_{\tau,a}^{(1)}F_x(\varepsilon, a) + \right. \\ &+ \frac{1-k}{2d} \sum_{i=1}^d \left(L_{\varepsilon,a}^{(2)} \left(F_{x+e_i}(a, \tau) + F_{x-e_i}(a, \tau) \right) + \right. \\ &\left. \left. + L_{\tau,a}^{(2)} \left(F_{x+e_i}(\varepsilon, a) + F_{x-e_i}(\varepsilon, a) \right) \right) \right] \quad (\text{F.12}) \end{aligned}$$

and so, for a stationary solution:

$$\begin{aligned} F_x(\varepsilon, \tau) &= \frac{1}{2} \sum_a \left[kL_{\varepsilon,a}^{(1)}F_x(a, \tau) + kL_{\tau,a}^{(1)}F_x(\varepsilon, a) + \right. \\ &+ \frac{1-k}{2d} \sum_{i=1}^d \left(L_{\varepsilon,a}^{(2)} \left(F_{x+e_i}(a, \tau) + F_{x-e_i}(a, \tau) \right) + \right. \\ &\left. \left. + L_{\tau,a}^{(2)} \left(F_{x+e_i}(\varepsilon, a) + F_{x-e_i}(\varepsilon, a) \right) \right) \right] \quad (\text{F.13}) \end{aligned}$$

where d is the space dimensionality of the system (and $q(x) = 2d$), and e_i are the lattice versors. When $x = 0$ instead, we have:

$$F_0(\varepsilon, \tau) = \delta_{\varepsilon,\tau} P(\tau) \quad (\text{F.14})$$

Remark that both the 1-point correlation function and the 2-point correlation function don't depend on C at equilibrium, as stated before.

F.4 Analytic Solution for 2-point Correlation Function

From now on we will assume that:

$$L^{(2)} = \mathbf{1} \quad (\text{F.15})$$

This assumption has an important physical meaning: each individual can give birth only to conspecific offsprings.

With this assumption we can write Eq.(F.12) as:

$$F_x^{n+1} - F_x^n = \frac{1-k}{d} [\Delta F_x^n + \mathbf{M}F_x^n] \quad (\text{F.16})$$

where

$$\mathbf{M} = \frac{dk}{1-k} [L^{(1)} \otimes \mathbf{1} + \mathbf{1} \otimes L^{(1)} - 2\mathbf{1} \otimes \mathbf{1}] \quad (\text{F.17})$$

and

$$\Delta F_x \equiv \sum_{i=1}^d [F(x + e_i) + F(x - e_i) - 2F(x)] \quad (\text{F.18})$$

Then, the stationary solution should satisfy:

$$\Delta F_x + \mathbf{M}F_x = 0 \quad (\text{F.19})$$

To solve this last equation we need to find a complete system of left and right eigenvalues of $L^{(1)}$, i.e. we need to find two system of vectors t_a^R , and t_a^L such that:

$$L^{(1)} t_a^R = \lambda_a t_a^R \quad (\text{F.20})$$

$$t_a^L L^{(1)} = \lambda_a t_a^L \quad (\text{F.21})$$

$$t_a^L \cdot t_b^R = \delta_{a,b} \quad (\text{F.22})$$

where λ_a are the eigenvalues of $L^{(1)}$. We also define:

$$v_{a,b}^A \equiv t_a^A \otimes t_b^A \quad (\text{F.23})$$

where $A = L, R$. Remark that:

$$v_{a,b}^L \cdot v_{c,d}^R = (t_a^L \cdot t_c^R) (t_b^L \cdot t_d^R) = \delta_{a,c} \delta_{b,d} \quad (\text{F.24})$$

Now, the vectors $v_{a,b}^A$ are the right and left eigenvectors of \mathbf{M} with eigenvalues:

$$\mu_{a,b} = \frac{dk}{1-k} (\lambda_a + \lambda_b - 2) \quad (\text{F.25})$$

Let's now define:

$$f_{a,b}(x) = v_{a,b}^L \cdot F_x \quad (\text{F.26})$$

thus:

$$F_x = \sum_{a,b} f_{a,b}(x) v_{a,b}^R \quad (\text{F.27})$$

Now, left multiplying Eq.(F.19) by $v_{a,b}^L$ one obtains:

$$\Delta f_{a,b}(x) + \mu_{a,b} f_{a,b}(x) = 0 \quad (\text{F.28})$$

This is just a numerical equation (not a vectorial one like before) and so it's easily solved going in Fourier Transform. Let's also decompose $f_{a,b}$ in:

$$f_{a,b}(x) = c_{a,b} f(x, \mu_{a,b}) \quad (\text{F.29})$$

where $f(x=0, \mu_{a,b}) \equiv 1$ by definition (or equivalently $c_{a,b} \equiv f_{a,b}(x=0)$).

