UNIVERSITÀ DEGLI STUDI DI PADOVA

TESI DI DOTTORATO DI RICERCA
DIPARTIMENTO DI INGEGNERIA IDRAULICA, MARITTIMA,
AMBIENTALE E GEOTECNICA (IMAGE)

SCUOLA DI DOTTORATO IN
SCIENZE DELL' INGEGNERIA CIVILE ED AMBIENTALE,
CICLO XXII

PATTERNS IN ECOLOGY AND GEOMORPHOLOGY OF RIVER BASIN ECOSYSTEMS

Supervisori: Chmo. Prof. A. Rinaldo

Dist. Prof. I. Rodriguez-Iturbe (Princeton University)

Direttore della Scuola: Chmo. Prof. S. Lanzoni

Dottorando: Matteo Convertino

27 Novembre 2009

PATTERNS IN ECOLOGY AND GEOMORPHOLOGY OF RIVER BASIN ECOSYSTEMS

Matteo Convertino

Advisors: Chmo. Prof. A. Rinaldo

Dist. Prof. I. Rodriguez-Iturbe (Princeton University)

Examiner: Chmo. Prof. L. Ridolfi (Politecnico di Torino)

A DISSERTATION
PRESENTED TO THE FACULTY
OF THE UNIVERSITY OF PADOVA
IN CANDIDACY FOR THE DEGREE
OF DOCTOR OF PHILOSOPHY

RECOMMENDED FOR ACCEPTANCE BY
"IMAGE" DEPARTMENT
GRADUATE SCHOOL IN CIVIL AND ENVIRONMENTAL
ENGINEERING SCIENCES, YEAR XXII

27 November 2009

Abstract

Some possible interactions between the hydrological, geomorphological, and ecological features and processes have been studied here at different spatial scales and resolutions in ecosystems. The ecohydrological framework is to perform an interdisciplinary research to detect with essential models the broad and complex patterns in biological, ecology, geomorphology and hydrology of river basins, in which water plays a key role.

Starting from the evidence of the invariance and universality of some geomorphological patterns in river basins, such as the drainage area and the Hack's lengths, the distribution of the distances between tributaries along the mainstream has been studied using digital terrain maps of real basins, that is directly related to the availability of water and of the channel and riparian area. Specifically it has been found that the probability of exceedence of of the alongstream distances of tributaries larger than a given threshold, has an exponential form in function of the boundary conditions and the Hurst exponent of the basin. Also theoretical constructs have been used such as optimal channel networks, Peano and Scheidegger networks at different growth stages, deriving exact analytical expressions for the probability of the distances between subbasins on a rectilinear mainstrem. The concepts of scale as extension, resolution and coarse-graining level have been introduced in order to study the properties of patterns resulting from ecological and geomorphological processes when varying these quantities.

Biodiversity patterns have been extensively simulated using two principal dispersal mechanisms with a neutral metacommunity model. The mechanism with exponential kernel dispersal and the mechanism with exponential plus a power-law kernel dispersal that ensures a long-distance dispersal to the species. Because the neutrality hypothesis all the individuals compete equivalently in the speciation and dispersal dynamics. The dependence on the scale, resolution and coarse-graining at which the ecosystem is simulated has been investigated for different biodiversity indicators, such as the local and regional species richness, the similarity in species richness between local communities, the species-area relationship and the probability distribution of cluster-size of conspecific individuals. Also empirical patterns of fishes and trees (with two classes of diameter resembling two distinct life-stages) of the Mississippi-Missouri River System have been analysed. The model is able to describe the probability distribution of the cluster-size further other patterns as already evidenced by previous studies. The influence of the topology of the ecosystem (2-D landscape and river network), of the shape, and of the environmental heterogeneities, have been analysed in detail. In general all the biodiversity patterns result dependent on the scale and on the resolution. Therefore the species-area relationship results invariant across coarse-graining levels.

The macroecological description of the spatial distribution of the species results reproduced by the neutral model optimally. An interesting relationship has been found between the exponent of the species-area relationship and the exponent of the power-law of the cluster-size for different topologies and dispersal cases. The exogenous (immigration and environmental variables) and endogenous factors (speciation and topology) have been disentangled in order to determine their contribution in the spatial patterns of fish and vegetation. The speciation-death and dispersal phenomena are the determinants in the clustering process without any need to include positive feedbacks between species in order to reproduce the spatial patterns. The environmental heterogeneities, for example dictated by the climate influence the structure of the power-law of the cluster-size of species only in conditions of dispersal limitation and saturation of the local communities. It has been put emphasis also on the species-individual relationship as a tool to better forecast the effect of climate change than the species-area curve, and to possible allometric relationships between the dispersal parameter and some species traits such has mass and lifespan. The relationship between the exponent of the species-area and the lifespan of species is in agreement with what found in literature.

The interactions between the geomorphic structure of ecosystems and the processes acting on them at different scale are of fundamental importance in the management of ecological and water resources. This is the reason why it is extremely important to understand the scale at which the relevant processes take place. It has been evidenced clearly the striking features of the geomorphic supports and external drivers on ecological and biological processes than on the hydrologic dynamics.

Keywords: biodiversity patterns, ecohydrology, scales, river basins, dispersal, neutrality, metacommunity, landscape ecology, geomorphology.

Sommario

Questa tesi tratta dello studio dei controlli idrologici e geomorfologici sulle caratteristiche e processi ecologici in ecosistemi fluviali. L' inquadramento ecoidrologico qui perseguito definisce un approcio interdisciplinare con lo scopo di modellare vasti e complessi patterns biotici and abiotici nei quali l' acqua gioca un formante ruolo chiave.

Partendo dall' evidenza dell' invarianza di alcune proprietà di patterns geomorfologici dei bacini fluviali, la tesi prima tratta della distribuzione delle distanze tra sottobacini lungo il corso d' acqua principale. Questa é una caratteristica importante per i processi ecologici perchè controlla lo sviluppo spaziale della capacità dell' habitat locale di ospitare individui delle specie, ed é stata studiata usando mappe digitali del terreno di bacini fluviali. La relativa disponibilità di acqua e l' area di canale e ripariale sono allo stesso modo investigate. Specificatamente é stato trovato che la probabilità di eccedenza delle distanze lungo il canale principale tra affluenti (maggiori di una data soglia) ha una forma esponenziale legata al vincolo geomorfico. Costrutti teorici sono stati usati allo stesso scopo, come le reti ottime di canali, le reti di Peano e Scheidegger, derivando espressioni analitiche esatte per caratteristiche geomorfiche rilevanti.

Patterns di biodiversità sono stati estensivamente studiati con simulazioni numeriche, con diverse textures ambientali definite dall' ambiente fluviale, e usando due principali meccanismi di dispersione in un modello neutrale a metcommunità. L' ipotesi di neutralità postula che tutti gli individui competono equivalentemente nelle dinamiche di speciazione e dispersione. Un meccanismo é con un kernel esponenziale che descrive certe caratteristiche ecologiche di interesse caratterizzate da una scala spaziale, mentre un' altro meccanismo é impiegando un kernel con dispersione a legge di potenza il quale assicura dispersioni a lunga distanza alle specie. É stata analizzata la dipendenza dalla scala, risoluzione e raffinamento del modello a cui l' ecosistema é simulato, di svariati indicatori di biodiversità, come la ricchezza locale e regionale di specie, la similarità in specie tra comunità locali nell' ecosistema, la relazione specie-area e la distribuzione di probabilità della dimensione dei cluster di individui conspecifici.

Patterns empirici di pesci e alberi del sistema fluviale del Mississippi-Missouri sono stati analizzati. Proprie simulazioni provano essere capaci di descrivere parecchi patterns biotici osservati, includendo la distribuzione di probabilità della dimensione dei clusters delle specie. Le influenze della topologia dell' ecosistema (paesaggio 2-D e rete fluviale 1-D), della forma, e delle disomogeneità ambientali, sono state analizzate in dettaglio. É stato trovato che in generale i patterns di biodiversità sono dipendenti dalla scale e risoluzione.

La tesi anche tratta della descrizione macroecologica della distribuzione spaziale delle

specie prodotta dal modello neutrale. É suggerito come l' esponente della rilevante relazione specie-area e l' esponente della legge a potenza della dimensione dei clusters sono relazionati per diverse topologie e scenari di dispersione. I fattori esogeni (immigrazione e variabili ambientali) ed endogeni (speciazione e topologia) sono stati studiati separatamente per determinare il loro contributo mutuo e individuale nella generazione dei patterns spaziali di pesci e vegetazione. I fenomeni di speciazione-morte e dispersione sono suggeriti essere i determinanti nel processo di clusterizzazione senza nessun bisogno di includere feedback positivi tra le specie. Le eterogeneità ambientali, ad esempio dettate dal clima influenzano sulla struttura a legge di potenza dei cluster delle specie solo in condizioni di dispersione limitata o di saturazione delle comunità locali. Relazioni specie-individui sono studiate come strumento per meglio predire l' effetto del cambiamento climatico, e possibilmente per inferire relazioni allometriche tra il parametro di dispersione e i tratti fisiologici delle specie come la massa e la durata di vita.

In conclusione le interazioni tra la struttura geomorfica degli ecosistemi e i processi che avvengono in essi a diverse scale sono di fondamentale importanza nel organizzazione e tutela delle risorse ecologiche e idrologiche. La tesi evidenzia il ruolo cruciale dei supporti geomorfologici e dei drivers esterni sui processi ecologici e biologici che sono controllati da dinamiche idrologiche.

Parole chiave: patterns di biodiversità, ecoidrologia, scale, bacini fluviali, dispersione, neutralità, metacomunità , ecologia del paesaggio, geomorfologia.

Acknowledgments

It is all the time not easy to write a thesis nor to acknowledge all the people that have shared such an important experience in work and in life. I was lucky to have wonderful advisors and colleagues that have been constantly an example for me. In particular Andrea Rinaldo and Ignacio Rodriguez-Iturbe for their support, teaching, contagious passion for research, and the opportunities given to me, regardless their location all around the world. I will be always grateful.

My experience at Princeton University was indubitably the key of many changes and choices in my life. I wish to say thank you in particular to Rachata Muneepeerakul (Chot) and Chitsomanus P. Muneepeerakul (Re) for their endless friendship. In particular I wish to thank Chot as a great colleague. His support made much more fruitful my experience at Princeton. I have to say thank to Sandro Azaele for his friendship and constantly rigorous advices in work. I can not forget also Megan Konar for our fun and mutual encouragement in work, Dario Pumo and Stefania Tamea. Thanks a lot to everyone. In general a great thanks to all the people of the CEE Dept. at Princeton University. I feel all the time like at home there. Thanks also to all the people know at Princeton University "spreaded" all around the campus.

I can not forget also the company and the teaching of my officemates in Padova, Gianluca Botter, Alessandro Frascati, Ludovico Nicotina, Enrica Belluco, and Chiara Venier. Thanks for everything to you also, it was a great pleasure to be with you. Thanks to all the people in the IMAGE Department that in various measure contributed to my growth.

Also I would like to thank Riccardo Rigon (University of Trento), always available to answer my doubts, and Filippo Catani (University of Florence) with which I am in debt about our work I would love to finish. Thanks also to Colin Stark at Columbia University, and to Luca Ridolfi (Politecnico di Torino) the "controrelatore" of my thesis for his patience and kindness.

A warm thanks to my family, my mother, my brother, my sister, among all the others. Their enduring patience and love will be all the time fundamental in my life. Finally but surely not the least important, I am very grateful to my girlfriend, Yanhua, that supported me with passion, smile, stubborness and tenderness, precision and advice. This is also for you, for all our lifelong plans.

I tried to give all the time my best, to learn the best. I hope I have done something worth and useful, lasting in time, for my career and life.

Contents

List of Tables

List of Figures

Nomenclature

Geomorphological parameters

A	Drainage area (pixels).
ϕ	Geomorphological diameter (m).
d	Geomorphological distance between links (m).
n	Total number of links in the space-filling network.
δ	Pixel side length at the finest CG resolution (= 100 m).
λ	Coarse-graining resolution parameter.

Ecological parameters

N	Total number of individuals in the basin, or habitat size ($= 1 \times 10^5$).
S_T	Total number of species in the basin at stationary state.
S	Average number of species encompassed in a circular area of radius r at a link.
H	Link habitat capacity ($\simeq N/n$ in the homogeneous case).
H^*	Number of individuals encompassed in a circular area of radius r at a link.
d_l	Dispersal length parameter (m).
θ	Diversification rate (number of new species introduced at each time step = $N\nu$).
ν	Per site diversification rate (θ/N).
b	Cauchy-kernel parameter.
K_{ij}	Fraction of individuals produced in j that go into i.

Acronyms

LSR	Local species richness (number of species at link i).
JSI	Jaccard similarity index.
RSR	Regional species richness ($= \langle S \rangle$).
CG	Coarse-graining.
CS	Cluster-size.
dbh	Diameter at breast height.
MMRS	Mississippi-Missouri River System.
DTA	Direct Tributary Area.
AARP	Average Annual Runoff Production.
AAP	Average Annual Precipitation.

Chapter 1

Introduction

"The most beautiful thing we can experience is the mysterious. It is the source of all true art and all science. He to whom this emotion is a stranger, who can no longer pause to wonder and stand rapt in awe, is as good as dead: his eyes are closed"
Albert Einstein

This thesis studies the interactions between hydrologic and ecological processes in river network ecosystems. Aiming at extending recent studies at the interface of hydrology, geomorphology and ecology under an integrated framework of analysis (*Muneepeerakul et al.*, 2008; *Azaele et al.*, 2009; *Bertuzzo et al.*, 2009; *Konar et al.*, 2009; *Rodriguez-Iturbe et al.*, 2009), the thesis addresses specific and related topics, ranging from biodiversity of freshwater fish in river networks to the distribution of riparian and not riparian vegetation, joined by the common thread of the study of the dynamics of hydrologically-controlled riparian ecosystems (*Camporeale and Ridolfi*, 2006) (and references therein).

The geomorphological environment where fluvial ecosystems thrive proves rather important. In fact, dendritic geometries abound in nature, covering a vast range of scales (*Ball*, 1999). Dendritic environments are observed in living and non-living systems, in networks of metabolic supply like in pulmonary systems, as well as in trails of foraging ants, plants, caves, and bird migration routes (*Solé and Goodwin*, 2002). River networks (*Rodriguez-Iturbe and Rinaldo*, 1997) produce scale-invariant structure from the scale of a few meters to thousands of kilometers, where what strikes is the apparent diversity of Nature's ways, and yet the deep symmetries one may reveal and expose. Indeed, countless processes essential to sustain human life and societies take place along dendritic structures, like metabolic networks (*West et al.*, 1999) or migratory patterns of birds (*Ostling*, 2005), a combination of innate cues and spatial memory (*Garber*, 1989; *Real*, 1994; *Burns and Thomson*, 2006). Other patterns are also dictated by dendritic pathways, like for instance human settlements that historically proceeded along river networks to follow the necessary water resources (*Ammerman and Cavalli-Sforza*, 1984a; *Campos et al.*, 2006).

Indeed, riparian ecosystems positioned along streams and rivers are shaped by dendritic networks. Their role is vital in their respective watersheds (*Perona et al.*, 2009), including nutrient filtering and biogeochemical processing (*Lowrance et al.*, 1984; *Peterjohn and Correll*, 1984; *Malanson*, 1993; *McClain*, 2003), shade and resource provision-

ing, and stream bank stabilization (*Gregory et al.*, 1991; *Malanson*, 1993; *Naiman and Décamps*, 1997). The spreading of water-borne diseases, such as cholera, occurs through water bodies linked by dendritic river networks whose features dominate the traveling fronts of the incipient outbreaks (*Bertuzzo et al.*, 2007).

The related dynamics have been previously studied extensively. However, previous studies were mostly confined within the framework of so-called mean-field or two-dimensional (2D) landscapes. This thesis will provide original insight into the role of the morphological and hydrological factors that characterize the fluvial ecosystems. River networks are perhaps the most widely appreciated instances of dendritic geometry in nature. They form complex landscapes characterized by well-studied scaling laws and deep symmetries that are shaped by climate, hydrology, geology, geomorphology, biology, and ecology (*Rodriguez-Iturbe and Rinaldo*, 1997). River networks and the associated transport processes provide a natural integrating framework for the hydrologic, ecologic, and geomorphologic dynamics that take place in the river basin and its varied ecosystems. The profound commonalities existing among all types of river basins and their drainage networks, together with the key role that the embedded dendritic constraints operating on the dynamics, allow for the establishment of key, general signatures whose statistical characterization is studied in this work.

The above basic issue has been addressed in the literature of epidemiology, hydrochory, ecology, and geography. However, most of the models that describe the relevant processes do not consider space in an explicit manner. Stochastic space-time evolution of living communities in explicitly spatial landscapes is a relatively new subject (see e.g., *Tilman and Kareiva*, 1997). Among the foremost of the mechanisms relevant to the spatial dynamics are the very configuration of local communities imposed by the river networks and the anisotropy in the spreading of species, pathogens, or other agents of interest along the waterways (*Campos et al.*, 2006; *Bertuzzo et al.*, 2007; *Muneepeerakul et al.*, 2007b). Such anisotropy proves essential. Even without any anisotropy, the network configuration per se exerts crucial effects on the spatial dynamics and patterns of the ecological processes occurring in the fluvial system (*Muneepeerakul et al.*, 2008). Hydrologic controls accompanying river basins play central roles in the spreading of species, water-borne disease, or populations in several other ways. For instance, they dictate the spatial distribution of streamflow rates and velocity, which regulate the amount of resources available for biotic and abiotic processes, thus resulting in the sustainable community sizes, as well as the severity of disturbances required to disrupt them. The combined effects of all these mechanisms are studied in that they give rise to interesting dynamics and patterns. Morphological constraints provided by the river basin are also vital, such as the features of the distances among subsequent tributaries, a subject studied in detail in this thesis.

It is also necessary to explore the consequences of external forcing on the dynamics of complex systems operating in dendritic networks. Precipitation events, and thus river flows, flooding, and other hydrologic disturbances are stochastic and seasonal (*Rodríguez-Iturbe and Porporato*, 2005). Due to external forcings, connectivity among stream reaches, the spatial subunits of river basins, changes over time and space. To investigate how external forcing alters the properties relevant to ecosystem dynamics, this thesis reexamines the dynamics and patterns of the biodiversity and disease systems

with transient connectivity. Transient connectivity is a feature investigated in spatial ecological models only recently.

In this thesis novel analysis are presented following the path of a few recent studies (*Muneepeerakul et al.*, 2008; *Azaele et al.*, 2009; *Bertuzzo et al.*, 2009; *Konar et al.*, 2009; *Rodriguez-Iturbe et al.*, 2009). Metacommunity models (as a set of local communities that are linked by dispersal of multiple potentially interacting species (*Giplin and Hanski*, 1991; *Wilson*, 1992; *Leibold*, 2004; *Economo and T.H.*, 2007)) are employed both for fishes and vegetation under the main hypothesis of neutrality (*Hubbell*, 2001). The environmental heterogeneity, e.g. the average annual precipitation, determines the number of individuals in each local community, while the structure of the dispersal determines how species spread in the ecosystems. The neutral assumption postulates that all individual organisms of the species are competitively equivalent. A main aim of the thesis deals with the exploration of the joint role of ecological neutrality and of the geomorphological support on the processes and dynamics operating on ecosystems. Insights provided here will lend themselves to issues of great practical importance such as integration of forest and fishes into large-scale resource management, and spatial conservation strategies.

A few preliminary definitions are in order. Broadly speaking, *biodiversity* refers to the extent of genetic, taxonomic and ecological diversity over all spatial and temporal scales (*Harper and Hawksworth*, 1994). In this thesis we focus on the ecological diversity over spatial scales, by assuming the stationary state of the relevant dynamical processes. Here we the scale of ecological patterns is considered with great care. Patterns in river basins are examined at different scales, resolution and coarse-graining level. As for the range covered, it is important to detect if a statistical similarity exists between the whole basins and all the subbasins (*Rodriguez-Iturbe and Rinaldo*, 1997). Scaling is an essential tool for it allows to extrapolate the behavior of each component since it is the same of the whole. As per the suitable resolution, it is fundamental to understand if the variables used to characterize the patterns vary in function of the quality of the available data. From a modeling viewpoint, it is also important to frame properly whether what is actually simulated reflects reality in a satisfactory manner at every level of resolution. A key issue in biodiversity, in fact, is to simulate the impact of artificial or natural changes and their effects. On one side it is essential to reduce the computational effort, while on the other it is necessary to provide reliable predictions.

Concept of scales are relevant to every ecosystem, whether fragmented by a network or comprised of a savanna ecosystem. A remarkable number of studies in ecology exists, especially for birds and vegetation, that treat the issue of patterns and processes at different scales also in relation to the geomorphology of the ecosystem in study (*Wiens*, 1989; *He et al.*, 1994; *Condit et al.*, 2000; *Gaston*, 2000; *Wu*, 2004; *Crawley and Harral*, 2001; *Gotelli*, 2002; *Harte et al.*, 2005; *Rahbek*, 2005; *Storch et al.*, 2007; *Nogués-Bravo et al.*, 2008; *Storch et al.*, 2008). The study has been done in the view of ecohydrology as a challenging multidisciplinary field (*Porporato and Rodriguez-Iturbe*, 2002), combining here ecology and geomorphology of ecosystems.

A few relevant questions about biodiversity patterns in this context are the following:

- How does landscape heterogeneity affect biodiversity patterns and processes?
- How does biodiversity relate to ecosystem function at different spatial scales?

- How do changes in land use and cover affect ecological processes?
- How can ecological information be transferred from small to large scales, and vice versa?

The research presented in this thesis concentrates on pinpointing geomorphological and hydrological controls on biodiversity patterns. The physical grounds and the effects of relevant ecological parameters in modelling ecological patterns are studied as well. The chosen analysis follows a distinguished path of studies searching for general scaling laws interpreting ecological processes (*Banavar et al.*, 1999; *Harte et al.*, 1999; *Ritchie and Olff*, 1999).

The study was first concentrated on the detection of novel scaling laws in river basins, particularly in the distribution of the distances within tributaries in river basins (*Convertino et al.*, 2007). The properties of the real river networks studied were tested against the properties of artificial river networks like Optimal Channel Networks (*Rinaldo et al.*, 1992; *Rodríguez-Iturbe et al.*, 1992; *Rinaldo et al.*, 1993; *Rigon et al.*, 1996; *Rodriguez-Iturbe and Rinaldo*, 1997), Peano and Scheidegger constructions, deriving for the latters exact results. The distance between tributaries is of fundamental importance since it is a proxy of the availability of water (and thus species' habitat) that affects riparian vegetation and the transport of species along the network.

The model employed for generating biodiversity patterns is a metacommunity model (*Giplin and Hanski*, 1991; *Wilson*, 1992; *Leibold*, 2004; *Economo and T.H.*, 2007). Note that neutrality is a concept that was introduced at the level of genes by *Kimura and Crow* (1964). It is quite interesting to see how neutrality can be a simple but powerful assumption to describe how nature works in some cases to upscale species properties from the individual to regions (*Vandewoestijne et al.*, 2004; *Bangert et al.*, 2008).

The study of species clustering in this context addresses the spatial distribution of species. Most importantly, the datasets used to test the models, in particular clustering schemes, refer to fishes (*Oberdorff et al.*, 1995; *Angermeier and Winston*, 1998; *Guégan et al.*, 1998; *Xenopoulos and Lodge*, 2006) and big-trees of the Mississippi-Missouri River System, first used by *Muneepeerakul et al.* (2008) and *Konar et al.* (2009) respectively. Also used here are the datasets of small trees employed to test the different spatial aggregation of trees at different life-stages *Plotkin et al.* (2000, 2002) and *Condit et al.* (1996). As for the case of the trees, the simultaneous fitting of biodiversity patterns is not the subject of this thesis as it is pursued elsewhere *Konar et al.* (2009).

The cluster process is also studied to possibly probe the dynamic origins of the power-law distribution that one would presume to observe for cluster sizes. Previous attempts exist in the literature *Garcia-Martin and Goldenfeld* (2006). Clustering of species occurs in many different ecosystems from coral reefs to forests, also at microscales (*Houchmandzadeh*, 2008). In this thesis we address whether power-law shapes of cluster-sizes exist, as well as the scale-invariance properties of biodiversity patterns from models and data. In order to understand the clustering the various determinants of the ecological patterns (i.e. environmental heterogeneity, neutral ecological drift, dispersal kernel structure, and topology of the ecosystem) have been disentangled. Then, biodiversity patterns have been analysed under several assumptions by: assuming a constant the number of individuals in each local community; varying the dispersal parameter on a

wide rage of values; changing the topology of the system; combining the the determinant factors (such as the dispersal kernel). *Borgono et al.* (2009) presents an extensive overview of mathematical models, also not kernel-dispersal based, addressing vegetation pattern formation.

Also another well-know pattern in ecology, the species-area relationship (*Arrhenius*, 1921; *MacArthur and Wilson*, 1963; *Condit et al.*, 1996; *Connor et al.*, 2000; *Hubbell*, 2001; *MacArthur and Wilson*, 2001; *Horner-Devine et al.*, 2004; *Garcia-Martin and Goldenfeld*, 2006; *Lewis*, 2006; *Guilhaumon et al.*, 2008; *Pigolotti and Cencini*, 2009; *Rosindell and Cornell*, 2009) has been analysed by studying it as a function of the determinants mentioned above. Clear but not trivial relationships between the slope of the species-area relationship and the scaling exponent of the power-law of the cluster-size seem to exist. The species-individual relationship has also been investigated as a pattern to better detect the environmental variation and changes that determine modifications in the abundance pattern. One possible issue, especially for large systems characterized by strong environmental heterogeneities, deals with populations densities (or abundance) and their correlation with area, possibly facing equilibrium theories of island biogeography (*MacArthur and Wilson*, 1963; *MacArthur and Wilson*, 2001).

Classic niche-based environmental controls and models foresee that species distributions are affected by dispersal, by the topology and by the shape of the ecosystem. As significant descriptor of the ecosystem, an important feature proves the geometrical diameter defined as the mean path between all the possible reachable sites (*Convertino et al.*, 2009). It measures the mean degree of connectivity within the whole network. Indeed, local dispersal plays a fundamental role in many processes and for a networked environment it crucially depends on local diameters (*Rodriguez-Iturbe et al.*, 2009).

Dispersal is a life-history trait that has profound consequences for populations. Viewed from an ecological perspective, dispersal influences the dynamics and persistence of populations, the distribution and abundance of species, and community structure. From an evolutionary perspective, dispersal determines the level of gene flow between populations and affects processes such as local adaptation, speciation and the evolution of life-history traits (*Dieckmann et al.*, 1998). In fact, it is difficult to imagine any ecological or evolutionary problem that would not be affected by dispersal.

The various consequences of dispersal are extensively discussed in the ecological, biological, zoological, and evolutionary literature (the Science Citation Index gave more than 1000 occurrences of *dispersal* in the abstract or title of papers for the year 2009 alone). Surprisingly, however, the question on why particular dispersal strategies evolve has received much less attention (*Dieckmann et al.*, 1998; *Bullock et al.*, 2002). One issue is that it is notoriously difficult to test dispersal theories verified in the field (*Ims and Andreassen*, 2005). There exists a serious gap between theory and data, and consequently our understanding of why particular organisms disperse in specific ways is still limited (*Dieckmann et al.*, 1998).

Moreover, research about biodiversity is of fundamental importance both theoretically to understand how species evolve, but also for their conservation. Not least it is of great importance to understand at which scales sampling campaigns should be performed and at which of them to act (*Edmunds and Bruno*, 1996; *Bissonette*, 1997; *Girdler and Barrie*, 2008). Human actions, motivated by social and economic driving forces, gen-

erate various pressures on biodiversity, such as habitat loss and fragmentation, climate change, land use related disturbance patterns, or species invasions that have an impact on biodiversity from the genetic to the ecosystem level (*Bangert et al.*, 2008). Each of these factors acts at characteristic scales, and the scales of social and economic demands, of environmental pressures, of biodiversity impacts, of scientific analysis, and of governmental responses do not necessarily match. However, management of the living world will be effective only if we understand how problems and solutions change with scale (*SFPRTD-EU*, 2009). To this end as well the aims of this thesis are deemed relevant.

Finally, using data from the literature (*Crocker*, 1938; *Gilbert and Singer*, 1975; *Edwards and Bohlen*, 1977; *Stoeckel et al.*, 1997; *Nathan et al.*, 2001, 2002a; *Stenzel et al.*, 2007; *Houchmandzadeh*, 2008; *Nathan et al.*, 2008b; *Blake et al.*, 2009; *Pigolotti and Cencini*, 2009; *Schurr et al.*, 2009), allometric relationships (*Nee et al.*, 1991; *Schmid et al.*, 2000; *Sutherland et al.*, 2000; *Burness et al.*, 2001; *Bowman et al.*, 2002; *Marba et al.*, 2007; *White et al.*, 2007; *Savage et al.*, 2008) are studied between the mean dispersal length of different taxa whose body size spans over seven orders of magnitude. The large range of dispersal lengths suggested by appropriate simulations, described in the thesis, require some speculation, possibly to relate long distance dispersals to the size of the species and of the ecosystem analyzed. Overall, in fact, it has been suggested that species tend to maximize their ability to disperse by evolving strategies maximizing their colonization and survivability (*Davies et al.*, 2004; *Pearson and Dawson*, 2004; *Bohrer et al.*, 2005; *Soons and Ozinga*, 2005; *Nathan*, 2006). Also an interesting perspective lies in previous effort to link global biodiversity patterns to evolutionary dynamics using metabolic theory (*Gillooly and Allen*, 2007).

The thesis is organized as follows:

- Chapter 2, describes the fishes and small/big-trees datasets used for comparative analysis with modeling results, one of the cores of the research. It should be noted that fish and big tree data were collected and originally analysed in *Muneepeerakul et al.* (2008) and *Konar et al.* (2009) respectively;
- Chapter 3, describes in detail the models used;
- Chapter 4, describes in details the numerical results about the coarse-graining dependence and invariance of some fundamental biodiversity patterns;
- Chapter 5, discusses the spatial clustering of individuals in function of the topology of the ecosystem, the environmental heterogeneities, and the parameters of the model. It has given particular importance to the cluster-size distribution of conspecific species (fishes and small/big trees) and to the species-area relationship;
- Chapter 6, presents the results about the probabilistic description of the distances between tributaries in river basin ecosystems and also in artificial networks providing in the latter cases exact solutions;
- Chapter 7, concludes the thesis with a brief presentation of the original contributions of the thesis and some perspectives for future research;

- Appendix A reports additional simulations of biodiversity patterns for a river basin with a fat-tail dispersal kernel and high speciation rate;
- Appendix B includes additional results about the probabilistic description of the distances between tributaries for the Scheidegger network at different growth stages.

The publications produced for peer-reviewed international journals during the doctorate program are reported in C. The present thesis in color is available at the University of Padova electronic archive "Padua@Research" (http://paduaresearch.cab.unipd.it/).

Chapter 2

Data

"The goal is to transform data into information, and information into insight"
Carly Fiorina

This Chapter reviews and organizes the data used for comparative purposes within the present thesis. Section 2.1 describes the main geomorphological properties of the networks employed in the research. Section 2.2 describes the datasets on fishes and trees of the Mississippi Missouri River System (in the following for brevity MMRS).

2.1 Geomorphology of the River Basins employed

In this thesis several different basins around the world have been analyzed. Here we give an overview of their foremost features. Further details are given in the chapters reporting the studies performed on them (chapters 4, 5, and 6).

The italian basins employed, depicted in Figure 2.1, are: a) a portion of the Tanaro river basin (Po basin, Piedmont region, Northern Italy, DTM 178 $\times$ 296 pixels) closed at Garessio at an area of about 530 km^2 and mean elevation of 1382 m a.m.s.l.; b) the Mis river basin, closed at Mis, and the Cordevole basin, closed at Sospirolo, are nested subbasins (Piave basin, Veneto region, Northern Italy) suitably extracted from the relevant DTM (255 $\times$ 271 pixels), the former with extension of about 218 km^2 (126 $\times$ 173 pixels) and the latter of about 691 km^2 with mean elevation of 1721 m. All the related river networks are suitably extracted from the relevant DTM at resolution 100 m. We also employ artificial networks for their flexibility and exact properties. Specifically, we employed Optimal Channel Networks whose geomorphological properties closely resemble the optimal features of real river basins, Peano networks and Scheidegger networks (*Rodriguez-Iturbe and Rinaldo*, 1997). The shape, dimensions and boundaries conditions of the artificial networks have been varied to study the influence of these parameters on geomorphological and biodiversity patterns. For an accurate description of the statistical properties of the networks employed, and for an updated list of pertinent references the reader is referred to *Rinaldo et al.* (2006). Figure 2.2 shows different patterns of the Tanaro river basin extracted with different geomorphological thresholds. The thresholding procedure is commonly employed in hydrology and geomorphology in order to detect

the relevant network structure to study hydrological and geomorphological processes at different resolution.

A major benchmark for the ensuing results is the Mississippi-Missouri river system. The whole extension of the MMRS is $2,980,000\ km^2$ and the average principal and transversal diameters are 1304.9 km and 2283.6 km respectively, so the shape of the MMRS is quite unique respect the elongated shape of the river basins. The mean extension of a direct tributary area (DTA) is 3900 km^2 then the mean diameter of a DTA is 62.45 km. The extension of the plot at which the trees-dataset is available, is 1/6 of an acre (then is 0.0675 $ha = 674.610m^2$). We acknowledge that due the very small size of the plots the sampling is not affected by the shape of the samples. As demonstrated by *Potts et al.* (2001) for plots with area less than 0.25 ha plot shape or aspect ratio of the plots have almost no real effect on the number of species observed.

For the MMRS, a direct tributary area (DTA) is a geographic region directly draining to a group of streams, i.e., not including areas upstream of it (*Muneepeerakul et al.*, 2008); the DTAs correspond to the U.S. Environmental Protection Agency and U.S. Geological Survey HUC8-scale subbasins defined by National Hydrography Database Plus (*USEPA-USGS*, 2006). Occurrence data for species and river network structure can be combined and analyzed for several biodiversity patterns that will be introduced later. The river network is modeled by an oriented graph constituted by nodes and edges (figure 2.3). An oriented graph is a directed graph in which the edges have a direction, in this way according tho the elevation drop that is the direction of the water from the source-area to the outlet. In the river network case (RN) each link of the network is between the center of mass of each pixel or DTA following the flow directions, then the neighbors are only the downstream and upstream links. Every link is a local community (LC) that hosts the individuals of species. Figure 2.3 shows the RN case of the MMRS; see also *Muneepeerakul et al.* (2008). In the 2-D landscape case instead every pixel is connected to all the eight adjacent neighbors. In topological unit each links has length equal to one. Considering for the diagonal directions the actual topological distance $\sqrt{2}$ does not make appreciable difference on the simulation of biodiversity and geomorphology patterns.

The DTAs for the trees analysis have been projected according the Lambert cylindrical equal-area criteria to consider the curvature of the Earth, since the spreading of the species can happen for very large distances (*Konar et al.*, 2009).

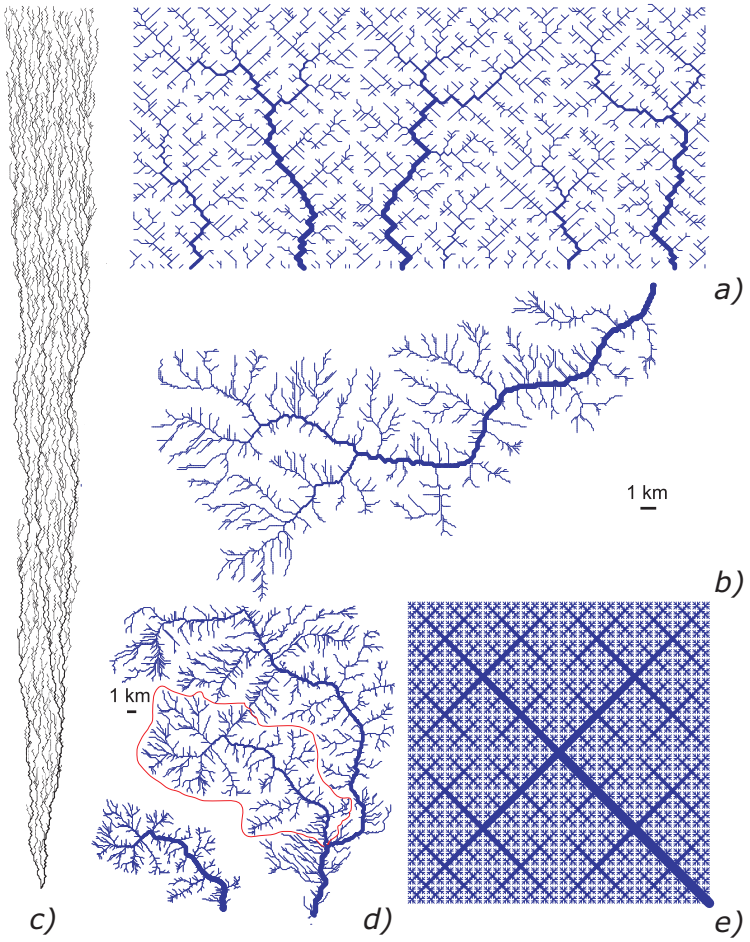

Figure 2.1: Sample networks used in the research: *a)* Multiple-outlet OCN where distances between tributaries are usually measured along the lower draining boundary – although in some instances we may resort to calculation of single-outlet OCNs to test mainstream distributions (see details in figure 6.1 chapter 6); *b)* The Tanaro river basin; *c)* Scheidegger network realized as the original construct and with the parallel side strongly longer than the shorter one in grant to obtain $H = 1/2$, it is a subbasin extracted from the basin-lattice (figure B.1 *a)* appendix B); *d)* Mis (detected by the red drainage-divide line), and Cordevole; *e)* Peano network at the seventh stage of the multiplicative growth process (258×258 pixels). For a detailed description of the specific construction rules and their references, see *Rinaldo et al.* (2006). OCN, real basins and Scheidegger network has been extracted with a threshold on area equal to $A_{th} = 10$. Each pixel of DTMs utilized have length equal to 100 m.

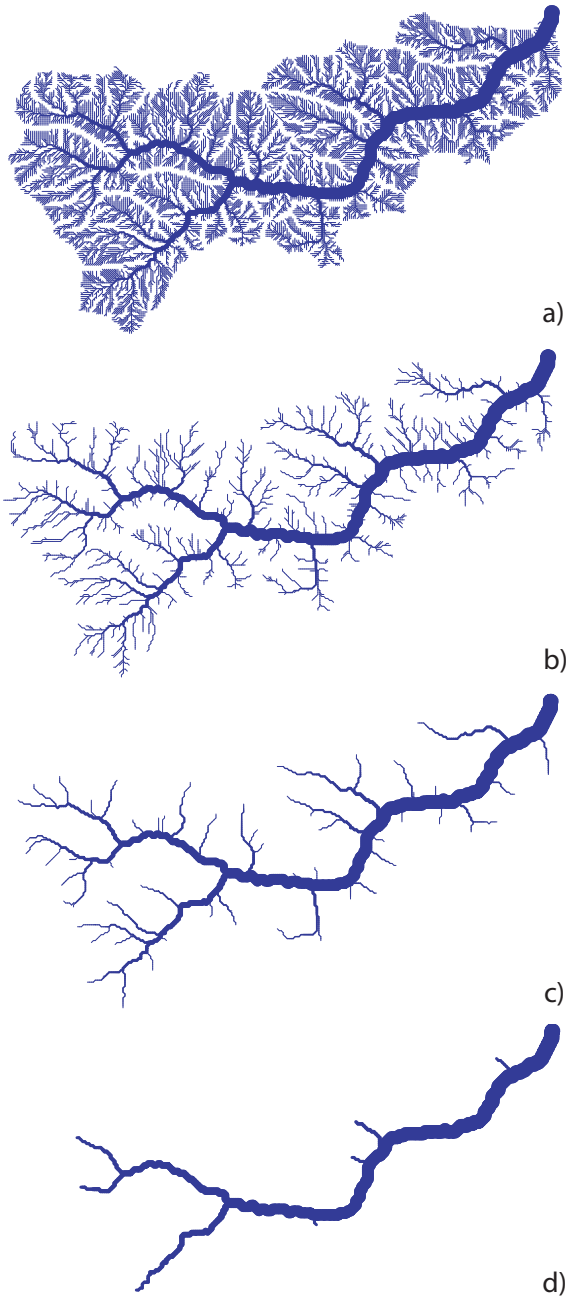

Figure 2.2: Tanaro river network extracted at different geomorphic thresholds A_th: a) space-filling network for $A \geq 1$ in which every pixel is channelized; b) network for $A \geq 10$; c) network for $A \geq 100$; b) network for $A \geq 1000$

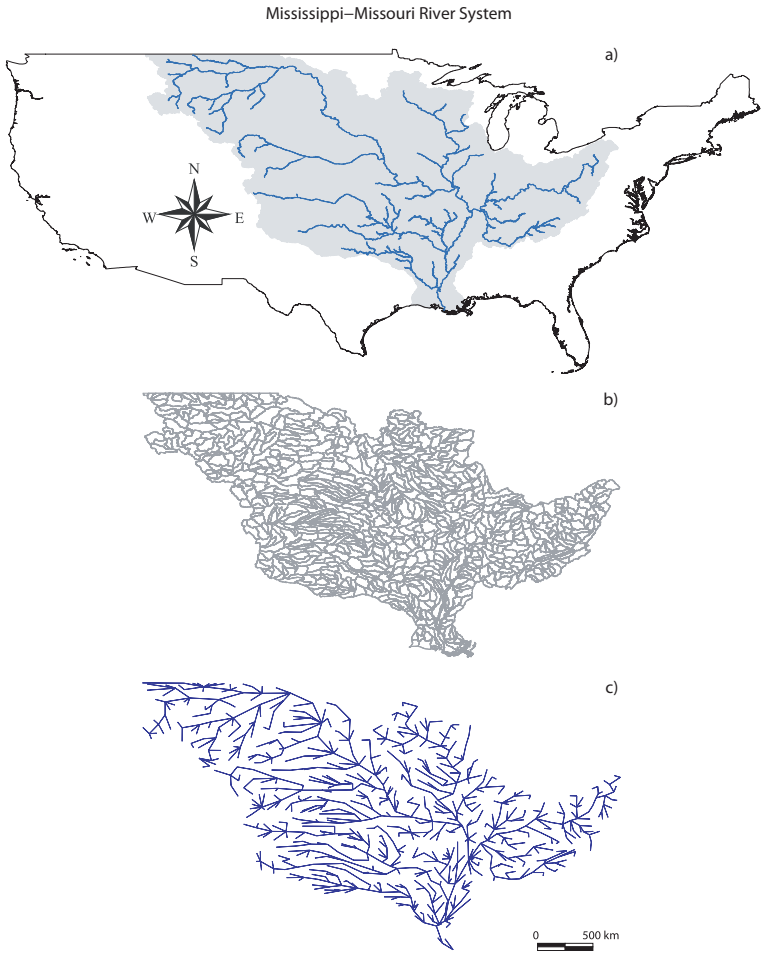

Figure 2.3: Representation of the MMRS: a) main branches of the real fluvial network of the MMRS and US boundaries (courtesy of *Bertuzzo et al.* (2009)); b) DTAs subbasin as classified by *USEPA-USGS* (2006); c) graph network of the MMRS as used in the model (*Muneepeerakul et al.*, 2008).

We now define some geomorphological quantities that will be used throughout the following sections. The diameter, ϕ_i, is defined as the average distance from the streamlink i to all other streamlinks j in the network, where the distance is measured through the network in the river network (RN) case or overall the landscape in the 2-D case. As defined above, it is a measure of the degree of connectivity associated with a local community in the ecosystem and it is crucial for the dispersal dynamics in a landscape. While the diameter is usually specified as the longest path between any pairs of links in the network (e.g., *Newman*, 2003) the definition of an ecological diameter as stipulated above seems more appropriate to our scope. Figure 2.4 shows the spatial patterns of the diameters for the river basins under study (OCN, Cordevole and Tanaro). Note that the minimum diameter, ϕ_{min}, is farther away from the outlet in the Tanaro river basin than in the OCN and the Cordevole. Moreover, link diameters do not necessarily change monotonically with their distance from the outlet. This is due to the elongation of the river basin, defined as the ratio between its longitudinal and transversal characteristic lengths, say $L_{\parallel}$ and $L_{\perp}$ respectively. The more elongated the network is, the closer to the middle of the basin the streamlink with minimum diameter, ϕ_{min}, tends to be. The elongation is a characteristic that is likely to be important in the functioning of river networks as ecological corridors. Notice that the average value of the diameter of the network, that is reported in each plot, corresponds to the average distance between streamlinks. Then, from a biodiversity viewpoint the elongated shape is a favorable feature. Figure 2.5 shows the patterns for the river network (RN) and 2-D cases within the Tanaro basin boundaries. The Tanaro basin has been taken as a prototype of optimal basin due to its geomorphological features, including elongation. Several biodiversity simulations have been carried out on it.

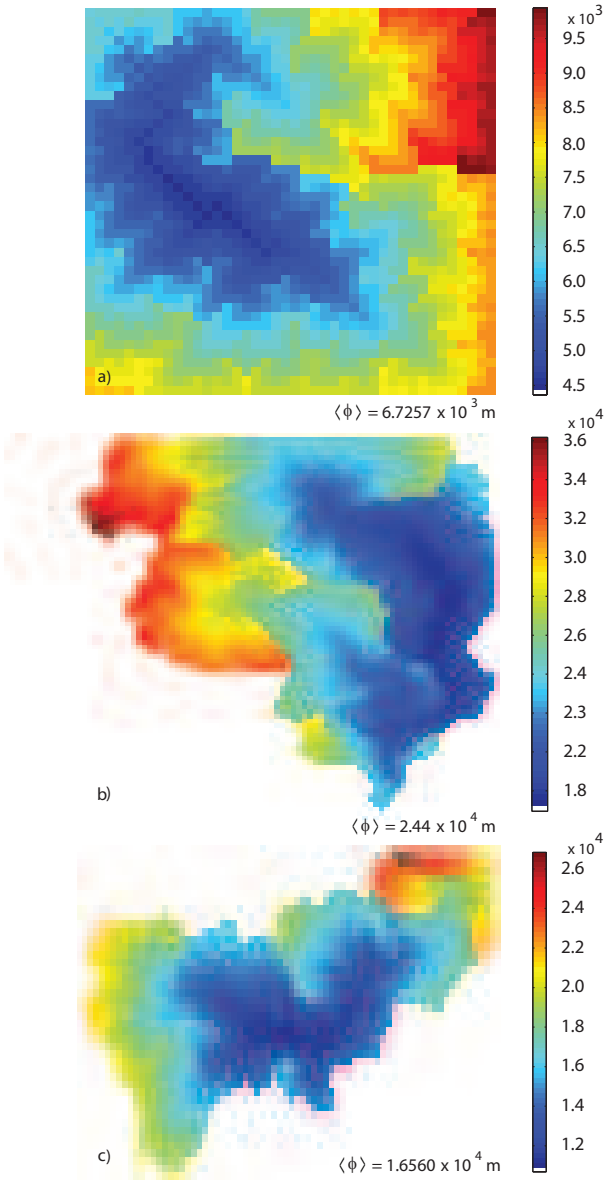

Figure 2.4: Spatial patterns of the streamlink-diameters for a) OCN, b) Cordevole and c) Tanaro 3) basins, shown at $CG = 150, 300$ and 400 respectively. The diameter ϕ_i, is defined as the average length from the streamlink i to all other streamlinks j in the network.

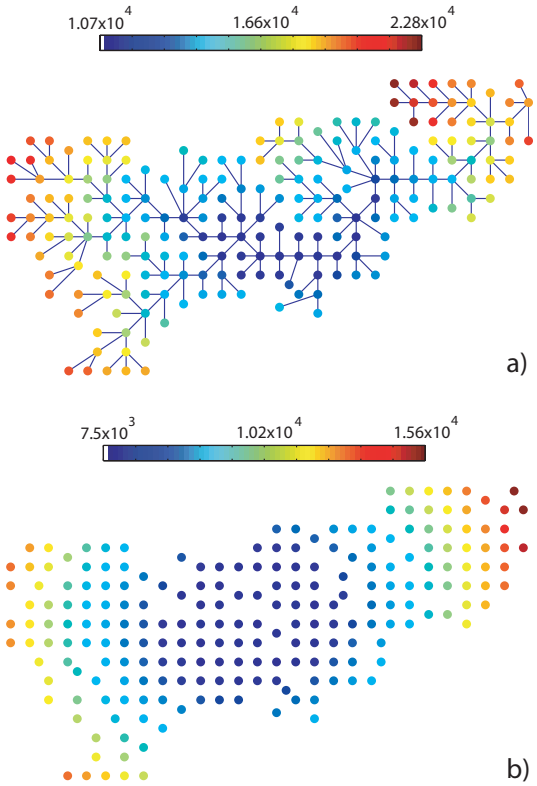

Figure 2.5: Diameter map for: a) Tanaro river basin; b) Tanaro 2-D landscape. In both domains the grain-length is 1000 m

The mean link length $\langle l \rangle$ for the RN in the MMRS is 110 km and in the 2-D counterpart is 52 km. Figure 2.6 shows the diameter in the MMRS RN without considering the curvature of the Earth (a), in the MMRS 2-D where the diameter is computed as simple euclidean distances (b), and in the MMRS 2-D where the diameter is calculated with the great-circle approach (c). In Table 2.1 the minimum, maximum and mean value of the diameter are reported according to the different topology and methods adopted, also determined by the euclidean distance and the flat-surface approximation. The latter (and other possible alternative methods) do not produce much variation for modeling the dispersal of species. Therefore the great-circle approach, widely accepted in geodesy, has been chosen for the ensuing calculations.

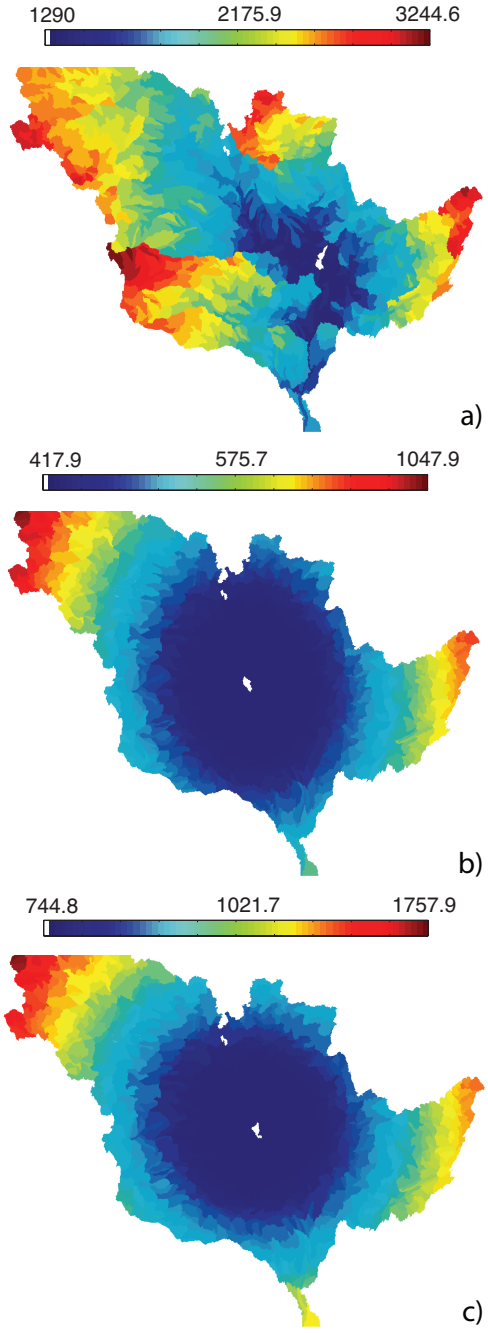

Figure 2.6: Diameter map for: a) MMRS river network without considering the curvature of the Earth; b) MMRS 2-D (the diameter is computed as simple euclidean distances) and c) MMRS 2-D landscape (the diameter is calculated with the great-circle approach). The distances are in km and the mean link-length is 110 km for the RN, 52 km in the 2-D case.

Table 2.1: d and ϕ as great-circle distances, pythagorean distances with parallel meridians, and euclidean distances

MMRS	min	max	$\langle . \rangle$
d between links			
RN	110	5060	2178.5
$2D_{gc}$	8.73	2928.3	1022.9
$2D_{sphere}$	8.72	2956.5	1024.3
$2D_{eu}$	5.26	1830.7	576.4
ϕ			
RN	1290	3244.6	2175.9
$2D_{gc}$	744.8	1757.9	1021.7
$2D_{sphere}$	745.12	1766.4	1023.1
$2D_{eu}$	417.9	1047.9	575.74

Another geomorphic quantity of ecological interest is the the distance, d, between links (alongstream) in the river network (RN), or between any pair of local communities in a 2-D landscape. The probability distribution functions for both d and ϕ for the MMRS are reported in Figure 2.7. Note that the bell-shape behavior of the distribution of d can be captured rather well by a Gaussian function. For the river network the alongstream distances are obviously larger than the euclidean distances in a 2-D landscape. This simple geometric fact has profound consequences on species dispersal. The diameter, ϕ, does not show an interesting probability distribution (figure 2.7 b)) for both the topology, RN and 2-D. Notice that $pdf(\phi)$ is the distribution of the mean values d_i of each link i. The great-circle approach for computing the distances instead provides clearly different results than assuming simple euclidean distances, underlying the importance of the Earth's curvature in modeling the dispersal of trees. Figure 2.8 shows probability distribution function of d obtained for the OCN, Cordevole and Tanaro river basins. In general, at approximatively the same total drainage area the mean distance between links, $\langle d \rangle$, is lower for more elongated basin than in more compact ones.

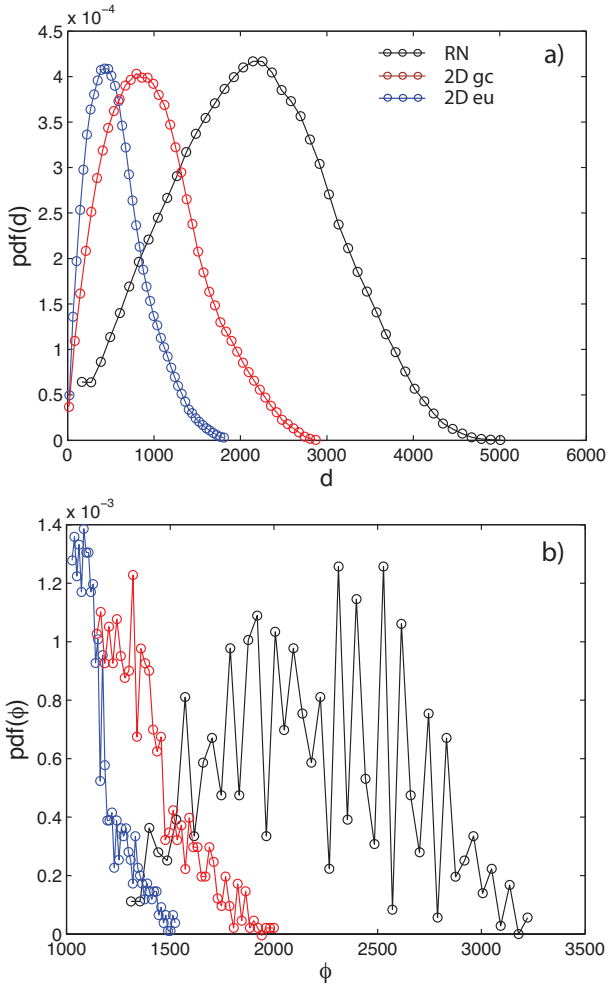

Figure 2.7: pdf(d) and pdf(ϕ) for the MMSR network, 2-D case with euclidean distance and 2-D case with geodetic distances.

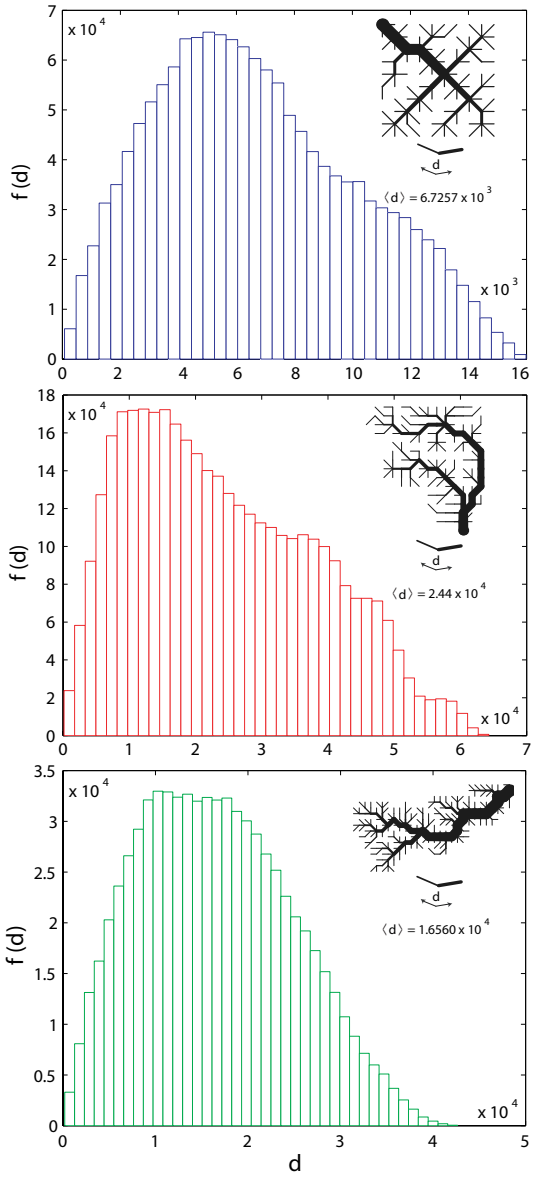

Figure 2.8: pdf(d) of the distance between links for the OCN, Tanaro and Cordevole.

Defining with A the drainage area, is possible to compute the exceedence probability that the are is greater or equal than a. The derived distribution $P(A \geq a)$ is a power-law with properties of scale-invariance and universality (*Rodriguez-Iturbe and Rinaldo*, 1997; *Convertino et al.*, 2007). Therefore, for example in the MMRS the scaling exponent of the power-law is very far from the universal value found in most of the basins throughout the world, i.e. $1 - \tau = -0.43$. In Figure 2.9 the distribution is show for the MMRS. The deviation from the universality in this case is due to the strong geological heterogeneity of the system that cover more than half of the USA. Nonetheless biodiversity patterns seem more influenced by the dissection exerted by the river network than the optimal geomorphological configuration. Therefore some optimal features of river basins, e.g. the elongation that can be represented by the elongation ratio $m = L_{\perp}/L_{\parallel}$, influence the biodiversity patterns. In chapter 4 and 5 a comparison between three different basins with different elongation has been performed. In Table 2.2 the longitudinal and transversal diameters of the employed networks have been reported. The MMRS and the Mis river basins constitute an anomaly among the elongated basins all around the world.

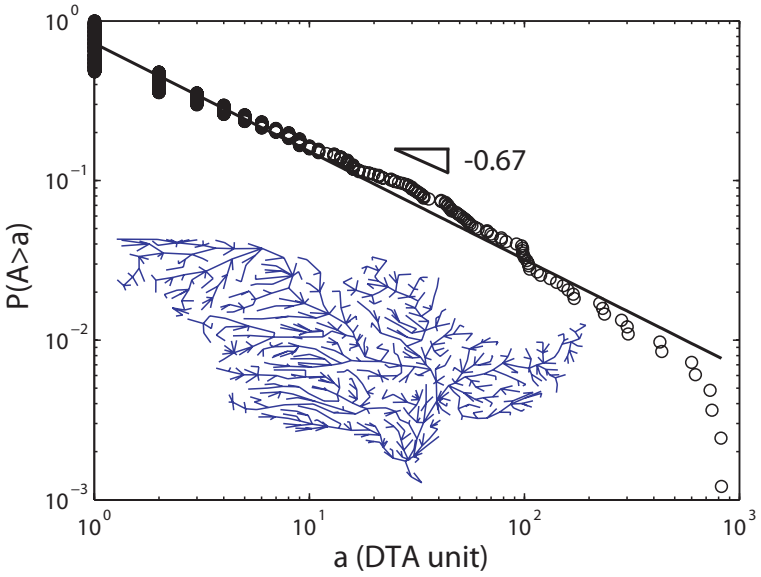

Figure 2.9: $P(A \geq a)$ of the MMRS in DTA units.

Table 2.2: $L_\perp$, $L_\parallel$ and elongation ratio $m = L_\perp / L_\parallel$ for the basins using in the research.

River Basin	$L_\perp$	$L_\parallel$	m
$Tanaro$	178	296	0.60
$Cordevole$	255	271	0.94
Mis	173	126	1.37
$MMRS$	2283.6	1304.9	1.75

In order to perform the analysis at the plot scale for the trees on the MMRS spatial plot-scale database (*USFS-FIDO*, 2009), three different grids have been superimposed on the spatial plot information whose unit-side length is 25, 50 and 100 km that imply a unit extension of 625, 2500, and 10,000 km^2 respectively. Since the average DTA area is 3900 km^2, the 50 km grid creates cells whose extension if more or less similar to the extension of a DTA. Figure 2.10 show the partition of the plot spatial distribution with the three different grids. This is useful to study how biodiversity patterns change with resolution, a subject which will be covered in detail in the following.

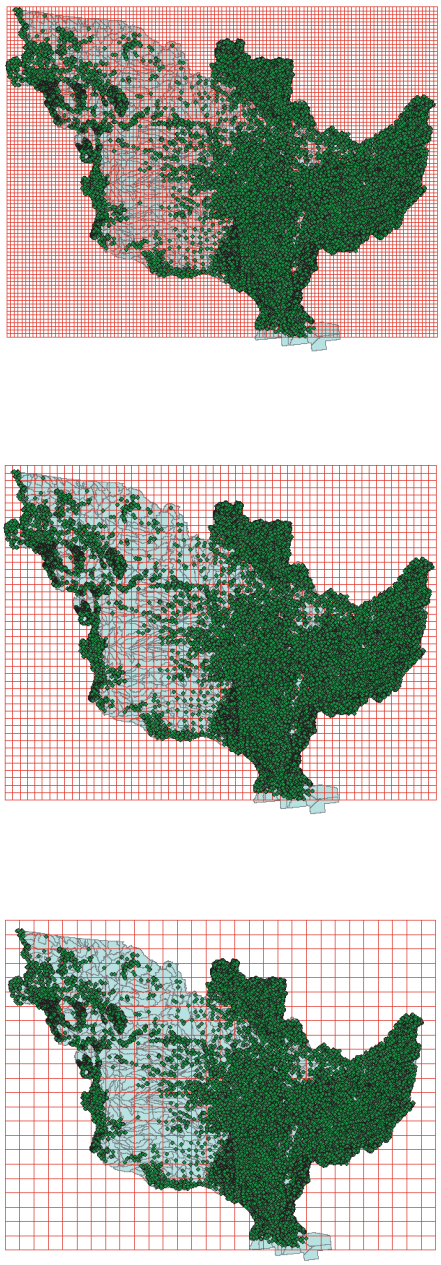

Figure 2.10: The figure shows the three grid employed in the resolution analysis performed on the tree data-plots of the MMRS: the resolution is $25,000$, $50,000$ and $100,000$ m (that is the side-length of the grid unit) for the top, middle and bottom plot respectively.

2.2 Trees and Fishes of the Mississippi-Missouri River Basin

In the analysis the 824 Direct Tributary Areas (DTAs) comprising the Mississippi-Missouri River System are populated with the occurrence data of 433 fish species (*Muneepeerakul et al.*, 2008), 233 big-trees species (*Konar et al.*, 2009) and 228 small-trees species. In Figure 2.14 is represented the network of the MMRS in which every DTA is connected to each other with a directed link according the flow-network in the case of fishes. For trees every DTA can be connected to all the others through euclidean distances.

For a detailed description of the model see chapter 3. The information about the fishes is given at the DTA scale. For the 2-D cases every local community can be connected with any other and a preferential direction does not exist. Figure 2.14 reports for each location the number of species that is defined as local species richness (LSR). The dataset of the fishes is the same used in *Muneepeerakul et al.* (2008) derived from the database compiled by NatureServe (*NatureServe*, 2004).

The tree dataset is from the US Forest Service (USFS) Forest Inventory and Analysis Database (FIADB) (*USFS-FIDO*, 2009). The information about the trees is given at the plot scale defined in section 2.1. Big trees are used in (*Konar et al.*, 2009) in order to test how climate change will impact their large scale patterns. The dataset of small trees has been introduced in this research for studying the spatial patterning of trees at different life-stages. Big trees are those whose diameter at breast height (dbh) is larger or equal than 5 inches, small trees those whose diameter is smaller than 5 inches.

Figure 2.11 represents the organization of the plots according *USFS-FIDO* (2009). USFS has created an inventory program to include all forested lands in the US, regardless of ownership or availability for forest harvesting (*Burkman*, 2005). The new program includes all forest land in all 50 States plus all of the territories and possessions of the US. It covers all public and private forest land such as reserved areas, wilderness, National Parks, defense installations, and National Forests (*Burkman*, 2005). A nationally uniform hexagonal-cells grid has been super-imposed by USFS over the existing set of sample locations, in order to provide a uniform basis for determining the annual set of measurement plots. For each hexagonal cell that is 6000 $ac = 24.28km^2$ there is one plot. The structure of the plot is represented in Figure 2.11 b) (in a three dimensional view) and c). An FIA plot consists of a cluster of four circular subplots spaced out in a fixed pattern. The plot is designed to provide a sampling frame in purpose of forest health and inventory (*Burkman*, 2005). Individual plots are remeasured in cycles of approximately $5-7$ years in the eastern U.S., and 10 years in the western U.S. Trees $1.0-4.9$ inches in diameter at breast height (dbh) are measured on 1/300-acre microplots, and trees 5.0-inches dbh and larger are measured on 1/24-acre subplots (*Bechtold*, 2003). Then one plot is 1/6 of an acre. As new trees grow into the microplots and subplots, they too are tracked until death *Bechtold* (2003). In each plots other variables are measured such as tree-volume or tree-basal area per year (growth), tree-height, woody debries, soil characteristic (*Burkman*, 2005). Growth is computed from sequential measurements of tree dbh over time. Tree dbh is usually located at 4.5 feet above ground line on the uphill side of the tree (*Burkman*, 2005). Crews identify which trees have survived since the previous inventory and which trees have grown into the plot, with a series of codes used to track tree history.

Figure 2.12 reports the patterns of the number of plots in each DTA, and the tree-density for both the dbh classes (big and small trees) considered. Here the tree-density is calculated as the ratio of the abundance of trees in each DTA and the respective area.

The information on trees at the DTA scale has been upscaled first by finding the number of plots in each DTA, after determining the density that is dividing the number of individuals in each DTA by the number of plots, then calculating the density per acre and inferring the abundance matrix multiplying the density to the forested area (FA in km^2). The forested area is given by the product of the forest fraction (USFS Forest Cover Types (*USFS*, 2009; *Konar et al.*, 2009)) multiplied by the DTA area computed using an Albers equal area conic projection of the USGS DTAs map given at geographical coordinates (*Bertuzzo et al.*, 2009). For the fishes database the abundance is not available therefore it is not possible to compare a priori the habitat capacity distribution of the model with data. For big trees and large trees the information about the abundance is instead available (*USFS-FIDO*, 2009).

Figure 2.13 reproduces the pattern of the average annual precipitation (AAP, (*USDA*, 2009; *Konar et al.*, 2009)), the average annual runoff production (AARP, (*Gebert et al.*, 1987; *Muneepeerakul et al.*, 2008)) and the forested area (FA, (*USFS-FIDO*, 2009; *USFS*, 2009; *Konar et al.*, 2009)) in km^2 that is the same for all the trees regardless the dbh class.

Figure 2.14 shows the local species richness (LSR) of fishes (*Muneepeerakul et al.*, 2008), big trees (*Konar et al.*, 2009) and small trees. As it will be discussed in chapter 3, LSR for fishes reflects closely the spatial patterns of the average annual runoff production (AARP, Figure 2.13, *Gebert et al.* (1987); *Muneepeerakul et al.* (2008)), while the LSR of trees partly reflects the forested area (FA) and the average annual precipitation (AAP, Figure 2.13, (*USDA*, 2009; *Konar et al.*, 2009)). Despite the great spatial similarity of the LSR patterns for big and small trees, visible at eyesight, it will be suggested that other, more quantitative, biodiversity patterns may differ considerably from corresponding reference patterns in the key drivers, thus adding interest to our search.

Figure 2.15 shows the pairwise similarity of the species using a metric called Jaccard Similarity Index (JSI) defined as the ratio between the number of common species and the total number of species in both local communities considered. The calculation is done among all the local communities of the ecosystem. In brief, $JSI = 1$ means total similarity in species composition while $JSI = 0$ means no similarity at all between any pairs of links at distance d. The JSI shows common features that seem to exist between taxa, whereas other biodiversity patterns such as the cluster-size distribution (chapter 5) suggest otherwise. Then quite interestingly the pairwise similarity of the species richness exhibits common features for trees and fishes, therefore some speculations can arise about common self-organizing optimal organization mechanisms of nature independently of taxa. We believe the dispersal have an overall main influence above all the other factors in determining the organization of species.

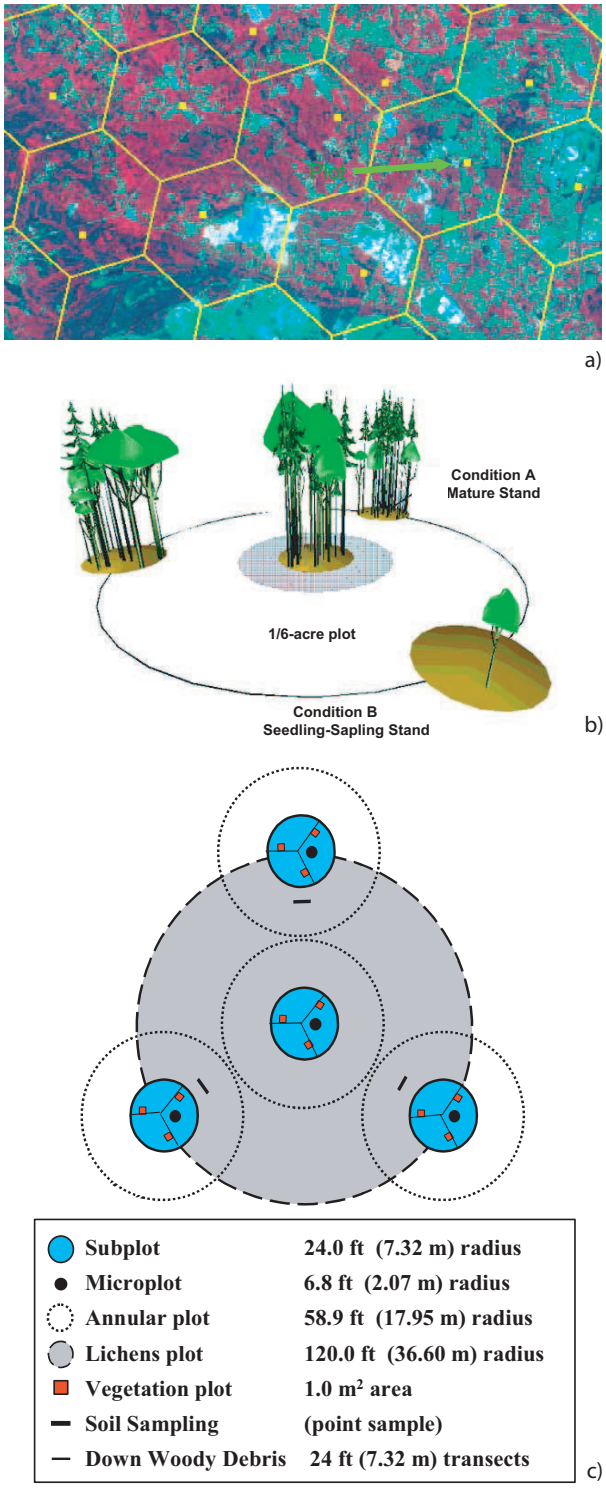

Figure 2.11: Organization of the trees-plot data in USA: a) hexagonal-cells grid in which every cell is 6000 *ac* and is sampled one plot (with permission of Mike Hoppus (NE-FIA), 2009); b) 3-D illustration of a plot that cover 1/6 *ac*; c) sampling organization in a plot composed by four microplots whose are is 1/24 *ac*.

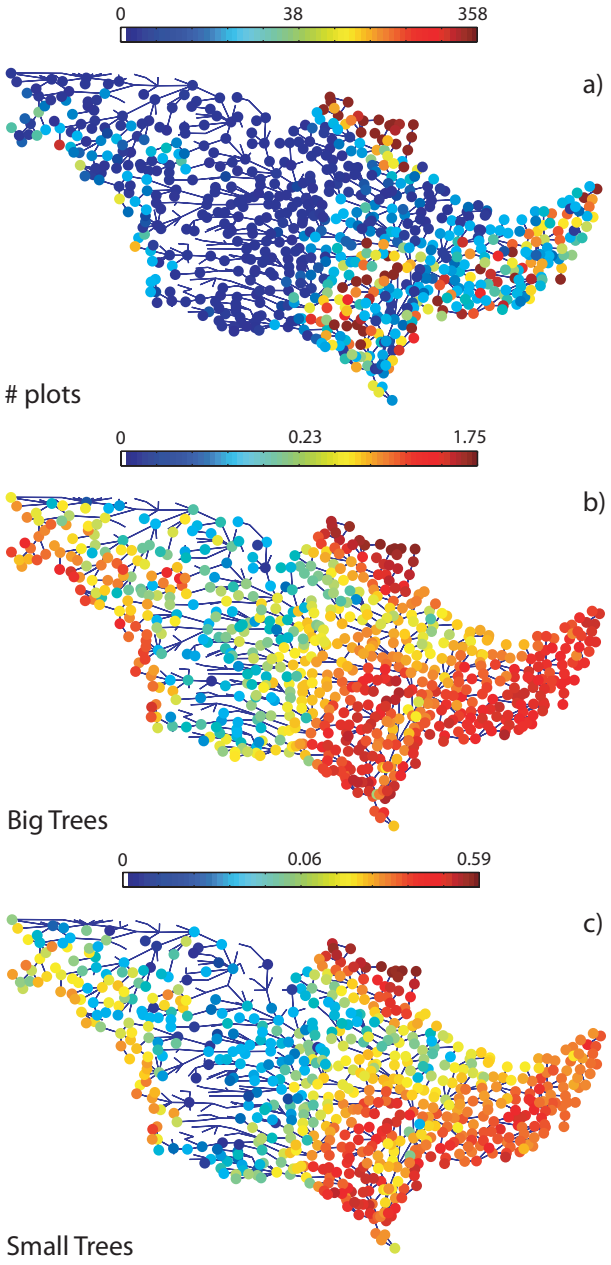

Figure 2.12: Data derived from FIA-USFS data: a) number of plots in each DTA; b) and c) tree-density per km^2 for big and small trees respectively.

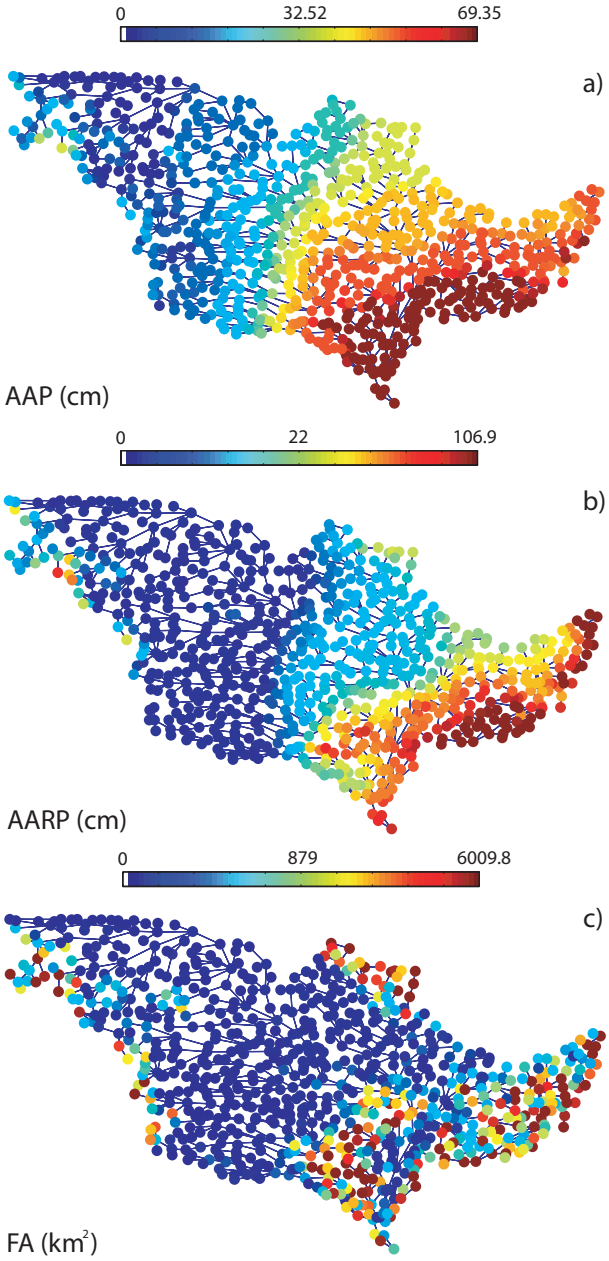

Figure 2.13: Patterns of AAP (*USDA*, 2009; *Konar et al.*, 2009) (a), AARP (*Gebert et al.*, 1987; *Muneepeerakul et al.*, 2008) (b), and FA (*USFS-FIDO*, 2009; *USFS*, 2009; *Konar et al.*, 2009) (c) in km^2.

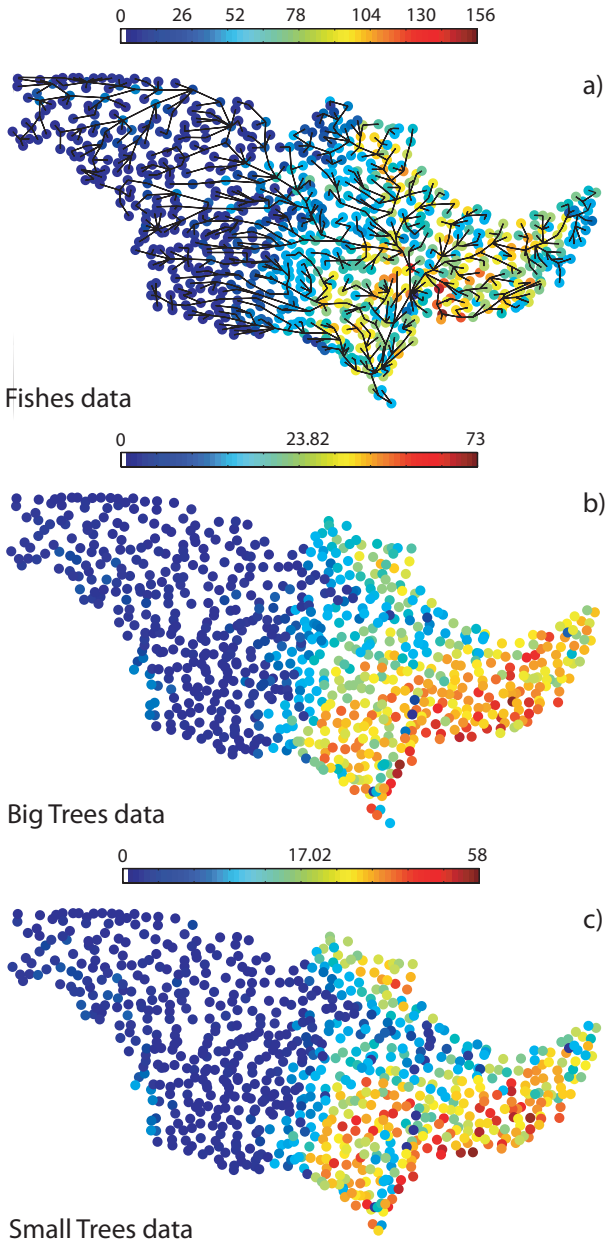

Figure 2.14: LSR patterns of fishes and trees data in the MMRS

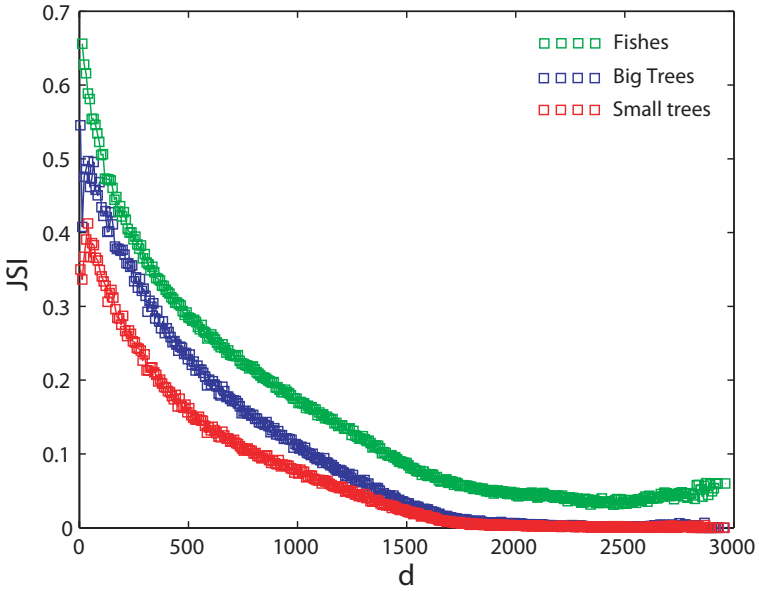

Figure 2.15: Jaccard Similarity Index for fishes and trees of the MMRS.

Figure 2.16 shows the most present species, the second most common species, the fiftieth most common species, and the one-hundredth most common species of the big-trees database (*USFS-FIDO*, 2009; *Konar et al.*, 2009). In figure 2.16 the choice of the rank for the species ranked in occupancy is arbitrary. For the occupancy, the first most common species seems to be placed according the availability of water that is higher in the east part of the MMRS. Only the most common species occupy most of the entire basin. The sharp decrease in species richness occurs a bit west than the 100 *deg W* meridian, which is also the transition of the sharp gradient in the mean annual precipitation, but probably there are also edge effects difficult to capture. The fiftieth most present species is located along a stripe from the east side of the Appalachian to the midwest-US (c), and the one-hundredth most common species is located in the south of the Great-Lakes region (d). It is fascinating how nature produces such intriguing spatial distribution.

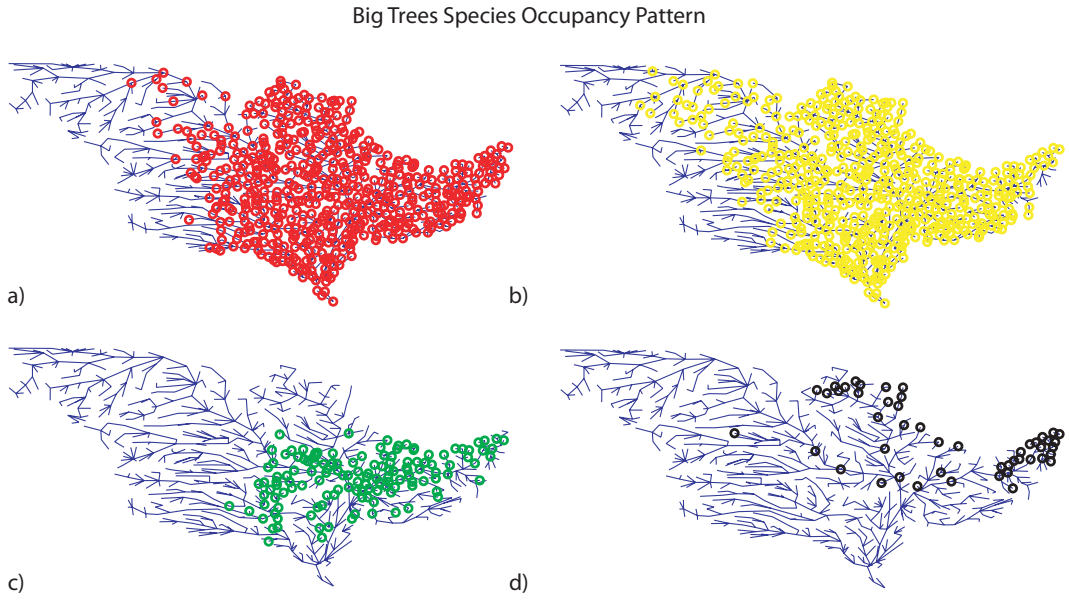

Figure 2.16: Big Trees Occupancy Patterns: a) red dots is the most present species; b) yellow dots is the second most common species; c) green dots is the fiftieth most common species; d) is the one-hundredth most common species.

Figure 2.17 shows the most abundant species, the second most abundant species, the fiftieth most abundant species, and the one-hundredth most abundant species of the big-trees database (*USFS-FIDO*, 2009; *Konar et al.*, 2009). The abundance distribution is quite different than the occupancy distribution. In particular, it seems that the abundance distribution captures the heterogeneities of the environment such as the geological heterogeneities. In (a) the most abundant species occupy the south-east part of the country, in (b) the second most abundant species is in a completely different geographic region, the west boundary of the basin, in (c) the fiftieth most abundant species is distributed in the midwest, and in (d) the one-hundredth most abundant species is concentrated in a small region in the east boundary of the MMRS (along the west side of the Appalachian approximatively).

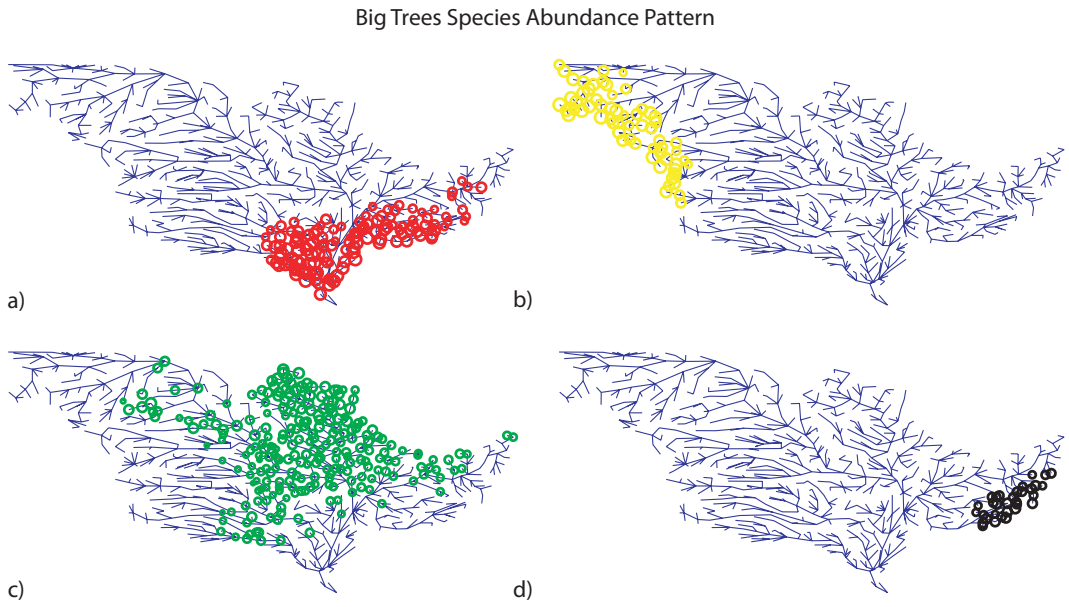

Figure 2.17: Big Trees Abundance Patterns: a) red dots is the most abundant species; b) yellow dots is the second most abundant species; c) green dots is the fiftieth most abundant species; d) is the one-hundredth most abundant species.

Figure 2.18 shows the relationship between abundance and occupancy for big and small trees species. The linear relationship indicates that the most abundant species are on average the most widespread. Considering the species one by one the one to one abundance-occupancy relationship is no more true as the pattern of Figures 2.16 and 2.17 shown. The scaling exponent in 2.18 different than 1 indicates that for species a perfect match between the number of individuals and the area that they occupy does not exist. This is more evident for big trees since the scaling shows an exponent larger than 1 than for small trees. Estimating the abundance from occurrence is one of the main debated issue in ecology (*He and Gaston*, 2007) that would be extremely important.

This is related as it will be shown in chapter 5 to the species-individual vs species-area relationship comparison. Briefly despite the strong heterogeneities, abundance is positively correlated with the area, against what the equilibrium theory of island biogeography (*MacArthur and Wilson*, 2001; *Connor et al.*, 2000) assumes.

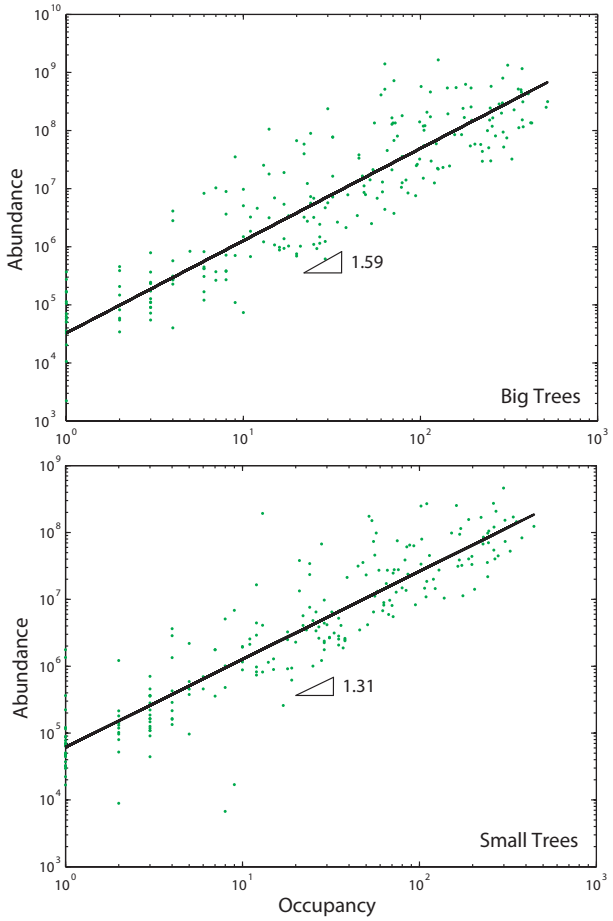

Figure 2.18: Occupancy-Abundance relationships for the Big Trees (top plot) and the Small Trees (bottom plot) of the MMRS.

Chapter 3

Model Description

"If people do not believe that mathematics is simple, it is only because they do not realize how complicated life is"
John Louis von Neumann

This chapter deals with the description of the mathematical models used in the ensuing analysis will be shown. In detail, section 3.1 defines the dispersal and reports an overview of the dispersal theory, section 3.2 describes the metacommunity models used in the ecosystems employed in the homogeneous case and incorporating the environmental heterogeneities. The cases of global and nearest-neighbor dispersal are also discussed. Observed features of the ecological patterns (of trees, fishes presence/absence data, clustering properties) will be compared with those deduced from the results of a neutral metacommunity model used to infer freshwater fish biodiversity in the Mississippi-Missouri fluvial system (*Muneepeerakul et al.*, 2008). In the following, for the thesis to be self-contained, we fully describe the framework adopted. Note that the particularization to fish communities will be relaxed to vegetation ecosystems without particular problems – key differences being easily pinpointed.

3.1 Dispersal Mechanisms

Biological dispersal refers to a species movement away from an existing population or away from the parent organism. Through simply moving from one habitat patch to another, the dispersal of an individual has consequences not only for individual fitness, but also for population dynamics, population genetics, and species distribution (*Hanski and Giplin*, 1997; *Hanski*, 1999). Many of the questions addressed by ecology and evolutionary biology require a good understanding of species dispersal patterns (*Broquet and Petit*, 2009).

The dispersal is a complex phenomena in nature that nowadays constitutes a large branch in ecology. The study of dispersal is often related to the study of the animal movements since animals are one of the most responsible agents in the spreading of seeds and fruit types of plants. The interplay between zoology and ecology generating biological changes has coined the discipline of movement ecology that consider also the interplay of dispersal with the landscape features and the environmental heterogeneities

(*Bartumeus and Levin*, 2008; *Damschen et al.*, 2008; *Getz and Saltz*, 2008; *Nathan et al.*, 2008a; *Revilla and Wiegand*, 2008). The study of dispersal started with the empirical experiment about the dispersal of pollens and seeds (*Crocker*, 1938).