And so we can write the solution of our model as:

$$F_x = \sum_{a,b} v_{a,b}^R c_{a,b} f(x, \mu_{a,b}) \quad (\text{F.30})$$

We'll now divide our derivation in 2 parts: in the tensorial part we'll derive an explicit expression for $c_{a,b}$ in terms of the vectors t_a^A , in the spatial part we'll study the behavior of $f(x, \mu_{a,b})$.

F.4.1 tensorial part

First of all we introduce the following notation:

$$(t_a^A)_i \quad (\text{F.31})$$

is the i -th component of the a -th eigenvector of type $A = L, R$ of the matrix $L^{(1)}$, as described above. If we denote similarly the components of the vectors v described above, we have:

$$(v_{a,b}^A)_{i,j} = (t_a^A)_i (t_b^A)_j \quad (\text{F.32})$$

Moreover, due to Eq.(4.8), we know for sure that there will be an eigenvalue $\lambda_0 = 1$. Their eigenvectors will be denoted with t_0^A , and will play a special role in our theory¹. Remark also that $\mu_{0,0} = 0$. From Eq.(4.8) we can also find that:

$$(t_0^L)_i = 1 \quad \forall i \quad (\text{F.33})$$

i.e. the left eigenvector with zero eigenvalue has all his component equal to 1². From Eq.(F.14) and Eq.(4.11) it's easy to calculate that:

$$F_x(i, j) = \delta_{i,j} (t_0^R)_i \quad (\text{F.34})$$

¹for instance $(t_0^R)_i$ will be the density of species i inside the plot considered

²Rigorously, there can be more than one eigenvector with a zero eigenvalue. In this case we will denote as t_0^L the one with all his components equal to 1, which is guaranteed to exist.

Remember also that from Eq.(F.29) we have that:

$$\begin{aligned}
c_{a,b} &= v_{a,b}^L \cdot F_0 \\
&= \sum_{i,j} \delta_{i,j} (t_0^R)_i (t_a^L)_i (t_b^L)_j \\
&= \sum_j (t_0^R)_j (t_a^L)_j (t_b^L)_j
\end{aligned} \tag{F.35}$$

and so we have an explicit formula for $c_{a,b}$.

F.4.2 spatial part

Now we want to solve the equation:

$$\Delta f(x, \mu_{a,b}) + \mu_{a,b} f(x, \mu_{a,b}) = 0 \tag{F.36}$$

with: $f(x = 0, \mu_{a,b}) = 1$. This can also be written as:

$$A_x f(x, \mu_{a,b}) - \delta_{x,0} [(A_x - 1) f(x, \mu_{a,b}) + 1] = 0 \tag{F.37}$$

where:

$$A_x = \Delta + \mu_{a,b} \tag{F.38}$$

Defining the Fourier Transform of $f(x, \mu_{a,b})$ as:

$$\hat{f}(p, \mu_{a,b}) = \sum_x e^{-ipx} f(x, \mu_{a,b}) \tag{F.39}$$

we have:

$$\hat{f}(p, \mu_{a,b}) l(p) - \int_{-\pi}^{\pi} \frac{d^d p}{(2\pi)^d} \hat{f}(p, \mu_{a,b}) l(p) = 0 \tag{F.40}$$

with:

$$l(p) = 2 \sum_{i=1}^d (\cos p_i - 1) + \mu_{a,b} \tag{F.41}$$

Then the solution is:

$$f(x, \mu_{a,b}) = g \int_{-\pi}^{\pi} \frac{d^d p}{(2\pi)^d} \frac{e^{ipx}}{l(p)} \tag{F.42}$$

where:

$$g = \int_{-\pi}^{\pi} \frac{d^d p}{(2\pi)^d} l(p) \hat{f}(p, \mu_{a,b}) \tag{F.43}$$

but imposing the condition at $x = 0$ one can also write:

$$g = \left(\int_{-\pi}^{\pi} \frac{d^d p}{(2\pi)^d} \frac{1}{l(p)} \right)^{-1} \quad (\text{F.44})$$