Dispersal was mostly developed as an autoecology theory since studies were performed considering empirically one individual species by one in relation to its environment rather than a global phenomena of the entire species population in the ecosystem. Dispersal has been measure for example for seeds dispersed by birds *Schurr et al.* (2009), seeds dispersed by wind *Crocker* (1938); *Nathan et al.* (2001, 2002a, 2008b), seeds dispersed by elephants *Blake et al.* (2009), for earthworms *Edwards and Bohlen* (1977), for Zebra mussel *Stoeckel et al.* (1997), for butterflies *Gilbert and Singer* (1975), for Snowy plovers *Stenzel et al.* (2007), and for elephants *Mennell and Scholes* (2008); *Blake et al.* (2009). Here because the neutral theory applied to a metacommunity the dispersal is assumed the same for all the individuals regardless species that is a strong assumption. Therefore the neutral theory is a null theory capable to test the importance of the neglected elements when they are not included in the description of the species that are considered more similar rather than more different (*Hubbell*, 2001; *Gewin*, 2006; *Hubbell*, 2006). If the ingredients not included in the neutral model are important the results do not reproduce the real biodiversity patterns, e.g. like in *Dornelas et al.* (2006); *Volkov et al.* (2007) for coral reefs and in *Graves and Rahbek* (2005); *Ostling* (2005) for Amazonian birds.

Dispersal mechanisms lead to pattern formation (*Borgono et al.*, 2009) and dispersal limitation is a root of the neutral theory coupled with the stochastic speciation and death of individuals. Niche theories are in some way opposed to dispersal theories and especially the neutral theory in which was introduces the concept of dispersal limitation *Hubbell* (2001). Niche theories put the main importance on the environmental heterogeneities that produce trade-offs and strong competition between species (*Gewin*, 2006) in a view of selective evolution. More in general in the ecological literature there is a separation between niche-assembly and dispersal-assembly theories (*Fargione et al.*, 2003). One of the main issue is the estimation of the dispersal length of the species since it is difficult to measure this quantity in the field. Therefore nowadays genetic and molecular biology techniques are able to provide reliable values of the dispersal length, as average and maximum value (*Broquet and Petit*, 2009). There is also a not clear definition of dispersal length since it is possible to distinguish between *natal dispersal length* and *breeding dispersal length*: the first is in general defined as the distance traveled by plant seeds, larvae, pollen or juveniles between the birthplace and the first breeding of fertilization site, while the latter is defined as the distance traveled by mature individuals between two successive breeding areas. The natal dispersal is usually the most used and it is proportional to the home-range size (*Bowman et al.*, 2002) that is the spatial extent or outside boundary of an animal's movement during the course of its everyday activities (*Burt*, 1943).

More in general dispersal refers to the movements of living or not-living organisms in their own environment: so we can speak about dispersal or individuals (*Ammerman and Cavalli-Sforza*, 1984b), dispersal of cells (*Marco et al.*, 2009), dispersal of epidemics (*Bertuzzo et al.*, 2008), dispersal of contaminants (*Botter et al.*, 2005), solutes and sediments in water streams and soils, hydrodynamic and geomorphological dispersal (*Rinaldo*

et al., 1991) to give some examples. Dispersal of animals and seeds is influenced by the vegetation patterns and it also constitutes itself the cause of the distribution of plants underlying a complex dynamic with feedback between their movement and the resulting patterns (*Hein et al.*, 2004; *Boyer et al.*, 2006).

The dispersal mechanism implemented is through a dispersal kernel that determines how the offsprings are dispersed in the network or in the 2-D open landscape. The kernel has a back-to-back non directional structure meaning that the species can move upstream or downstream with equal probability in the river network case, and in every direction with equal probability in the 2-D landscape case. The model was originally developed in (*Muneepeerakul et al.*, 2007b) and modified in (*Muneepeerakul et al.*, 2008). Here the version used is the latter one (*Muneepeerakul et al.*, 2008) both for modeling the fishes and trees in the MMRS, and for the numerical studies about the influence of the parameters on biodiversity patterns. The dispersal kernel is basically the probability density of moving a particular distance from the starting point where the individuals starts its spreading. For the fish case the dispersal kernel "fat tailed" means that the long distance tails of the distribution decay slowly such that the variance of the dispersal kernel does not converge and is therefore undefined (or infinite). This means that is possible to have long distance dispersal events (LDD), but still the mean dispersal distance could be very small. In this case long distance dispersal is defined based on the overall shape of the dispersal kernel. The choose of the exponential-Cauchy distribution appears then the most appropriate. As for trees the kernel has an exponential decay and this pose some questions about the possible existence of different optimal survival strategies for trees than the long distance dispersal (LDD) or it is also plausible that fishes are free to explore long distance dispersal while seeds are trapped by the pre-existing canopy producing causing the less effective short distance dispersal. Long distance dispersal has been in fact proved to be an optimization strategies of populations. This suggests that species would adopt a fat-tail dispersal strategy that should maximize their colonization and survivability (*Alerstam et al.*, 2003; *Davies et al.*, 2004; *Pearson and Dawson*, 2004; *Bohrer et al.*, 2005; *Soons and Ozinga*, 2005; *Nathan*, 2006; *Schurr et al.*, 2009). It will be shown how biodiversity patterns, then species, are less sensible to the topology of the ecosystem when they spread with a fat-tail dispersal, then they are more robust to endogenous or external perturbations in the connectivity. The definition of long distance dispersal is not unique. Both absolute and proportional definitions of LDD are used in the literature (*Cain et al.*, 2000; *Nathan*, 2005). The absolute definition identifies LDD events as those dispersal events that are longer than a specified threshold distance (e.g. 1 km). The alternative to this absolute definition is a proportional definition, which defines LDD events as those that exceed a certain high quantile of dispersal distance (e.g. the 99 % quantile as the distance exceeded by only 1 % of all seeds). It is important to note that the proportional definition actually identifies extreme dispersal events rather than LDD events: in a species with highly restricted seed dispersal, the 99 % quantile of dispersal distance might be located at only a few meters, a distance that typically will not be considered as "long" (*Schurr et al.*, 2009). Actually the threshold distance for the absolute definition should be chosen based on the question studied: when studying the dynamics of a plant species in a fragmented landscape, such an LDD threshold could be the typical distance between neighboring habitat fragments (e.g. the average intra-

species patch distance). More in general the LDD events are function also of the size of the organism rather than the size of the ecosystem in study.

Moreover the dispersal can be density dependent, the dispersal range can be influenced by environmental constraints or dispersal barriers, and it can be operated by animals or referred to plant-dispersal mechanisms (*Schurr et al.*, 2009). In the latter case it is called seed dispersal and it can be by gravity, wind (*Nathan et al.*, 2002a,b), ballistic (autochory), animals and water (*Schurr et al.*, 2009). *Seidler and Plotkin* (2006) shows how the conspecific spatial aggregation of trees is correlated with the mode of seed dispersal. Seed dispersal has many consequences for the ecology and evolution of plants. Dispersal is necessary for species migrations, and in recent times dispersal ability is an important factor in whether or not a species transported to a new habitat by humans will become an invasive species (*Lensink and Neubert*, 2003). Dispersal is also predicted to play a major role in the maintenance of species diversity. Dispersal of seeds away from the parent organism has a central role in two major theories for how biodiversity is maintained in natural ecosystems, the Janzen-Connell hypothesis (*Janzen*, 1970; *Hyatt et al.*, 2003) and recruitment limitation (*Harms et al.*, 2000).

3.2 Neutral Metacommunity Model on River Networks and 2-D Landscapes

This study makes use of the neutral metacommunity model developed by *Muneepeerakul et al.* (2008) for freshwater fish biodiversity in the MMRS. In this framework, the system is made up by different local communities with a certain habitat capacities, interconnected through the ecological corridors provided by the river network or in the 2-D landscape. No life-history optimal survival strategies (*Löbel and Rydin*, 2009; *Löbel*, 2009) are accounted in the model.

Variables and parameters employed in the model are summarized in tables 3.1, 3.2, 3.3, 3.4, 3.5 that will be explained in the following. Every local community is assumed to be always saturated at its capacity, i.e., no available resources are left unexploited. At each time step, a random individual (taxon unit) in a random local community (LC), i.e. a link in the space-filling network or a pixel in a 2-D landscape, dies and its site becomes available for colonization by an other taxon unit. With probability ν, the per-site diversification rate, the site will be colonized by a species not already present in the system. Note that the per-site diversification is a rate per birth due to speciation, to external introduction of non-native species, or to immigration (and re-immigration) of a new species from outside the system. With the remaining probability $1 - \nu$, the site is colonized by a propagule (if trees) or fish unit already existing in the system. The colonization process is modeled through the dispersal of taxon units "produced" by the local communities. The per-site diversification rate, ν, is given by the ratio θ/N, where θ is the overall diversification rate (or biodiversity parameter (*Hubbell*, 2001)), that is, the average number of new species introduced at each generation, and N is the total number of individuals in the system. In the model each spatial unit, i.e. a local community (a link in the river network ,e.g. between two DTAs in the MMRS, or pixel in a 2-D

landscape) represents a local community (fish or tree unit in the MMRS) interconnected with all the others through the ecological corridors provided by the river network or Euclideanly in a 2-D landscape topology. The whole set of local communities forms the metacommunity. In the heterogeneous case, e.g. for the MMRS each DTA has a different taxon habitat capacity, H, defined as the number of "taxon units" (fish or tree units in the MMRS) sustainable by the resources present in that particular site (DTA in the MMRS). A taxon-unit can be thought of as a sub-population of individuals of the same taxon belonging to the same species. A taxon (plural taxa) is a population, or group of populations of organisms which are usually inferred to be phylogenetically related and which have characters in common that differentiate the unit (*ICZN*, 1999). Then a metapopulation is composed by all the taxon-units belonging to the same species, and the metacommunity is composed by all the metapopulations of all the species present in the ecosystem. The model does not need to assume any description about the availability of resources since the the dynamics of birth and death is purely stochastic.

We investigate a number of biodiversity patterns, to be described in the next section (chapter 4), after the system has reached the stationary state where the total number of species in the system, S_T, fluctuates within a narrow range and the ecological patterns observed no longer show systematic variations. A statistically stationary value of S_T is achieved on average at the same simulation time regardless of the ecological scenario implemented, the basin and the resolution. As for the simulations reported in chapter 4 performed at different coarse-graining level of resolution (chapter 4), we proceed as follows for the three basins employed:

- generation of the networks at four different CG-resolutions (that is the grain-size of the ecosystem (see chapter 4, section 4.2));
- run the model to the stationary state for the four CG-resolutions;
- analysis of macroecological patterns emerging from each resolution. The probability P_{ij} that an empty site in local community i is colonized by an individual from local community j is given by:

$$P_{ij} = \frac{K_{ij} H_j}{\sum_j K_{ij} H_j} \,, \tag{3.1}$$

  where K_{ij} is the dispersal kernel (see below) and H_j is the habitat capacity of local community j, assumed to be uniform and constant for all links without environmental heterogeneities. In the homogeneous case H_j is given by the ratio N/n between the total number of individuals N (kept constant under coarse graining) and the number of links n that depends on the coarse graining level adopted. Therefore n decreases from the finest to the coarsest resolution in the uniform case. Table 3.1 gives the values of n and H employed for all the river networks and the different coarse-graining levels used in this paper. See Table 3.2 for the numerical values of the parameters assumed in the simulated ecological scenarios called a for the exponential kernel and b for the exponential-Cauchy kernel.

The coarse-graining dependence study has not been performed for the MMRS for

which the simulations have been run with the network at the DTA level (see figure 2.3 in chapter 2). Figure 3.1 reports the trend of the total number of species S_T toward the stationary state of a simulation in the river network case. The stationary state is quite insensible to the number of species the beginning of the simulation (in this case $S_{T=0} = 1000$). The fluctuations in time are due to what is called "ecological drift" according the unified neutral theory (*Hubbell*, 2001). When the model reaches a stationary state (e.g. here after 400 generations) it is run other 10 cycles and the results presented are the average over those 10 cycle. At each time step the dynamic of birth-death and dispersal take place among the species. Table 3.1 reports n and H that are the number of links and the habitat capacity respectively, employed for the simulations of biodiversity patterns reported in chapter 4. The coarse-graining resolution is the grain size of the river network.

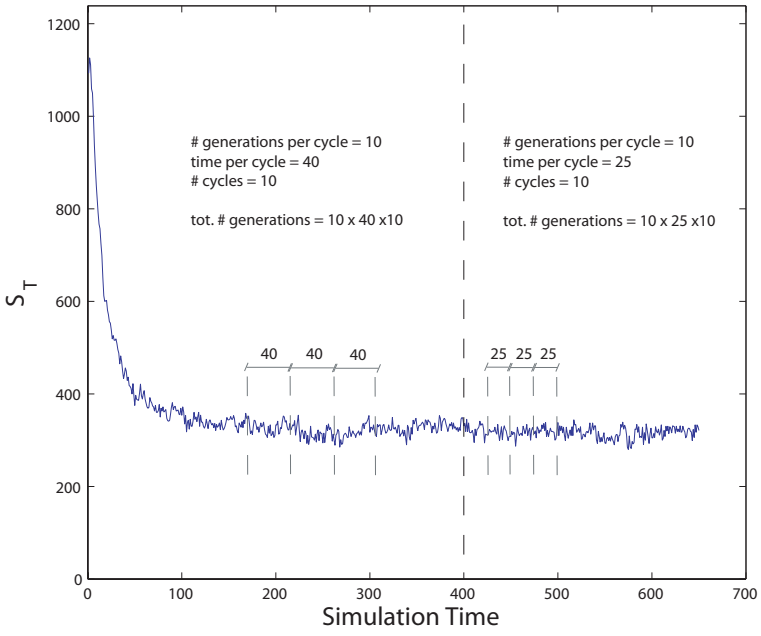

Figure 3.1: Behavior of S_T vs Simulation Time for the Tanaro river network. The qualitative behavior of the curve is the same for all the river networks, therefore it strongly depends on the parameter of the model.

Table 3.1: Variables and parameters used in the neutral metacommunity model at the various coarse-graining levels. The geomorphological parameters are reported in upper white rows, while the ecological parameters are listed in the bottom grey rows. The analysis of results is done after the system reaches stationarity when the stochastic variable S_T, i.e. the number of species in the whole basin, is stable.

	OCN				Cordevole				Tanaro			
CG	150	200	300	550	300	500	750	1500	400	600	800	1000
n	1711	924	383	96	2755	1036	478	114	1208	546	308	217
H	58	108	261	1042	36	97	209	877	82	183	325	461

The dispersal kernel K_{ij} represents the probability that a propagule produced at a site j arrives at i after dispersal. The general kernel structure implemented in this thesis is given by:

$$K_{ij} = C_j \left[exp\left(-\frac{L_{ij}}{d_l}\right) + \frac{b^2}{L_{ij}^2 + b^2} \right] , \tag{3.2}$$

where C is a normalization constant determined numerically such that, $\sum_i^N K_{ij} = 1$ for every local community j (i.e. individuals do not travel out of the system, which is assumed to be isolated), where N is the total number of communities. The kernel used herein is a combination of a back-to-back exponential (with mean dispersal length d_l, when $b = 0$) and Cauchy distributions which adds a fat-tail character to K_{ij} (*Muneepeerakul et al.*, 2008). Then the model has at most four parameters for calibration: i) the mean habitat capacity $\langle H \rangle$ that defines the total number of taxa units in the system $N = \langle H \rangle \cdot n$; ii) the average number $\theta = \nu \cdot N$ of new species introduced at every N deaths and iii) the two shape parameters of the kernel, d_l and b.

Equation 3.1 is written for the local-community j as a whole, and it manifests clearly the neutral assumption made at the individual level since the probability P_{ij} is the same for all the local communities. Here it is proposed a formulation of the probability at the species-level that is different for the present species. The neutral theory assumes no species-specific competitive advantages and therefore the probability that an empty site is colonized by a certain species depends only on the relative abundance of individuals of that species present in the site. From the above assumptions, the probability P_{ij}^s that an empty unit in site i will be colonized by a propagule of the species s arriving from site j can be expressed as:

$$P_{ij}^s \;=\; (1-\nu)\, \frac{K_{ij}\, H_j}{\sum_{k=1}^{n} K_{ik}\, H_k} \;\; \frac{n_j^s}{H_j} \tag{3.3}$$

where H_k is the habitat capacity of site k and n_j^s is the abundance of taxa units of the species s present in the site j. The first ratio in the right-hand side of equation (3.3) represents the probability that the empty site in local community i is taken by a taxa-unit arriving from local community j, whereas the second ratio expresses the probability that the propagule belongs to the species s. To determine which species from local community

j succeeds, it is drawn an individual at random from local community j. This action of random draw is what equation 3.3 describes by multiplying with n_j^s/H_j. The neutral assumption assumes equivalence at a per capita level, not at a species level. It means that every individual is the same regardless of species. For example, if species A has 1000 offsprings (recruits) and species B has 10 offsprings, then species A's success probability is 100 times higher; if species B has 10000 offsprings and species A has 100 offsprings, then species A's success probability is 100 times lower. That is, species identity does not matter, only relative abundance of offsprings does.

Note that an increment in the distance L_{ij} in equation 3.2 occurs when traveling along the network, or Euclideanly in a 2-D landscape, one moves from one local communities to an another. In the thesis we often use the term "mean dispersal length" for the dispersal parameter d_l, but the meaning is not as the real mean length that species experience because the dispersal. The reader should interpret d_l as a parameter of the metacommunity model useful to capture the biodiversity patterns. The ecological distance L_{ij} is defined as the length of the shortest path from link j to link i through the network. Generally the ecological distance, L_{ij}, for the network is defined as:

$$L_{ij} = ND_{ij} + w_u NU_{ij} \ , \tag{3.4}$$

where ND_{ij} and NU_{ij} are the downstream and upstream distances comprising the shortest path from link j to link i throughout the network. The parameter w_u represents the directional bias in the spreading of individuals along or not the network. Here the transport is assumed to be unbiased ($w_u = 1$) implying equal probability to travel either downstream or upstream (*Muneepeerakul et al.*, 2008b) so the ecological distance is merely the geometrical distance $L_{ij} = ND_{ij} + NU_{ij}$ along the network. Distances in the ecosystem are measured in meters, rather than in pixel or link units, allowing for proper comparisons across different coarse-graining levels. The middle point of each streamlink is used for calculating the distances used in the kernel, as well as for the calculation of the diameter and the distance between links. In the case $b = 0$ the structure of the kernel is exponential and this has been used for modeling the biodiversity pattern of big and small trees. As for the trees the fitting of the biodiversity patterns as in (*Muneepeerakul et al.*, 2008) is much well obtained in *Konar et al.* (2009) (that include permeable boundaries, ν in function of the distance, and a different kernel for the immigration and speciation), but as for the cluster-size pattern the simple model here used is sufficient. As for the trees case in the MMRS the distances calculated according the great-circle approach (Chapter 2) have been used, and they determined by the following:

$$L_{ij} = R_m \, arccos(cos(DTA_{2i})cos(DTA_{2j})cos(DTA_{1i} - DTA_{1j}) + sin(DTA_{2i})sin(DTA_{2j})) \ . \tag{3.5}$$

where $R_m = 6371.01 \ km$ is the average radius of the Earth thinking is as a spheroid, DTA_2 is the latitude and DTA_1 is the longitude of each DTA. In table 2.1 of chapter 2 are also reported the distances calculated using the flat-surface pythagorean distance $L_{ij} = \sqrt{R_e^2 \, (DTA_{2i} - DTA_{2j})^2 + R_p^2 \, (DTA_{1i} - DTA_{1j})^2 \, cos((DTA_{2i} + DTA_{2j})/2)^2}$, where R_e and R_p are the equatorial and polar radii, but they do not differ so much

compared to the more exact great-circle distances. The euclidian approach ($L_{ij} = (\Delta DTA_{2ji}^2 + \Delta DTA_{2ji}^2)^{1/2}$) leads clearly to relevant mistakes in the estimation of the distances. If $b > 0$, the kernel is a combination of the exponential and the Cauchy components and exhibits a fat-tail that produces a long-distance dispersal. This has been used for modeling the biodiversity patterns of fishes in the MMRS (*Muneepeerakul et al.*, 2008) and for studying their spatial clustering in the present thesis. In the first case Equation 3.2 decays exponentially instead. In the latter case it has a power-law decay that implies higher probability to have the same species at larger distances L_{ij}. Notice that, it is under the latter hypothesis that the model of *Muneepeerakul et al.* (2008) better describes the pattern of fish biodiversity in the Mississippi-Missouri river system.

As for the details of interactions depending on the fine spatial scales in streamlinks or in local communities in a 2-D landscape, consistently with a neutral metacommunity approach (*Hanski*, 1999), we assume that they may be well-summarized by using finite local habitat capacities. Ultimately, this rests on the limited availability of local resources and turns out to be a plausible assumption under stationary conditions (*Kirley*, 2005; *Azaele*, 2006; *Azaele et al.*, 2006; *Wang et al.*, 2008). On the other side, here we are describing an isolated ecosystem which evolves on temporal scales smaller than those characteristic of the actual speciation. Therefore, one expects that the temporal variations in the diversification rates are negligible. Nonetheless on chapter 4 the focus is on the effect of resolution change solely, and on chapter 5 on the effect of scale change also.

In the following, we shall refer to two different ecological scenarios implied by different structures of the dispersal kernel: (a) the exponential case, and (b) the exponential-Cauchy case. The different sets of parameters are given in table 3.2 and the results are reported in chapter 4.

Table 3.2: Values of the parameters used in the neutral metacommunity model for the two ecological scenarios as defined in the text. For brevity we refer to (a) as the exponential case, and to (b) as the exponential-Cauchy case.

Case	d_l	b	θ	N
a	500	0	40	1×10^5
b	500	300	40	1×10^5

3.2.1 Habitat Capacity in the MMRS

For modeling the dispersal of fishes and trees in the MMRS the habitat capacity has been taken proportional to the average runoff production and the DTA area in the fishes case, and proportional to the product of the forested area (FA) and the average annual precipitation (AAP) in the trees case. Runoff depends on precipitation, evapotranspiration and infiltration; an AARP map for the MMRS is reported in figure 2.13 of chapter 2. Obviously the latter assumption is somewhat crude and does not take into account all the factors and processes that could determine resources and habitat availability at local scale in river networks (*Alexander et al.*, 2000; *Battin et al.*, 2003, 2008). Nonetheless, in particular at regional scales, we shall assume that runoff production and average annual precipitation coupled with the forested area, are the main drivers of the resource availability for freshwater fish and trees respectively, and we speculate that they are able to capture the broad spatial gradients of habitat capacity across wide and heterogeneous regions with strong climatic and environmental gradients as the one considered herein. In presence of strong gradients (see figure 2.13 in chapter 2), this model would be unlikely to replicate observed patterns of biodiversity in a system where habitat capacity were determined by allochthonous resource inputs that were independent of, rainfall, runoff (e.g., closed-canopy stream systems) and forested area (*Bertuzzo et al.*, 2009). The habitat capacity in the fish case is the same adopted my *Muneepeerakul et al.* (2008), that is

$$H_i = C_H \, \frac{AARP_i \; DTA_{ai}}{\frac{\sum_{k=1}^{N} AARP_k \; DTA_k}{N}} \; , \tag{3.6}$$

rounded to the nearest integer. DTA_a denotes watershed area, N (= 824) the total number of DTAs, C_H the estimate (due to rounding) of average habitat capacity, and $AARP$ the average annual runoff production. The results confirm that the weight of the drainage are is important to capture better the biodiversity patters.

The habitat capacity in the trees case is maximum value between the real LSR derived from data and the following,

$$H_i = C_H \, \frac{FA_i \; AAP_i}{\frac{\sum_{k=1}^{N} FA_k \; AAP_k}{N}} \; , \tag{3.7}$$

where FA is the forested area thats is derived by the product between the DTA_a and the forest fraction, AAP is the average annual precipitation, C_H is the average local community size of big and small trees that has been taken the same. Previous attempt have beed done considering H_i for trees proportional to the evapotranspiration that is $ET = AAP - AARP$ but with no better results. *Konar et al.* (2009) assumes H_i proportional to the forested area and C_H function of the average habitat capacity. In *Konar et al.* (2009) is also varied H_i in function of climate change scenario to investigate the effect of climate change on biodiversity patterns.

Table 3.3 reports the actual and modeled habitat capacity for fishes and trees of the MMRS. H_{data} is available only for trees since the abundance information is not available for fishes. H_{hete} is the average of the local community habitat capacities found by using

equation 3.6 and 3.7, then N_{hete} is the number of the number of individuals given by the sum of all the modeled habitat capacities. It is interesting to notice that N_{data} that is the real number of individuals is much lower for small trees than big trees so many trees in the MMRS are with high value of dbh (diameter at breast height). It is important to note also that N_{data} is different than N_{hete} because the model is not able to capture the real number of individuals or better the abundance pattern as reality is. The biodiversity patterns are captured reproducing well the occupancy pattern.

Table 3.3: Local-community mean habitat capacity and number of individuals in the MMRS, data and heterogenous model.

Case	H_{data}	H_{hete}	N_{data}	N_{hete}
$Fishes$	–	530	–	436,730
$Big\ Trees$	823	400	678,082	330,415
$Small\ Trees$	205	400	168,755	330,054

Table 3.4 reports the dispersal parameters needed for equation 3.2. Values of distances are in km. N_{homo} refers to the homogeneous case without considering the environmental variability and it is the result of the product between the number of links in the MMRS, $n = 824$, and the mean habitat capacity H.

The choice of adopting different values of dispersal parameter for big and small trees in a biological and ecological point of view can be arguable. Indeed the seed dispersal mechanisms may be assumed to be the same (*Seidler and Plotkin*, 2006). We have considered the same species (except five more species for big trees) so the difference in the spatial organization can only be attributed to external factors, like e.g. canopy gaps (*Duncan*, 1993) or life-stage differences (*Li et al.*, 2008, 2009), captured in spatial patterns by using a different dispersal parameter. For a review of seed dispersal near and far across temperate and tropical forest *Clark et al.* (1999) propose a kernel that is able to reproduce an exponential and also fat-tail kernel. Here for trees it has been used only a back-to-back exponential kernel. Moreover the spatial organization of small and big trees is clearly different, as suggested e.g. by (*Condit et al.*, 1996) and (*Plotkin et al.*, 2000, 2002). The difference in the spatial organization lead to a species-area relationship and to a different probability of the cluster-sizes. This difference can emerge from the different dynamics, e.g. different probability of death (higher for small trees).

Species-level variation in regeneration speed is reflected by variation in the range of tree diameters within the clusters (*Plotkin et al.*, 2002). *Plotkin et al.* (2002) found that e.g. is very rare to find large trees in the center of clusters, then spatial biodiversity patterns clearly reflects different mechanisms beyond the dispersal. The diversity of tree life-history traits and the diversity of responses to abiotic influences are reflected by the diversity of their spatial arrangements (*Plotkin et al.*, 2002; *Seidler and Plotkin*, 2006). Therefore here the model for big and small trees differs only for the dispersal parameter and this results to be sufficient to describe the patterns of trees at the two different diameter classes.

In Table 3.5 the parameters of the model have been reported for topological distances,

Table 3.4: Parameters of the NMM for fishes and trees of the MMRS. The mean link length is assumed to be $\sim$ 110 km for the river network and $\sim$ 52 km for the 2-D landscape.

Case	d_l (km)	b (km)	θ	H	N_{homo}	N_{hete}	S_T
$Fishes$	81.67	4.50	66	530	436,720	436,730	433
$Big\ Trees$	25	0	30	400	329,600	330,415	233
$Small\ Trees$	20	0	30	400	329,600	330,054	228

with the purpose to have a direct comparison with the paper *Muneepeerakul et al.* (2008). The mean dispersal length in topological distances is given by $d_l = -1/ln(a)$, that multiplied by the mean link length gives the d_l in actual distance (km). The parameter b of the fat-tail kernel in real distances is found multiplying it by the mean link-length in km. Figure 3.2 shows the trend of S_T of the simulations for the Tanaro in the RN and 2-D case. The magnitude of the ecological drift after reaching the stationary state is the same in both cases. Therefore the 2-D configuration allows a faster reaching of the stationary state, because the multiple direction allowed to the dispersal.

3.2.2 Global and Nearest-Neighbor Dispersal

Here we report the kernel structure for the exact global and nearest neighbor dispersal implemented in this study. In chapter 5 the results about simulation for the Tanaro RN and 2-D cases with these two kernels have been reported. The global and nearest-neighbor kernel dispersals allow to study the extreme cases of habitat-capacity limitation and dispersal limitation respectively. The former allows the dispersal in the whole ecosystem instead the latter allows the dispersal only between the first neighbors. In chapter 4 some results for dispersal parameter lower than the grain-size (that is the nearest neighbor dispersal distance) have been investigated. All the other cases for mean dispersal length $d_l > \lambda\delta$, where $\lambda\delta$ is the length of the pixel local-community at the chosen grain-level, and smaller then the size of the ecosystem (e.g. the mean diameter $L_{\|}$), presents biodiversity patterns in between the global and nearest-neighbor ones.

The global kernel dispersal is given by:

$$K_{ij} = 1/n \ , \tag{3.8}$$

for every pair of links ij where n is the total number of links.

Instead the nearest neighbor kernel dispersal is given by:

$$K_{ij} = \frac{1}{n_n + 1} \ , \tag{3.9}$$

where n_n is the number of neighbors around the link i, and the j-s are the neighbors along the network or in a 2-D landscape.

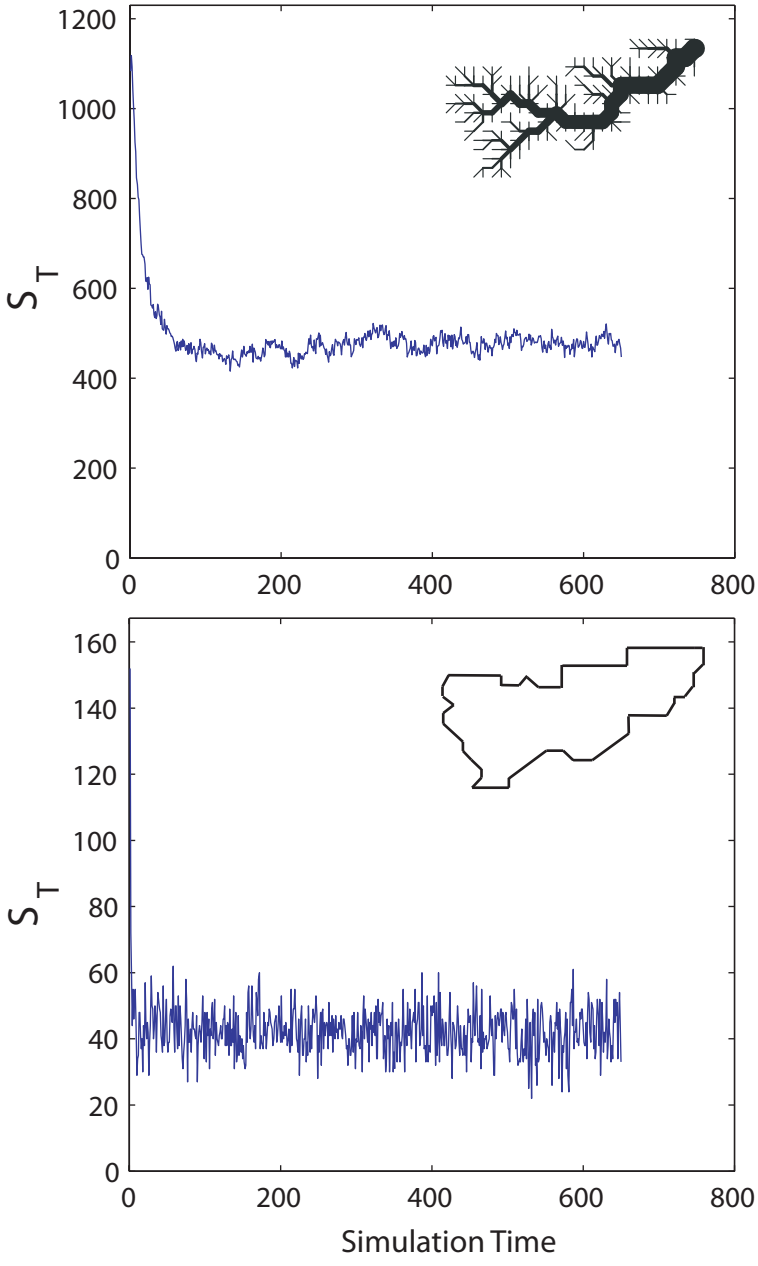

Figure 3.2: Behavior of S_T vs Simulation Time for the Tanaro river network (top plot) and the 2-D landscape (bottom plot) with the same boundaries for the networked basin.

Table 3.5: Parameters of the NMM for fishes and trees of the MMRS using topological distances.

Case	Fishes	Big Trees	Small Trees
$d_l(topo)$	0.74	0.23	0.18
a	0.26	0.0123	0.0041
b	0.03	0	0

Chapter 4

On Neutral Metacommunity Patterns of River Basins at Different Scales of Aggregation

"When I hear of the destruction of a species, I feel just as if all the works of some great writer have perished"
Theodore Roosevelt

Neutral metacommunity models for spatial biodiversity patterns are implemented on river networks acting as ecological corridors at different resolution. Coarse-graining elevation fields (under the constraint of preserving the basin mean elevation) produces a set of reconfigured drainage networks. The hydrologic assumption made here implies uniform runoff production such that each link has the same habitat capacity. Despite the universal scaling properties shown by river basins regardless of size, climate, vegetation or exposed lithology, we find that species richness at local and regional scales exhibits resolution-dependent behavior. In addition, we investigate species-area relationships and rank-abundance patterns. The slopes of the species-area relationships, which are consistent over coarse-graining resolutions, match those found in real landscapes in the case of long-distance dispersal. The rank-abundance patterns are independent on the resolution over a broad range of dispersal length. Our results confirm that strong interactions occur between network structure and the dispersal of species, and that, under the assumption of neutral dynamics, these interactions produce resolution-dependent biodiversity patterns that diverge from expectations following from universal geomorphic scaling laws. Both in theoretical and applied ecology studying how patterns change in resolution is relevant for understanding how ecological dynamics work in fragmented landscape, and for sampling and biodiversity management campaigns especially in consideration of climate change.

4.1 Introduction

Understanding local and regional patterns of species distributions is a major goal of ecological and biological research (*Levin*, 1992; *Nilsson et al.*, 1994; *Matthews*, 1998;

Levine, 2000a,b, 2003; *Rodriguez-Iturbe et al.*, 2009). Ecological systems are complex and seemingly unpredictable due to the large number of interacting individuals of different species involved, but patterns in the spatial distribution of organisms provide important clues about the underlying mechanisms that structure ecological communities. Although many mechanisms governing the biodiversity of species have been described, we are far from a complete understanding of the factors controlling the probabilistic properties of ecological patterns. Also it is not clear either what ecological descriptors can be assumed as invariant across different scales of observation and resolutions, or the interactions between geomorphological features and ecological dynamics, in this context hinged to river network structure. Similarity across spatial scales remains an open issue in ecological research where most studies focus on variables across a specific level of observation. Frequently, predictions from numerical models are available at resolutions coarser than those typical of ecological investigations (*Levin*, 1992). Many of these studies refer to one particular species (*Green et al.*, 2004; *Green and Bohannan*, 2006) with the extension to multi-species ecosystems being far from trivial. In this chapter, we address these issues for ecosystems in fluvial dendritic networks that obey the neutral theory (*Hubbell*, 2001). Biodiversity patterns, especially those of vegetation, may have effects of such ecosystem functions as productivity, nutrient cycling, and sediment transport; these functions in turn affect hydrological processes like evapotranspiration and runoff production. Furthermore, understanding the linkage between geomorphology and ecology in fragmented fluvial ecosystems is an relevant topic (*Benda et al.*, 2004), especially towards biodiversity preservation in response to climate change or anthropic interventions. Indeed an important problem in theoretical and applied landscape ecology is understanding whether and how biodiversity patterns change as a function of what is actually measured. Here the null-hypothesis is the resolution-invariance of biodiversity patterns facing the established such features of fluvial landscapes (*Rodriguez-Iturbe and Rinaldo*, 1997).

Space fragmentation plays a larger role in river basins than in savannas ecosystems since in the first case the drainage network acts as a stronger controlling factor for the spreading of organisms (e.g., *Nilsson et al.*, 1989; *Gascon et al.*, 2000; *Honnay et al.*, 2001; *Levine*, 2003; *Rodriguez-Iturbe et al.*, 2009). Interactions have also been documented among hydro-geomorphological features of non-dendritic landscapes and ecological patterns. For examples in savannas where a preferential connectivity is not exerted by a network, *Caylor and Shugart* (2006); *Scanlon et al.* (2007a); *Borgono et al.* (2009) documented the existence of power-laws on the size of the vegetation clusters, an important feature explained through local interaction mechanisms acting under a global hydrologic constrain. Power-law clusters of halophytic vegetation have also been documented by *Marani et al.* (2004) for tidal environments where hydrologic controls are provided by tidal networks whose geomorphological features (*Marani et al.*, 2002, 2003) differ from fluvial ones. As for river basins in water-controlled ecosystems it appears there may exist mechanisms leading to the optimal organization of vegetation patterns around the channel network structure under the controlling influence of water stress (*Caylor et al.*, 2004, 2005). Other empirical studies have suggested the importance of the river network on the biodiversity of freshwater fish, and in general of riverine habitats (*Poole*, 2002; *Ward et al.*, 2002; *Wiens*, 2002; *Fernandes et al.*, 2004; *Benda et al.*, 2004;

Fisher et al., 2007; *Muneepeerakul et al.*, 2008). It has been found, for instance, that fragmentation within river networks may produce long-term genetic variations among species and communities (*Colwell*, 2000; *Gascon et al.*, 2000). As ecological corridors, river networks also affect the spreading of human population and of water-borne disease (*Rodriguez-Iturbe et al.*, 2009). Also in the case of river basins, *Muneepeerakul et al.* (2008b) made clear the crucial effect that directionality, a key feature for certain species living in the drainage network ecosystem, has on biodiversity patterns.

With the above premises, the present chapter specifically aims at investigating, under general coarse-graining of the networked landscape: the resolution-invariance of species-area relationships; the interplay between the network shape and the dispersal of species; and the behavior of relevant macroecological descriptors, like species richness at local and regional scale and pairwise species similarity, through different scales of spatial aggregation. The embedded constraint in river basins is that the dispersal of species may happen only along the directed spanning structure of the drainage network that exhibits clear scale and resolution invariant properties over several orders of magnitude (*Rodriguez-Iturbe and Rinaldo*, 1997). In general patterns of species richness in real landscapes may differ from the neutral ones addressed in this study (*Chave et al.*, 2002; *Purves and Pacala*, 2005), therefore it is accepted that even in the presence of strong species-specific interactions neutral models may capture broad features of biodiversity patterns.

The metacommunity model is built here around four main assumptions:

1. *Space-filling network.* The simulations are performed on space-filling networks (i.e. all the pixel are channelized), implying there is no differentiation between hillslope and channel sites. The only geomorphological constraints are thus provided by drainage directions;

2. *The definition of streamlink.* The link connecting a pixel i to another pixel j where no tributaries occur between them is defined as streamlink. If a confluence is encountered a new streamlink originates. We assume that each streamlink (whose length in reality is in turn dependent on a threshold for the size of tributary), constitutes one local community (LC);

3. *Uniform habitat capacity.* All the local communities have the same habitat capacity, the number of individuals in each local community. The implicit assumption here is that the climatic variation over the studied region is negligible. Then the hydrologic assumption is that the runoff production is the same for all the links, which implies the habitat capacity is uniform in space;

4. *The neutral assumption.* Neutrality implies that each individual is competitively equivalent at per capita level regardless of species (*Hubbell*, 2001). The neutral theory is a valid approximation considering species at the same tropic level that equally interact like particles. Neutral theory is capable of generating patterns of abundance, diversity, species turnover, and geographic range similar to those observed in natural ecosystems that in fact do not feature neutral relationships among species (see e.g., *Bell*, 2001; *Purves and Pacala*, 2005; *Mouillot and Gas-*

ton, 2007; *Muneepeerakul et al.*, 2008).

4.2 On Coarse-Graining River Basins

Both real river basins as well as theoretical constructs are studied in this chapter, namely, the Cordevole and Tanaro basins in Northern Italy, and an optimal channel network (OCN) in a square-lattice domain (*Rodríguez-Iturbe et al.*, 1992). The latter is an artificial network obtained through a selection process minimizing total energy dissipation (a property of steady-state landforms) from which one obtains scaling forms that are in excellent agreement with real networks extracted from DEMs (e.g., *Rodriguez-Iturbe and Rinaldo*, 1997). The river networks extracted from the real basins and the OCN are plotted in Figure 4.1.

Details on the Cordevole and Tanaro basins are given in *Convertino et al.* (2007). Digital Elevation Models (DEMs) are employed in order to derive elevation maps at different scales of aggregation, or resolutions, from which the space-filling networks are extracted (see Figure 4.1). Note that a space-filling network is obtained when all pixels in the landscape domain are assumed to be channelized. In the case at hand, the DEM of Cordevole river basin has a drainage area of 691 km^2 (255 $\times$ 271 pixels at a resolution of 100 m), and the Tanaro river basin has a drainage area of 530 km^2 (178 $\times$ 296 pixels at a resolution of 100 m). The planar optimal channel network (OCN) is generated via a simulated annealing process which minimizes the total dissipation of energy expenditure in the transport of water. The chosen dimensions are 66 $\times$ 66 pixels, corresponding to a drainage area of approximatively 44 km^2 where an arbitrary resolution of 100 m is assumed for comparative purposes with the real river basins. The associated 3D landscape of the OCN is derived using the slope-area relationship $|\overrightarrow{\nabla z_i}| \propto A_i^{\gamma-1}$ with $\gamma = 0.5$, where $|\overrightarrow{\nabla z_i}|$ is the elevation gradient and z_i the elevation of each pixel with drainage area A_i (e.g., *Rodriguez-Iturbe and Rinaldo*, 1997).

The dependence on the resolution of observation of the biodiversity patterns is studied by observing how they vary upon coarse-graining in the description of the river basin. Figure 4.2 illustrates the coarse-graining procedure applied to the Tanaro river basin.

An example of invariance of geomorphological properties across different resolutions is shown in Figure 4.3 for the three basins, where the probability of exceedence of the drainage area $P(A \geq a)$ at a randomly chosen point is plotted at different scales of aggregation. In Figure 4.3 the scaling exponents are very close to the typical value for river basins $1 - \tau = 0.43$: it is the reported the average over resolution and the standard errors (Table 4.1) are found by reduced major axis regression. The value a_{max}, i.e. a at the outlet, decreases because a_i is characterized here by the number of links upstream of site i. The resolution-parameter λ is given by the coarse-graining resolution divided by the finest resolution in meters, i.e. $CG/100$.

The distribution of A is maintained across different resolutions, although the accuracy of the estimation of the slope in the scaling region decreases under coarse-graining. The elevation of an arbitrary pixel i at a certain resolution λ is given by the mean value of the elevations z_i of pixels at the finest resolution encompassed in the area $\lambda^2\delta^2$ of the new

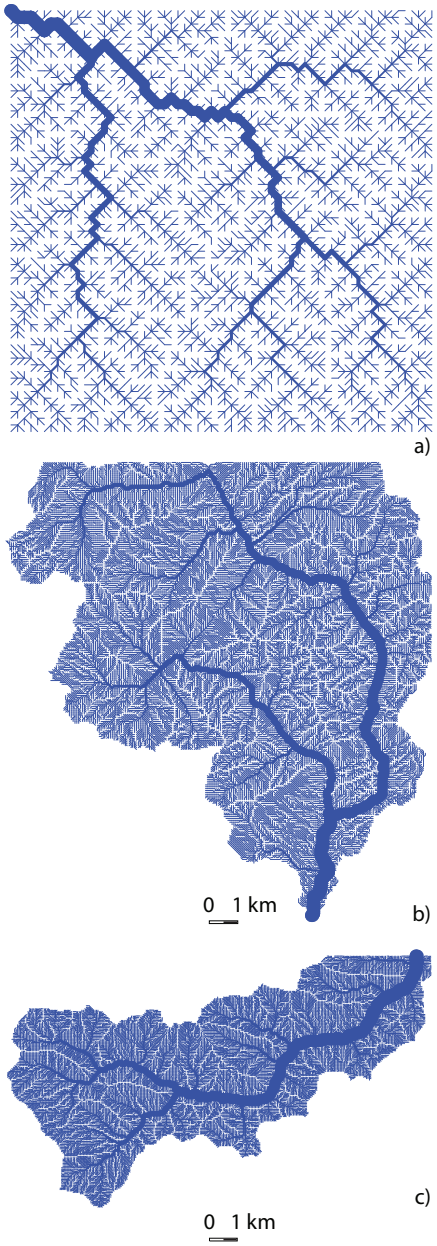

Figure 4.1: Space-filling networks derived from DEMs at their original finest resolution: a) OCN with single outlet and no-flux on sides, whose overall dimensions are 66 × 66 pixels (each of which is arbitrarily assumed to correspond to an area of 100 × 100 m^2; b) Cordevole, main tributary of Piave basin, Northern Italy, suitably extracted from the DEM, 255 × 271 pixels, at 100 m resolution; c) Tanaro, Po basin, Northern Italy, extracted from the DEM, 178 × 296, pixels at 100 m resolution. The space-filling character is enforced by a threshold area for channelization $A \geq A_{th} = 1$ in pixel units.