Then we are left to evaluate the integral:

$$I(x) = \int_{-\pi}^{\pi} \frac{d^d p}{(2\pi)^d} \frac{e^{ipx}}{\sum_{i=1}^d (1 - \cos p_i) + \lambda_{a,b}} \quad (\text{F.45})$$

with $\lambda_{a,b} = -\mu_{a,b}/2 \geq 0$ ³ Using the formula:

$$\frac{1}{a} = \int_0^{\infty} dt e^{-at} \quad (\text{F.46})$$

one have:

$$\begin{aligned} I(x) &= \int_0^{\infty} dt \int_{-\pi}^{\pi} \frac{d^d p}{(2\pi)^d} e^{ipx} e^{-\left[\sum_{i=1}^d (1 - \cos p_i) + \lambda_{a,b}\right]t} \\ &= \int_0^{\infty} dt e^{-(d+\lambda_{a,b})t} \prod_{i=1}^d \int_{-\pi}^{\pi} \frac{dp_i}{(2\pi)} e^{ip_i x_i + t \cos p_i} \\ &= \int_0^{\infty} dt e^{-(d+\lambda_{a,b})t} \prod_{i=1}^d \int_0^{\pi} \frac{dp_i}{(2\pi)} e^{t \cos p_i} \cos(p_i x_i) \\ &= \int_0^{\infty} dt e^{-(d+\lambda_{a,b})t} \prod_{i=1}^d I_{x_i}(t) \end{aligned} \quad (\text{F.47})$$

where $I_{x_i}(t)$ is a modified Bessel I function of order x_i .

For $d = 1$ we have that:

$$\int_0^{\infty} dt e^{-(1+\lambda)t} I_x(t) = \frac{\left(1 + \lambda + \sqrt{\lambda(2 + \lambda)}\right)^{-x}}{\sqrt{\lambda(2 + \lambda)}} \quad (\text{F.48})$$

and so:

$$f(x, \mu_{a,b}) = \left(1 - \mu_{a,b}/2 + \sqrt{\mu_{a,b}/2 (\mu_{a,b}/2 - 2)}\right)^{-x} \quad (\text{F.49})$$

For $d \geq 2$ this solution does not have a closed form, but it is very easy to calculate numerically.

³it is very easy to show that the matrix $L^{(2)} - 1$ is definite nonpositive, and as such all his eigenvalues are negative or zero, and the same is true for all $\mu_{a,b}$.

Bibliography

- [1] Thomas Malthus , *An Essay on the Principle of Population as it Affects the Future Improvement of Society with Remarks on the Speculations of Mr. Godwin, M. Condorcet, and Other Writers*. London, printed for J. Johnson, in St. Paul's church-yard (1798).
- [2] see www.ctfs.si.edu for the project web page.
- [3] the project web page is at www.tideproject.org
- [4] J.Chave, S.Levin. Scale and scaling in ecological and economic systems. *Environmental and Resource Economics*, **26**, 527-557 (2003).
- [5] J.Maynard Smith. *Models in Ecology*. Cambridge University Press (1974).
- [6] R.Condit, S.P.Hubbell, R.B.Foster. Changes in tree species abundance in a neotropical forest: Impact of climate change. *Journal of Tropical Ecology*, **12**, 231-256, (1996).
- [7] R.H.MacArthur, E.O.Wilson *The Theory of Island Biogeography*. Monographs in Population Biology. (Princeton Univ. Press, Princeton, NJ, 1967).
- [8] Brown, J.H., and A.C. Gibson, *Biogeography*. St. Louis: Mosby (1983).
- [9] S.P.Hubbell, *The Unified Neutral Theory of Biodiversity and Biogeography* (Princeton Univ. Press, Princeton, NJ, 2001).
- [10] A.S.Corbet, The distribution of butterflies in the Malay Peninsula, *Proceedings of the Royal Society, London, Series A* **16**: 101-116 (Chap.2) (1941).