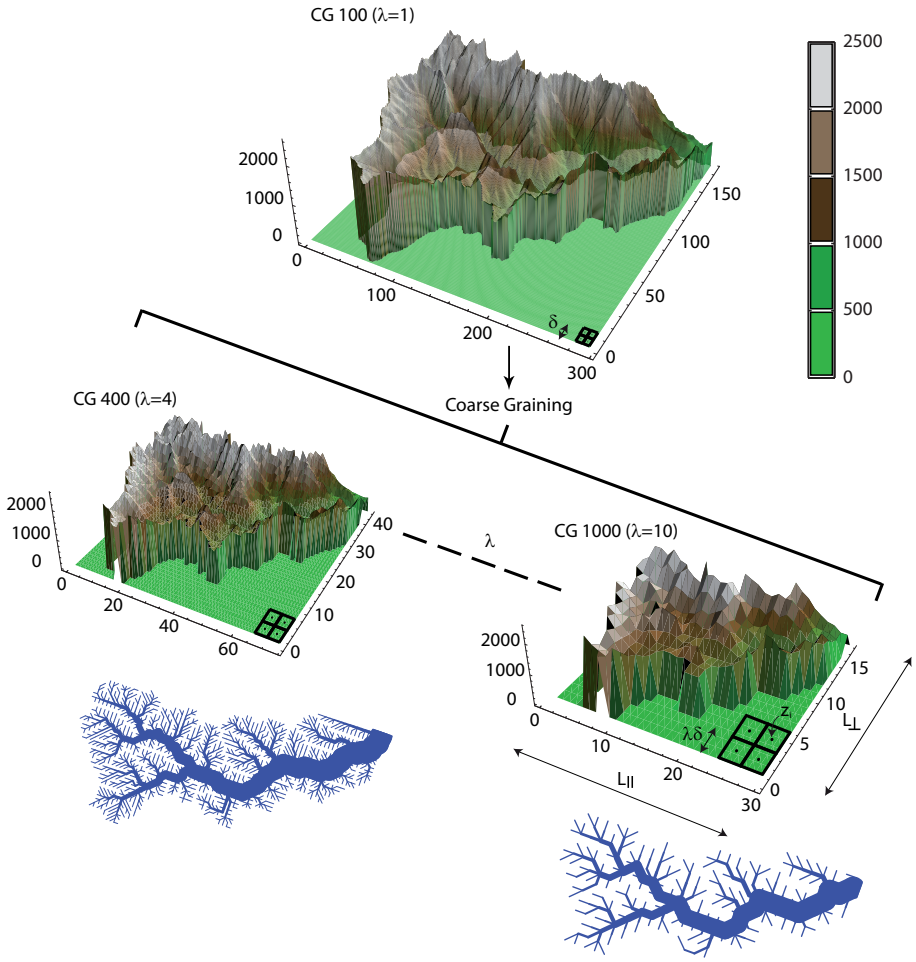

Figure 4.2: Illustration of coarse-graining (CG) technique applied to derive different elevation fields of the Tanaro basin starting from the most detailed DEM (top plot for $\lambda = 1$), whose extracted network is shown in Figure 4.1. Although coarse-graining brings changes into how streamlinks are connected, which in turn modify the slope-area relationship of the original landscape, the obtained CG-DEMs still statistically preserve the slope-area scaling relationship. $L_\perp$ and $L\|$ are the transversal and longitudinal diameters of the basin. All CG-landscapes are derived from the original topography with $\lambda = 1$ by averaging the elevations of λ^2 neighboring pixels. $\delta = 100$ m is the pixel size at the original resolution. Extracted networks and CG-DEMs are shown for $CG = 400$ ($\lambda = 4$) and $CG = 1000$ ($\lambda = 10$) (bottom part of the Figure). The mean elevation is preserved across aggregation scales.

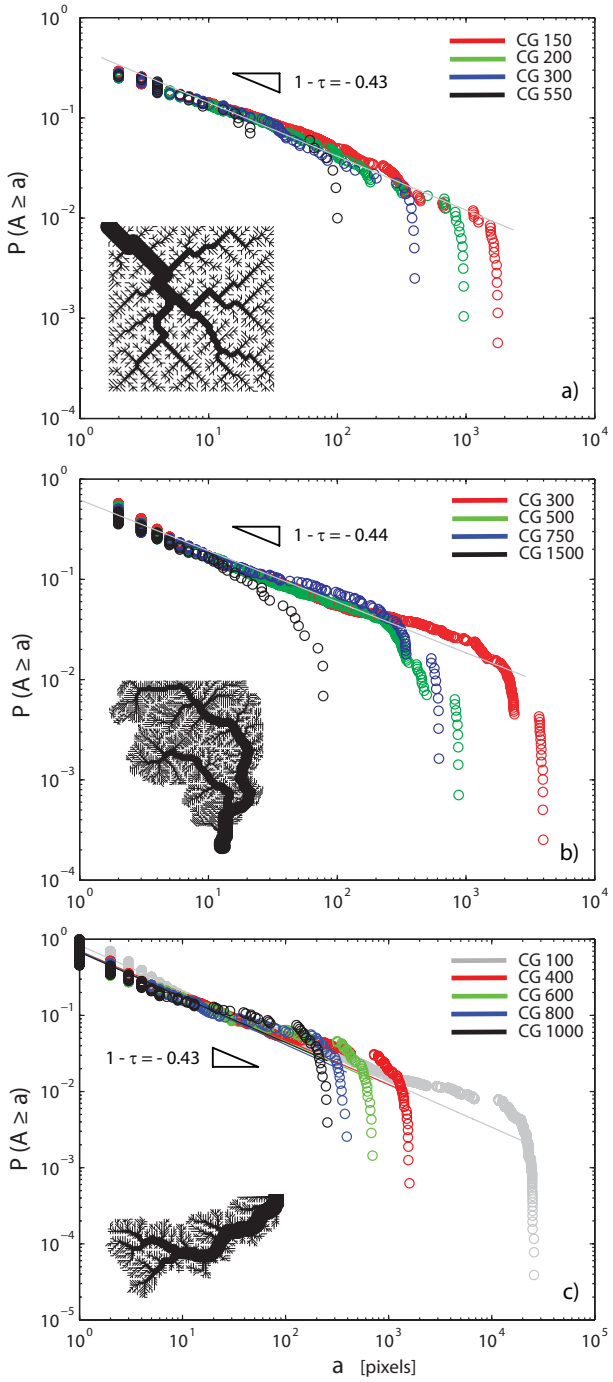

Figure 4.3: 1) Probability of exceedence $P(A \geqslant a)$ of the drainage area in pixel units at any randomly chosen point, A, for the OCN (a), Cordevole (b), and Tanaro river basin (c), using different levels of coarse-graining CG.

pixel i, where δ is the side-length at the finest resolution and λ the resolution parameter (see Figure 4.2). The invariance law of the cumulative drainage areas throughout the coarse-graining operation (Figure 4.3) is a direct consequence of the fractal structure of the river basin (*Rodriguez-Iturbe and Rinaldo*, 1997). It is important to note that across CG-resolutions the potential energy of the landscape is kept constant, and although the general directions of the fluxes are similar across CG-resolutions, the connectivity of the network experiences changes.

4.3 Biodiversity Patterns of River Basins at Different Scales of Resolution

This section focuses on the dependence of different biodiversity measures on the spatial resolution for the networks used in the analysis (OCN, Cordevole, Tanaro). The value of the habitat capacity and the number of local communities are given in Table 3.1 in Chapter 3. The two different ecological scenarios with parameters given in Table 3.2 (Chapter 3) are implemented for all river networks at different coarse graining levels. All results presented here are the ensemble average over ten snapshots after the system has reached a statistically stationary state. For the biodiversity patterns analyzed (Figure 4.4, 4.5, 4.6, 4.7, 4.8, and 4.9), the lowest coefficient of determination is 0.78, computed by using the standard deviation averaged over ten realizations at stationary state. The deviation of each single-realization curve from the average over the ten snapshots is small and for this reason we have reported only the mean curve for the patterns analyzed. The coefficient of determination, R^2, has been reported for the SAR because it is very consistent upon coarse-graining.

First, we have analyzed the rank-abundance (RA) curves resulting from the two ecological scenarios, namely the exponential and exponential-Cauchy dispersal kernel case for the three basins (Figures 4.4 a), c), e) and b), d), f) respectively). Here, the abundance of a species is simply the number of individuals belonging to that species. In both cases, the plots show overlapping curves from different coarse-grain levels. In addition, we investigated the RA curves as a function of the dispersal parameters d_l and the diversification rate θ. The resolution-invariance of the RA curves is not observed only for very local dispersal. Under such circumstances, in fact, the influence of the dispersal, and therefore that of coarse graining, is stronger on the total number of species. It is interesting to note that the number of species at the steady state, S_T, significantly depends on the dispersal only when d_l values are small independently of the diversification rate; this is in agreement with *Chave et al.* (2002) and with the results shown in Figure 4.5 a) and b). The number of species at stationary state S_T is quite insensitive to river basin, resolution or ecological scenario. This is suggested to stem from the fact that the total species richness depends more critically on N and θ as long as the dispersal is sufficiently non-local. If the dispersal is extremely local, S_T tends to increase (Figure 4.5 a)).

Figure 4.5 a) and b) show the dependence of S_T on d_l and θ for the exponential and the exponential-Cauchy kernel cases, respectively, for the Tanaro basin. Two values of θ have been considered: $\theta = 4$ (red lines), and $\theta = 40$ (blue lines) for $CG = 400$ (dashed

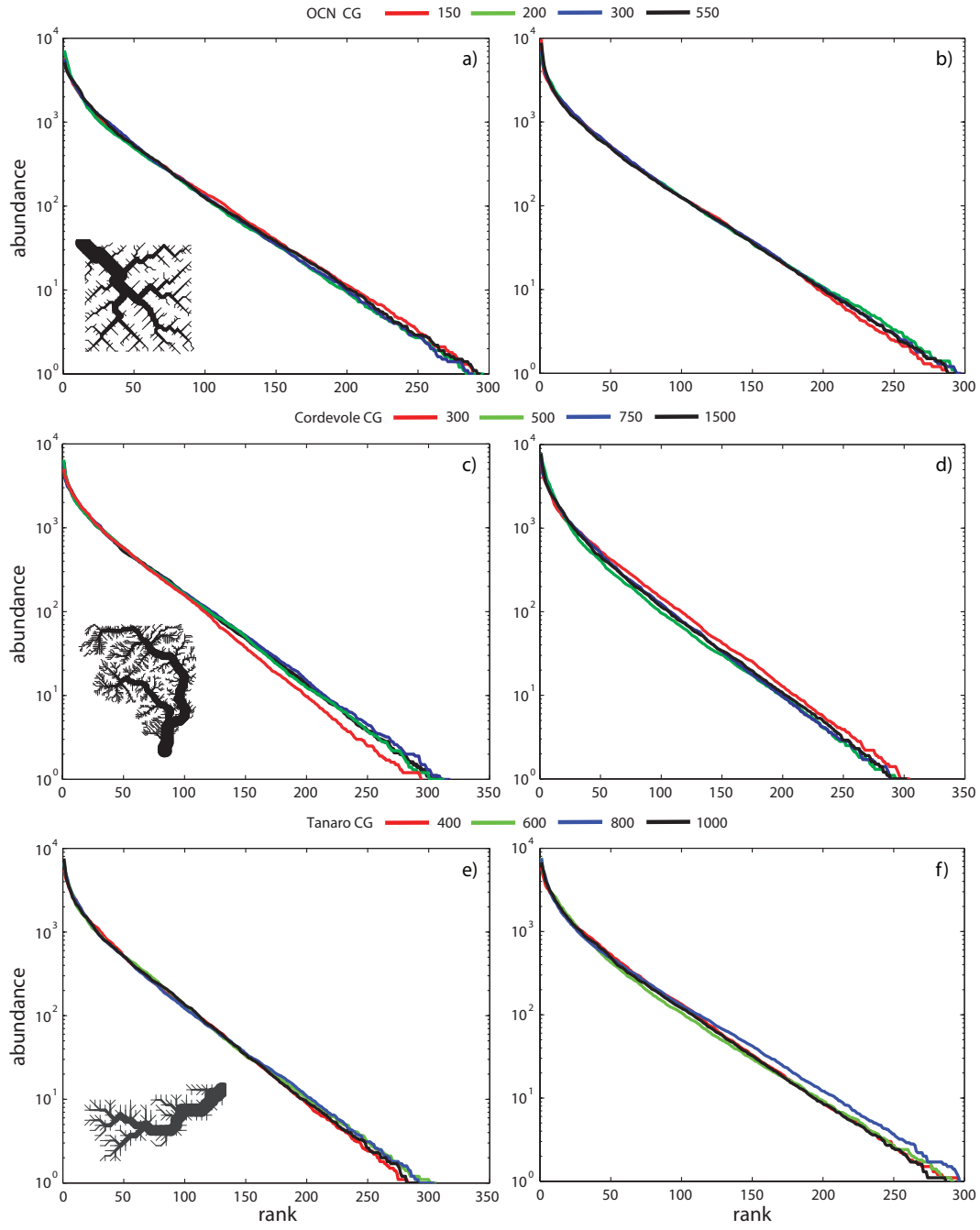

Figure 4.4: Rank-Abundance curves associated with the exponential (left plots) and the exponential-Cauchy kernel dispersal (right plots) for the OCN (a) and b)), the Cordevole basin (c) and d)), and the Tanaro basin (e) and f)). The resolution-invariance of the *RA*-curves is observed for both the sets of parameters used in the simulations (Table 3.2 in Chapter 3). S_T that is the maximum rank is very invariant (~ 300) regardless the basin, the resolution, and the dispersal mechanism because the latter is sufficiently non-local. This is because the total species richness depends critically on N and θ.

lines) and $CG = 1000$ (solid line). The resolution dependency of RA is marked by the gap between the $CG = 400$ and $CG = 1000$ resolution curves that shows a meaningful variation of S_T only for very small values of d_l.

In the case of the exponential kernel (Figure 4.5 a)) there is a clear threshold value of the dispersal length, i.e. about $d_l = 250$, that determines the transition for the rank-abundance to hold independently of the resolution and the diversification rate θ ($d_l > 250$). It is in this sense we refer to the resolution-independency of the RA pattern. Below $d_l = 250$, the total number of species at stationarity state, S_T, results sensitive to the coarse-graining resolution; the RA pattern for $d_l < 250$ exhibit a plateau that means the presence of many common species with the same abundance. This happens because the dispersal length is so small that the movement of species is very limited: for $d_l = 1$, S_T tends to coincides with the number of local communities, n, in which on average there is only one species. In the case of the exponential-Cauchy kernel (Figure 4.5 b)) the resolution-invariance is consistent for all values of the dispersal parameter and the resolution. The difference of S_T for $\theta = 4$ and 40 is similar to the exponential kernel dispersal case, but for very small d_l (owing to the fat-tail character of the dispersal) the species are able to move along the whole network thus maintaining the resolution-invariance of the RA pattern. Where in fact S_T tends to be equal to the number of local communities, a drop is produced in the rank-abundance curves (see Figure 4.5 c)) that results in changes in the maximum value of the rank between coarse-grained resolutions (the number of local communities n decreases as the CG-resolution increases). In the case of global dispersal (i.e. a dispersal kernel $K_{ij} = 1/n$, see Chapter 3) the invariance of the rank-abundance plot is indeed observed.

Figure 4.5 c) reports the rank-abundance for non-trivial values of d_l and different θ at resolution $CG = 1000$ of the Tanaro basin, so $d_l = 1000$ in this case is equal to the the mean distance between two neighboring communities. Increasing the diversification rate, e.g. for $\theta = 40$ (solid and black-crosses grey lines, Figure 4.5 c)), the increase in the number of species is much stronger than when assuming lower values of the dispersal parameter (solid and dotted black lines) at the same CG resolution. Thus for higher values of θ the number of species at stationary state, S_T, is larger. For intermediate values of d_l and θ, the RA patterns are not affected by the resolution. We also note that the effect of coarse-graining (CG) is intrinsically present for the discretization of the dispersal kernel structure, K_{ij} vs L – this is particularly evident when d_l is small and the CG resolution is coarse. For the Cordevole and the OCN the results shown in are analogous to these shown for the Tanaro in Figure 4.5. In Figure 4.5 two values of θ have been considered as example: 4 (red lines), 40 (blue lines) for $CG = 400$ (dashed lines) and $CG = 1000$ (solid line). The resolution dependency of RA is marked by the gap between the $CG = 400$ and $CG = 1000$ resolution curves that shows a meaningful variation of S_T for very small values of d_l. Plot c) shows the rank-abundance curves for the Tanaro basin in the exponential dispersal for the resolution $CG = 1000$, $\theta = 4, 40$ and different values of the dispersal length $d_l = 150, 250, 500, 1000$ with different linestyles according to the legend. For the Cordevole and the OCN the results are analogous to these shown for the Tanaro. Table 3.2 (Chapter 3) reports the values of the parameters for the simulated ecological scenarios (a) and (b).

Figure 4.6 illustrates the dependence of the regional species richness RSR on area

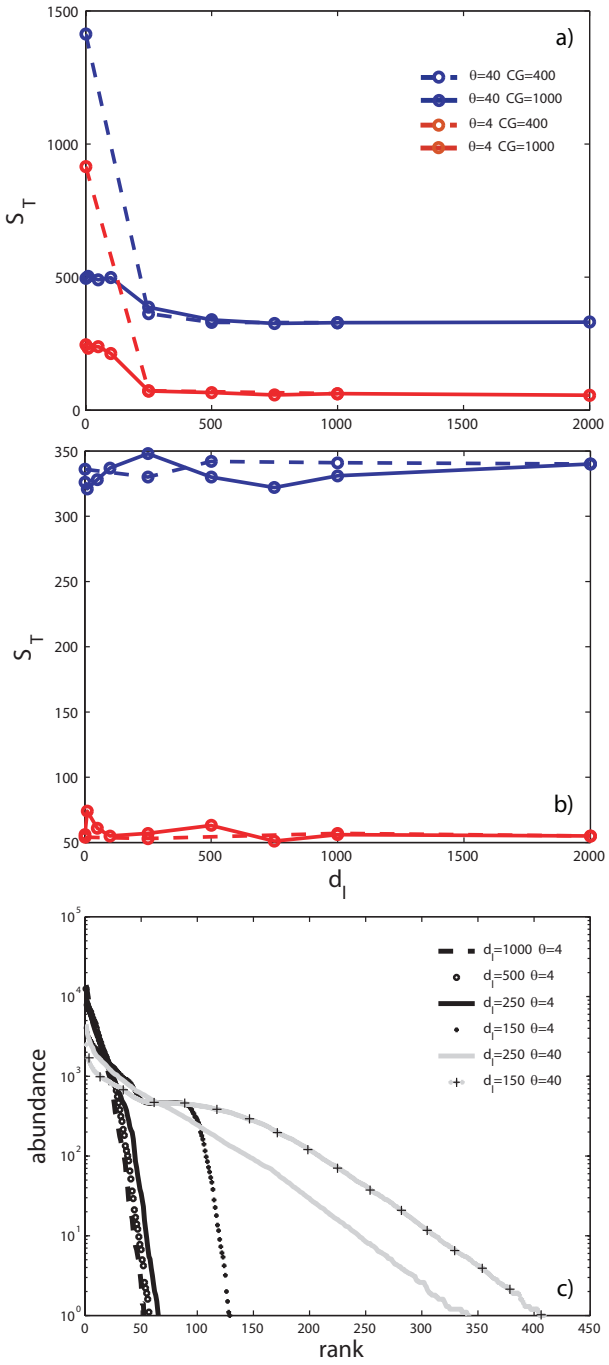

Figure 4.5: Plots a) and b) show the dependence of d_l and θ on the total number of species at stationary state, S_T, for the exponential and the exponential-Cauchy kernel case respectively for the Tanaro basin.

Table 4.1: Values of the exponent z from the species-area relationship $RSR \sim (H^*)^z$ computed by taking the average number of species in circular areas with different radii randomly placed all over the basin, in ecological scenarios (a) and (b) and for the three basins considered. The slope z is derived by the linear and Jackknife model (*Warton et al.*, 2006), and the coefficient of determination, R^2, is reported. τ is the scaling exponent of $P(A \geq a)$, where A is the drainage area, and its variability is the standard error found by reduce major axis regression bootstrapping over cases and resolutions (*Warton et al.*, 2006).

Exponents	OCN	$Cordevole$	$Tanaro$
τ	1.43 ± 0.01	1.44 ± 0.01	1.43 ± 0.02
Scenario a			
z	0.25	0.49	0.43
R^2	0.98	0.86	0.87
Scenario b			
z	0.22	0.28	0.28
R^2	0.99	0.87	0.98

at various CG scales of aggregation. RSR is computed as the number of species S found over circular areas with different radii placed randomly over the basin. The RSR values are then binned depending on the habitat capacity H^* (i.e. the number of sites) encompassed by the circular area and the average value is taken as representative of the ensemble. A species-area relationship (SAR) is expressed as $RSR \sim (H^*)^z$. Figure 4.6 (left plots, a), c), e)) and (right plots, b), d), f)) show the $RSR = \langle S \rangle$ as function of H^* for ecological scenarios (a) and (b) (Table 3.2 in Chapter 3) respectively. We note that for both types of kernels the different levels of coarse-graining exhibit a collapse of RSR vs H^*. The species-area relationship $RSR \sim (H^*)^z$ computed independently of the network structure, holds reasonably well for the case of dispersal with an exponential kernel and remarkably well when the dispersal includes a fat-tail component for all the three basins. The values of the exponent z are estimated for all cases and networks and are given in Table 4.1.

The slopes are computed using the linear and Jackknife regressions (*Warton et al.*, 2006) and the coefficient of determination is derived. It is interesting to notice that the value of z estimated in the case of the exponential kernel dispersal for the Tanaro and Cordevole river basins, is considerably higher than those usually found in biogeographical species-area relationships for ecosystems in typical 2D landscapes (*MacArthur and Wilson*, 2001). This is not the case when the dispersal includes a fat-tail component where the obtained values of z are in the range of those commonly observed. This suggests that species would adopt a fat-tail dispersal strategy should they maximize their colonization and survivability (*Davies et al.*, 2004; *Pearson and Dawson*, 2004; *Bohrer et al.*, 2005; *Soons and Ozinga*, 2005; *Nathan*, 2006). This is consistent with the invariance of S_T vs d_l in case b rather than in case a. This can indicate that the long-distance dispersal tends to diminish the impact of the network structure on the species-area relationship,

and it is evidenced also by the more straight-convex character of the SAR (left plots vs right plots in Figure 4.6). Figure 4.6 c) and e) show a double trend not present in Figure 4.6 d) and f). As suggested by *Hubbell* (2001), the SAR is constrained by the geometry of the biogeographic area in which the processes of origination, dispersal and extinction act. The upturn of the SAR occurs in correspondence to a correlation length, independent of θ, that represents the spatial scale of the geometric constraint imposed by the landscape (*Hubbell*, 2001). We do not observe a significant difference in the value of z in the two different dispersal cases for OCNs (Figure 4.6 a) and b)) perhaps due to the shape of the basin.

We further investigate the effects of coarse graining on the local species richness, LSR, defined as the number of species present in each local community at stationarity. LSR represents a measure of local biodiversity. Figure 4.7 shows the probability of exceedence of the normalized local species richness, LSR/H^z, for the OCN (Figure 4.7 a) and b)), Cordevole (Figure 4.7 c) and d)), and Tanaro river basin (Figure 4.7 e) and f)) where H is the habitat capacity of each local community and z is the exponent of the species-area relationship (see Figure 4.6). The probability $P(LSR/H^z \geq lsr/H^z)$, however, does not collapse onto a unique curve. In the exponential-Cauchy dispersal scenario (right plots) the overlapping of the curves is more evident than in the exponential dispersal scenario (left plots), as observed for the species-area relationship (Figure 4.6), due to the higher homogeneity of the LSR in the latter case. This is more evident for the two real river basins, Cordevole and Tanaro, particularly for the latter probably because of its more pronounced elongation. Higher values of LSR are observed in the exponential-Cauchy kernel dispersal than in the exponential kernel dispersal.

Figure 4.8 shows the behavior of the LSR as function of the diameter ϕ for the OCN (Figure 4.8 a) and b)), Cordevole (Figure 4.8 c) and d)), and Tanaro river basin (Figure 4.7 e) and f)), in the exponential dispersal case (left plot) and exponential-fat tail dispersal case (right plots). The LSR generally increases from the finest to the coarsest resolution (due to the decrease in the number of links n while the total number of individuals N remains constant) and ordinarily decreases when the diameter ϕ becomes larger. High values of the LSR are expected for streamlinks with small diameters because they are well connected to all the other links in the basin and thus, on the average, species can travel to links with small diameters more frequently than to their large-diameter counterparts. This is particularly true in the case of exponential kernels because, if the spreading of species shows a fat-tail component, they can move over long distances resulting in more regular patterns. This is suggested by Figure 4.8 where the values of the LSR are quite uniform in the three basins for the ecological scenario (b). Large LSR exists at the outlet, along the main streams of the network, and in sources-areas it is quite big. Note that high values of LSR across all coarse-graining resolutions exist in links where the diameter, ϕ, is small. From the LSR pattern the structure of the network emerges in the case of the exponential dispersal kernel – Figure 4.8 a), c), e) – and the gradient of species richness along the mainstream towards the source areas of the basin is relatively sharp. For the case of the exponential-Cauchy dispersal kernel – Figure 4.8 b), d), f) – the LSR pattern is relatively homogeneous and the alongstream gradient in species richness is not as steep as in case (a). The insets of the plots are the LSR patterns properly color-coded (top of Figure 4.8). We find

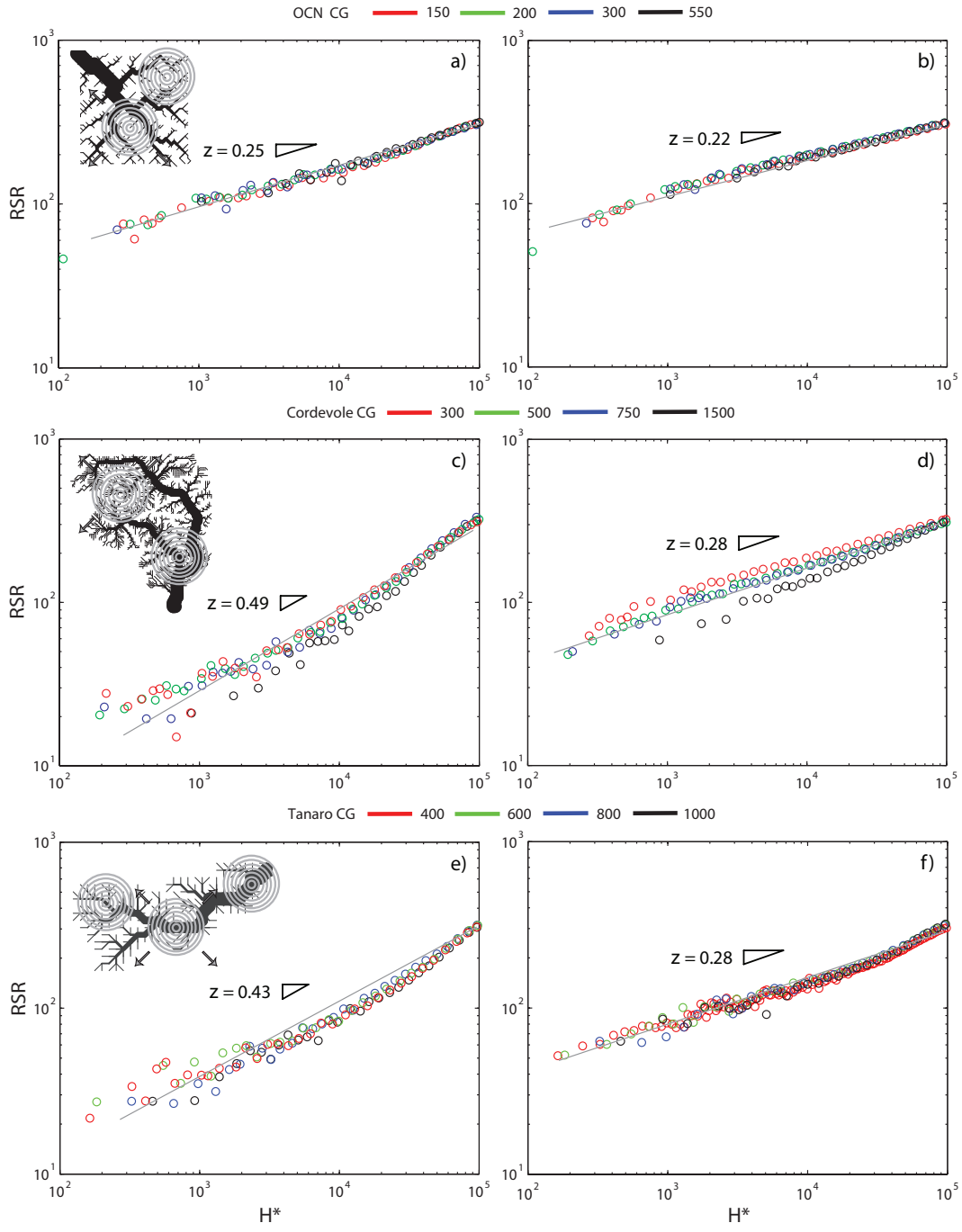

Figure 4.6: SAR derived between $RSR = \langle S \rangle$ vs H^* for the three basins (OCN (a) and b)), Cordevole (c) and d)), and Tanaro (e) and f))) at different resolutions for two types of dispersal kernel, namely, exponential (left plots) and exponential-Cauchy kernels (right plots); the circles represent the binning averages. In the insets it is illustrated the approach for computing the species richness at regional scale. The estimated exponents are derived by linear and Jackknife regressions (*Warton et al.*, 2006)) and the coefficient of determination, R^2, is calculated (Table 3.2 in Chapter 3).

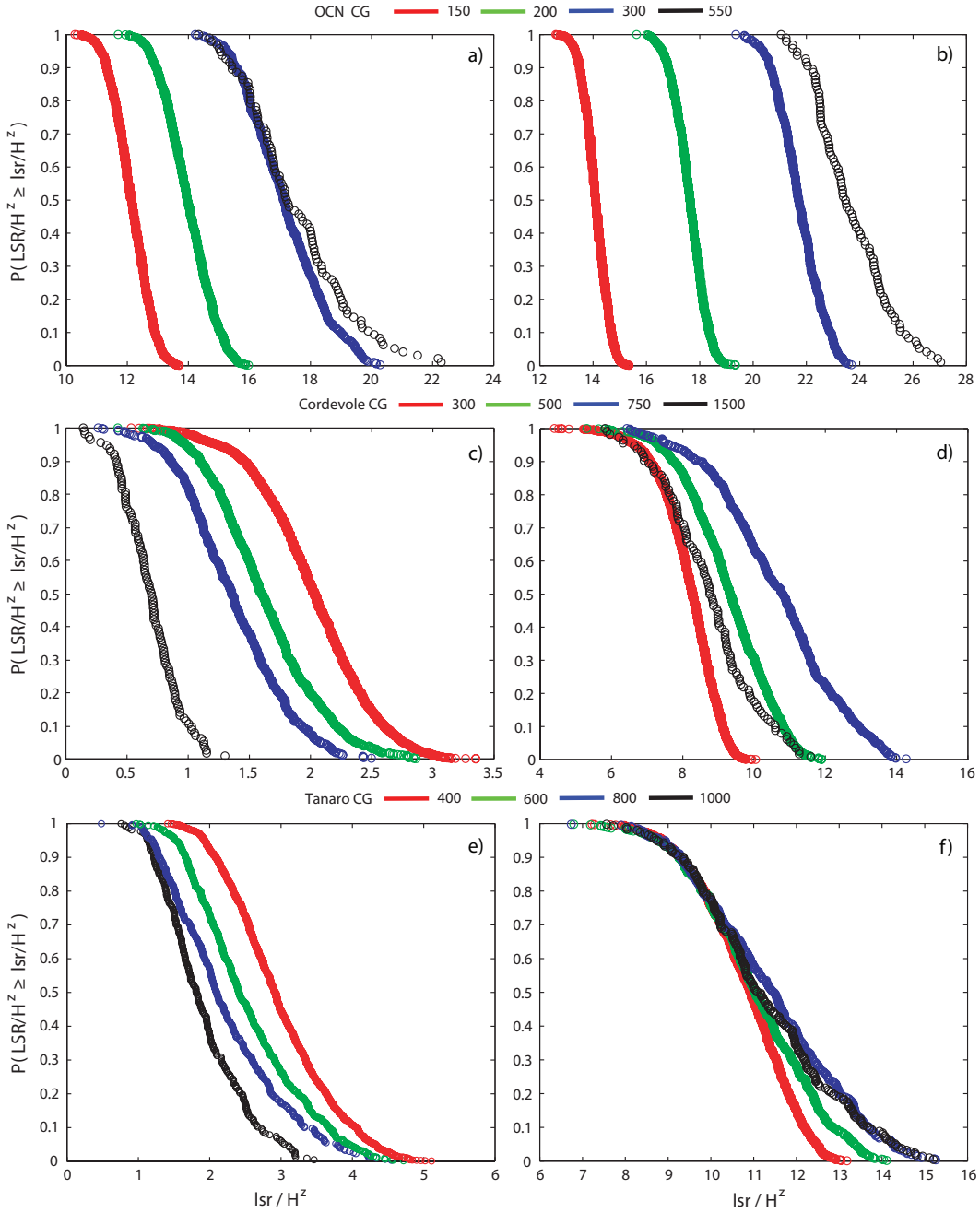

Figure 4.7: Rescaled probability of exceedence of Local Species Richness, $P(LSR/H^z \geq lsr/H^z)$, at different resolutions for the exponential (left plots) and exponential-Cauchy kernel (right plots) for the three river basins: OCN, a) and b), Cordevole, c) and d), and Tanaro, e) and f). The distributions do not collapse onto the same curve and do not exhibit a power-law behavior. The exponents z are those derived by fitting the species-area relationship in Figure 4.6.

that the average value of LSR in the whole basin decreases with the diameter for every resolution. The OCN shows the largest values of the maximum and mean LSR regardless of resolution; the mean and maximum values of LSR are larger for the Tanaro than for the Cordevole owing to the larger diameter. It is not observed a relevant difference in the LSR vs ϕ pattern between the two dispersal cases for the OCN (Figure 4.8 a) and b)).

As a descriptor of the spatial structure of biodiversity - β-diversity - we use the Jaccard Similarity Index, JSI, defined as:

$$JSI_{ij} = \frac{LSR_{ij}}{(LSR_i + LSR_j - LSR_{ij})} , \tag{4.1}$$

where LSR_{ij} is the number of species that are common to both links i and j, and LSR_i is the total number of species in link i. We examined the JSI index as a function of the alongstream distance, d, between links. It ranges from 0 if two local communities share no species, to 1, if their species composition are identical. Figure 4.9 reports the decay with distance of the JSI index for the OCN (Figure 4.9 a) and b)), Cordevole (Figure 4.9 c) and d)), and Tanaro river basin (Figure 4.9 e) and f)), in the exponential dispersal case (left plots) and exponential-fat tail dispersal case (right plots). As expected, the pairwise similarity of species among links at a given distance d is greater for the ecological scenario (b) where the Cauchy component of the dispersal kernel plays an important role. In Figure 4.9 c) and e) for Cordevole and Tanaro respectively, JSI values in case (a) for all resolutions basically coincides. The decay of JSI is faster in the case of exponential kernel (a) implying that local communities at large distances tend to have very different species composition. In all cases the value of $JSI = 1$ for $d = 0$ is not included. For the OCN there is a relevant difference for the pairwise similarity pattern between resolutions even in the exponential dispersal case, but the decay of JSI vs d is faster than in case (b). This behavior could be related to the limited extension and the shape of the OCN employed. We finally note that simulations with different values of the total number of individuals, N, and speciation rates, θ, do not produce significant differences in the computed biodiversity patterns and the above remarks.

4.4 Conclusions

The following conclusions are worth mentioning:

- By employing a metacommunity model for species that act neutrally, the study of the macroecological descriptors (species richness at local and regional scale, and the pairwise species similarity) in river basins has shown the validity across resolutions of a species-area relationship dependent on the different network structures and the diverse ecological scenarios. In the case characterized by a kernel whose structure is the combination of an exponential at short distances, and a fat-tail that controls the dispersal at large distances, the exponent of the species-area (SAR) relationship assumes quite different values from the case with only the exponential but they match those found in real landscape. For both the exponential dispersal and the exponential-Cauchy dispersal kernels a unique regime appears for the SAR;

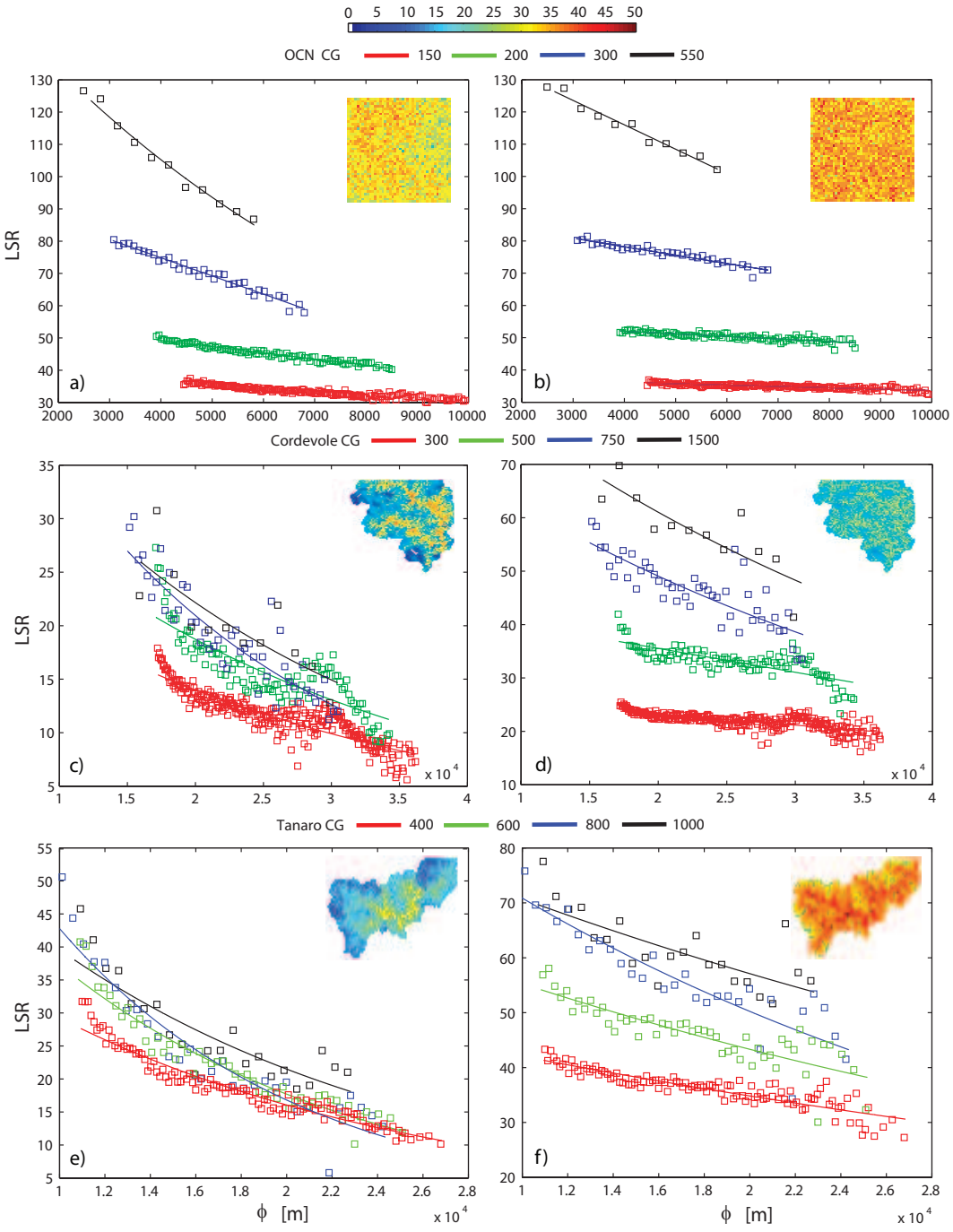

Figure 4.8: *LSR* vs the diameter ϕ at different resolutions for the exponential (left plots) and exponential-Cauchy kernel (right plots) for the three river basins: OCN, a) and b), Cordevole, c) and d), and Tanaro, e) and f). The upper legend specify the color for each resolution, and the color-bar refers to the insets in each plot showing the *LSR* pattern related to each ecological case.

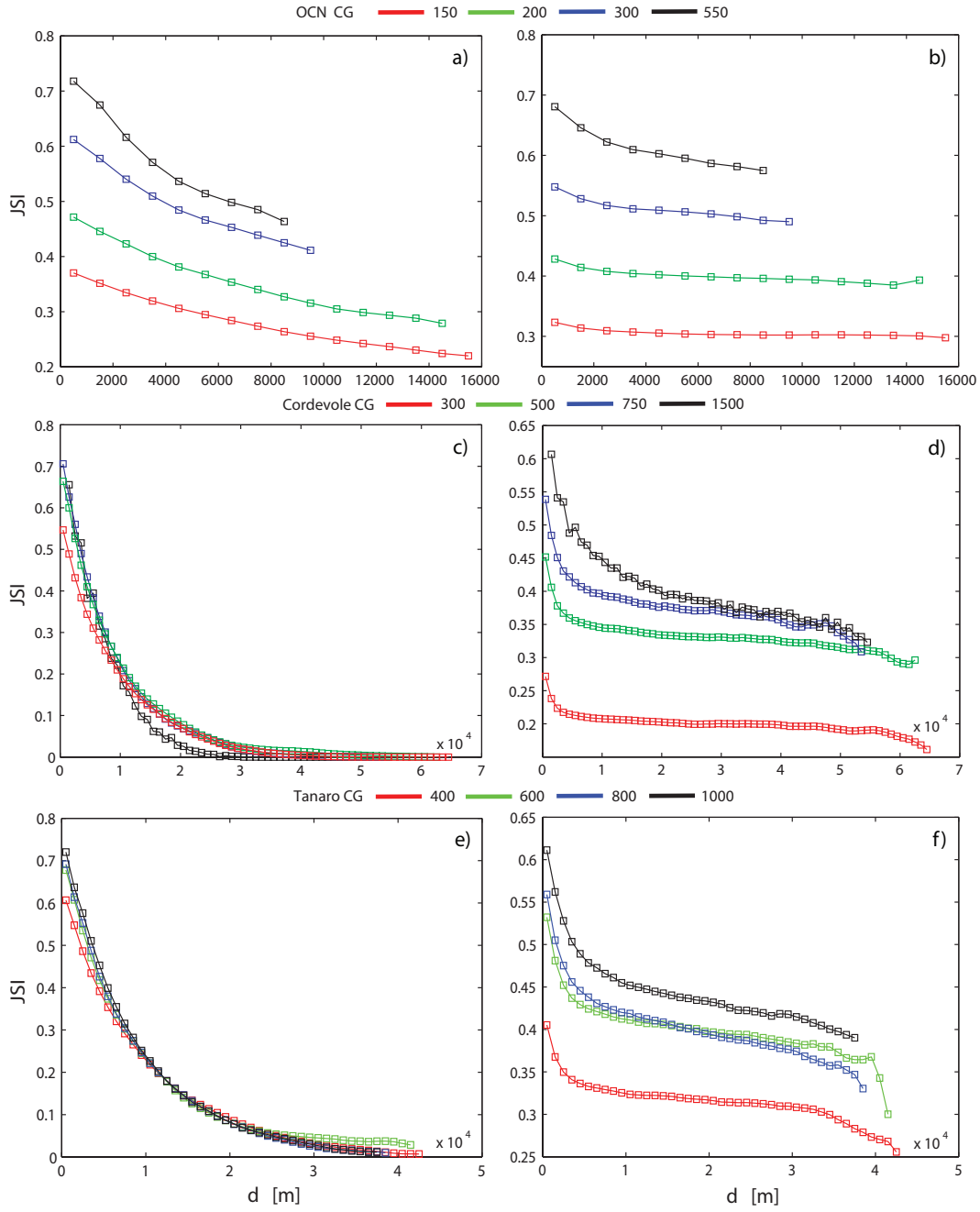

Figure 4.9: Plots of the average JSI vs d at different coarse-graining resolutions for the exponential (left plots) and exponential-Cauchy kernel (right plots) for the three river basins: OCN, a) and b), Cordevole, c) and d), and Tanaro, e) and f).

- The results suggest that the local species richness (LSR) and the species' similarity index (JSI) depend on the coarse-graining resolution used in the analysis. Thus, the robust invariance of the geomorphological characters of river basins does not imply universality in the resulting ecological patterns seen at different resolutions under the hypothesis of neutral dynamics. However, future work should be focused on possible scaling relations of these ecological indicators, not only in function of the resolution but also in function of the scale of analysis;

- The rank-abundance curves shows a resolution-invariant character regardless the specific value of the diversification rate and for a broad range of dispersal lengths. In the exponential dispersal case there exists a limited region for $d_l < 250$ in which the total number of species varies significantly with the resolution and the RA pattern have many common species with the same abundance. In the exponential-Cauchy dispersal case the invariance of RA patterns is granted for all values of the dispersal length. Moreover the number of species at stationary state S_T is quite insensitive to different river basins and ecological scenarios.

Thus we may reinforce earlier suggestions about the importance of the network structure on biodiversity patterns.

Chapter 5

Neutral Metacommunity Clustering: River Basin and Landscape Species Patterns

"It is not the strongest of the species that survives, nor the most intelligent that survives. It is the one that is the most adaptable to change"
Charles Darwin

Moving from the analysis of the spatial distribution of species of real data of fishes and vegetation in the Mississippi Missouri River System, this chapter studies the clustering of species that has been simulated on different real and artificial networks and on a 2-D non-fragmented landscape by a neutral metacommunity model. Power-law distribtions of the cluster-size have been proved for fishes, big trees and small trees, whose slope varies underlying different dynamics in the spatial distribution of species. The analysis on real data for the MMRS evidences that the probability of exceedence of the cluster size, $P(CS \geq c)$, is not invariant across different scales, resolutions and coarse-graining levels. A scale/grain analysis has been performed for the species-area and the species-individual relationship, the latter as a tool to detect the signature of environmental heterogeneities in biodiversity patterns. A link is suggeested to exist between the species-area relationship and the probability distribution of the cluster-size.

5.1 Introduction

Here it is proofed that the neutral metacommunity model adopted is capable to reproduce the cluster-size distribution for different taxa, fishes and trees of the MMRS. For the fishes the model developed in *Muneepeerakul et al.* (2008) has been used that reproduces simultaneously other four relevant biodiversity patterns, namely the pdf of the local species richness (LSR), the trend of LSR in function of the distance to the outlet, the rank-occupancy curve, and the Jaccard similarity index (JSI) that is similarity of LSR between local communities. As for the trees the fitting of these patterns is obtained in *Konar et al.* (2009) (whose model include permeable boundaries, ν in function of the

distance, and a double exponential kernel for the immigration and speciation), but as for the cluster-size pattern the simple model for trees described in chapter 3 is deemed sufficient. It is shown that the only relevant dependence on the cluster-size distribution is exerted by the dispersal parameter d_l. The aim is to understand wheter the clusterization process of the species depends mainly on the birth-death and dispersal stochasticity of the neutral model, the topology of the ecosystem, or on the environmental heterogeneities.