- [11] R.A.Fisher, A.S.Corbet, C.B.Williams, The relation between the number of species and the number of individuals in a random sample of an animal population. *Journal of Animal Ecology* **12** 42-58 (1943).
- [12] F.W.Preston, The commonness, and rarity, of species. *Ecology* **29**: 254-283 (1948).
- [13] S.Engen R.Lande Population Dynamic Models Generating the Log-normal Species Abundance Distribution. *Mathematical Biosciences* **132**, 169-183 (1996).
- [14] S.Pigolotti, A.Flammini, A.Maritan. *Physical Review E*, (2004).
- [15] R.Durrett, S.Levin. Spatial models for species-area curves. *J.Theor.Biol.*, **179**:119-127 (1996).
- [16] J.Harte, A.Kinzig, J.Green. Self-similarity in the distribution and abundance of species. *Nature*, 424:1006-1007 (2003).
- [17] J.Harte, S.McCarthy, K.Taylor, A.Kinzig, and M.L.Fisher, *Oikos* **86**, 45 (1999).
- [18] R.Condit *et al.*. Beta-Diversity in Tropical Forest Trees, *Science* **295**, 666 (2002).
- [19] R.Condit *et al.*. Spatial Patterns in the Distribution of Tropical Species, *Science* **288**, 1414 (2000).
- [20] R.Condit *Science* **290**, 671a (2000).
- [21] J.Duivenvoorden, J.C.Svenning, S.J.Wright. *Science* **295** 636 (2002).
- [22] Data freely available at <http://ctfs.si.edu/datasets/bci/>
- [23] J.Chave, E.G.Leigh Jr. A Spatially Explicit Neutral Model of β -Diversity in Tropical Forests, *Theoretical Population Biology* **62**, 153-168 (2002).
- [24] Cronk, J.K. & Fennessy, M.S. (2001). *Wetland Plants: biology and ecology*. Boca Raton, Florida: CRC Press

- [25] Ungar, I.A., *Ecophysiology of Vascular Halophytes*, CRC Press (August 23, 1991).
- [26] J.van de Koppel *et al.* *American Naturalist* **165** (2005).
- [27] Marani, M., Silvestri, S., Belluco, E., Camuffo, M., D'Alpaos, A., Defina, A., Lanzoni, S., Marani, A., Tortato, M., Rinaldo, A. (2003). Patterns in tidal environments: salt-marsh channel networks and vegetation. In: Proceedings of the IEEE International Geoscience and Remote Sensing Symposium, Toulouse.
- [28] Marani, M., Lanzoni, S., Silvestri, S., Rinaldo, A. (2004). Tidal landforms, patterns of halophytic vegetation and the fate of the lagoon of Venice. *Journal of Marine Systems*, 51, 1-4, 191-210.
- [29] Marani, M., Belluco, E., Ferrari, S., Silvestri, S., D'Alpaos, A., Lanzoni, S., Feola, A., Rinaldo, A., Analysis, synthesis and modelling of high-resolution observations of saltmarsh ecogeomorphological patterns in the Venice lagoon, *Estuarine Coast. Shelf Sci.*, in press, 2005a.
- [30] Thomson, A.G., Fuller, R.M., Eastwood, J.A. (1998). Supervised versus unsupervised methods for classification of coasts and river corridors from airborne remote sensing. *International Journal of Remote Sensing*, 19, 17, 3423-3431.
- [31] Silvestri, S., Marani, M., Marani, A. (2003). Hyperspectral remote sensing of salt marsh vegetation, morphology and soil topography. *Physics and Chemistry of the Earth*, 28, 1-3, 15-25.
- [32] Silvestri, S., & Marani, M. (2004). Salt marsh vegetation and morphology, modelling and remote sensing observations. In S. Fagherazzi, M. Marani, & L.K. Blum (Ed.), *The Ecogeomorphology of Tidal Marshes. Coastal and Estuarine Studies*, 59 (pp. 266), hardbound, ISBN 0-87590-273-1, AGU Code CE0592731.
- [33] A.J.Lotka. *Elements of Physical Biology*. Baltimore: William and Wilkins (1925).
- [34] V.Volterra. Variazione e fluttuazioni del numero di individui in specie animali conviventi. *Mem. Accad. Nazionale Lincei*, 2:31-113 (1926).