In general, aggregation patterns of vegetation benefit from a separate large-tree/small-tree treatment, that permits also to test the theories of *Janzen* (1970) and *Connell* (1971). Recent observations have shown that even at microscales, microbes tends to cluster-size according a neutral dynamic, see (e.g., *Houchmandzadeh*, 2002, 2008). Power-law distributions of the cluster-size of patches have been analysed diffusely in previous studies, both numerically and by using empirical data. *Scanlon et al.* (2007b) found power-law of the clsuter-size in savannah ecosysystems, *Kéfi et al.* (2007) found power-laws for the number of patches in Mediterranean arid ecosystems that is function of the grazing pressures and aridity, *Wu* (2004) commented the qualitative trend of the power-law of the cluster-size in function of the scale, and *He and Hubbell* (2003) characterized the distribution of species in the Barro-Colorado island usin the percolation theory. In general clustering is a widespread characteristic of many biotic system, for example *Boer et al.* (2008) found aggregation in clusters of forest-fires confirming the model of *Bak et al.* (1990) as reviewed also in *Zinck and Grimm* (2009), *Vandermeer et al.* (2008) found clusters of ant colonies in tropical agrosystems, and *Marco et al.* (2009) found power-law of cluster-size of cancer cells in analogies to the species distribution in ecosystems. Various theories and description have been proposed for the clustering phenomena. (*Pascual et al.*, 2002) used lattice-based models with antagonist ecological interactions concluding that the clustering is a self-organized process arising from the interaction of species, *He and Hubbell* (2003) provided an elegant formulation using the percolation theory at different scale but without specifying any determinant mechanism for the phenomena, *Plotkin and Muller-Landau* (2002) modeled the clustering of conspecific individuals with a negative binomial distribution, that implies a lognormal species-abundance distribution, in place of the Poisson. Moreover *Garcia-Martin and Goldenfeld* (2006), also trying to link the species-area relationship (SAR, the relationship between the number of species and their habitat area) to the aggregation of species, argued that the power-law SAR is a robust consequence of a the skewed species-abundance distribution with higher rarity, together with the observation that individuals of a given species tend to cluster.

As in *Houchmandzadeh* (2002); *Houchmandzadeh and Vallade* (2003); *Houchmandzadeh* (2008); *Marco et al.* (2009) in this chapter it is found that ecological systems display considerable patchiness and this is cused solely by the dynamics of birth, death and dispersal. These causes seem to be common to all aggregation phenomena. Of course it would be absurd to neglet the existence of the influences of the landscape structure and environemtal heterogeneities on the spatial distribution of species (*Young et al.*, 2001; *Houchmandzadeh*, 2002) but their importance is relevant only in particular conditions. The topology of the cluster-size probability distribution has been studied also in function of the parameters of the model.

This is in agreement with *Etienne and Alonso* (2007) that claim that stochasticity

produced by the neutral drift and dispersal can explain species coexistence. Then the unified neutral theory of biodiversity (*Hubbell*, 2001, 2006) is a robust ecological theory that adds an explanation of the reason for which at the macroecological scale is possible to detect real patterns without assuming interaction between species. *Marco et al.* (2009) evidenced the central role of local speciation in each local community and disperal of species to other local communities. It is shown here how dispersal limitation or very large dispersal that leads to saturation of the local communities and null correlation of LSR, determine the breaking of the power-law structure of the cluster-size distribution.

Other studies in literature motivate the ubiquity and emergence of clustered patterns to niche features, e.g. soil specialization and species traits (*Russo et al.*, 2007). Qualitative studies also exist that use different types of null models (*Hausdorf and Henning*, 2007), despite there is some consensus in considering the neutral model different than standard null models (*Gotelli*, 2006). *Houchmandzadeh and Vallade* (2003) elegantly provide also the analystical solution for the neutral clustering problem in 1-D (that can be for example the case of the bed of a river or along the shore of an island), in 2-D and 3-D.

The clustering phenomena has been analysed as a function of the topology of the ecosystem and in particular comparing the dissected ecosystem of river basins vs the open structure of 2-D landscape with the same shape. The effect of the branching structure of rivers has also been analysed. The comparison between ecosystems with the same boundaries seems much more fair than what *Muneepeerakul et al.* (2007a,b) proposed in which the comparison was performed using ecosystems with different shape and number of local communities. Emphasis about the importance of dendritic ecosystems has been reported, among the others, by *Fagan* (2002), *Morrissey and Kerckhove* (2009) for the maintenance of gene flow, *Gastner et al.* (2009) studying the transition from connected to fragmented vegetation patterns due to external changes, and by *Lowe et al.* (2006); *Lowe* (2008) and *Grant et al.* (2007) for linking animal dynamics and traits to the dispersal in riverine ecosystems.

The clustering analysis is performed on: 1) different river networks with different elongation ratio (OCN, Cordevole and Tanaro) (Chapter 2) to study the effect of the shape of the network; 2) different dispersal kernels, namely exponential and exponential-Cauchy kernel, varying also the speciation rate (Chapter 3); 3) homogeneous and heterogeneous conditions that imply constant and variable habitat capacity respectively (Chapter 3). All the above analysis are also performed for different topologies, river network (RN) and 2-D landscape (2D), in order to understand better the key determinants for the spatial distribution of species. Also in this chapter the species-area relationship, introduced in chapter 4, has been analysed in function of the probability distributions of the cluster-size.

5.2 Cluster-size Distribution and Species-Area relationship in the MMRS

The definition of *cluster* varies accordingly to the topology of the ecosystem in which the spatial aggregation of species takes place. In this way the description of the patterns tries

to understand the responsible processes of the clsutering dynamic. The cluster-size (CS) is defined as the number of adjacent local communities (DTAs (direct tributary areas) in the MMRS or grid-cells) in which a species is present. The adjacency criteria is function of the dispersal mode. Considering the river network a DTA or grid-cell is considered part of a cluster if it is immediately adjacent along the network with topological distance equal to one (upstream or downstream). In a 2-D landscape the four-site von Neumann adjacency criteria is sufficient in the definition of a cluster, so every neighbor of each site belongs to the cluster if the species considered is present (see Figure 5.1).

It has to be noted that here the employed clustering analysis differs from the usual one found in literature in most of the previous studies. In each DTA or grid-cell there can be present more than one individual of the same or of different species, that act neutrally, due to the metacommunity character of the model. In this way we refer to "neutral metacommunity clustering". In most of the already existing studies only one individual is present for each grid of the domain, both in empirical studies due to the structure of the data (*Scanlon et al.*, 2007b, e.g.,) and in modeling studies (e.g. in voter models (*Pascual et al.*, 2002)). The cluster size is evaluated then for each metapopulation resulting from the metacommunity model. Here a metapopulation include all the individuals of the same species in all the local communities of the ecosusystem. Conspecific aggregation appear ubiquous both for fishes and vegetation in the MMRS, and in simulated species-patterns.

The different dynamics of small and big trees is evidenced by some past studies also for different patterns, e.g. for the species-area relationship that is itself related to the clustering process (*Plotkin et al.*, 2000, 2002). The species-area relationship is linked to the spatial distribution of species, which accordingly to the literature is essentially governed at multiple scales by birth, death and migration process. The taxonomic and regional uncertainty in the derivation of the species-area is reported in (*Guilhaumon et al.*, 2008). In this view it is evident the central role of the dispersal kernel. The species-area relationship, $S \sim cA^z$ (in the following SAR), was first postulated by *Arrhenius* (1921) and subsequently *MacArthur and Wilson* (1963); *MacArthur and Wilson* (2001) considered the power law in light of a dynamic equilibrium of species exchanges between islands. Figure 5.1 shows the criteria to determine the cluster-size of the species in a river network and in a 2-D landscape picking up two generic species A and B and some DTA from the MMRS (see Figure 2.3 in Chapter 2).

Figure 5.2 shows $P(CS \geq cs)$ that is the probability of exceedence of the cluster-size (CS) for fishes (*Muneepeerakul et al.*, 2008), big trees (*USFS-FIDO*, 2009; *Konar et al.*, 2009) and small trees of the MMRS. The probability of exceedence of the cluster-size has been computed as in *Clauset et al.* (2007). The power-law holds for three orders of magnitude and ϵ is the slope of the assumed probability distribution,

$$P(CS \geq cs) \sim cs^{-\epsilon}\, F(\frac{cs}{CS_{max}})\,, \tag{5.1}$$

where F is a scale-function. The values of ϵ for the data, model and the homogeneous case are reported in Table 5.3. The slope is lower for big trees than for fishes (higher ϵ) meaning that is more probable to find larger clusters of big trees. Moreover ϵ of small trees is lower than for big trees meaning smaller clusters for the former ones. The model

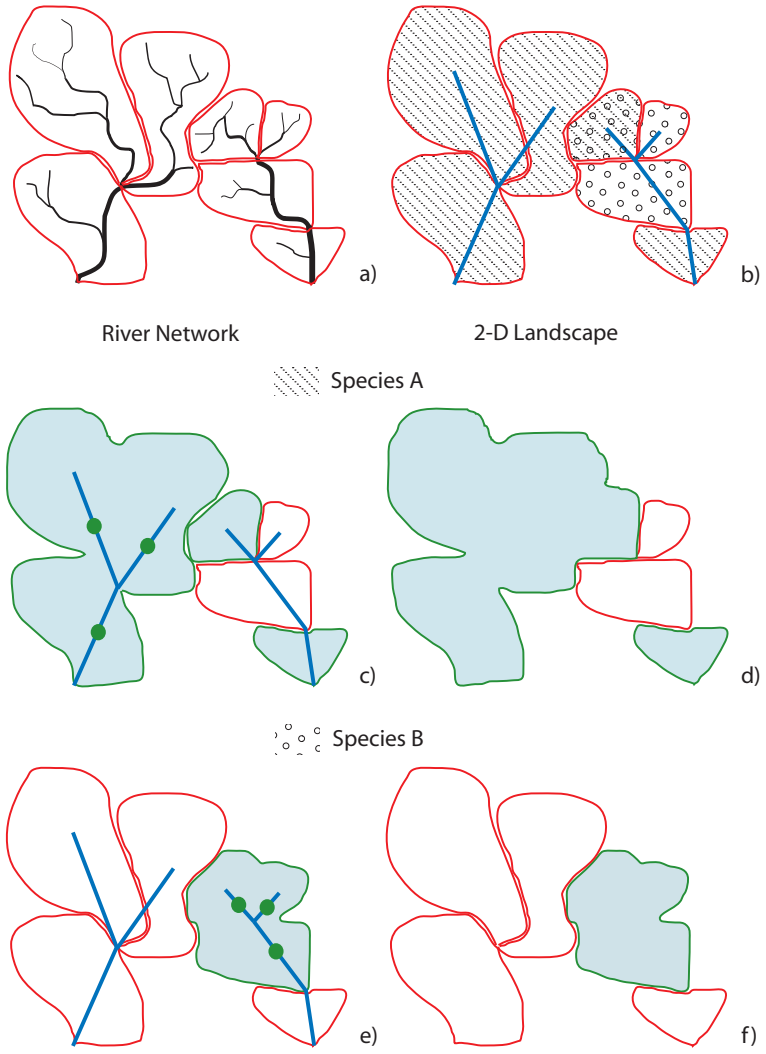

Figure 5.1: Criteria to determine the cluster size of the species in a River Network and in a 2-D landscape. a) is the structure of the DTA and part of the real network of the MMRS; b) is the modeled network with links that connect each DTA; c) represents the clusters for the species A in the RN case; d) represents the clusters for the species A in the 2-D landscape; e) represents the clusters for the species B in the RN case; f) represents the clusters for the species B in the 2-D landscape;

is capable to reproduce optimally the power-law of the cluster-size as evidenced by the overlapping of the black cuve (model) on the red curve (data) in figure 5.2. Also in figure 5.2 $P(CS \geq cs)$ is reported for the case in which the habitat capacity is constant for all the local communities (see chapter 3). The distribution of CS in the homogeneous condition does not vary significatively for fishes, instead it becomes milder for trees. The maximum value of the cluster-size is larger as the probability of observing large clusters. CS_{max} is larger for fishes than trees and the homogeneous condition determine larger CS_{max} for all the three taxa. The SAR for fishes and trees of the MMRS is shown in figure 5.3 for real data and in the homogeneous case. Here in the species-area pattern the influence of the environmental heterogeneities is stronger than on the $P(CS \geq cs)$ since z varies a lot in the homogeneous case respect the real one. Further consideration about the effect of the environmental heterogeneities will be discussed in section 5.5. With uniform habitat capacity z results smaller than in the real case. Both in figure 5.2 and 5.3 the computation for fishes and trees is at the DTA level.

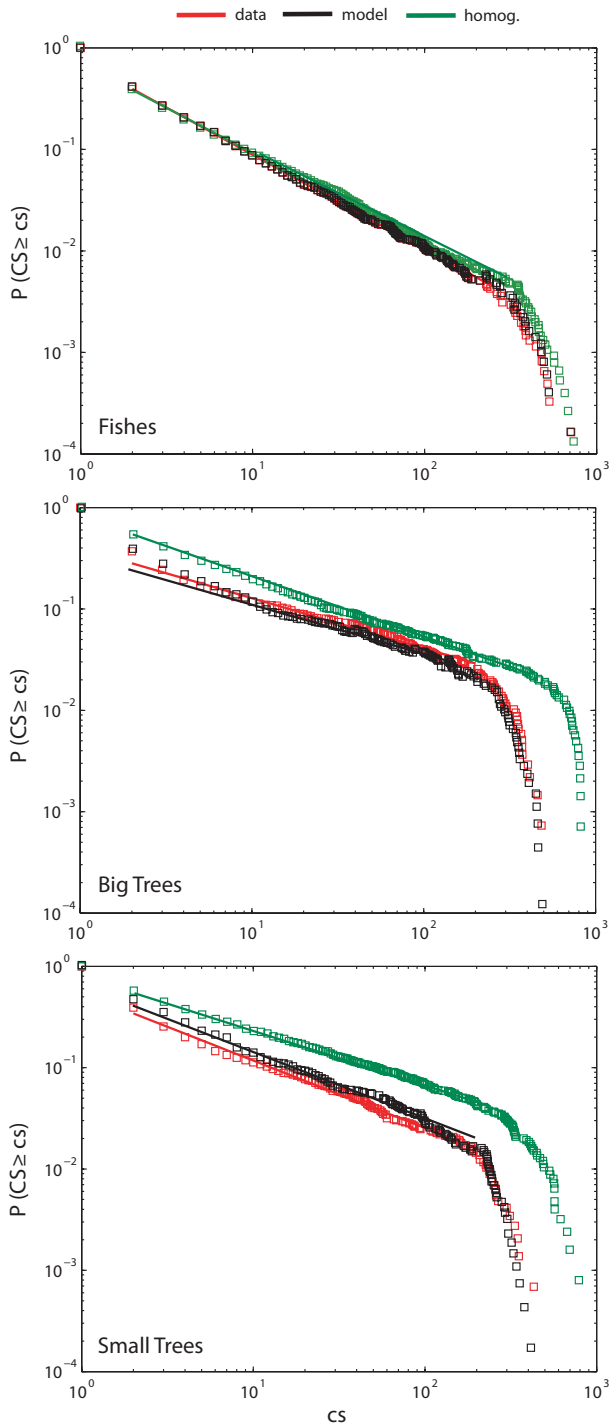

Figure 5.2: $P(CS \geq cs)$ of fishes and trees of the MMRS, data at the DTA level and model results.

Figure 5.4 disentangles the structure of $P(CS \geq cs)$ of big-trees of the MMRS in function of their occupancy. The most present species are able to determine the structure of the power-law over the entir range of cluster size as shown in plot b), where the probability of exceedence of the cluster size for the 10 most common species is reported; these species shape the entire distribution so they have the largest clusters but also the smallest ones. The least common species (rare) have only small clusters (plot d)) and the species with intermediate occupancy shape the middle part of the distribution. Then in order to estimate ϵ the information needed is only about few most present species, that can be e.g. very useful for reducing the effort of sampling campaigns. It seems then only the information about the commonest species is relevant to detect the properties of the clustering pattern. Collectively species are spatially organized all together forming a power-law of the probability of occurrence of the cluster size. Therefore this does not mean fractal scale-invariant patterns of the biodiversity patterns as shown in chapter 4.

$P(CS \geq cs)$ can be derived also for each single species (intraspecies distribution), without merging the cluster-sizes of all the species, but as expected is not possible to defend a power-law of the cluster size due to the limitedness of the number of clusters for each of them. Therefore the range of the cluster-size, at least for the ten most common species, ranges over three orders of magnitude but the number of clusters is not sufficient to defend a power-law. $P(CS \geq cs)$ is then the interspecies distribution if nothing is said contrarily.

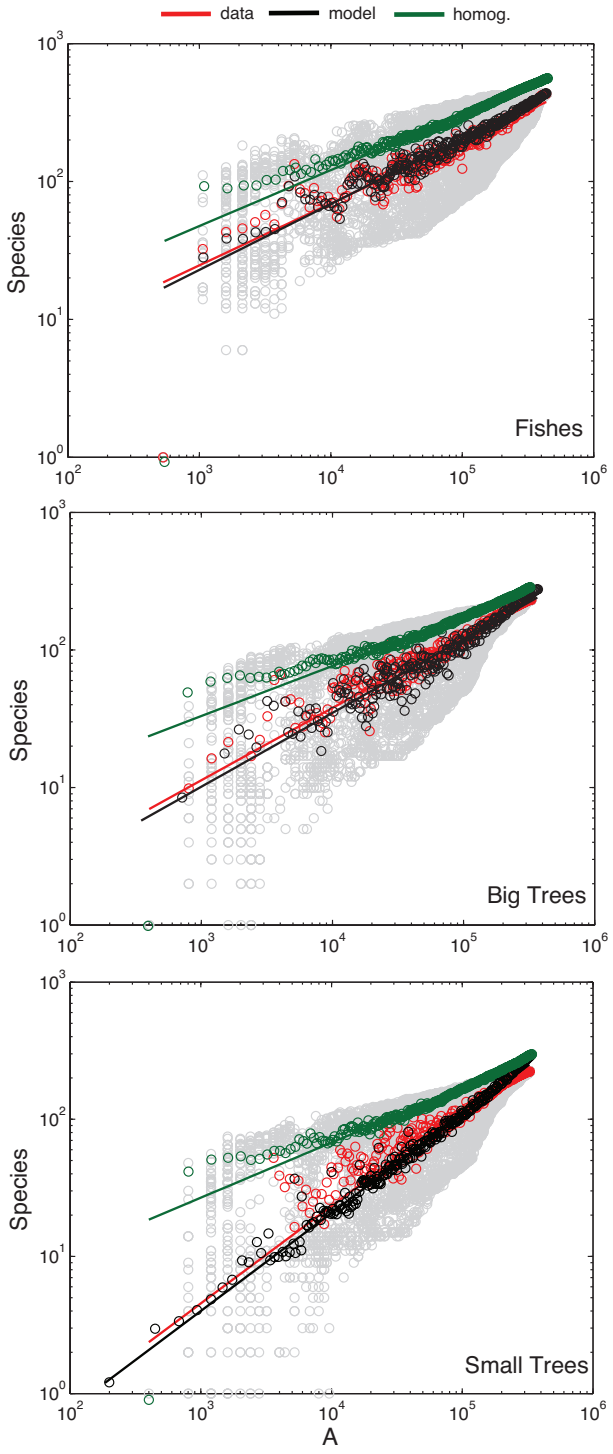

Figure 5.3: SAR of fishes and trees of the MMRS, data and model results. The grey dots represents the actual values of the Species in function of the Area. The scatter of the model results is less pronounced than data as appears also from the binned dots.

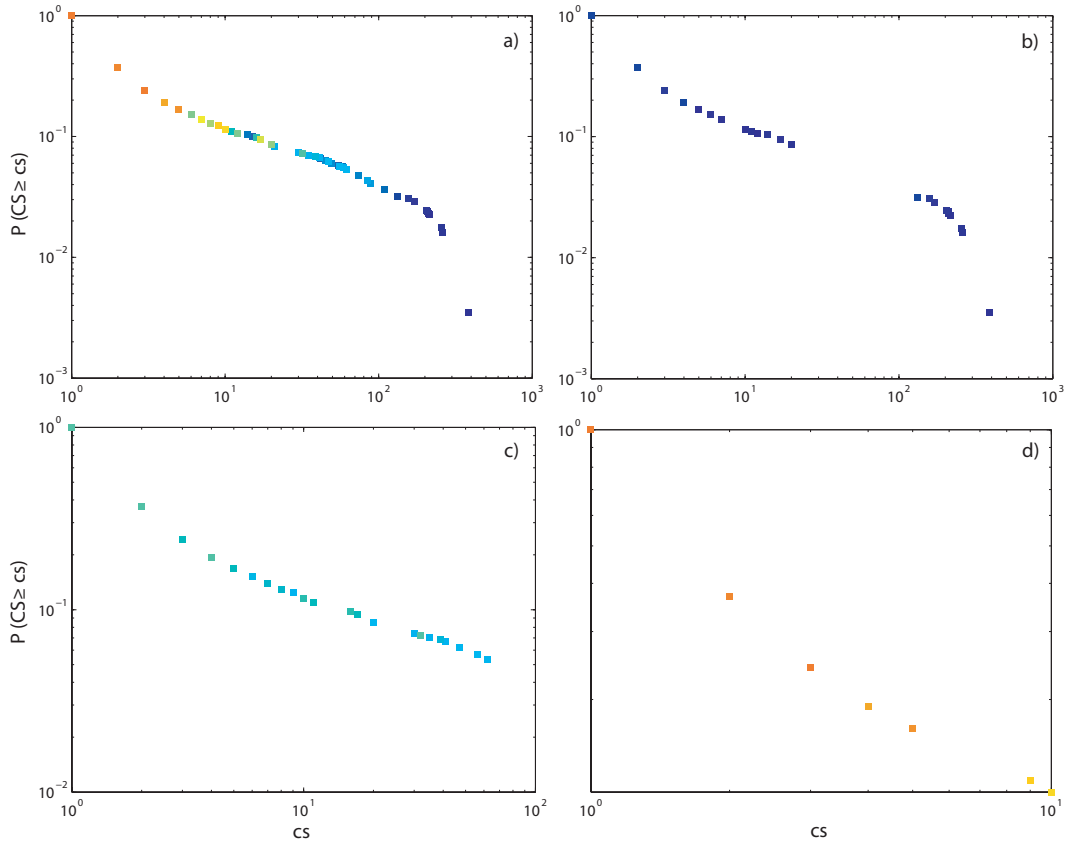

Figure 5.4: a) $P(CS \geq cs)$ for big trees of the MMRS: the color is proportional to the occupancy of the species (blue is most present, red is least present). Cluster size distribution in the probability of exceedence for: b) the ten most common species; c) the 20-30th most common species; and c) the ten least present species.

5.3 Scale and Resolution of the Cluster-Size Distribution and SAR

In this section the dependence of $P(CS \geq cs)$ and the SAR is analysed in function of the scale (extent), resolution (grain-size), and coarse-graining level at which the ecosystem is simulated.

Figure 5.5 shows a preliminary study about the scale invariance of $P(CS \geq cs)$ and the SAR in the MMRS using the big trees dataset (*USFS-FIDO*, 2009; *Konar et al.*, 2009) at the plot-scale. In order to test the scale invariance the dataset has been analysed in the whole MMRS and in the southern portion of the basin that includes the outlet (a region comprising part of the states of Texas, Louisiana, Mississippi, Arkansas, and Oklahoma) where the spatial cover of plots is quite homogeneous in order to avoid bias due to the limitedness of data. In order to compute the cluster-size distribution according the method described in section 5.2, different grids with a decreasing resolution have been overimposed on the big-trees dataset (see chapter 2). Since the dispersal of trees happens Euclideanly on a 2-D not dissected ecosystem the only-adjacency criteria (von Neumann neighboring) has been used: a pixel of the grid is a unit of a cluster if in the plots relaying into it the species considered is present. For the analysis in figure 5.5 the finest resolution of the grid (25,000) has been choosen. As it is shown also in table 5.1, in general both ϵ and z decrease as the extension becomes smaller. Then ϵ and z for the big trees dataset in the outlet portion of MMRS are smaller than for the dataset in the whole basin. Further investigations need to be done in order to test better the relation between $P(CS \geq cs)$ and the scale of analysis choosing e.g. sub-regions of the MMRS with uniform but different climate (testing the influence of the dominant climate regime) and with different plot-density (testing the influence of the dishomogeneity of data). In table 5.1 and in figure 5.6 the behavior of ϵ, $\langle CS \rangle$, and z is reported in function of the grain-size (resolution) of the adopted grid on the MMRS big-trees dataset. It is evident how all the three clustering variables increase with decreasing the resolution, that is from the finest to the coarsest grain-size grid. At the coarsest grid (here 100,000) the probability to observe large clusters is higher as well as to have an higher mean value of LSR.

Since the average DTA area is 3900 km^2, the 50 km grid creates cells whose extension if more or less similar to the extension of a DTA of the MMRS. Then $P(CS \geq cs)$ and the SAR should lie in between the mentioned patterns of the grids at scales 50 and 100. Therefore only z at the DTA scale (table 5.3) seems close to z derived at the 50 plot-scale, not ϵ nor $\langle CS \rangle$. Then this puts an hint about the correctness of the method to upscale the information at the plot-scale to larger scales (e.g. at the DTA level) maintaining invariant the biodiversity properties or limiting the deviations from the mean values within a certain tolerance.

Figure 5.7 shows $P(CS \geq cs)$ for the Tanaro basin in the scenario of exponential kernel dispersal (a) for the river network (RN), and b) for the 2-D landscape case), and for the exponential-Cauchy kernel dispersal (c) for RN, and d) for 2-D) at different coarse-graining scales. The cases implemented are reported in table 3.2, chapter 3. $P(CS \geq cs)$ results not invariant across coarse-graining scales. ϵ decreases going from the coarsest to the finest resolution while $\langle CS \rangle$ increases. The probability to have a large number of

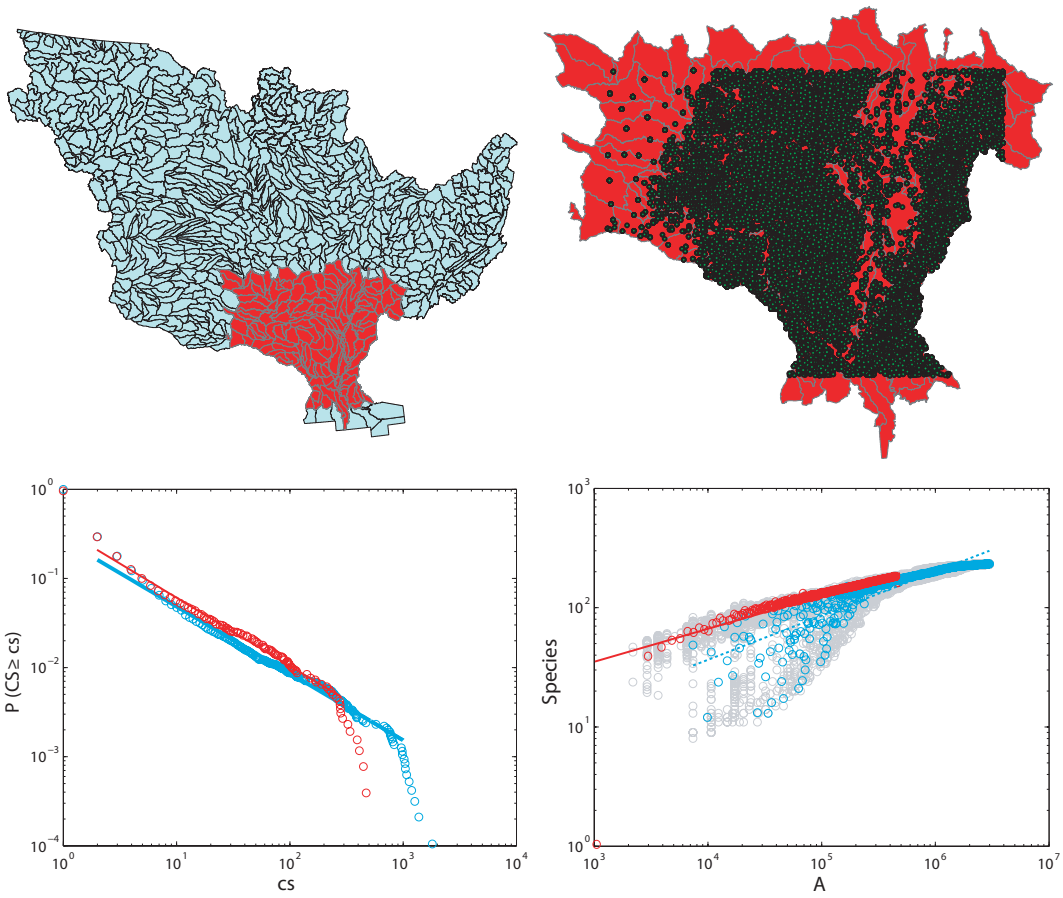

Figure 5.5: $P(CS \geq cs)$ and SAR for the MMRS at different scales, the whole basin and the southern part. The analysis is performed at the DTA level.

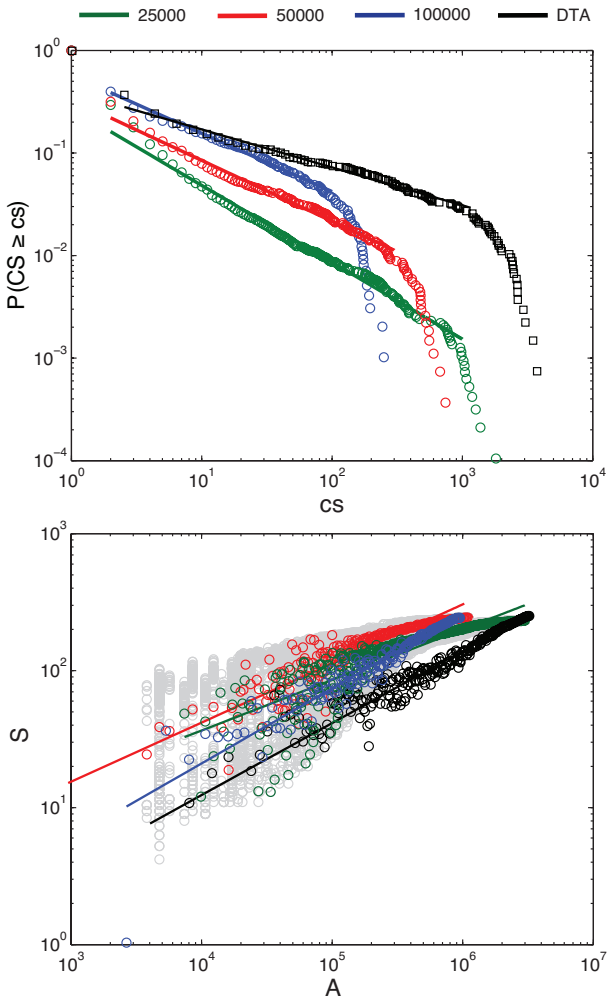

Figure 5.6: $P(CS \geq cs)$ and SAR for the big tree plots-data of the MMRS at different grid scales.

Table 5.1: Clustering parameters for BT of the MMRS at different resolutions at the plots scale.

Variables	ϵ	$\langle CS \rangle$	$z(A)$
Big Trees			
$25,000$	-0.75	6.25	0.37
$50,000$	-0.59	9.82	0.43
$100,00$	-0.56	10.98	0.55
25000 *outlet*	-0.77	5.54	0.29

clusters increases when the spatial resolution increases since the number of pixels in the grid become higher. In figure 5.7 a) the curves seems collapsing on a same curve due to dispersal-limitation effects because the disperal paramter is very low in the exponential kernel dispersal scenario (chapter 4).

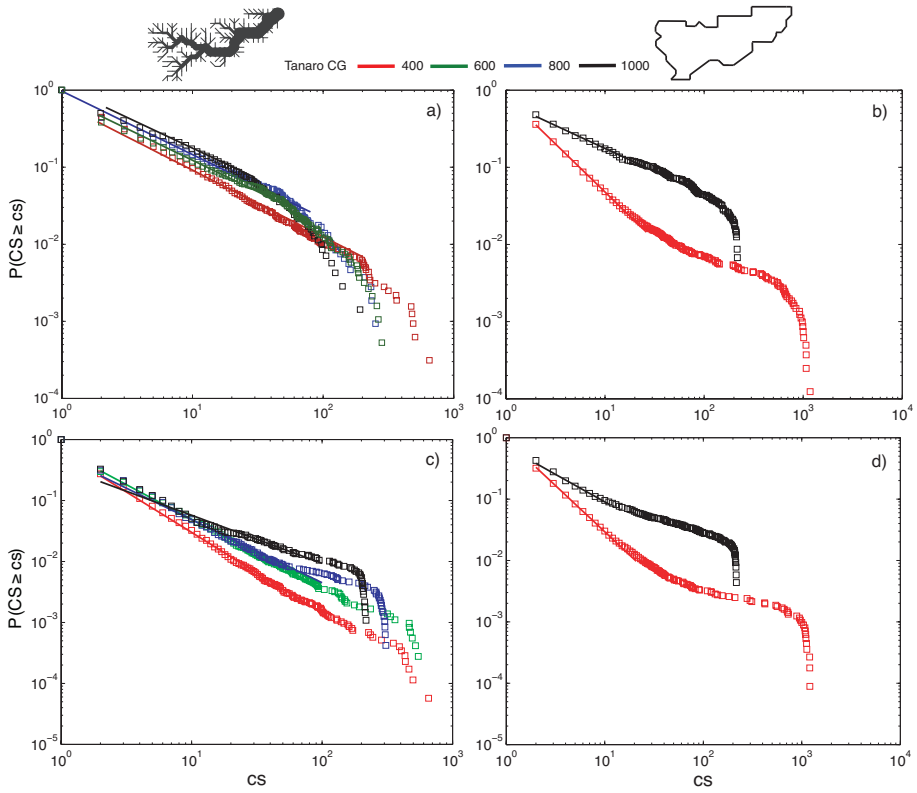

Figure 5.7: $P(CS \geq cs)$ for the Tanaro basin with exponential kernel dispersal (a) for RN, b) for 2-D), and with exponential-Cauchy kernel dispersal (c) for RN, d) for 2-D) at different coarse-graining scales. In b) and d) the 600 and 800 coarse-graining curves (not reported) lie in between the red and black curves.

In summary the findings are:

1. $P(CS \geq cs)$ and the SAR depend on the grain-size (resolution);
2. $P(CS \geq cs)$ is generally robust but variant across scales (extent) as the the SAR results dependent on the scale;
3. $P(CS \geq cs)$ is in general variant across coarse-graining (model) while the SAR is invariant.

5.4 Relationship between SAR and Cluster-size Distribution

In this section the attention is focused in linking the biodiversity patterns of the species-area relationship and the the exceedence probability of the cluster-size. Specifically the aim is in relating their scaling exponents, z and ϵ respectively, for different topologies and shapes of the studied ecosystem (figures 5.8 and 5.9). Also the behavior of z, ϵ and $\langle CS \rangle$ has been investigated in function of the dispersal parameter d_l.

In general z gets higher for smaller d_l (Figure 5.10). This is in fact true comparing z for the exponential and for any other exponential-fat tail dispersal kernel scenario. Because in general a decay of z is observed in function of d_l, the smaller value of z for big trees than for small trees both in the homogeneous and heterogeneous case is justified (Figure 5.8). Also z for fishes is smaller than z for big trees mostly because the dispersal parameter is higher with the additional determinant of the fat-tail kernel that adds a long-dispersal character to the dispersal. Table 3.4 in chapter 3 shows the values of the dispersal parameter adopted for fishes, big trees and small trees.

The z-values for the exponential-Cauchy dispersal scenario are lower than the z-s for the pure exponential dispersal scenario both in the RN and 2-D case. In figure 5.8 and 5.9 we show a clear trend between z and ϵ distinguishable for different topologies (RN and 2-D) and ecological scenario (a) and b)). We have also reported the z-ϵ points for the fishes, big and small trees of the MMRS (F cross, BT and ST squares respectively) which follow the same trend of the simulated cases but shifted above.

In figures 5.10 and 5.11 z, ϵ and $\langle CS \rangle$ for the Tanaro, Cordevole and OCN are analysed in function of the dispersal parameter d_l.

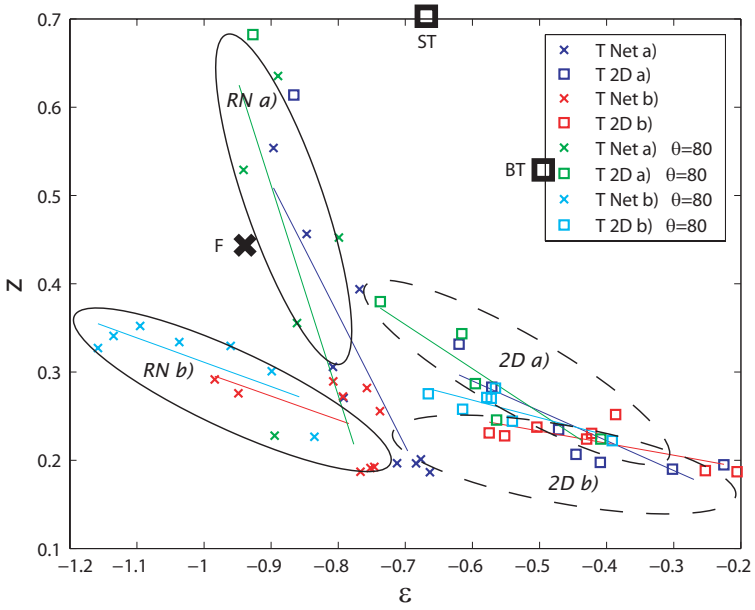

Figure 5.8: Relationship between z and ϵ for the Tanaro river network (crosses) and the 2-D landscape (squares) with the same boundaries. The scenario a) is with the exponential kernel and the scenario b) is for the exponential-Cauchy kernel. Both scenarios have been analyzed also for $\theta = 80$. The dashed-line ellipses refer to the 2-D case and the solid-line ones to the network case.

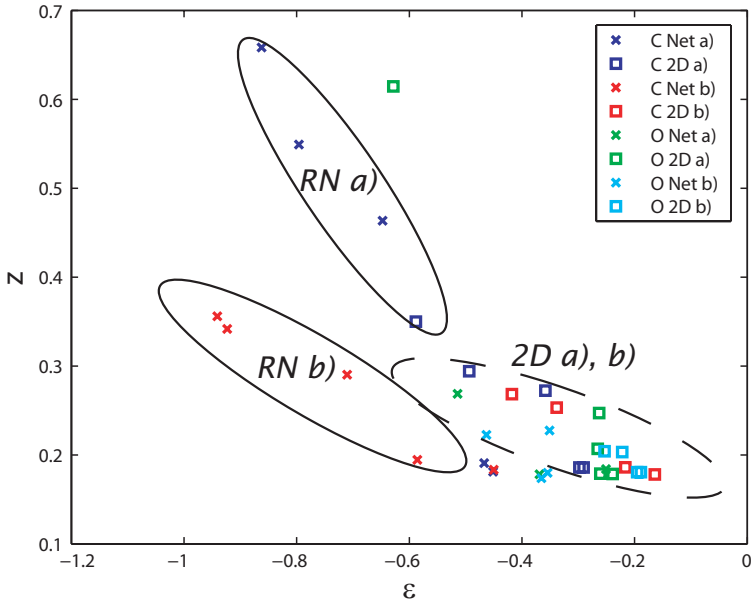

Figure 5.9: Relationship between z and ϵ for the Cordevole and OCN river network cases (crosses) and the 2-D landscapes (squares) with the same boundaries. The scenario a) is with the exponential kernel and the scenario b) is for the exponential-Cauchy kernel. The dashed-line ellipses refer to the 2-D case and the solid-line ones to the network case. Cordevole and OCN show lower values of z but higher values of ϵ.

5.4.1 The Species-Individual Relationship

We have investigated the species-individual relationship (in the following SIR) that joined with the SAR is a useful pattern to evidence the role of the environmental heterogeneities which can be relevant in the spatial arrangement of the species in an ecosystem.

Contrarily to *Condit et al.* (1996) we have found that for different dbh classes (where dbh stand for diameter breast height that is the diameter measured usually at 4.5 feet above the ground line on the uphill side of the tree (*Burkman*, 2005)) the SIR shows different exponents as the SAR. *Condit et al.* (1996) shows instead the SIR is a unique curve for all the dbh classes. This contrasting results can doubtly depend on the wide range of climatic zones across the MMRS vs the unique predominant climate in the region analysed in *Condit et al.* (1996). In *Condit et al.* (1996), where tropical trees are particularly analyzed, the different organization of the sampled dataset can be instead the determinant reason for the discording results: in particular the dbh classes assume trees with diameter above an increasing threshold, excluding all the others, in a way that the sampled populations is smaller and smaller. Rather in your study it has been tried to differentiate different life-stages dividing the tree-population in young and old trees in function of a rough criteria that is the dbh value.

According to *Brown and Gurevitch* (2004) the SIR is capable to detect additional biodiversity information than the SAR, the latter very used in the past as a tool to detect the effect of climate change on the habitat-loss of species (*Thomas et al.*, 2001, 2004; *Lewis*, 2006). Therefore the estimates of the climate changes effects using the SAR have be proofed to be overestimated (*Lewis*, 2006) due to the bias between sampled area and individuals, then the SIR could be a better instruments to forecast the future climate impact.

Figure 5.12 shows the species-individual relationships for fishes and trees of the MMRS, data and model results. H^* is the number of individuals in the sampled area that, in the case of heterogeneous habitat capacity, does not coincide with the number of local communities as in chapter 4. Then SAR and SIR are different only in the heterogeneous case that is proposed for the fishes and trees of the MMRS: in red is the SIR for real data, in green is the for the heterogeneous combined case (fishes modeled in a 2-D landscape and trees in RN). In table 5.3 and table 5.4 the scaling slope of the SIR is reported. It is to notice that $z(A)$ of the SAR is always greater than $z(H^*)$ meaning that the number of individuals H^* scales with the area with an exponent greater than one. In this way the overestimation of the habitat-loss, due to climate change for example, using the SAR is diminuished using the SIR. Moreover the effect of climate change potentially impacts the nummber of individuals rather then the habitat area then the species-individual relationship is a more correct tool to be used. $z(H^*)$ for the fishes simulated in 2-D landscape with an Euclidean dispersal is smaller than in the real case, while for trees modeled in RN it is larger than in the 2-D case.

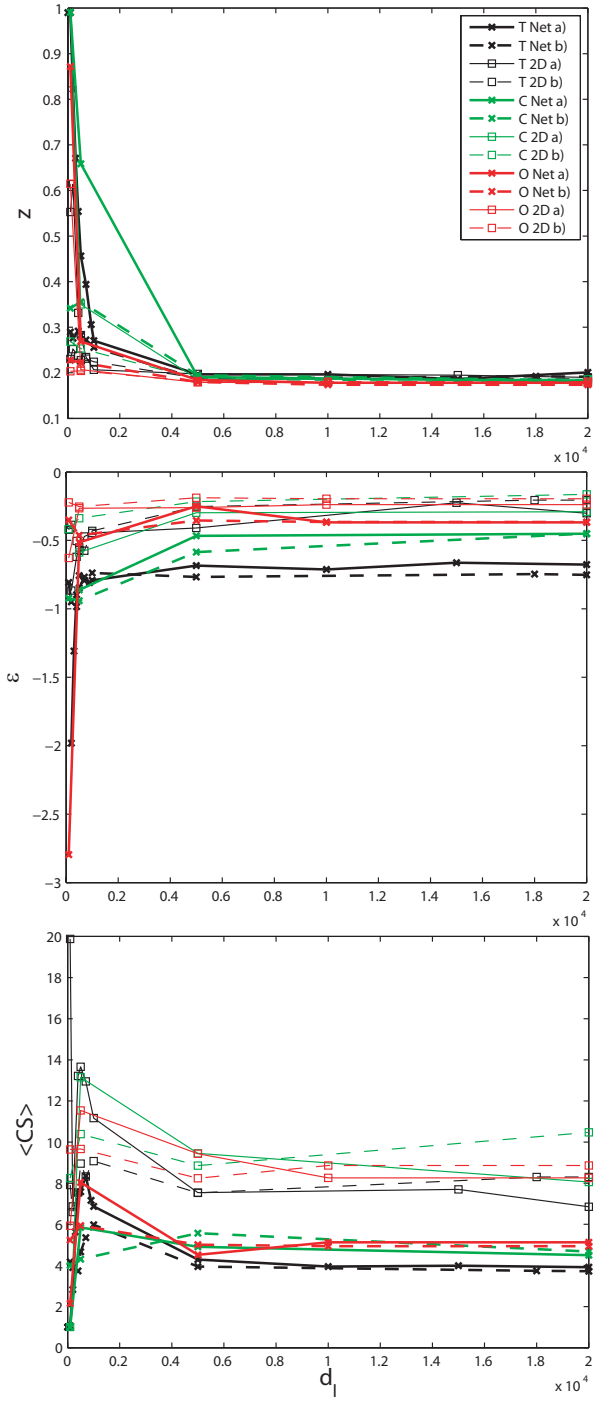

Figure 5.10: Clustering properties z, ϵ and $\langle CS \rangle$ in function of the mean dispersal length d_l and kernel dispersal for the Tanaro, Cordevole and OCN.

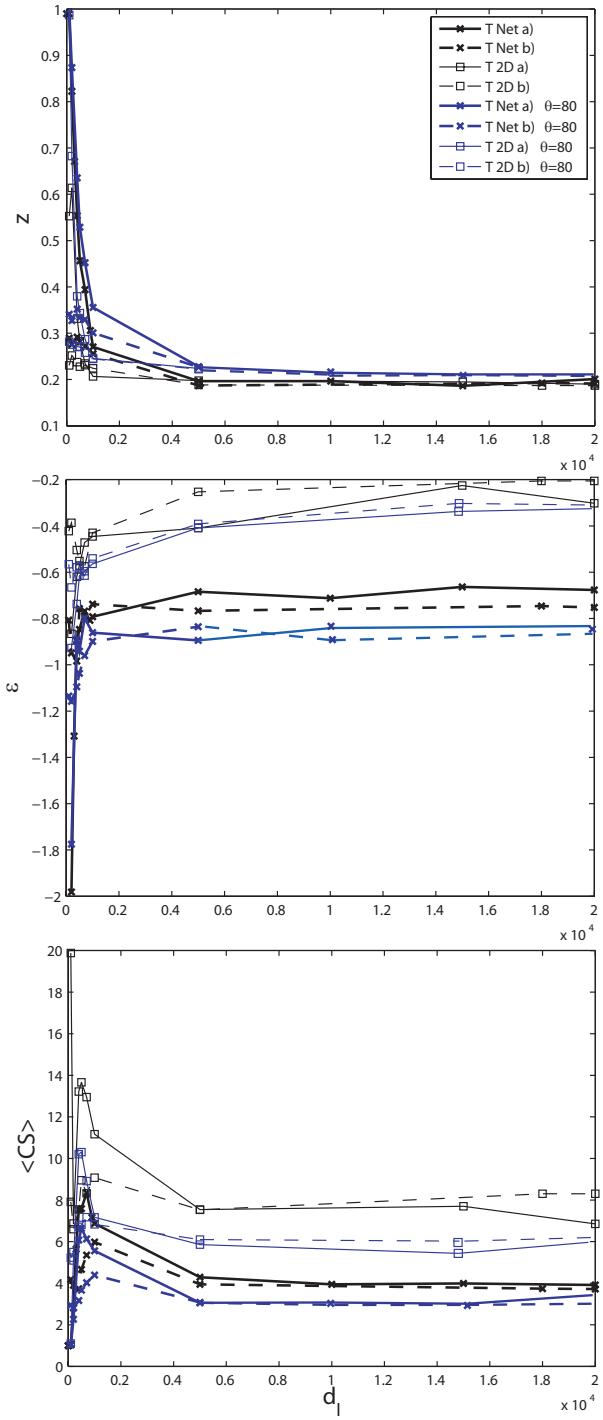

Figure 5.11: Clustering properties z, ϵ and $\langle CS \rangle$ in function of the mean dispersal length d_l, kernel dispersal and θ for the Tanaro river basin

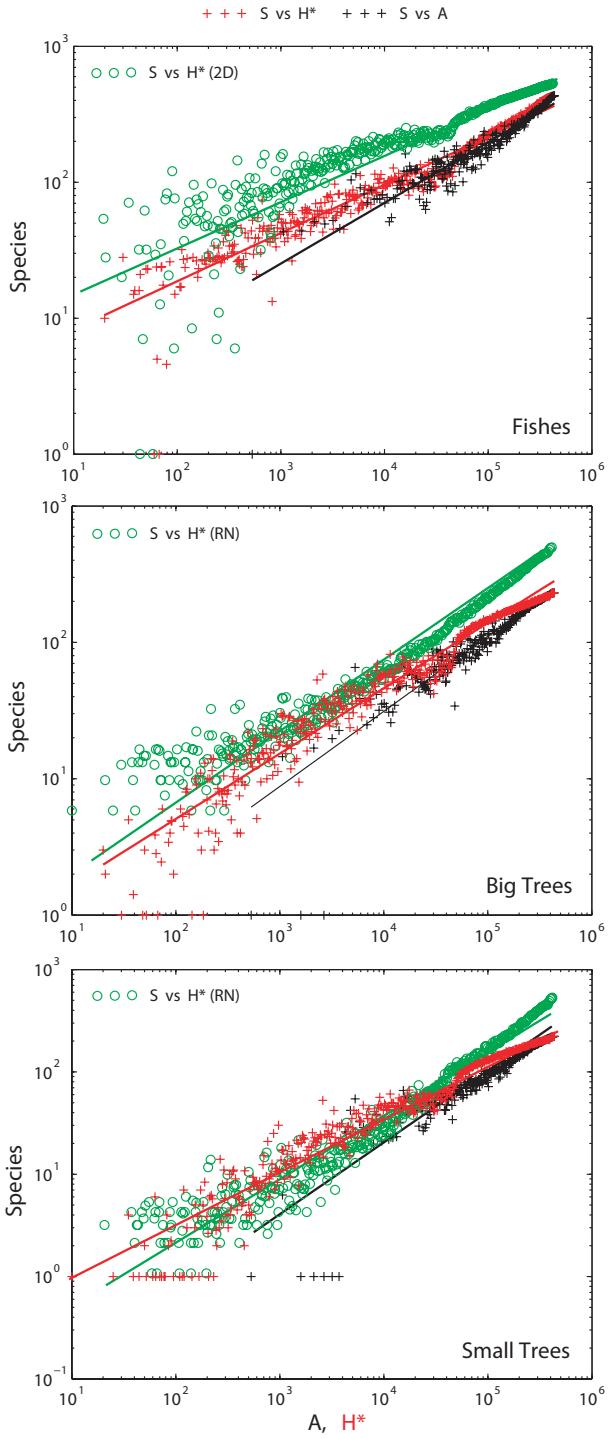

Figure 5.12: Species-Individual relationships for fishes and trees of the MMRS, data and model results

5.5 Effect of Landscape Topology, Shape and Environmental Heterogeneities

In this section the role of the topology (RN vs 2-D), shape, and environmental heterogeneities of an ecosystem, has been studied on the spatial aggregation of species.

The homogeneous condition is simulated assuming an habitat capacity that is uniform and constant for each local community, i.e. the average habitat capacity H in table 3.4 of chapter 3. In the homogeneous case the exponent ϵ of the exceedence distribution of the cluster-size is greater than in the heterogeneous case meaning a milder power-law, then larger clusters. Conversely z of the species-area relationship is smaller than in the heterogenous case. Moreover $\langle CS \rangle$ and LSR are higher in the homogeneous than in the heterogeneous case, both for fishes and vegetation regardless the dbh class.The effect of the environmental heterogeneities is stronger in the SAR (figures 5.3 5.17) than in $P(CS \geq cs)$ (figure 5.16). Figures 5.13, 5.14, and 5.15 reproduce the LSR patterns of fishes, big trees and small trees of the MMRS in the homogeneous condition, combined condition (fish-dispersal in 2-D and trees-dispersal in RN), and homogeneous combined conditions. Table 5.2 reports the values of LSR (mean $\langle . \rangle$, min and max) for data (see Figure 2.14 chapter 2) and model, homogeneous, combined and homogeneous combined scenario. The homogeneous condition produces a wider range of cluster sizes and the observed ϵ is larger than in the real case. Big trees are more aggregated than small threes and fishes ($\epsilon_{BT} > \epsilon_{ST} > \epsilon_F$) despite the maximum value of CS is bigger for the fishes.

In general then it is not true that the river network enhances the power-law behavior of the cluster-size dsitribution of species. The same robust power-law of the cluster-size is observed for big/small trees whose spatial patterns, as for the fishes in the river network, is almost entirely created by the birth-death and dispersal dynamics on the 2-D landscape. Changing the topology produces a drastic change on the power-law distribution of the cluster-size. If individuals spreads in a 2-D MMRS topology with the same parameters of fishes, $P(CS \geq cs)$ shows a milder behavior and the probability of having large clusters is higher while the probability of having small clusters is lower or equal than in the real RN MMRS case. This happens because the ratio between the dispersal parameter of fishes and the mean diameter of the 2-D topology is larger than the ratio between the same dispersal parameter and the mean diameter of the RN topology of the MMRS. Instead if individuals spread with the same dispersal parameters of trees in a RN MMRS topology, the effect is the contrary than the former for fishes in a 2-D topology: the probability of having large clusters is smaller while the probability of having small clusters is higher or equal than in the real 2-D case. In the combined case of trees the dispersal is quite small in relation to the mean diameter of the RN of the MMRS. The ratio $d_l/\langle \phi \rangle$ is defined here as "ecological ratio" and it is extremely important for the distribution of species in an ecosystem. d_l is the dispersal parameter (chapter 3) and $\langle \phi \rangle$ is the mean diameter of the ecosystem 2. In the fish case in a 2-D topology the communities tends to be saturated instead in the trees case in a RN topology the species tend to do not move or to be limited in a small region around the place where they are introduced. The former is an habitat-capacity limited case while the latter is a dispersal limited case.

In figure 5.16 is possible to see how the homogeneous condition affects fishes and

trees weakly, in general producing a power-law that is milder than the real one observed. Moreover taking out the environmental variability in the combined case (i.e. fishes in 2-D and trees in RN) the effects are much bigger than not considering them in the real counterpart (fishes in RN and trees in the 2-D topology) since the combined case is affected by *habitat-capacity limitation* for fishes, and by *dispersal limitation* for trees. This effect is more evident for small trees because the dispersal parameter is lower than for big trees. In figure 5.16 the blue distributions show the breaking of the power-law structure for fishes in 2-D and for trees in the RN with homogeneous conditions. For fishes the power-law scaling behavior of the cluster-size distribution is seriously reduced to two orders of magnitude, while for big and small trees in the RN the behavior of the cluster-size distribution is about exponential and no more power-law type. In the combined case the influence of the environmental heterogeneity is much bigger since the spatial arrangement of species changes consistently changing the shape of the cluster-size distribution. Then ecosystems in situation of stauration or limited dispersal are more fragile when changing in the exogeneous variables happens and this can lead to potential habitat-loss and species-range reduction.

Then the major responsible factor of the clustering of species is imputable to the long-distance dispersal coupled with the ecological drift. We conclude than the effect of the environmental heterogeneity is relevant when the dispersal is low compared to the mean diameter of the ecosystem (e.g. that is the case of fishes in the 2-D MMRS landscape and of trees in the RN MMRS landscape). Tables 5.3 5.4 reports the values of ϵ, $\langle CS \rangle$, $z(A)$, and $z(H^*)$ for the fishes and trees of the MMRS, data and model results.

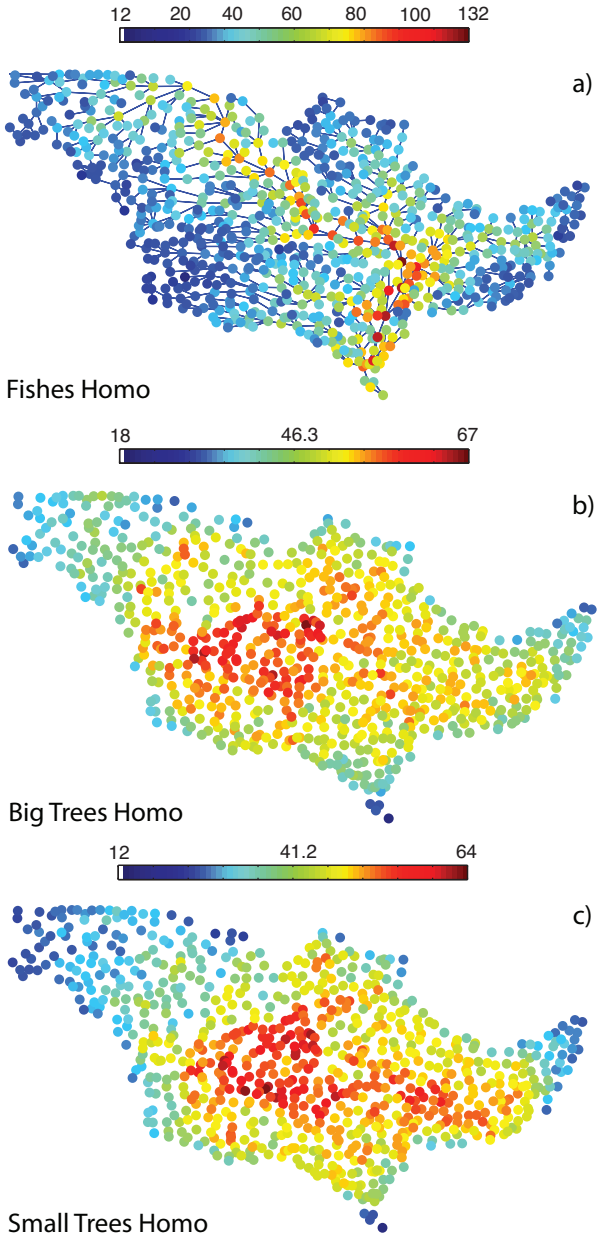

Figure 5.13: Simulated LSR patterns of fishes and trees in the MMRS with homogeneous conditions

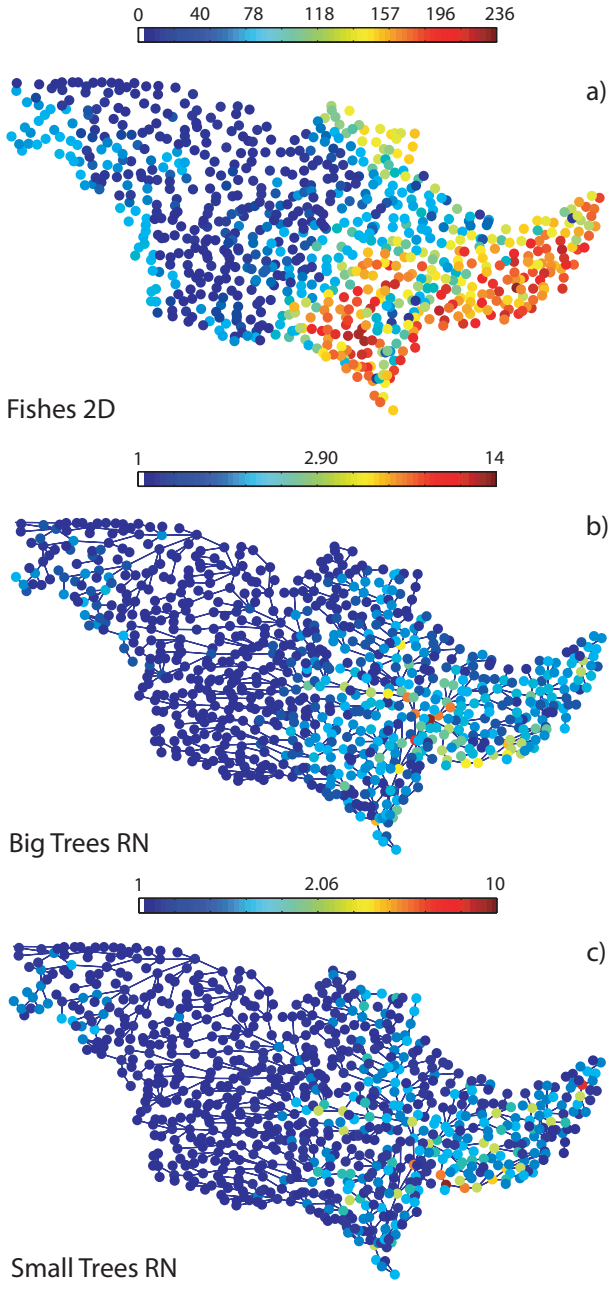

Figure 5.14: Simulated LSR patterns for the combined heterogeneous conditions: a) 2-D with fishes dispersal parameters; b) and c) river network with big/small trees dispersal parameters respectively.

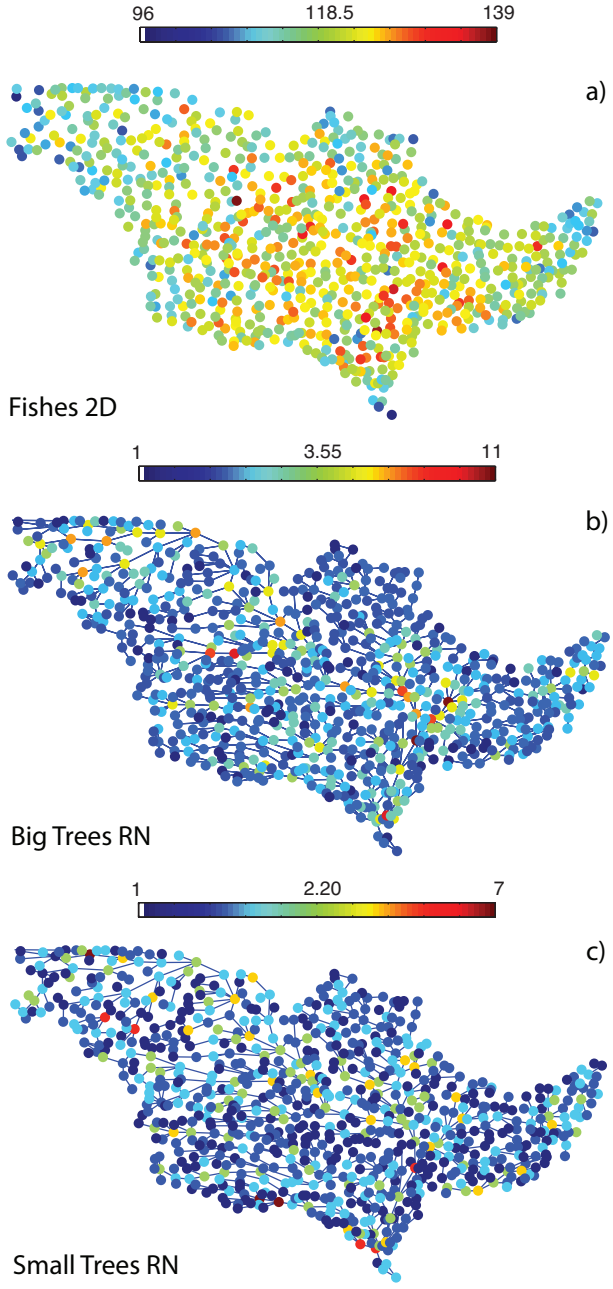

Figure 5.15: Simulated LSR patterns for the combined homogeneous conditions: a) 2-D with fishes dispersal parameters; b) and c) river network with big/small trees dispersal parameters respectively.

Table 5.2: LSR of the MMRS, data, heterogeneous and homogeneous model results.

LSR	$\langle . \rangle$	max	min
Fishes			
$data$	46.5	156	0
$RNhomog$	67.3	132	32
$2Dheterog$	80.5	236	0
$2Dhomog$	118.5	139	96
Big Trees			
$data$	24	73	0
$2Dhomog$	46.3	67	18
$RNheterog$	2.90	14	1
$RNhomog$	2.20	11	1
Small Trees			
$data$	17	58	0
$2Dhomog$	41.2	64	12
$RNheterog$	2.06	10	1
$RNhomog$	2.20	7	1

What is then the meaning of small values for ϵ more extensively? Small values of the slope of the power-law cluster-size distributin are in general associated with high values of z, meaning that for the same area the number of species found is larger, implying higher local species richness on average. In the homogenous case ϵ is higher and z is lower than the real case but the number of species found in the same area is larger, therefore in general there is not a direct relation between $\langle LSR \rangle$ and z. The number of species for small areas found in the homogenous case is higher than in the heterogenous case. Then the increase of S with A is milder because the individuals of the species are more uniformly distributed. The SAR is shown in figure 5.3 for real data and in the homogeneous cases.

In figure 5.17 the SAR is reported for the combined and the homogeneous combined cases. In the combined case z is lower for fishes and higher for trees than in the real case, as ϵ is higher and lower respectively both for the two taxa than in the real case. The combined homogeneous case here is not so different than the combined heterogeneous case, evidencing the robustness of the species-area relationship. Figure 5.18 proposes z, ϵ, and $\langle CS \rangle$ in function of the dispersal parameter for fishes and trees of the MMRS, data and model results. Table 5.5 presents a qualitative overview of the effects of the changing in the topology and in the environmental condition (taking out the heterogeneituies) on some biodiversity descriptors.

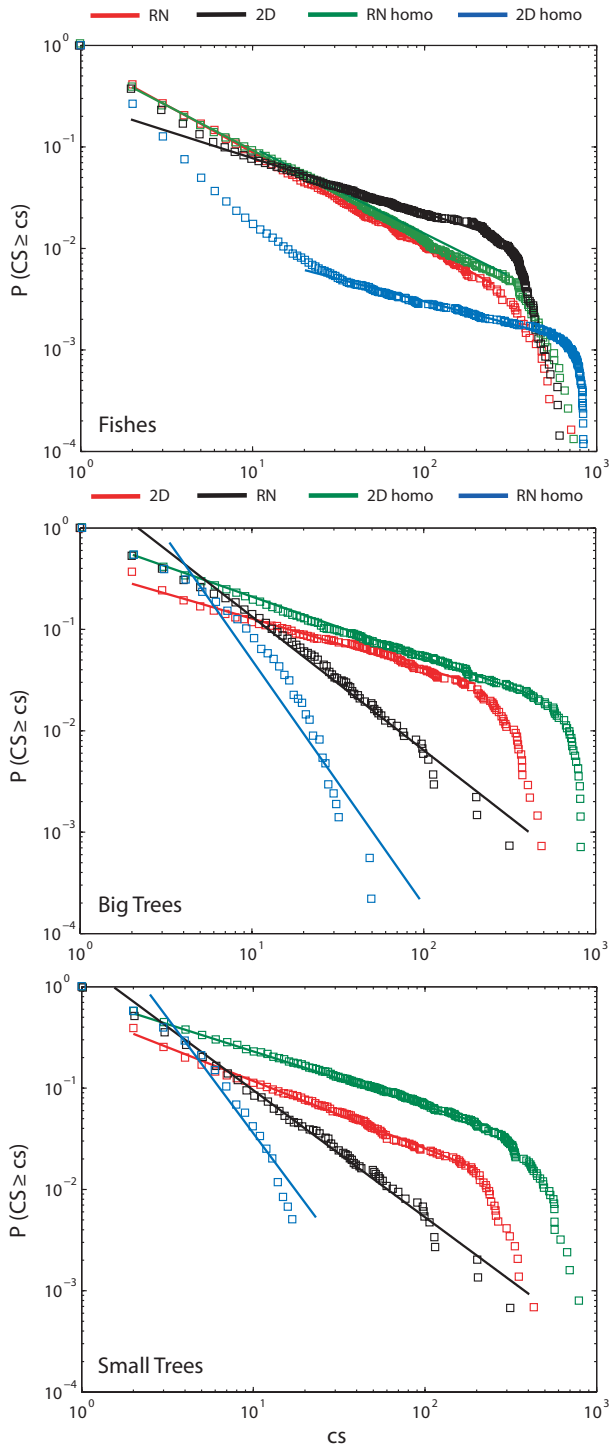

Figure 5.16: $P(CS \geq cs)$ of fishes and trees of the MMRS, data at the DTA level and homogeneous, heterogenous and combined (fishes in 2-D, trees in RN) model results.

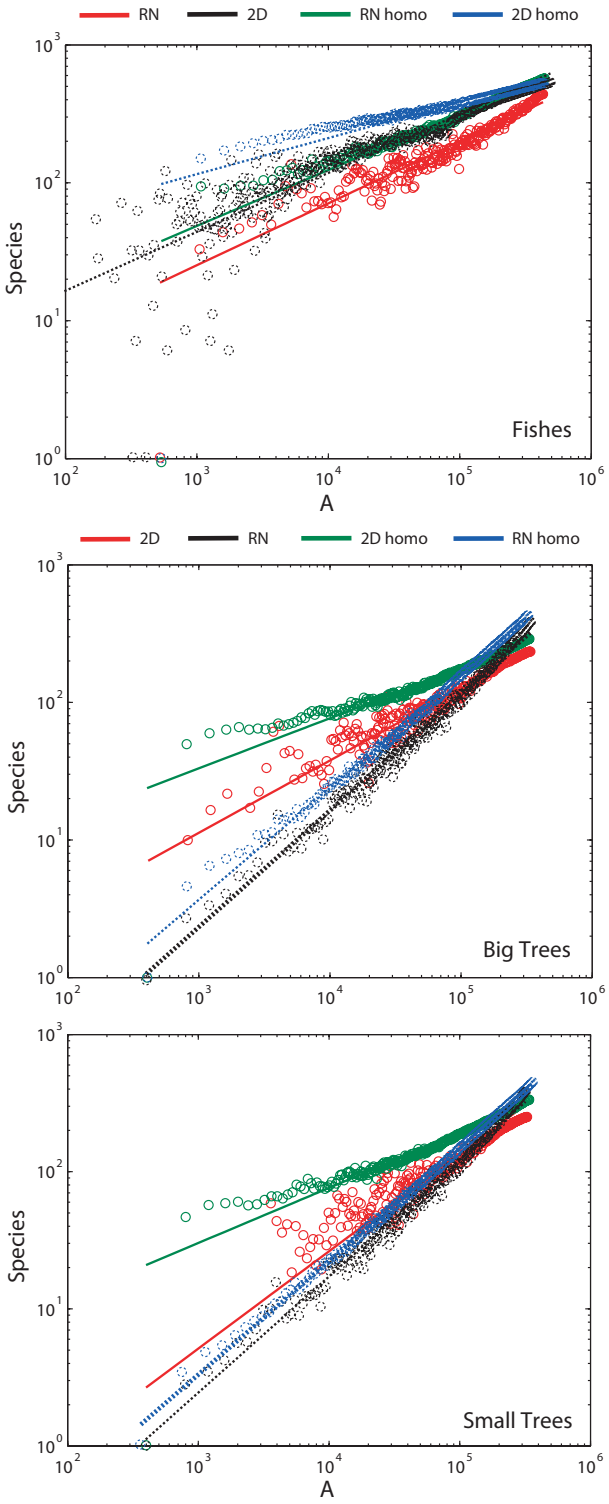

Figure 5.17: SAR of fishes and trees of the MMRS, heterogeneous, homogeneous and combined model results.

Table 5.3: Clustering Properties of the MMRS, data, heterogeneous and homogeneous model results.

Variables	ϵ	$\langle CS \rangle$	$z(A)$	$z(H^*)$
Fishes				
data	-0.93	6.27	0.45	0.35
model	-0.92	6.67	0.46	0.36
homogeneous	-0.88	7.00	0.40	0.40
Big Trees				
data	-0.49	14.03	0.53	0.40
model	-0.50	14.80	0.53	0.42
homogeneous	-0.53	26.04	0.36	0.36
Small Trees				
data	-0.67	9.56	0.72	0.42
model	-0.66	10.20	0.73	0.43
homogeneous	-0.55	27.16	0.40	0.40

Table 5.4: Clustering Properties of the MMRS, heterogeneous and homogeneous combined model results (fishes dispersal in 2-D and trees dispersal in RN.

Variables	ϵ	$\langle CS \rangle$	$z(A)$	$z(H^*)$
Fishes 2-D				
heterogeneous	-0.54	9.7	0.41	0.31
homogeneous	-0.33	9.35	0.26	0.26
Big Trees RN				
heterogeneous	-1.32	6.07	0.85	0.58
homogeneous	-1.80	5.04	0.81	0.81
Small Trees RN				
heterogeneous	-1.25	4.97	0.89	0.61
homogeneous	-2.28	2.98	0.88	0.88

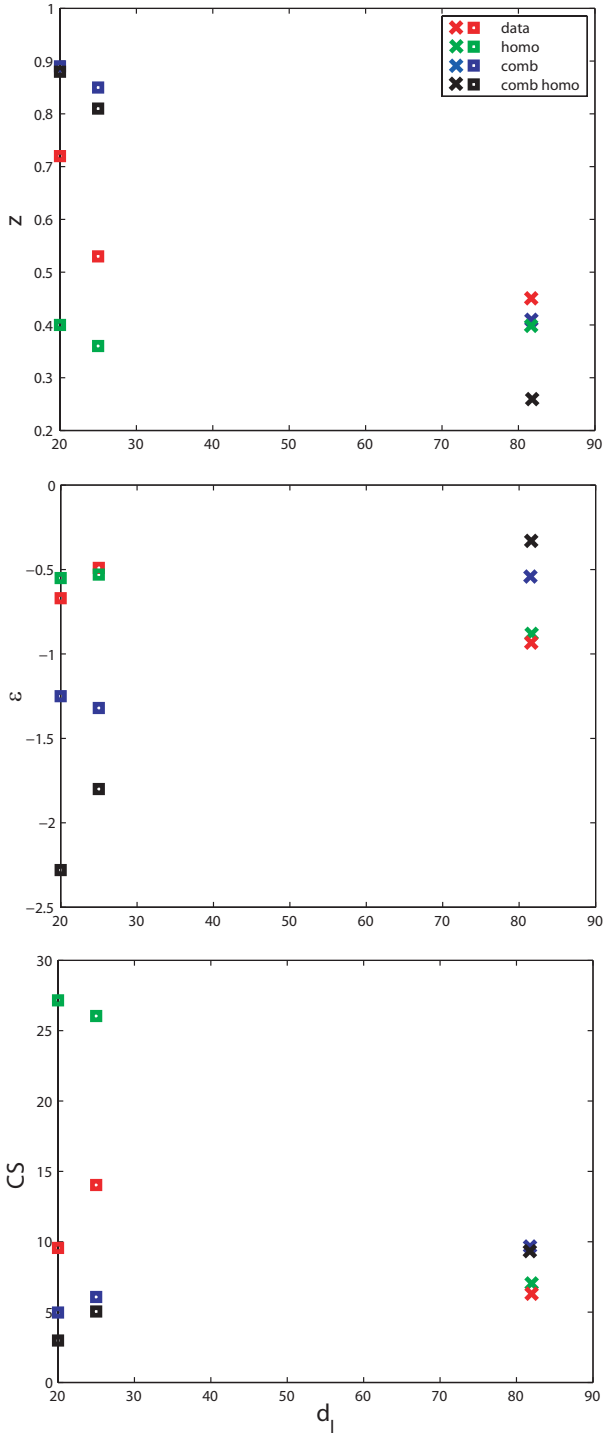

Figure 5.18: z, ϵ and $\langle CS \rangle$ in function of d_l, for the MMRS data and model results.

Table 5.5: Sensitivity of the clustering properties of fishes and trees in the MMRS changing topology and in the homogeneous conditions.

	F	BT	ST
Topology			
ϵ	$\uparrow$	$\downarrow$	$\downarrow$
$\langle CS \rangle$	$\uparrow$	$\downarrow$	$\downarrow$
z	$\downarrow$	$\uparrow$	$\uparrow$
$\langle LSR \rangle$	$\uparrow$	$\downarrow$	$\downarrow$
S_T	$\uparrow$	$\uparrow$	$\uparrow$
Homogeneous condition			
ϵ	$\uparrow$	$\approx\downarrow$	$\uparrow$
$\langle CS \rangle$	$\uparrow$	$\uparrow$	$\uparrow$
z	$\downarrow$	$\downarrow$	$\downarrow$
$\langle LSR \rangle$	$\uparrow$	$\uparrow$	$\uparrow$
S_T	$\uparrow$	$\uparrow$	$\uparrow$

In figures 5.19 and 5.20 we show the importance of the branching structure of the river network that is the cause of the spatial dissection of the ecological habitat in riverine ecosystems. The change in topology (RN to 2-D or viceversa) allows to study the effect of the *fragmentation* exerted by the river network using the same dispersal parameters. Here the term *spatial fragmentation* means a dissection of the ecosystem that is not necessarily related to a loss of individuals as mostly reported in the ecological literature, where *fragmentation* is associated to habitat-loss. Then here the meaning is only as a generic geometrical change of the geometry. The difference consists solely in the way in which the species spread in the ecosystem: along the network in the RN case, and everywhere in the space (Euclidean paths) in the 2-D case. Considering the 2-D case if the dispersal parameters are kept the same as the RN case, the dispersal is facilitated due to the lower diameter of the 2-D topology (see Figure 2.5, Chapter 2). For this reason in the following the 2-D case has been studied in which $d_{lRN}/\phi_{RN} = d_{l2D}/\phi_{2D}$, so the *dispersal ability* is the same in the fragmented and non-fragmented landscape, but still the geometry (2-D) is different than the RN case. In this way it is possible to detect also the importance of the branching structure of the river network i.e. the dendritic effect. The branching degree of a river network is for example captured by the fractal dimension of the rivulets (e.g. see Chapter 6 and *Rodriguez-Iturbe and Rinaldo* (1997)). We have performed the experiment only for the exponential kernel dispersal scenario (case a) Table 3.2 in Chapter 3). Then considering the actual values of the diameters and the equivalence above of the ecological ratios between topologies, the dispersal parameter of the 2-D case with an equivalent dispersal ability of the RN case is $d_{l2D\,eq} = 315$ for the choosen value of the model. In particular we have found the following relationships true:

- $z_{2D} < z_{2D\,eq} < z_{RN}$;
- $\epsilon_{2D} > \epsilon_{2D\,eq} > \epsilon_{RN}$;

- $\langle CS \rangle_{2D} > \langle CS \rangle_{2D\ eq} > \langle CS \rangle_{RN}$;
- $\langle LSR \rangle_{2D} > \langle LSR \rangle_{2D\ eq} > \langle LSR \rangle_{RN}$.

For the regional species richness (RSR) that is the number of species in an area A and the Jaccard similarity index (JSI) that is the similarity in local species richness (LSR) between local communities, the same inequality of $\langle LSR \rangle$ holds for the three cases analysed above (topologies RN, $2-D$, and $2-D_{eq}$ that is the $2-D$ topology with the same ecological ratio of the RN case), instead for the γ-diversity (i.e. the total number of species S_T) the inequality is like for z among cases. Table 5.6 reports ϵ, $\langle CS \rangle$, LSR, JSI and RSR for the Tanaro RN, 2-D and 2-D RN dispersal-equivalent case.

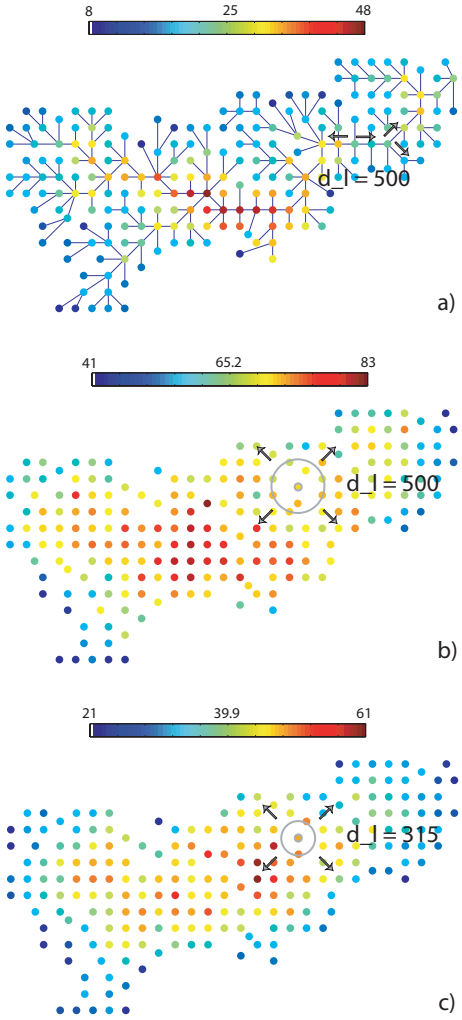

Figure 5.19: LSR pattern in the exponential kernel dispersal for: a) Tanaro RN case; b) Tanaro 2-D landscape; c) Tanaro 2-D equivalent case with the same $d_l/\langle\phi\rangle$ ratio of the RN case. The grain size is 1000 m.

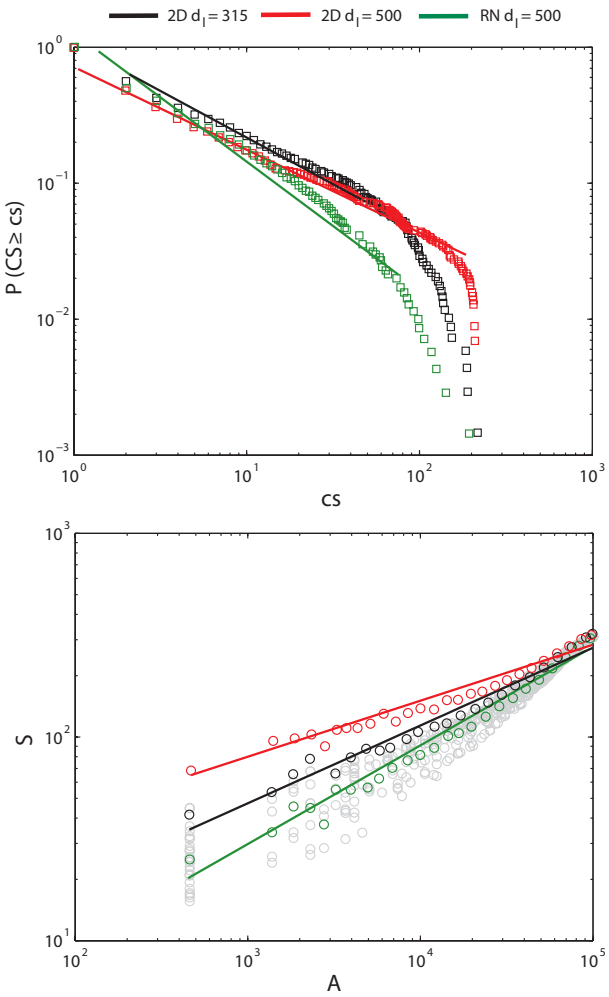

Figure 5.20: $P(CS \geq cs)$ for the Tanaro basin at $CG = 1000$ with exponential kernel dispersal in the network case (green), 2-D case with the same dispersal parameter (red) and in the 2-D case with the same $d_l/\langle\phi\rangle$ ratio of the RN case.

Table 5.6: ϵ, $\langle CS \rangle$, LSR, JSI and RSR for the Tanaro RN, 2-D and 2-D RN dispersal-equivalent case

	RN	$2D$	$2D_{eq}$
ϵ	-0.946	-0.637	-0.697
$\langle CS \rangle$	7.587	13.663	12.662
z	0.483	0.279	0.382
$\langle LSR \rangle$	25	65.23	39.91
LSR_{min}	8	41	21
LSR_{max}	48	83	61
$\langle JSI \rangle$	0.17	0.45	0.28
JSI_{min}	0.0026	0.25	0.0704
JSI_{max}	0.7977	0.7347	0.7279
$\langle RSR \rangle$	47.8	89.76	66.05
RSR_{min}	8	41	21
$RSR_{max} = S_T$	322	284	311

Here the effect of the shape of the ecosystem, that is dictated by the shape of its boundaries, is investigated in detail. The three basins used in the scale-invariance study in chapter 4, namely the optimal channel network (OCN), and the two real river basins Cordevole and Tanaro. Regardless the topology (RN or 2-D) the exponent z of the SAR is higher for the Tanaro than the Cordevole and the OCN. Also the Tanaro exhibits the lowest value of ϵ meaning that is less probable to find large clusters in elongated basins. This is also proofed by the lowest value of $\langle CS \rangle$ for the Tanaro. Cordevole and OCN present an elongation ratio ratio taht is higher than the Tanaro. z for the Cordevole is higher than for the OCN. The differences in the clustering patterns of the river basins analysed is imputable to the elongation of the river network that is higher in the Tanaro than for the Cordevole. The OCN that is square has null elongation, that is elongation ratio ratio equal to one equivalently. Further investigations will explore deeper the elongation effect on the biodiversity patterns adopting OCNs with different aspect ratio and plotting the biodiversity and clustering patterns in function of the elongation ratio. The total number of species, S_T or γ-diversity, seems to be not affected by the shape of the river basin nor by the topology (RN or 2-D). The LSR assumes the largest value in the OCN since it has the smallest dimension among the compared basins that means a lower number of local communities which on average have an higher number of individuals. Then a proper normalization based on the total extension of the basin would make fair the comparison between the networks in study. Table 5.7 reports the results of the analysis of the clustering features (ϵ, $\langle CS \rangle$, z) and of some macroecological variables (S_T, LSR, JSI, RSR) in function of the river network: O, C and T are OCN, Cordevole and Tanaro respectively.

Table 5.7: Analysis of the clustering features and of the macroecological variables in function of the river network: O, C and T are OCN, Cordevole and Tanaro respectively.

	RN	2D
ϵ	$T \ll C \ll O$	$T \ll C \ll O$
$\langle CS \rangle$	$T \ll C \lesssim O$	$T \ll C \lesssim O$
z	$T \gg C \gg O$	$T \gg C \gg O$
S_T	$T \approx C \approx O$	$T \approx C \approx O$
LSR	$T \gg C < O$	$T \gg C < O$
JSI	$T \gtrsim C < O$	$T \gtrsim C < O$
RSR	$T \gtrsim C < O$	$T \gtrsim C < O$

5.6 Effect of the Dispersal

In this section the role of the dispersal is further analyzed on shaping the spatial distribution of species that affects all the biodiversity patterns. The clustering patterns are particularly considered in order to detect, possibly, some transitions between clumped, random and uniform distributions of species in the ecosystems analysed.

5.6.1 Global and Nearest Neighbor dispersal

The results presented here are obtained using the model described in section 3.2.2 of chapter 3 for the RN and 2-D topologies in the Tanaro boundaries. The global and the nearest-neighbor dispersal are implemented. The global dispersal (indicated with G) case represents the extreme case in which all the local communities tend to be saturated because the species can travel in the whole ecosystem, then for this reason the global dispersal scenario is referred as the "habitat capacity limitation" case. The nearest neighbor dispersal case (NN) is not the extreme case of the "dispersal limitation" scenario since the the mean dispersal length is equal to the grain-size that can be a very large value. Instead the dispersal parameter can be smaller than the grain-size representing "local dispersal" scenarios as for example shown in chapter 4. Basically the nearest neighbor dispersal case can potentially reproduce a local dispersal situation but not surely and extreme case of saturation for limited dispersal in which species do not move from their habitat capacity. The definition of "local dispersal" then need more specification because it has to be related to the scale and resolution of the ecosystem studied. It is important to notice that both the G and NN dispersal kernel cases do not depend on the dispersal parameter d_l (equations 3.8 and 3.9 in chapter 3).

Figure 5.21 shows $P(CS \geq cs)$ and the SAR for the the global and nearest-neighbor dispersal cases. The differences arises between the RN and 2-D topologies in both the dispersal cases. The power-law of the cluster-size is well developed in the RN global dispersal case and in the 2-D nearest neighbor dispersal case. $P(CS \geq cs)$ in the 2-D global dispersal case has a limited scaling range because an habitat-capacity limitation effect since the diameter of the Tanaro 2-D landscape is much smaller than of the RN topology counterpart; on the other side $P(CS \geq cs)$ in the RN nearest-neighbor dispersal

case is well-developed without suffering any dispersal limitation effect and the number of species S_T tends to be equal to the number of local communities. The species-area relationship for the global dispersal case is substantially the same for the RN and 2-D topologies meaning the same spatial organization of species. Then in the global dispersal scenario the LSR patterns are expected to be very similar and the difference is on the number of clusters of a given dimension CS, that leads to a different power-law of the cluster size. In the NN dispersal case z for the RN topology is larger than for the 2-D case meaning a more homogeneous pattern and larger LSR for the latter topology. Figure 5.22 shows the occupancy-rank and rank-abundance curves for the global and the nearest neighbors scenarios for the Tanaro basin in the RN and 2-D topology. The occupancy-rank is concave downward in both topologies for the global dispersal meaning the same spatial organization and many common species (the initial plateau), while it is definitely concave upward for the RN NN dispersal case meaning few common species. The rank-abundance curve instead does not vary so much among cases.

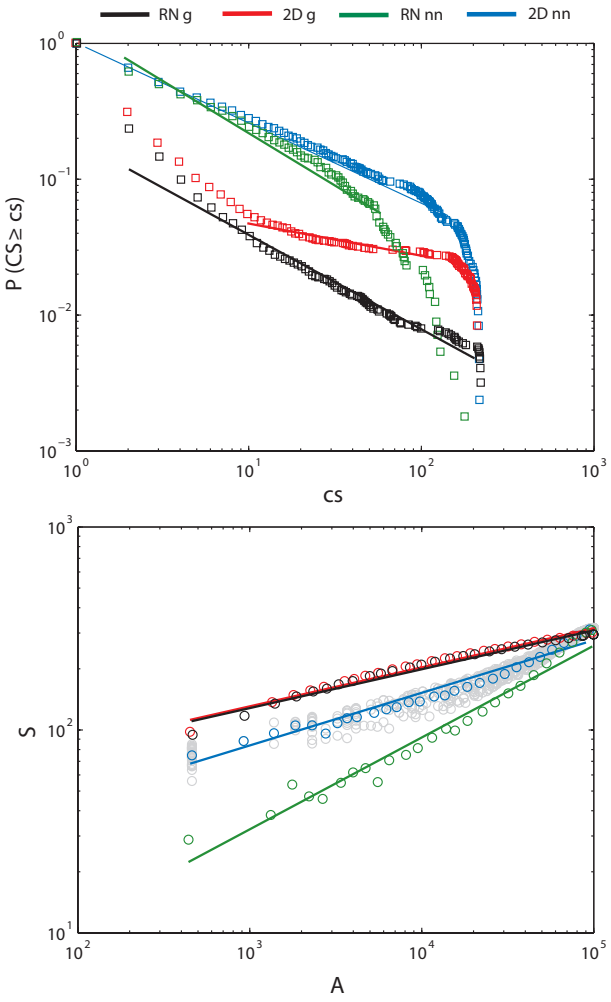

Figure 5.21: $P(CS \geq cs)$ and SAR with the Global (g) and Nearest Neighbor (nn) kernel dispersal for the Tanaro river network at $CG = 1000$

Figure 5.23 reports the JSI for each case analysed. It is evident how the average distance between local communities (d is the distance between local communities) is larger in the RN topolgy. The similarity between local communities is larger in the 2-D than in the RN case both for the NN and G dispersal case. The JSI for the global dispersal case is basically constant for every d but the average value is higher for the 2-D topology than for the RN, while in the nearest-neighbor dispersal case the decay of the JSI is much faster for the RN case. Figure 5.24 shows the pdfs of LSR, JSI and RSR in the NN and G dispersal cases. It appears that the differences between the RN and the 2-D landscape topologies are evident only in the nearest-neighbor case in which the 2-D landscape presents higher LSR, JSI and RSR. LSR, JSI and RSR are also higher on average in the global dispersal case than in the nearest-neighbor case but the range of values is much more limited. The spatial patterns of LSR are reproduced in Figure 5.25. The RN and 2-D topologies are undistinguishable in the global dispersal case, instead in the NN dispersal case the LSR assumes high values along the main network (like to reproduce riparian vegetation) in the RN topology, and high values in the middle of the landscape in the 2-D topology. For the 2-D topology the decay of LSR from the local community with the lowest diameter decays circularly as a consequence of the Euclidean dispersal in all the directions. Table 5.8 reports the numerical values of the clustering features and some biodiversity properties of the G and NN dispersal cases investigated.