- [35] C.B.Huffaker. Experimental studies in predation: dispersion factors and predator-prey oscillations. *Hilgardia*, 27(4):343-383 (1958).
- [36] M.Kimura. *The neutral theory of molecular evolution*. Cambridge University Press (1983).
- [37] G.Bell. Neutral Macroecology. *Science* **293** 2413 (2001).
- [38] D.Tilman. *Plant Strategies and the Dynamics and Structure of Plant Community*. Princeton (1988).
- [39] B.J.McGill. A test of the unified neutral theory of biodiversity. *Nature*, **422**:881-885 (2003).
- [40] J.Harte. Tail of Death and Resurrection. *Nature* **424**, 1006 (2003).
- [41] D.Y.Zhang, K.Lin *J.Theor.Biol.***188**, 361-367 (1997).
- [42] I.Volkov, J.R.Banavar, S.P.Hubbell, and A. Maritan. Neutral theory and relative species abundance in ecology. *Nature*, **424**:1035-1037 (2003).
- [43] A.J.McKane, D.Alonso, R.V.Sole. Analytic solution of Hubbell's model of local community dynamics. *Theoretical Population Biology*, (2003).
- [44] T.J.Pfaff. *Stochastic Processes and their applications* **104**, 325-347 (2003).
- [45] R.Durrett, S.Levin. Spatial Aspect of Interspecific Competition. *Th.Pop.Bio.* **53**, 30-43 (1998).
- [46] J.Chave, H.C.Muller-Landau, S.A.Levin.*American Naturalist* **159** 1-23 (2002).
- [47] J. Silvertown, S. Holtier, J. Johnson, and P. Dale, *Journal of Ecology* **80**, 527 (1992).
- [48] R. Holley and T. M. Liggett, *Ann. Probab.* **3**, 643 (1975).
- [49] P.L.Krapivsky, *Physical Review A* **45**, 1067 (1992).
- [50] L. Frachebourg and P. L. Krapivsky, *Phys. Rev. E* **53**, R3009 (1996).

- [51] E.Ben-Naim, L. Frachebourg and P. L. Krapivsky *Phys. Rev. E* **53**, 3078 (1996).
- [52] E.Ben-Naim, P. L.Krapivsky, F.Vazquez, S.Redner. *Physica A***330**, 99-106, (2003).
- [53] T. M. Liggett, *Interacting Particle Systems* (Springer, New York, NY, 1985).
- [54] T. M. Liggett, *Stochastic Interacting Systems: Contact, Voter and Exclusion Processes* (Springer, Berlin, Germany, 1999).
- [55] R. Durrett and S. A. Levin, *Phil. Trans. R. Soc. Lond. B* **343**, 329 (1994).
- [56] M.Vallade, B.Houchmandzadeh. *Physical Review E* **68** 061902 (2003).
- [57] B.Houchmandzadeh. *Physical Review E* **66** 052902 (2002).
- [58] B.Houchmandzadeh,M.Vallade. *Physical Review E* **68** 061912 (2003).
- [59] T.Zillio, I.Volkov, J.R.Banavar, S.P.Hubbell, A.Maritan. Spatial Scaling in Model Plant Communities. *Physical Review Letters* **95**, 098101 (2005).
- [60] M.Abramowitz, I.A.Stegun, *Handbook of Mathematical functions* (Dover Publications Inc., New York, 1965)
- [61] C.W.Gardiner. *Handbook of stochastic methods*. Springer-Verlag. Berlin (1985).
- [62] S.Karlin, H.M.Taylor. *A first course in stochastic processes. (2nd ed.)*. New York, NY, Academic Press, (1975).
- [63] S. Chandrasekhar. Stochastic processes in physic and astronomy. *Rev.Mod.Phys.* **15**(1), (1943).
- [64] W.Feller. *An introduction to probability theory and its applications(2nd ed.)*. New York, Wiley, (1957).
- [65] N.G.van Kampen. *Stochastic processes in physics and chemistry*. Amsterdam, North-Holland, (1981).
- [66] H.Risken. *The Fokker-Planck equation: methods of solution and applications (2nd ed.)*. Berlin, Springer-Verlag, (1989).