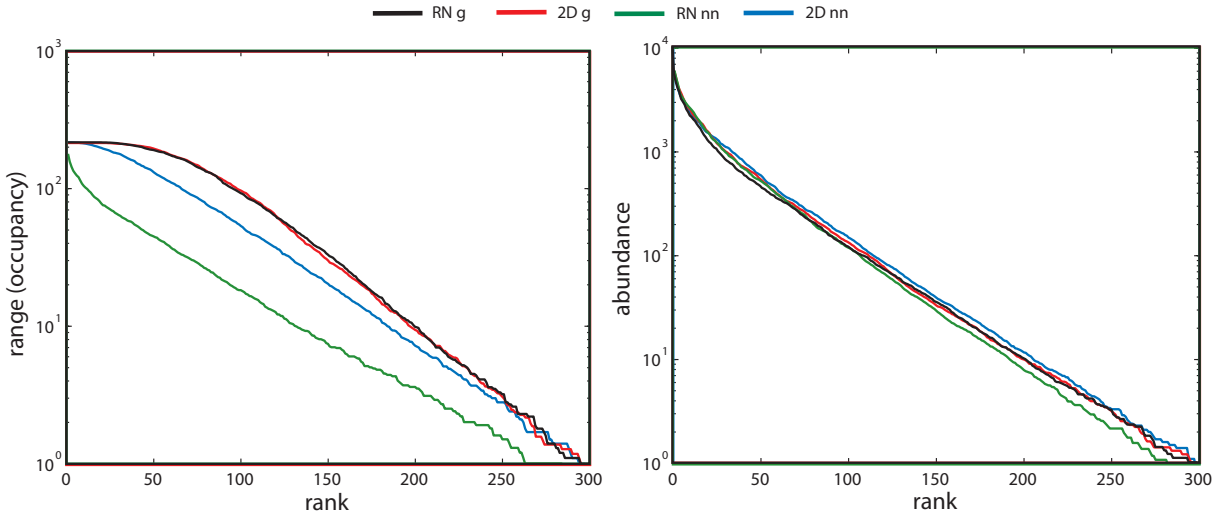

Figure 5.22: Rank-Occupancy and Rank-Abundance curves for the global and the nearest neighbors models in the Tanaro river basin.

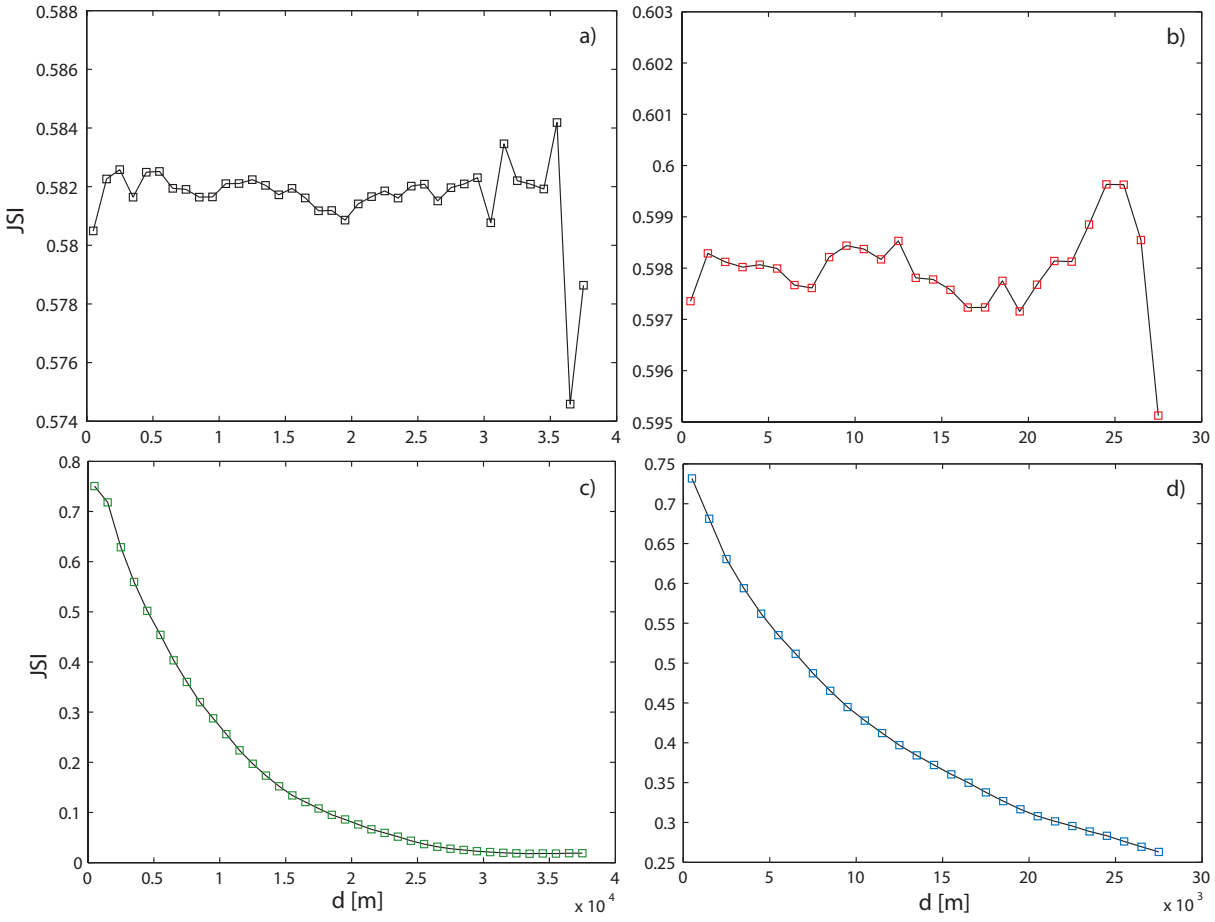

Figure 5.23: JSI for the global and nearest neighbor kernel dispersal simulations of the Tanaro river network. a) RN global; b) 2-D global; c) RN nearest neighbors; d) 2-D nearest neighbors.

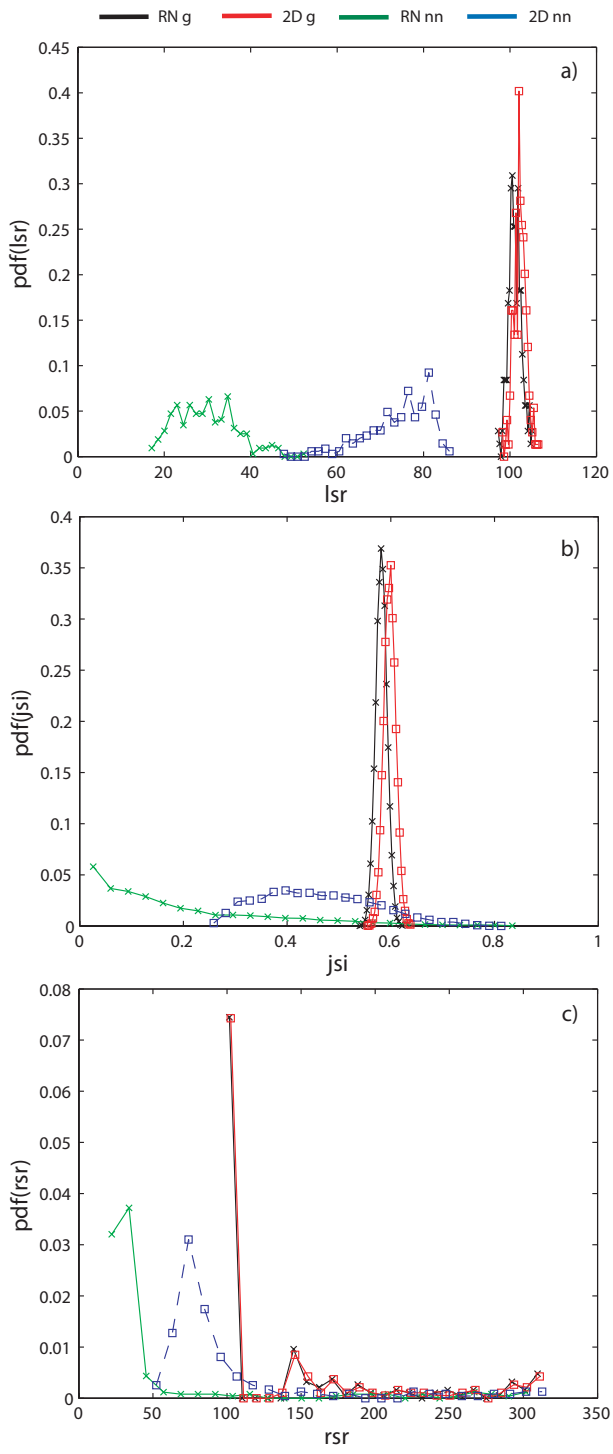

Figure 5.24: pdfs of LSR, JSI and RSR for the global and nearest neighbor model for the Tanaro river basin and associated 2-D landscape

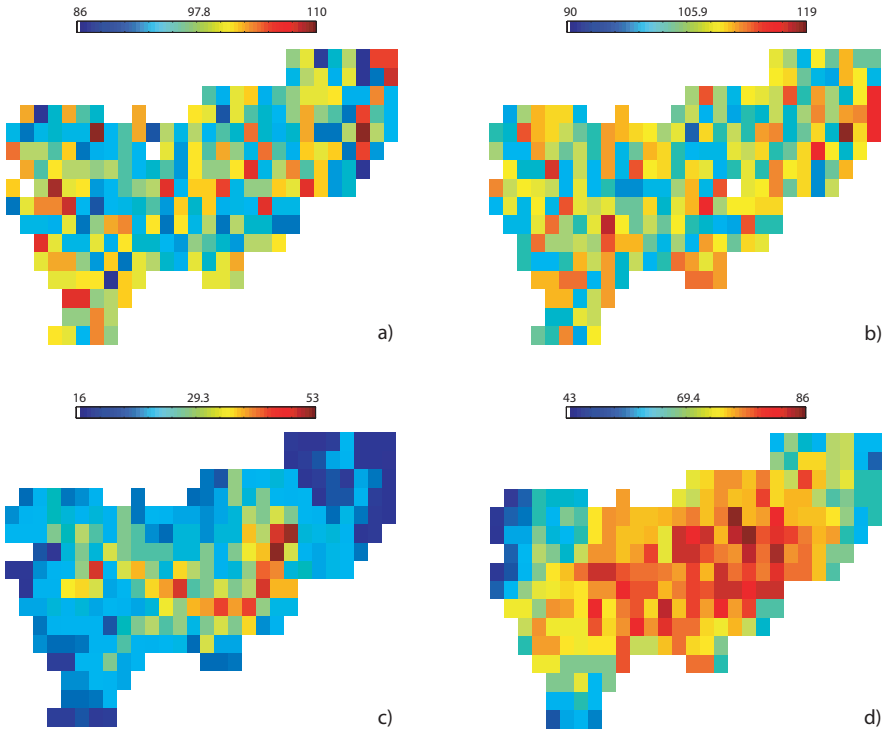

Figure 5.25: LSR patterns for Global and Nearest Neighbor Dispersal experiments for the Tanaro river network. a) RN global; b) 2-D global; c) RN nearest neighbors; d) 2-D nearest neighbors. Not the same color range for the patterns.

Table 5.8: Biodiversity and clustering features of RN and 2-D landscapes in the global and nearest neighbor kernel dispersal

	Global		Nearest Neighbor	
Topology	2D	RN	2D	RN
ϵ	-0.24	-0.70	-0.59	-0.80
$\langle CS \rangle$	7.64	3.78	20.52	11.03
z	0.19	0.19	0.26	0.46
LSR_{max}	119	110	86	53
LSR_{min}	90	86	43	16
$\langle LSR \rangle$	106	98	69	35

5.6.2 On the Role of the Mean Dispersal Length

In this section is investigated the behavior of the biodiversity pattern in function of the main dispersal parameter d_l. Figure 5.26, 5.27, 5.28, 5.29, 5.30, 5.31, and 5.32 report the biodiversity patterns of the power-law of the cluster-size, the species-area relationship, the range-rank curve, the rank-abundance curve, the pdf of the LSR, JSI and RSR respectively, for the ecological scenario a and b (see chapter 3 table 3.2) in the Tanaro basin at coarse-graining level $CG = 1000$ varying the dispersal parameter d_l in a wide range of values. Table 5.33 reports the qualitative behavior of clustering features and some biodiversity patterns in function of the dispersal parameter d_l, the fat-tail Cauchy dispersal parameter b, and the biodiversity parameter θ. In appendix A the same biodiversity patterns here studied are reported for the case with exponential-Cauchy kernel dispersal and biodiversity parameter $\theta = 80$. Table 5.9 compares the clustering features, ϵ, $\langle CS \rangle$, and z between the exponential and exponential-Cauchy kernel dispersal scenarios. $P(CS \geq cs)$ of the river network has a more straight/concave downward trend than $P(CS \geq cs)$ of the 2-D landscape considering the same dispersal scenarios. Also $P(CS \geq cs)$ for the 2-D landscape presents in general an higher cutoff so the scaling region is smaller than in the RN case, especially in the exponential-Cauchy case due to the habitat-capacity limitation phenomena. The exponential distribution of the power-law of the cluster-size is observed only for very small values of dispersal length in the ecological scenario a both for $\theta = 40$ and $\theta = 80$. It is also important to note that for values of dispersal parameter greater than the grain size for which the ecosystem is modeled (in this case 1000) there is a change in the power-law behavior of the cluster-size distribution (reddish lines in figure 5.26). This abrupt change also happens for other biodiversity patterns. The SAR in figure 5.27 seems to be a more robust patterns than $P(CS \geq cs)$. z decreases for increasing d_l, and for values of the dispersal parameter larger than the grain-size of the simulated ecosystem it becomes quite stable around small values not found in reality. In Figures 5.10 and 5.11 the values of z (of the SAR of figure 5.27) and ϵ (of $P(CS \geq cs)$ in figure 5.26) in function of d_l are reported. The range for z is between 1 when $d_l = 1$ that means basically no dispersal (in this case S_T is equal to the number of local communities and a perfect linear 1:1 correspondence between abundance and occupancy occurs) and 0.20 when d_l is very large compared

to the grain size (e.g. here for $d_l = 20,000\ m$). The range of the common z-values observed in nature is between 0.1 and 0.6. In figures 5.28 and 5.29 the range-rank and the abundance-rank patterns are analysed. For the RN, S_T that is the maximum range, generally decreases while increasing d_l, on the contrary for the 2-D landscape. For low values of d_l (below the grain-size) the shape of the range rank is concave upward while the abundance-rank presents a plateau meaning that there is a multitude of species with the same abundance. For values of d_l greater than the grain-size the range-rank curve is concave downward and the abundance-rank assumes the common shape with the upward concavity implying that most species are rare (phenomena called rarification in *Muneepeerakul* (2007)). Figure 5.30, 5.31, and 5.32 show the pdf of LSR, JSI and RSR respectively. All these patterns present a lager mean value while increasing d_l and there is an abrupt jump of the pdfs for dispersal values larger than the grain-size. This happens also to $P(CS \geq cs)$. The pdf of the JSI is well fitted by a gaussian function with the only exception for low values of the dispersal in the exponential dispersal case in which pdf(JSI) assumes the shape of a gamma distribution approximatively. Actually for extremely local values of mean dispersal length the shape of the pdfs of the JSI looks like a Pareto distribution as the pdf of the RSR independently of d_l.

Finally figures 5.33 and 5.34 report S_T (γ-diversity), and the average values of LSR, JSI and RSR, in function of d_l for the RN and 2-D topologies in the exponential kernel dispersal case (table 3.2 in chapter 3). In table 5.11 there is a final comparison of the clustering and other macroecological variables until here cosnidered between the RN and 2-D landscapes. The results hold for both the dispersal kernels used.

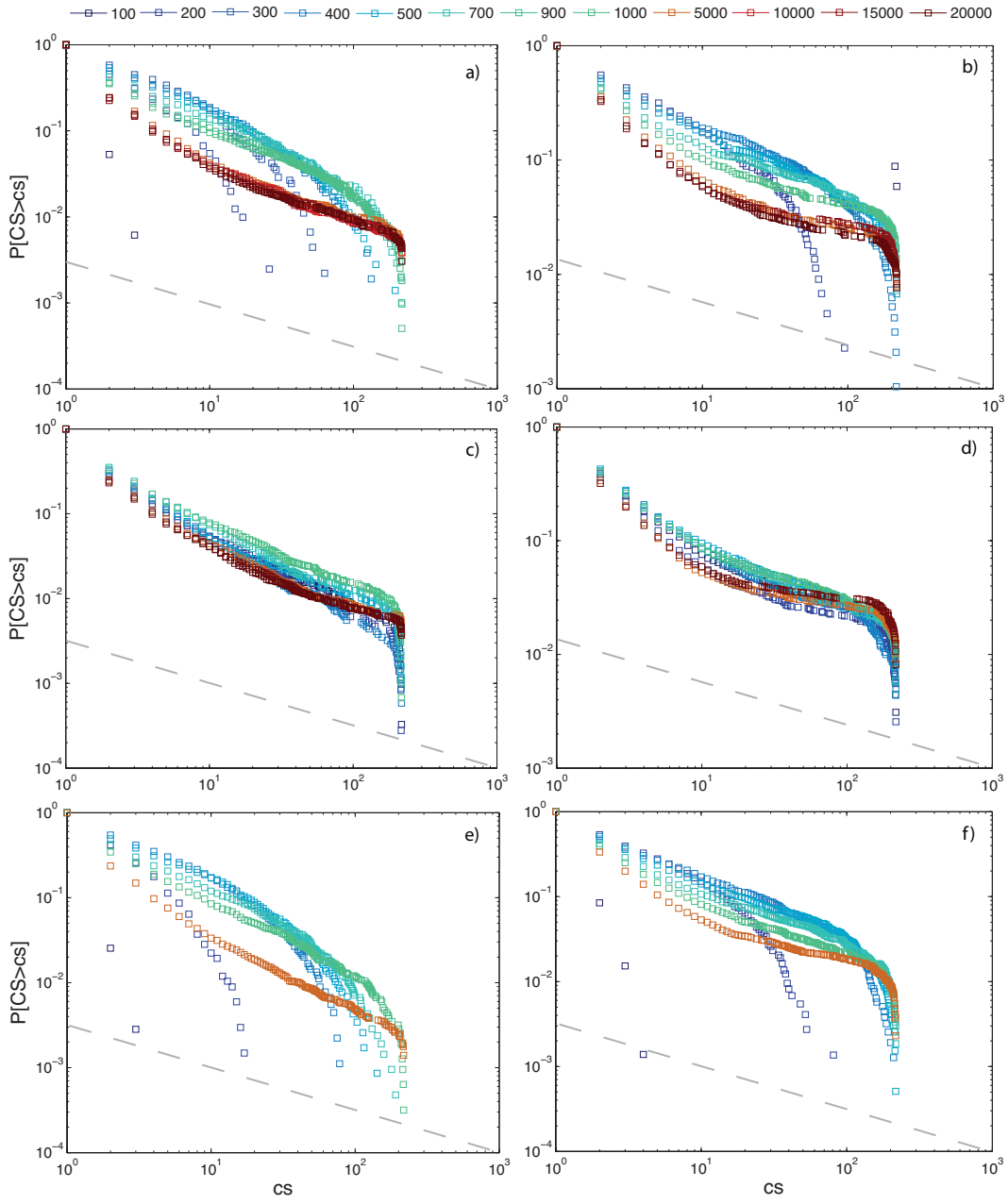

Figure 5.26: Numerical results of $P(CS \geq c)$ for the Tanaro river basin for the exponential and exponential-Cauchy dispersal increasing d_l (see legend): a) RN exp., b) 2-D exp., c) RN exp-Cauchy, d) 2-D exp-Cauchy, e) RN exp. $\theta = 80$, f) 2-D exp. $\theta = 80$. The slope of the grey line is -0.5.

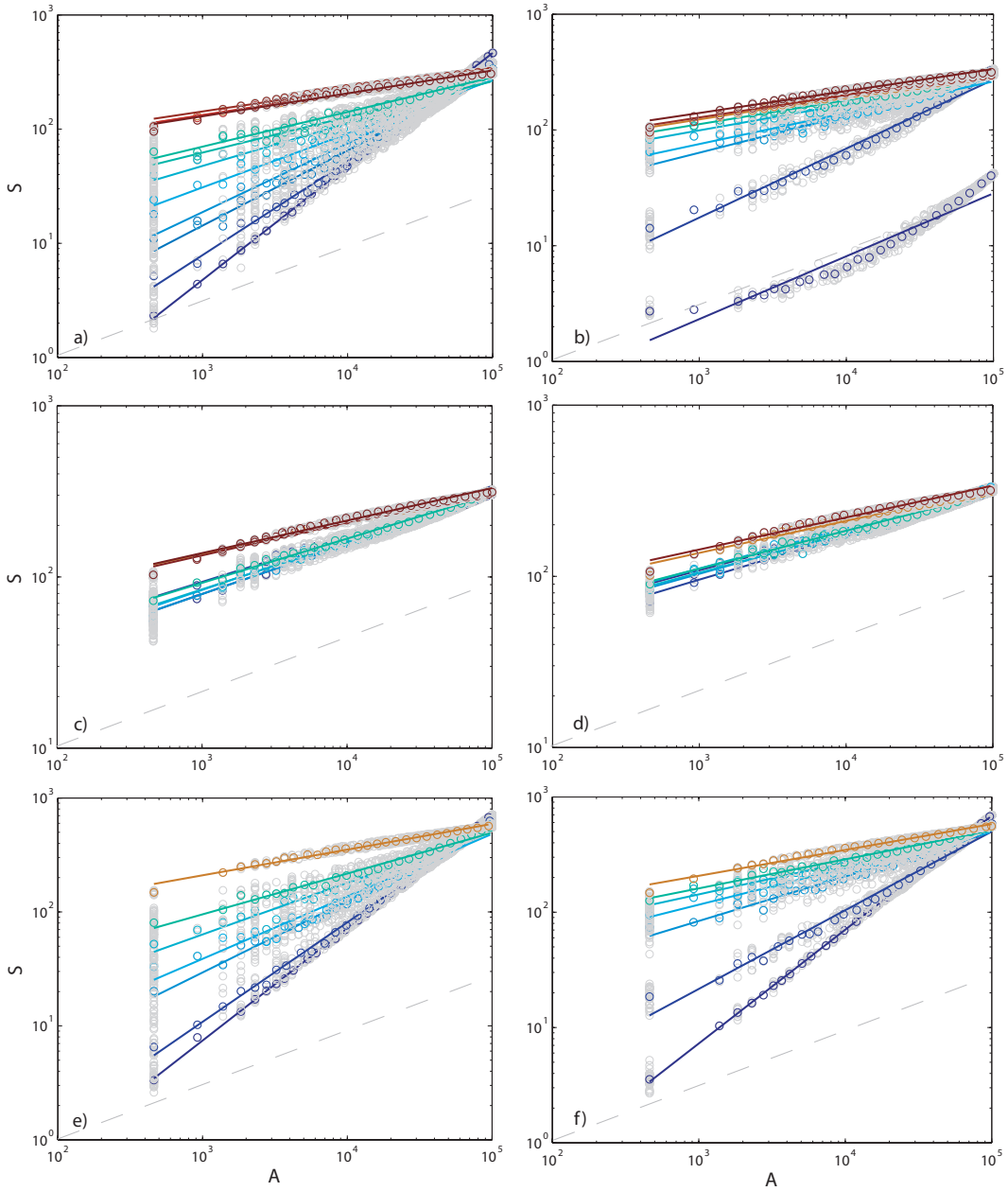

Figure 5.27: Numerical results of the SAR for the Tanaro river basin for the exponential and exponential-Cauchy dispersal increasing d_l (see legend): a) RN exp., b) 2-D exp., c) RN exp-Cauchy, d) 2-D exp-Cauchy, e) RN exp. $\theta = 80$, f) 2-D exp. $\theta = 80$. The slope of the grey line is -0.5.

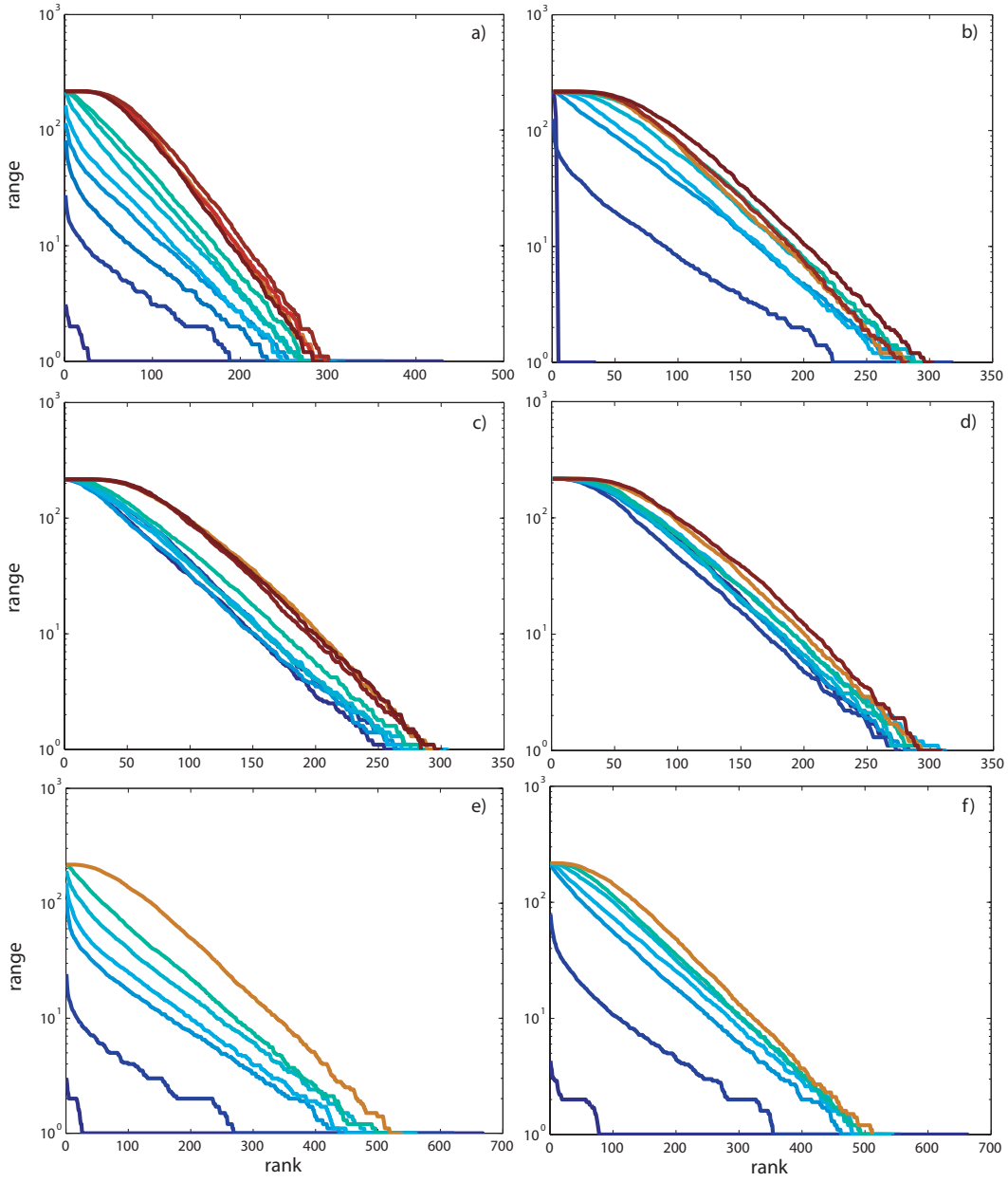

Figure 5.28: Numerical results of the Range-Rank for the Tanaro river basin for the exponential and exponential-Cauchy dispersal increasing d_l (see legend): a) RN exp., b) 2-D exp., c) RN exp-Cauchy, d) 2-D exp-Cauchy, e) RN exp. $\theta = 80$, f) 2-D exp. $\theta = 80$.

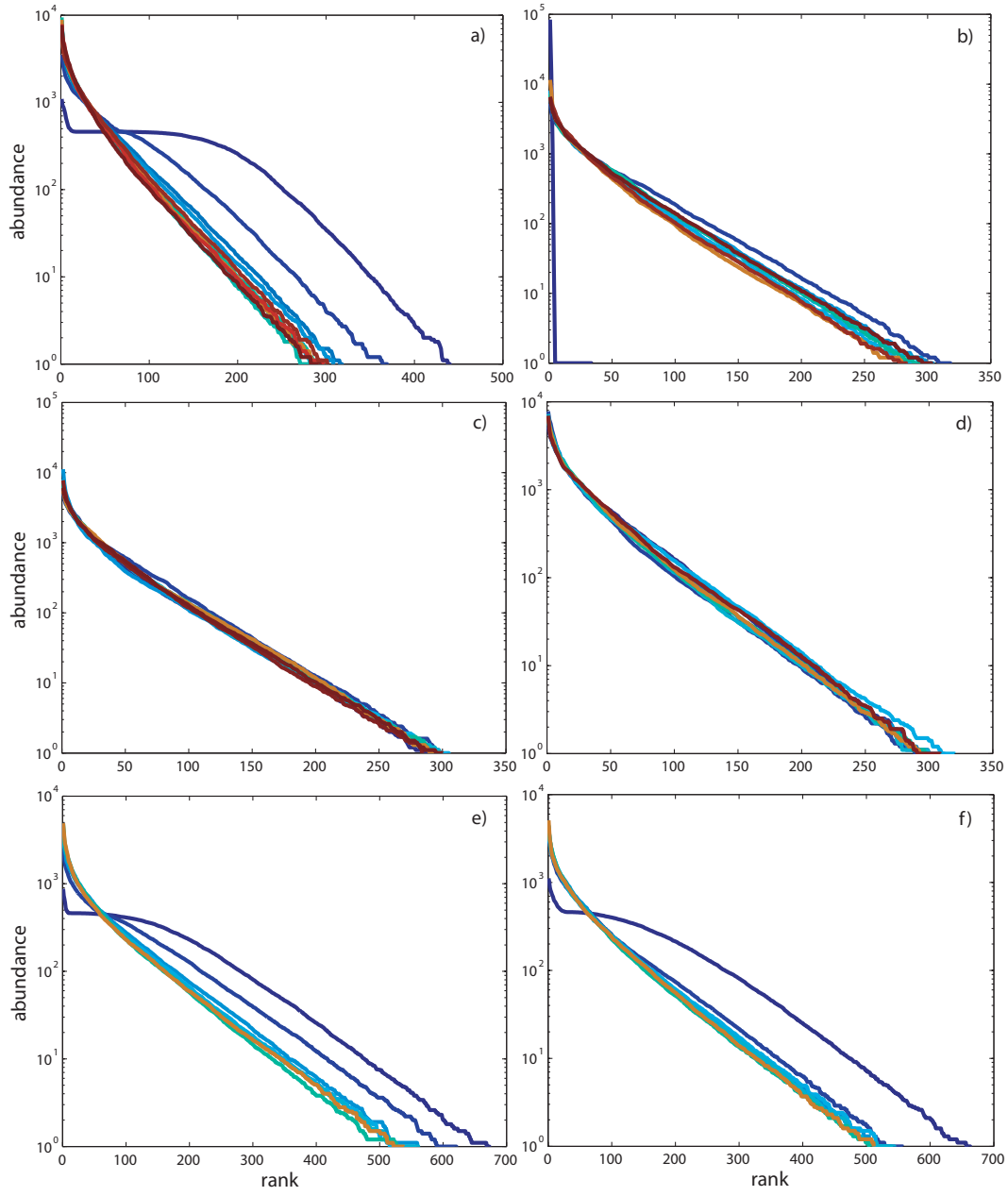

Figure 5.29: Numerical results of the Abundance-Rank for the Tanaro river basin for the exponential and exponential-Cauchy dispersal increasing d_l (see legend): a) RN exp., b) 2-D exp., c) RN exp-Cauchy, d) 2-D exp-Cauchy, e) RN exp. $\theta = 80$, f) 2-D exp. $\theta = 80$. The slope of the grey line is -0.5.

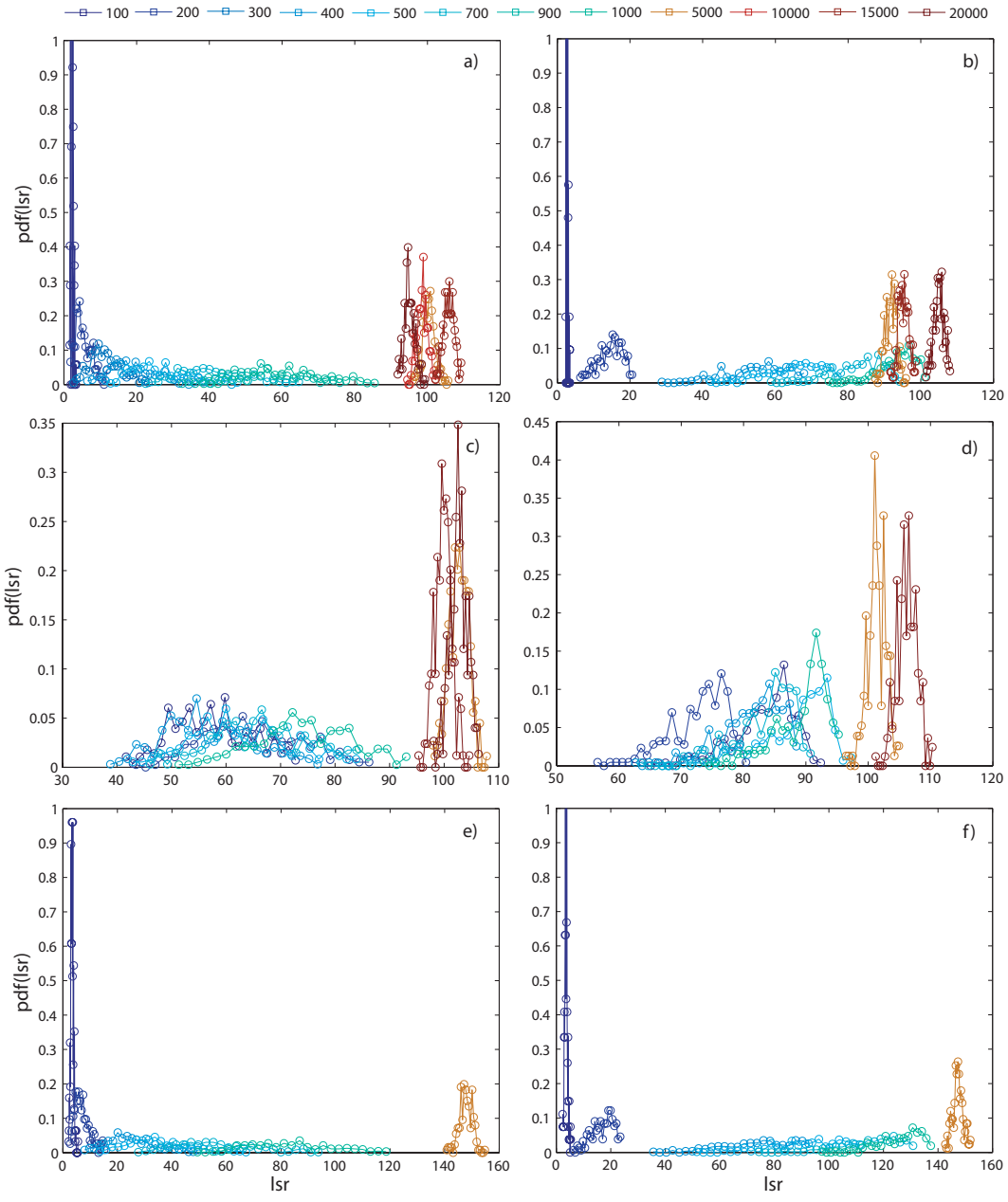

Figure 5.30: Numerical results of the pdf(lsr) for the Tanaro river basin for the exponential and exponential-Cauchy dispersal increasing d_l (see legend): a) RN exp., b) 2-D exp., c) RN exp-Cauchy, d) 2-D exp-Cauchy, e) RN exp. $\theta = 80$, f) 2-D exp. $\theta = 80$.

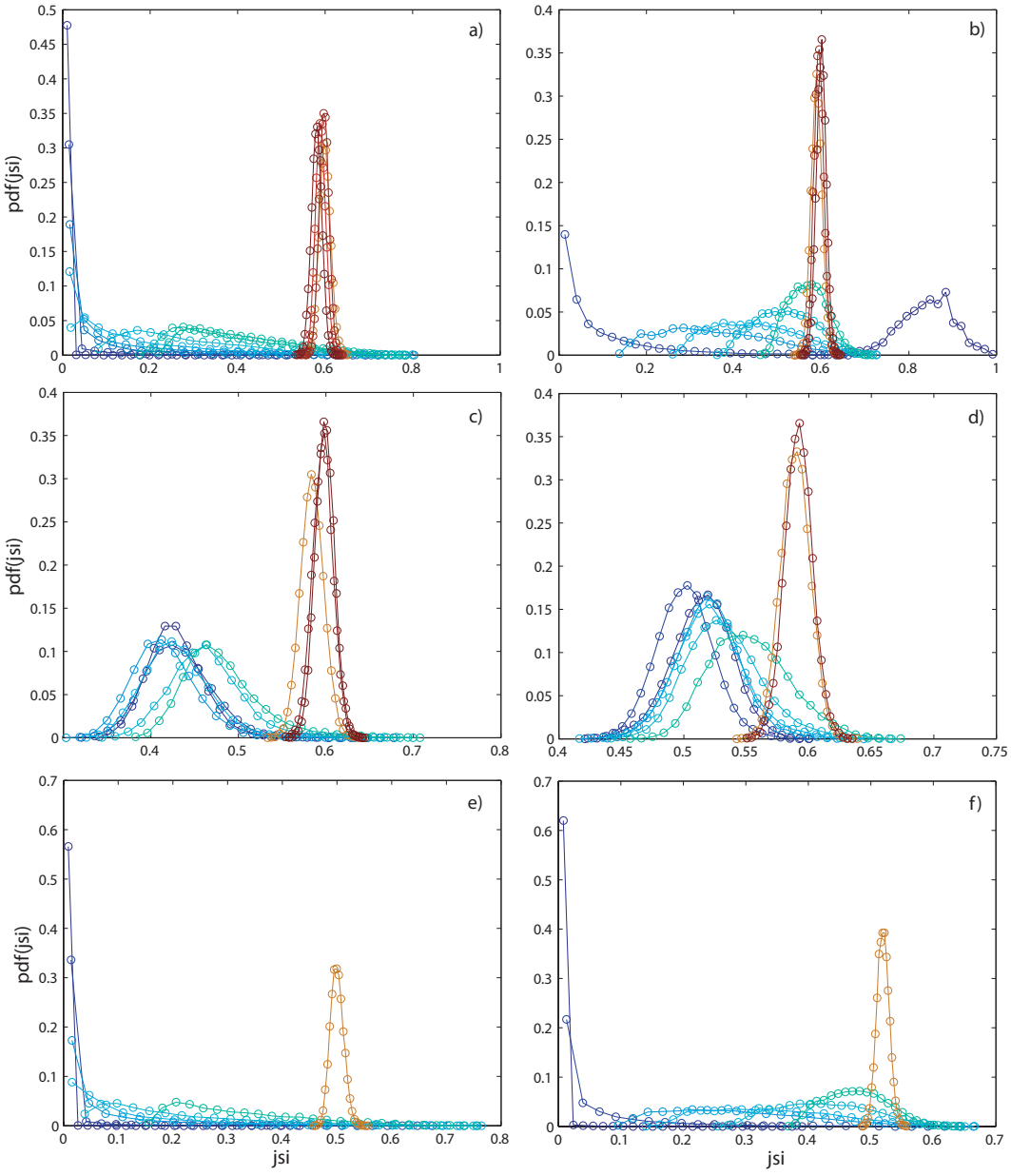

Figure 5.31: Numerical results of the pdf(jsi) for the Tanaro river basin for the exponential and exponential-Cauchy dispersal increasing d_l (see legend): a) RN exp., b) 2-D exp., c) RN exp-Cauchy, d) 2-D exp-Cauchy, e) RN exp. $\theta = 80$, f) 2-D exp. $\theta = 80$.

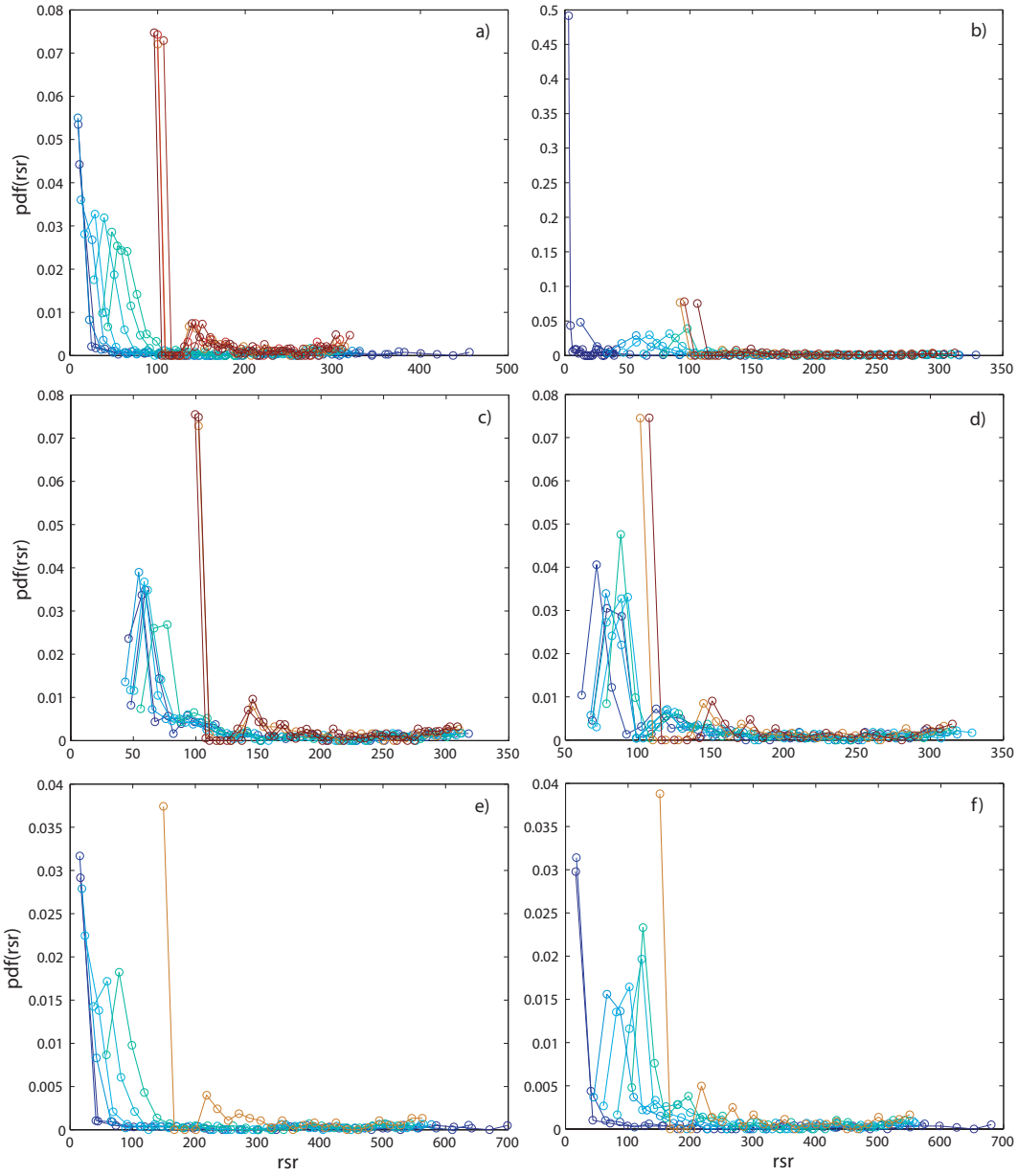

Figure 5.32: Numerical results of the pdf(rsr) for the Tanaro river basin for the exponential and exponential-Cauchy dispersal increasing d_l (see legend): a) RN exp., b) 2-D exp., c) RN exp-Cauchy, d) 2-D exp-Cauchy, e) RN exp. $\theta = 80$, f) 2-D exp. $\theta = 80$.

Table 5.9: Comparison of the clustering features in the the RN and 2-D landscapes in the exponential and exponential-Cauchy kernel dispersal

	RN	2D
ϵ	$b \leqslant a$	$b \geqslant a$
$\langle CS \rangle$	$b \lesssim a$	$b \geqslant a$
z	$b \leqslant a$	$b \leqslant a$

Table 5.10: Analysis of the clustering and macroecological properties increasing the dispersal parameter and speciation. The results are mainly based on the simulation on the Tanaro river network.

	RN	2D
Parameter d_l		
ϵ	$\uparrow$	$\uparrow\uparrow$
$\langle CS \rangle$	$\approx\uparrow$	$\approx\uparrow$
z	$\downarrow$	$\downarrow$
S_T	$\downarrow$	$\downarrow$
$LSR\ JSI\ RSR$	$\uparrow\uparrow\uparrow$	$\uparrow\uparrow\uparrow$
Parameter b		
ϵ	$\downarrow$	$\uparrow$
$\langle CS \rangle$	$\approx\downarrow$	$\approx\uparrow$
z	$\approx\downarrow$	$\approx\approx\downarrow$
S_T	$\approx\uparrow$	$\approx\uparrow$
$LSR\ JSI\ RSR$	$\uparrow\uparrow\uparrow$	$\uparrow\uparrow\uparrow$
Parameter θ		
ϵ	$\downarrow$	$\downarrow$
$\langle CS \rangle$	$\downarrow$	$\downarrow$
z	$\uparrow$	$\uparrow$
S_T	$\uparrow$	$\uparrow$
$LSR\ JSI\ RSR$	$\uparrow\downarrow\uparrow$	$\uparrow\downarrow\uparrow$

Table 5.11: Comparison of the clustering and macroecological variables between the RN and 2-D landscapes. The results hold for both the dispersal kernels used.