- [67] F.Bowman. *Introduction to Bessel functions*. Dover Publications Inc. (1958).
- [68] N.N.Lebedev *Special functions and their applications*. Dover Publications Inc. (1972).
- [69] J.R.Banavar, J.L.Green, J.Harte, A.Maritan. Finite Size Scale in Ecology. *Physical Review Letters* **83** 4212 (1999).
- [70] A.Rinaldo, A.Maritan, K.K.Cavender-Bares, S.W.Chisolm. Cross-scale Ecological Dynamics and Microbial Size Spectra in Marine Ecosystems. *Proc.R.Soc.Lond. B* **269**, 2051-2059 (2002).
- [71] J.Harte, E.Conlisk, A.Ostling, J.L.Green, A.B.Smith. A Theory of Spatial Structure in Ecological communities at Multiple Spatial Scales. *Ecological Monographs*, **75**, 179-197 (2005).
- [72] S.M.Bhattacharjee, F.Seno. *J.Phys A* **34**, 6375 (2001).
- [73] C.Wills, R.Condit, R.B.Foster, S.P.Hubbell. *Proc.Natl.Acad.Sci, USA* **94** 1252-1257 (1997).
- [74] J.Molofsky, R.Durrett, J.Dushoff, D.Griffeath, S.Levin. *Th.Pop.Biol.* **55** 270-282 (1999).
- [75] D. H. Janzen, *Am. Nat.* 104, 501 (1970).
- [76] J. H. Connell, in *On the role of natural enemies in preventing competitive exclusion in some marine animals and in rain forest trees*, edited by P. J. Den Boer and G. R. Gradwell (PUDOC, Wageningen, The Netherlands, 1971), *Dynamics of Populations*, pp. 298-312.
- [77] K.E.Harms, S.J.Wright, O.Calderon, A.Hernandez and E.A.Herre, *Nature* (London) **404**, 493 (2000).
- [78] S.P.Hubbell, personal communication.

Acknowledgments

*I don't know half of you half as well as I should like;
and I like less than half of you half as well as you deserve.*
(J.R.R.Tolkien)

Three years are a long time, and the amount of people that I've meet during this wonderful experience that is the PhD is fairly big. So I apologize in advance with all the people that I will forget to mention here.

My first and greater thanks goes to my supervisor Amos Maritan, who taught me how to become a scientist. I owe to him most of the useful thing I know about research, both on the technical and the human side. And, thanks to him, now I have a well-furnished trick bag that will be useful in the future.

Then I have to thanks our collaborators. Jayanth R.Banavar and Igor Volkov of the Penn State University: they taught me, among other things, respectively how to write articles and the usefulness of double-checking the results. Steven P.Hubbell was a source of ideas and informations. Marco Marani and Enrica Belluco from the Università degli Studi di Padova were very kind, patient and industrious collaborators, and I can't really ask any better.

Now, for non-research-related acknowledgements: Giacomo, powerful linux guru (I'm your humble padawan), that still bear me after three years in the same flat. You really have the patience of a saint. Demian and his flying mole, the former introduced me to the Mac ("you will be assimilated!"), the latter showed me the craziness of the former. Many hours of insightful discussion I owe to you.

My parents Luciano and Ornella, and my sister Lisa, that patiently wait for me to come home only for seeing me some hours. You've been fantastic.

Trieste, who provided the magnificent scenery of Adriatic sea and Carso mountains. And lots of wind. You're really not bad when one learns how to know you, girl.

Gigi, who was my friend and *compagno d'arme* during my stay in Trieste. You made me understand history (at least a little bit) and that is more than any teacher can claim. All the people at Accademia di Studi Medievali di Gradisca, especially Vinicio and Lorenzo. Even in history the path of truth goes through experimentation. Sometimes painful, sometimes harsh, but always damn funny experimentation.

Caterina, which exuberance lead me to experiences that I won't have tackled alone, namely Tango. My Tango teachers Hugo, Matteja, Giovanni and Marco for fun hours and patience. All the dancers in the group, especially Lisa with his smooth passo, Elisa with his slow cruse and Marina for teaching me the closes. Hope to dance with you again in the future.

Renata Carlon and all the people at the Ki Aikido Dojo in Mestre. I'll probably never be able to do katatedori kokyunage the correct way, nor to really understand the weight underside trick. But these are details, the important things I remember all.

Daniela, and all the scouts in Maerne. Ideas can change, opinions can divide. But scout once, scout forever. I really hope to camp with you again.

Lara, for many hours of chatting and emotional support.

All the people in Acquario and all the sect of Black Engineers at Università degli Studi di Padova. All the friends in Sissa. All the friends scattered around the world.

And last but by all means not least, Eva, my love, who has patiently suffered the distance and has traveled many times the Death Quadrilateral (Trieste, Udine, Mestre and Padova) only to see me. I owe to you more than I can say.

*In the last analysis, there are only atoms.
There's just one science, Physics;
everything else is social work.*

(James Watson, Nobel Laureate 1962
for DNA)