	Topology
ϵ	RN $\preccurlyeq$ 2D
$\langle CS \rangle$	RN $\preccurlyeq$ 2D
z	RN $\gtrsim$ 2D
LSR	RN $\approx\preccurlyeq$ 2D [1]
JSI	RN $\approx\preccurlyeq$ 2D
RSR	RN $\approx\preccurlyeq$ 2D [2]
S_T	RN $\succcurlyeq$ 2D [2]

[1] RN $\approx\preccurlyeq$ 2D is more evident in the exponential-Cauchy kernel dispersal case

[2] S_T is lower for the RN than the 2-D only in the nearest neighbor dispersal case

5.7 Interrelationships between Taxa

In this section some conclusions about the validity of the dispersal values used are traced. The relationship between the LSR of trees and fishes has been also analysed. Can these relationships tell us about the interactions between organisms of different taxa? Also combining to the data of fishes and trees of the MMRS other data from literature possible allometric scaling relationships between dispersal length, mass, lifespan of species and the slope z of the species-area relationship have been investigated. Figure 5.35 shows the relationship between the LSR of fishes and trees of the MMRS. The relationship is quite strong between trees (slope = 0.77) and also quite surprisingly it holds between fishes and big trees, meaning that despite the different topologies in the dispersal behavior of different taxa and their sensitivity to different environmental stressors, the pattern of LSR is not completely different. In table 5.36 there is a collection of data, i.e. mean dispersal length d_l, mass m, lifespan T and exponent z of the SAR for different taxa (see references in the caption of the table), including also those of fishes and trees of the MMRS. The motivation of this analysis is to put in evidence how large is the range of dispersal values among taxa, in part to justify the large range of d_l used in the simulations. As second motivation, not less important than the former, even though the dispersal parameter d_l used in the metacommunity model is not the real average dispersal value that fishes and seeds of trees adopt in their dispersal (see chapter 3), it is interesting to compare the values used in the simulations with the values of other taxa trying to establish allometric relationships. It has been mentioned that species tend to

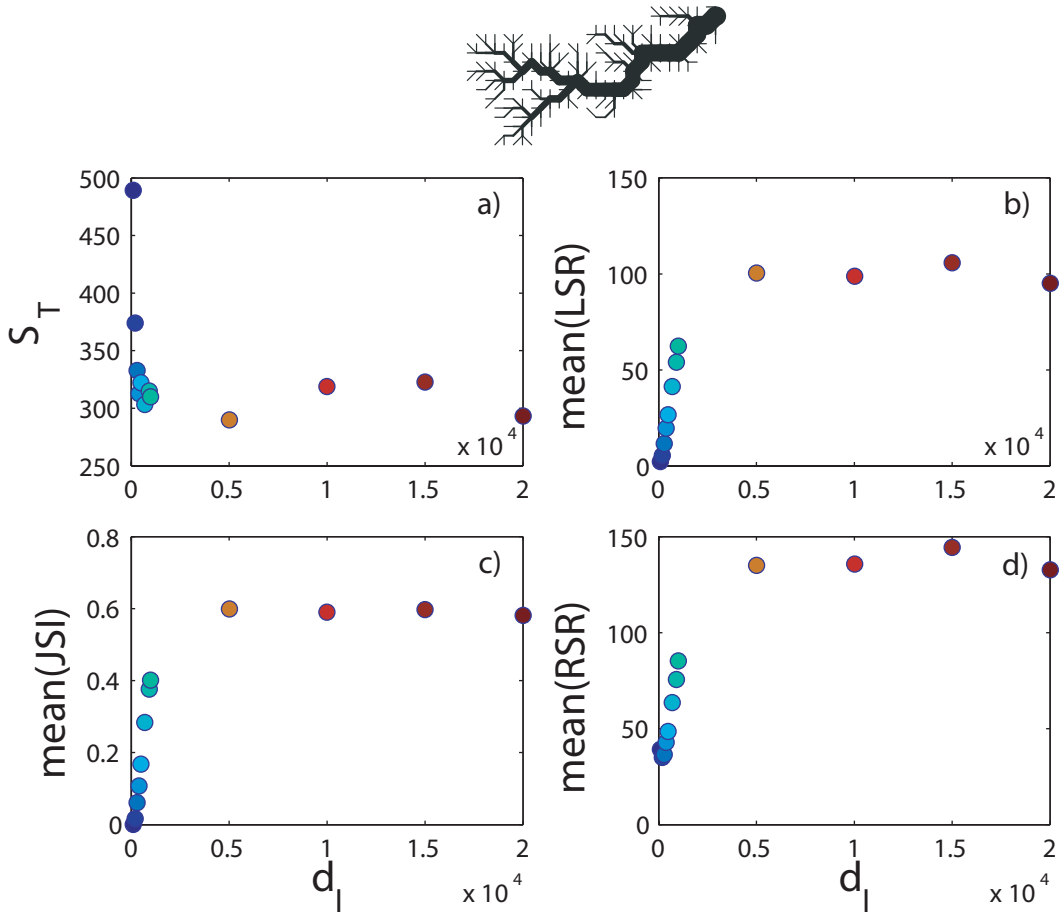

Figure 5.33: Behavior of S_T, $\langle LSR \rangle$, $\langle JSI \rangle$, and $\langle RSR \rangle$ in function of d_l for the Tanaro RN case with exponential kernel dispersal (case a) table 3.2, chapter 3).

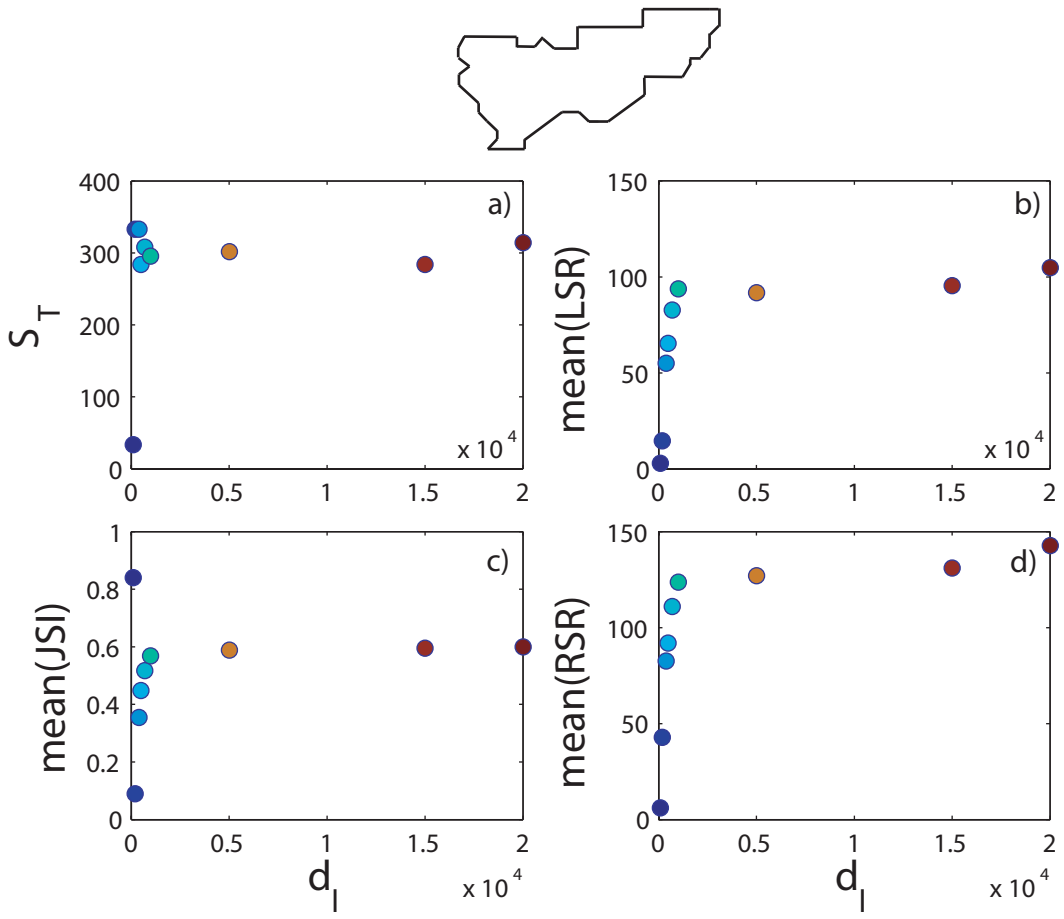

Figure 5.34: Behavior of S_T, $\langle LSR \rangle$, $\langle JSI \rangle$, and $\langle RSR \rangle$ in function of d_l for the Tanaro 2-D case with exponential kernel dispersal (case a) table 3.2, chapter 3).

adopt long-distance dispersal strategies. Therefore long-dispersal should be defined in function of the size of the organism (that is a surrogate of the mass or body-size) rather than compared to the size of the ecosystem that varies also in function of the study performed. Since size should scale like $length^3$ there might be other factors at play in the allometric scaling unless the scaling exponent is equal to this factor. Thinking about the long-distance dispersal it is ensured by a dispersal kernel that is "fat tailed". The dispersal kernel is the probability density of moving a particular distance from your starting point and "fat tailed" means that the long distance tails of the distribution decay slowly such that the variance of the dispersal kernel does not converge and is therefore undefined (or infinite). This means that mathematically at least, is possible to have long distance dispersal events, but still the mean dispersal distance could be very small. In this case long distance dispersal is defined based on the overall shape of the dispersal kernel. The choice of the exponential-Cauchy distribution appears then the most appropriate even though there are many types of dispersal kernels that ensures long-distance events (*Clark et al.*, 1999). In Figure 5.36 it is shown how the relationships between d_l vs mass and vs lifespan do not hold. This can be both for the limited data used or for the normal fluctuations due to the large variation of d_l for the same mass m and lifespan T of species (e.g. seeds with equal mass transported by animals are in general dispersed at smaller distances than if dispersed by wind). Indeed we find quite nice relationships between mass and lifespan and more interestingly between the z of the species-area curve and the lifespan. The scaling exponent found between $1/z$ vs $ln(T)$ is -2.00 that is close to the exponent found in literature (*Pigolotti and Cencini*, 2009). In the log-log plot the allometric relationship between z and T presents an exponent equal to 0.26.

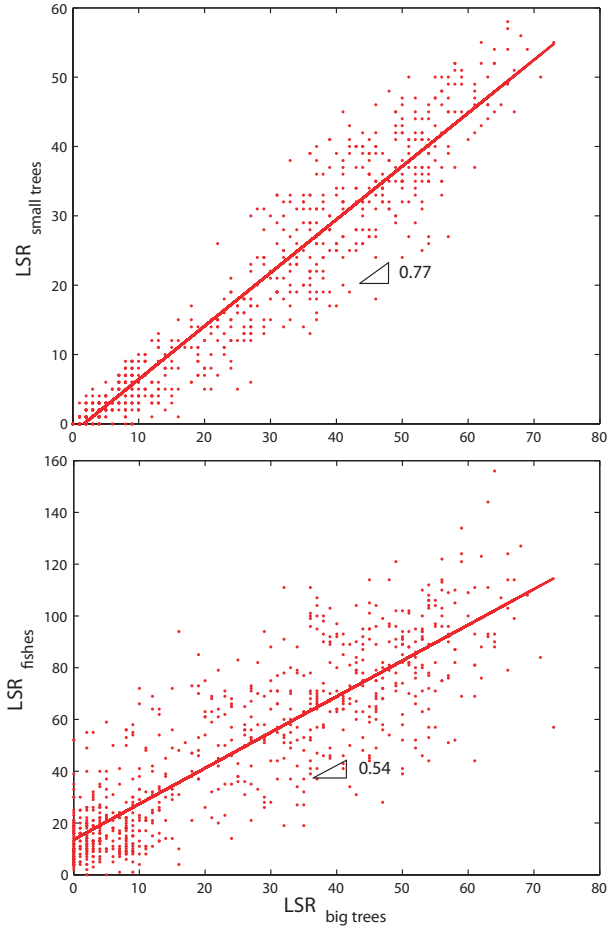

Figure 5.35: LSR of small trees and fishes in function of LSR of big trees.

Table 5.12: Mean dispersal length d_l, mass m, lifespan T and exponent z of the SAR for different taxa. For unicellulars see *Houchmandzadeh* (2008); *Pigolotti and Cencini* (2009), for seeds dispersed by birds *Schurr et al.* (2009), for seeds dispersed by wind *Crocker* (1938); *Nathan et al.* (2001, 2002a, 2008b), for seeds dispersed by elephants *Blake et al.* (2009), for earthworms *Edwards and Bohlen* (1977), for Zebra mussel *Stoeckel et al.* (1997), for butterflies *Gilbert and Singer* (1975), for Snowy plovers *Stenzel et al.* (2007), for elephants *Mennell and Scholes* (2008); *Blake et al.* (2009).

Case	d_l (km)	m (g)	T $(days)$	z
Unicellulars (Dictyostelium discoideum)	2×10^{-5}	–	$4 \div 5$	0.05
Seeds (birds dispersal)	$0.1 \div 1$	$10 \div 12$	$180 \div 36500(191625)$	$0.3 \div 0.7$
Seeds (wind dispersal)	$0.02 \div 0.04$	$10 \div 12$	$180 \div 36500(191625)$	$0.3 \div 0.7$
Seeds (elephants dispersal)	$5 \div 10$	$10 \div 12$	$180 \div 36500(21900)$	0.25
Earthworms	0.02	11.2	$365 - 1890$	0.08
Zebra mussel (MMRS)	$100 \div 1600$	$1 \div 5$	$3 \div 90$	–
Fishes (MMRS)	82	$2000 \div 5000$	$2920 \div 5475$	0.45
Butterflies	$0.02 \div 1000$	$0.003 \div 1.5$	$20 \div 182$	0.1
Snowy plover	$10 \div 50$	$30 \div 57$	$730 \div 1095$	0.125
Mangabeys	$1.2 \div 2.2$	$3000 \div 20 \times 10^3$	$6570 \div 10950$	–
Elephants	$40 \div 57$	$2300 \times 10^3 \div 7000 \times 10^3$	$25000 \div 29200$	–

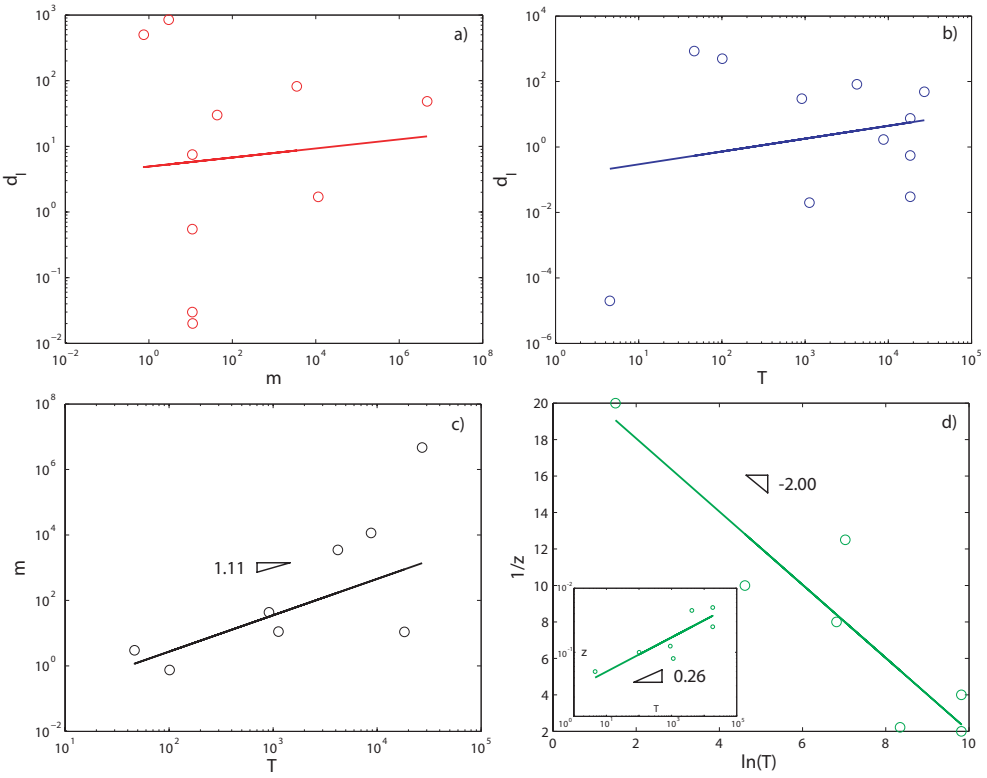

Figure 5.36: Relationships between the dispersal length, lifespan, mass and the exponent z of the species-area relationship for the taxa listed in Table 5.36. a) d_l vs m; b) d_l vs T; c) m vs T; d) $1/z$ vs $ln(T)$ (in inset z vs T).

5.8 Conclusions

In the present chapter it has been proofed the following:

- A robust but not trivial relationship between z and ϵ dependent on the ecosystem topology and shape and on the dispersal kernel structure;
- Purely random distributions of organisms have a flat pair correlation function that is like a constant JSI (see chapters 2 and 4) in the ecological literature. In a metacommunity $P(CS \geq cs)$ is more sensible to the shape of the occupancy-rank pattern rather then the rank-abundance pattern although they are related. The abundance-distribution like Preston's lognormal (*Preston*, 1948) is not the *conditio sine qua non* the clustering is not observed. $P(CS \geq cs)$ is influenced by the dispersal mechanism (that include the value of the dispersal parameter) combined with the topology. The environmental heterogeneities that affect the number of individuals in each local communities have an influence when the dispersal is extremely local (so when the speciation is the prevalent mechanism) or when the dispersal is too much global that means random spreading of species. Moreover the power-law structure of $P(CS \geq cs)$ does not mean fractal distribution of species. The correlation function is not needed to be power-law type as demonstrated by the global dispersal case in a river network. The power-law structure of the SAR is less sensible to the dispersal parameter until a minimum clustering is guaranteed;
- The heterogeneity of the environmental conditions leads to a lower value of ϵ, that means steeper power-law distributions of the cluster size, then smaller clusters than in an ideal homogeneous scenario. Conversely z of the species-area relationship is higher in the real case than in the homogeneous case. Moreover $\langle CS \rangle$ and $\langle LSR \rangle$ are lower in the case with heterogeneities than with the homogeneous conditions. In the homogeneous condition the power-law of CS is still observed;
- The power-law structure of the cluster-size is very dependent on the topology and on the shape of the ecosystem. The slope ϵ varies but the shape of $P(CS \geq cs)$ is almost the same both for RN and 2-D. The power-law of CS shows that in the 2-D case the number of small clusters is lower (higher lower bound of CS so bigger unreliable region due to the lower number of clusters) than in the RN counterpart, then the scaling region of $P(CS \geq cs)$ in the 2-D case is slightly narrower. The finite size-effects are the same in the RN and 2-D case.
- $z_F < z_{BT} < z_{ST}$, $\epsilon_F < \epsilon_{BT} > \epsilon_{ST}$, and $\langle CS \rangle_F < \langle CS \rangle_{BT} > \langle CS \rangle_{ST}$. Aggregation is stronger for larger dbh classes (big trees) if as a measure of aggregation of species ϵ and $\langle CS \rangle$ are considered: larger clusters are observed for large-diameter trees on the contrary, of what has been found for tropical forest e.g. by (*Li et al.*, 2009). Rare species are more strongly nested than abundant species;
- The elongation produces high values of z, low values of ϵ, and $\langle CS \rangle$. Therefore the elongation enhances the LSR due to low values of the mean diameter in elongated networks;

- The dependence of the power-law of the cluster-size on the scale, resolution and coarse-graining level on which the ecosystem is modeled. Table 5.13 reports a synthesis of the study about the scale and resolution in variance of biodiversity patterns and geomorphological variables. The SAR is only invariant across coarse-graining levels. The dependence on the resolution of the JSI and of the means RSR is because the former is a pairwise comparison on spatial patterns that are not the same across resolution, while the latter is an average measure among areas of increasing dimension diminishing the resolution. It is then very interesting how in geomorphology many quantities, like the drainage area A and the upstream lengths L are independent of scale, resolution and coarse-graining level, and result also universal meaning that their average value is very robust for the majority of river basins all around the world.

Table 5.13: Synthesis paper of the macroecological variables in function of the scale, resolution and coarse-graining: (I) means invariant and (V) means variant (or dependent). In grey the geomorphological variables.

	Scale	Resolution	CG (model)
$P(CS \geq cs)$	V	V	V
SAR	V	V	I
$\langle LSR \rangle$	V	V	V
JSI	V	V	V
$\langle RSR \rangle$	V	V	I
A	I	I	I
L	I	I	I

Chapter 6

The Probabilistic Structure of the Distance Between Tributaries of Given Size in River Networks

"Philosophy [nature] is written in that great book which ever is before our eyes – I mean the universe – but we cannot understand it if we do not first learn the language and grasp the symbols in which it is written. The book is written in mathematical language, and the symbols are triangles, circles and other geometrical figures, without whose help it is impossible to comprehend a single word of it; without which one wanders in vain through a dark labyrinth"
Galileo Galilei

We analyze the distribution of the distances between tributaries of a given size (or of sizes larger than a given area) draining along either an open boundary or the mainstream of a river network. By proposing a description of the distance separating prescribed merging contributing areas, we also address related variables, like mean (or bankfull) flowrates, channel and riparian areas width, which are derived under a set of reasonable hydrologic assumptions. The importance of such distributions lies in their ecological, hydrologic and geomorphic implications on the spreading of species along the ecological corridor defined by the river network and on the propagation of infections due to water-borne diseases, in particular in view of exact theoretical predictions explicitly using the alongstream distribution of confluences carrying a given flow. Use is made here of real river networks, suitably extracted from digital elevation models, Optimal Channel Networks (OCNs) and exactly solved tree-like constructs like the Peano and the Scheidegger networks. The results obtained redefine theoretically in a coherent and general manner – and verify observationally – the distribution function of the above distances and thus provide the general probabilistic structure of tributaries in river networks. Specifically, we find that the probability of exceedence of the alongstream distance d of tributaries of size larger than a has the explicit form $P(\geq d) = \exp\left(-Cd/a^{H/(1+H)}\right)$ where C is a constant that depends on the choice of boundary conditions and $H \leq 1$ is the Hurst exponent.

6.1 Introduction

The role of the structure of river networks in modeling hydrochory (the spreading of species along ecological corridors), human-range expansions (i.e. predicting how populations migrate when settling into new territories), or the spreading of water-borne infections has been recently recognized through quantitative models of dispersion along fractal networks coupled with proper reactions at their nodes (*Méndez et al.*, 2004; *Méndez et al.*, 2004; *Campos et al.*, 2006; *Muneepeerakul et al.*, 2007b,a; *Bertuzzo et al.*, 2007). An essential ingredient therein is the fact that settlers, infective agents or colonizing species do not occupy all the territory isotropically, but rather follow river patterns as general corridors and pathways. Among the key results, it was argued in a quantitative manner that landscape heterogeneities play an essential role in several physical and biological processes of great social, biological or physical importance that involve river basins as substrate (*Ammerman and Cavalli-Sforza*, 1984a; *Holmes*, 1993; *Fort and Méndez*, 1999; *Campos et al.*, 2006; *Muneepeerakul et al.*, 2007b,a; *Bertuzzo et al.*, 2007).

One interesting corollary of any analysis of migration/spreading fronts is the central role of the network structure acting as the substrate for wave propagation. Specific and reliable structural models thus need to be employed, as well as the probabilistic description of key random variables affecting the process. Empirical representations of natural forms in the river basin and their mathematical models, possibly addressing their fractal nature, have been thoroughly studied in the past (see e.g., *Rodriguez-Iturbe and Rinaldo*, 1997). We are specifically concerned here with features that might affect spreading processes on networks, like for instance the type and strictness of self-similar properties appearing at different scales of observation, or the looplessness of the structure (*Rinaldo et al.*, 1998, 2006). The type of self-similarity observed for riverine trees is particularly important. In fact, whether or not one would observe the same distribution of confluences regardless of the stream chosen as receiving water body (see Figure 6.1), is central to the generality of the distributions studied herein, and deeply related to the similarity of the parts and the whole embedded in the fractal form (*Rodriguez-Iturbe and Rinaldo*, 1997). Moreover, similarity is allowed only in a finite range of scales, as it needs proper finite-size corrections because upper and lower cutoffs in the aggregation structure – which reflect, respectively, the drainage density defining where channels begin, and the loss of statistical significance of areas approaching the maximum basin size (see e.g., *Maritan et al.*, 1996; *Rodriguez-Iturbe and Rinaldo*, 1997). Quite useful in this context is the use of Optimal Channel Network models (*Rodríguez-Iturbe et al.*, 1992; *Rodriguez-Iturbe et al.*, 1992; *Rinaldo et al.*, 1992; *Rinaldo et al.*, 1993), i.e. the robust determination of network structures that minimize total energy dissipation, that mimic quite efficiently the fluvial landforms over arbitrary ranges of scales and boundary conditions. Note that this property has also been shown to be an exact attribute of the steady state of the general (small-gradient) landscape evolution boundary value problem (*Banavar et al.*, 2000, 2001).

For general numerical calculations of the structure of distances of confluences of tributaries of a given area we shall adopt here (Figures 6.1 and 2.1 in chapter 2): i) the topology and geometry of real rivers, chosen from our archives and suitably extracted from digital terrain maps of variable resolution (see e.g., *Rodriguez-Iturbe and Rinaldo*,

1997); and ii) local-minimum Optimal Channel networks (OCNs) (*Rodríguez-Iturbe et al.*, 1992; *Rodriguez-Iturbe et al.*, 1992; *Rinaldo et al.*, 1992; *Rinaldo et al.*, 1993; *Rodriguez-Iturbe and Rinaldo*, 1997). They hold fractal characteristics that are obtained through a specific selection process from which one obtains a rich structure of optimal scaling forms that are known to closely conform to the scaling of real networks. To derive exact results, instead, we shall resort to Peano and Scheidegger networks. The former is a deterministic fractal (*Mandelbrot*, 1983) whose main topological and scaling features have been determined analytically (*Marani et al.*, 1991; *Colaiori et al.*, 1997), and the latter a random aggregation model of directed network (*Scheidegger*, 1967). The former has has been recently employed (*Bertuzzo et al.*, 2007) to prove that geometrical constraints imposed by a fractal network imply strong corrections on the speed of migration fronts. It was noted therein that it is not surprising that Peano and OCNs lead to similar results, because the speed of the front depends chiefly on topological features that are indeed quite similar for real, OCN and Peano networks. In fact,the wave speed is affected mostly by the gross structure encountered by the front while propagating along the network, chiefly the bifurcations. Hence topology, rather than the fine structure of the sub-paths, dominates the process. Because of this, we also address in detail the Scheidegger network (introduced in Figure 2.1, chapter 2), an exactly treatable tree-like pattern constructed by self-avoiding unbiased random aggregation pattern with injection (*Scheidegger*, 1967; *Huber*, 1991; *Takayasu et al.*, 1991). Indeed its topology and aggregation is known exactly to be rather different from real rivers and OCNs or Peano's construction. Optimal Channel Networks (OCNs) used in our computations are local minima (that is, dynamically accessible) of global energy dissipation, and are obtained by a simulated annealing optimization procedure. With relaxed constraints like periodic boundary conditions, both for single and multiple outlets, carefully annealed OCNs exhibit a value for the exponent of the aggregated area close to the mean-field one, which is known exactly to characterize the ground state and Peano (*Rodriguez-Iturbe and Rinaldo*, 1997).

Ecological implications of the distribution of tributaries relate to the availability of flow, slope and riparian area available in defining an ecological corridor (*Campos et al.*, 2006; *Bertuzzo et al.*, 2007; *Banavar et al.*, 2007; *Muneepeerakul et al.*, 2007a). Riparian zones, in fact, play many important roles in regulating ecosystem function within streams, surrounding environments, and upland areas. They are usually related to the stream width or to topographically concave areas, features that inevitably relate to landscape-forming discharges and hence to the distribution of tributary areas. Spreading of species or infections on river networks also requires flow specifications defining, for instance, the bias affecting dispersion of water-borne agents or propagules. Moreover, although models of biodiversity in two-dimensional landscapes had been extensively studied (see e.g., *Tilman and Downing*, 1994; *Hubbell*, 2001), only recently *Muneepeerakul et al.* (2007a) have shown that the combined effects of directionality of dispersal produced by the network landscape significantly alter various biodiversity patterns of vegetation communities obeying the neutral model. Be it metapopulation models coupled with the strictly hierarchical competition-colonization trade-off model to study the resulting biodiversity patterns in the river basin (*Muneepeerakul et al.*, 2007a), or patterns of infection or migration spreading (*Bertuzzo et al.*, 2007), it is clear the fundamental role played by the network substrate and of its properties on several ecological processes. About

the purported connection with riverine ecology, suffice here to say that the distribution we seek defines analytically the distribution of the distance of mean flowrates, channel width or riparian areas alongstream. All are, in fact, proxies of cumulative basin area, and hence of tributaries' contributing area, which conditions such distribution.

Three sample basins in Northern Italy of different size are investigated in detail here as source of real-world data. It is easily overlooked, in fact, that the very definition of the confluence of streams with a tributary depends on specifications like the threshold area for channelization, or the size of the confluent basin that defines a meaningful tributary. On this basis alone we argue that a tenet of geomorphology, i.e. that the link length distribution is the same random variable regardless of the link magnitude (see e.g., *Leopold et al.*, 1964), not uniform at all, needs be revisited. A link, in fact, is only defined as the alongstream distance separating two consecutive tributaries of any size. As we shall see, this concept is neither theoretically or empirically wrong, but it is strongly correlated to an arbitrary threshold in tributary area being assigned. Even operationally in the extraction of a river network, the concept is thus of fundamental significance.

This chapter is organized as follows. Section 2 proposes our new theoretical framework for the quantities of interest. Section 3 discusses the theoretical and empirical resources mobilized for the validation of the theory, specifically treating tributaries draining on mainstreams (in the case of real rivers and Peano networks), or tributaries draining onto a draining multiple-outlet line (in the case of OCNs and Scheidegger networks). Section 4 proposes results and discussion of the comparative analysis of theory and data from various sources. A set of conclusions, Section 5, then closes the chapter.

6.2 Theoretical framework

Figure 6.1 illustrates the definition of the random variable that we address throughout this chapter. The distance separating confluences of tributary areas larger than a threshold a is shown along: *a.1)*, *a.2)* the baseline of multiple outlet OCNs, and *b)* the mainstream of a real river network. Note that a river basin, and any subbasin nested into it, is generally an anisotropic system defined by a longitudinal typical length $L = L_{\|}$ (which we will identify with the linear size of the system), and a typical transverse length $L_{\perp} \leq L_{\|}$, both measured along the principal inertia axes (Figure 6.1). We shall assume that the width scales as $L_{\perp} \sim L_{\|}^{H}$ (with $0 \leq H \leq 1$) and call the basin self-affine if $H < 1$ and self-similar if $H = 1$. H is called the Hurst coefficient because of the analogies with fractional Brownian motion contexts and is typically in the range $H = 0.75 \div 1.00$ ((*Rodriguez-Iturbe and Rinaldo*, 1997), pp. 174-175). As a consequence the maximum characteristic area A_{max} is postulated to scale as $A_{max} = L_{\|} L_{\perp} \sim L^{1+H}$ (*Maritan et al.*, 1996; *Rinaldo et al.*, 1998).

In Figure 6.1, the lower open boundary border of the OCN is sampled for tributary area in this case – here d_f, $H \sim 1$ (owing to the arrangement chosen, whereas single-outlet OCNs usually yield lower H and higher d_f values see *Rigon et al.* (see e.g., 1996) and Table 6.1). In such a case the length l of each tributary subbasin is roughly perpendicular to the draining side. In more details figure 6.1 *a.1)* shows an OCN with cylindrical boundary conditions (i.e. periodic in the transverse direction), and a no-flux condition

along the upper side. The overall dimensions are 780×78 pixels in the case shown herein ($m = \frac{L_\perp}{L_{||}} = 10$). The threshold area for channelization is chosen at 10 pixels; *a.2)* the same OCN drawn with threshold area at 1000 pixels, which indeed shows tributary subbasins with area greater than $a = 1000$. The random distances among them are highlighted; *b)* A real river network, the Tanaro basin (globally ~ 8000 km^2) closed at Garessio (Piemonte, Northern Italy), suitably extracted from digital terrain map (DTM), of dimensions 178×296 pixels where we have chosen for computational reasons relatively large pixels (area 100×100 m^2). Note the particular arrangement of tributaries draining from both sides. The inset illustrates the notation where $L = L_{||}$ and $L_\perp \sim L^H$ are respectively the longitudinal and transverse length scales along the principal axes of inertia, H is the Hurst parameter, l the mainstream length.

In all cases we can define, along either mainstreams or multiple-outlet boundaries or else anywhere within the basin, the random contributing area at a site, say A, whose probability distribution is here computed through $P(A \geq a)$, that is, the relative proportion of areas, sampled anywhere, larger than the given value a (Figure 6.2) (*Rodríguez-Iturbe et al.*, 1992; *Rodriguez-Iturbe et al.*, 1992; *Maritan et al.*, 1996; *Rodriguez-Iturbe and Rinaldo*, 1997) whose scaling form is:

$$P(A \geq a) \;\sim\; a^{-(\tau-1)}\; F\left(\frac{a}{A_{max}}\right) \;\sim\; a^{-(\tau-1)}\; F\left(\frac{a}{L^{1+H}}\right), \tag{6.1}$$

where: $A_{max} = L^{1+H}$ (*Maritan et al.*, 1996) is the largest area that defines the upper cutoff for the distribution (the length scale L is defined later), and $\tau = 1.43 \pm 0.02$ with remarkably small scatter from a surprisingly large archive of real river basins regardless of climate, vegetation, exposed lithology or mainstream sinuosity of the different river basins (Figure 6.2) (*Rodríguez-Iturbe et al.*, 1992; *Maritan et al.*, 1996; *Rigon et al.*, 1996; *Rinaldo et al.*, 1998). Obviously the probability density function (pdf) $p(a) = -dP(A \geq a)/da$ follows as:

$$p(a) \;=\; p(a|L) \;=\; a^{-\tau} f\left(\frac{a}{L^{1+H}}\right), \tag{6.2}$$

where f follows from direct integration of Equation (6.1) and the notation $p(a|L)$ stresses the conditional character of the pdf on the cutoff size L^{1+H} (*Maritan et al.*, 1996). Further relationships involving scaling exponents (*Rinaldo et al.*, 1998) are Hack's law $l \sim A_{max}^h$ (where $h \sim 0.57$ is the Hack's exponent with implications on the type of self-similarity operating within the river basin (*Rigon et al.*, 1996, 1998) and $l \sim L^{d_f}$ (with $d_f \sim 1.0$, see Table 6.1), where l is the length of the longest stream in the drained area itself function of the contributing area value a (see Figure 6.1), from which one derives linkages that constitute stringent comparative tools, e.g.

$$\tau \;=\; 2 \;-\; \frac{d_f}{1+H} \;=\; 2 - h, \tag{6.3}$$

which we shall usually simplify by assuming $d_f \sim 1$, hence leading to $2-\tau \sim 1/(1+H)$. Moreover, the relation $\tau + h \sim 2$ naturally conforms to the common values $\tau \sim 1.43$, $h \sim 0.57$ (*Maritan et al.*, 1996). As shown in Figure 6.2 the above relationships are verified

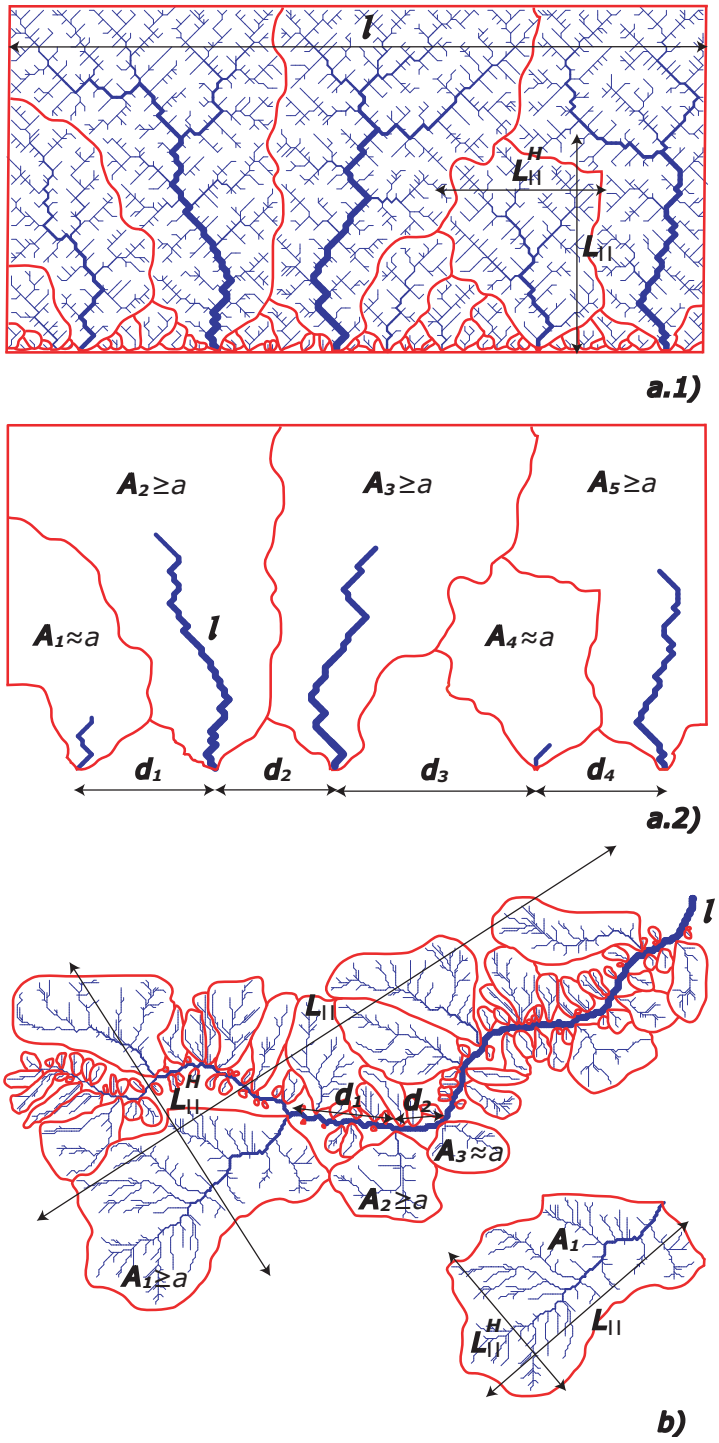

Figure 6.1: Samples of networks on which calculations are performed.

computationally or exactly derived for all the constructs shown in Figure 2.1, Chapter 2 and in Figure 6.1, including the Scheidegger construction which exhibits the exact value $\tau = 4/3$ for the distribution of areas along the baseline (*Huber*, 1991; *Takayasu et al.*, 1991) and a slow convergence to the exact asymptotic values for area at any point (Figure 6.2) (see e.g., *Maritan et al.*, 1996; *Rigon et al.*, 1996; *Rodriguez-Iturbe and Rinaldo*, 1997). A summary of the relevant scaling exponents derived in this study is reported in Table 6.1 jointly with comparative results to be discussed later.

The constraint of conservation of total area suggests (*Maritan et al.*, 1996; *Rodriguez-Iturbe and Rinaldo*, 1997) that the distribution of areas sampled along a given line boundary where multiple outlets occur, say $p_b(a|L)$, or along a mainstream draining from both sides, say $p_{ms}(a|L)$, differs from Equations 6.1 and 6.2 because tributary area sampled anywhere within the tributary is a different random variable, sampled from a different population from those constrained in relation to their possible spatial locations. Indeed if at i sites one collects the areas A_i and must enforce the constraint $\sum_i A_i = A_{max}$ (where A_{max} is the total area), the resulting population is different from that leading to Equation 6.1 because the analog areas A_i sampled anywhere do not add to total area (*Rodriguez-Iturbe and Rinaldo*, 1997). If we assume that the distribution at the mainstream sites is given by (*Maritan et al.*, 1996; *Rodriguez-Iturbe and Rinaldo*, 1997):

$$p_{ms}(a|L) \;=\; a^{-\tau_{ms}} f_{ms}\left(\frac{a}{L^{1+H}},\right) \tag{6.4}$$

the area conservation constraint, i.e.

$$l\langle a\rangle_{ms} \;\sim\; L^{1+H}, \tag{6.5}$$

(where l is the length of the mainstream) leads, by explicitly calculating the mean and power counting, to the relationship

$$\tau_{ms} \;=\; 1 + \frac{d_f}{1+H}, \tag{6.6}$$

and ultimately to the inequality:

$$\tau \;\leq 1.5 \;\leq\; \tau_{ms}, \tag{6.7}$$

indicating there are significant differences between τ and τ_{ms}. For mainstream boundaries, draining areas of overall size L^{1+H} we have normally $\tau \sim 1.43$ and $\tau_{ms} \sim 1.5$, although in particular cases like Scheidegger's (where $H = 1/2$, $d_f = 1$ and $\tau = 4/3$ exactly (*Huber*, 1991) one has $\tau_{ms} = 5/3$ (see Table 6.1). Figure 6.2 shows the relevant numerical computations, where finite-size effects are particularly evident. In Figure 6.2 we have computed either $\tau - 1 = \beta$ or $\tau_b - 1 = \beta_b$ from the estimate of the slope of the double logarithmic plot. Note that the less constrained arrangement of the periodic boundary conditions calls for values of τ closer to the exact value 3/2 of the ground state; *b)* Sample statistics of the Tanaro river basin. Note that $\tau_{ms} - 1 = \beta_{ms} \sim 0.52$ for areas along the mainstream and $\tau \sim 0.44$ for the entire basin, as we expected. Note also the noteworthy lower cutoff imposed by the extent of unchanneled sites owing to

Table 6.1: Summary of computed and predicted scaling exponents, respectively for: Peano's network (Figure 2.1 *e)*, Chapter 2); Scheidegger's constructions (Figure 2.1 *c)*, Figure B.1); multiple-outlet (*mo*) OCNs (Figure 6.1 *a.1)*, *a.2)*, and Figure 2.1 *a)*, Chapter 2) where areas are sampled at the lower boundary; three real river basins (Tanaro, Mis and Cordevole respectively, Figure 6.1 *b)* and Figure 2.1 *b)* (Chapter 2) for the first one, Figure 2.1 *d)*, Chapter 2, for the latter ones); a single-outlet (*so*) OCN where areas are sampled along the mainstream. The computed scaling coefficients are: H (row 1), d_f (row 2), τ (row 3), τ_{ms} (row 8) and τ_b (row 10) estimated numerically by fitting scaling laws (for τ, τ_b and τ_{ms} see Figure 6.2 and checked against exact solutions when available (see Equations 6.1 to 6.9); the scaling exponent $\phi + 1 = \tau$ (row 4) estimated from the relationship $\langle a \rangle_a \propto a^{\phi}$ (Equation 6.11, Figure 6.3); the third independent estimate of $\tau = 2 - \frac{d_f}{1+H}$ (Equation 6.3, row 5) is then directly computed from the computational estimates of H, d_f; the scaling exponent ψ computed from the relationship $\langle d \rangle_a \propto a^{\psi}$ (row 6, Figure 6.7); the scaling relationship $\psi = \frac{H d_f}{1+H}$, derived theoretically (Equation 6.19), is compared from the independent estimates of H, d_f (row 6); the scaling exponent τ_{ms} of the distribution of areas sampled along the mainstream (row 7) is computed, when appropriate, from our data. Note that we have not computed it only for multiple outlet OCNs; the theoretical prediction $\tau_{ms} = 1 + d_f/(1+H)$ (Equation 6.6, row 8) is analyzed from our data; the scaling exponent τ_b is computed via the proper area distributions sampled along a line boundary where multiple-outlets occur (row 9). This calculation is possible only for multiple-outlet OCNs (Figure 6.1 *a.1)*, *a.2)*, and Figure 2.1 *a)*, Chapter 2) and for the particular Scheidegger-like constructions shown in Figure B.1; the theoretical prediction $\tau_b = 2 - 1/(1+H)$ (Equation 6.9) is compared against our data (row 9). Variabilities of measured exponents are standard errors found by Reduced Major Axis Regression (RMA or SMA) bootstrapping over cases, and deriving slopes by the linear and the Jackknife models (*Warton et al.*, 2006).

Scaling exponent	*Peano*	*Scheidegger*	*OCN* (*mo*)	*Tanaro*	*Mis*	*Cordevole*	*OCN* (*so*)
H	1.00	1/2	0.97	0.75	0.78	0.82	0.85
d_f	1.00	1.00	1.01	1.03	1.02	1.02	1.01
τ	3/2	4/3	1.45	1.44	1.43	1.43	1.45
$\phi + 1$	3/2	1.33 ± 0.02	1.46 ± 0.02	1.42 ± 0.03	1.45 ± 0.03	1.44 ± 0.03	1.45 ± 0.02
$2 - \frac{d_f}{1+H}$	3/2	4/3	1.48	1.41	1.43	1.44	1.45
ψ	1/2	0.33 ± 0.02	0.50 ± 0.01	0.43 ± 0.02	0.45 ± 0.02	0.46 ± 0.02	0.47 ± 0.02
$\frac{H d_f}{1+H}$	1/2	1/3	1/2	0.44	0.45	0.46	0.46
τ_{ms}	3/2	1.66 ± 0.02	–	1.52 ± 0.03	1.55 ± 0.03	1.54 ± 0.03	1.53 ± 0.02
$1 + \frac{d_f}{1+H}$	3/2	5/3	–	1.58	1.57	1.56	1.54
τ_b	–	1.33 ± 0.02	1.50 ± 0.02	–	–	–	–
$2 - \frac{1}{1+H}$	–	4/3	1.49	–	–	–	–

the relatively large extraction threshold chosen – a step quite suitable for 8000 km^2 networks; *c)* Sample statistics of the Scheidegger network with $L_\perp \sim L^{1/2}$ and $L_{||} \sim L$. The Scheidegger model requires $\tau_b = \tau$, and might be treated as a single- or multiple-outlet construction, with directed, subvertical drainage directions and periodic boundary conditions (see also Appendix B). Here we show the results for different ratios of sides (see Legend) starting with the lattice characterized by $m = \frac{L_\perp}{L_{||}} = 1/10$ which allows subbasins properly characterized by $H = 1/2$. Statistics for the single-outlet Peano basin are also shown for which, as for Sheidegger model, is true the equivalence $\beta_b = \beta$.

Finally note that yet another random variable could be derived sampling area along a straight boundary perpendicular to the main axis $L_\parallel$ of the basin. In such a case, independently on the size of the boundary as long as it exceeds L^H, one has

$$p_b(a|L) \;=\; a^{-\tau_b} f_b\left(\frac{a}{L^{1+H}},\right) \tag{6.8}$$

where

$$\tau_b \;=\; 2 \;-\; \frac{1}{1+H}, \tag{6.9}$$

with the usual meaning of the symbols. For straight multiple-outlet boundaries in Scheidegger constructs (of the type discussed in Appendix B) we find the exact result $\tau_b = \tau = 4/3$. Notice that if $d_f = 1$, $\tau = \tau_b$ and if $H = 1$ all area scaling exponents equal 3/2.

An interesting feature that can be derived from Equation 6.2 and its variations (4) and (8) is $\langle a \rangle_a$, i.e. the mean area of all tributaries whose size is larger than a given threshold area a. This is tantamount to averaging the size of all incoming tributaries larger than a, resulting in the obvious fact $\langle a \rangle_a \geq a$. Indeed this quantity matters here, because we might want to define sites with flow, width or riparian area larger than a threshold value to define viable minimum values, e.g. for biological conservation. One thus has:

$$\langle a \rangle_a \;=\; \frac{\int_a^\infty dx\; x\; p(x|L)}{\int_a^\infty dx\; p(x|L)} \;=\; \frac{\int_a^\infty dx\; x^{-z} f(x/L^\phi)}{\int_a^\infty dx\; x^{-\tau} f(x/L^\phi)}, \tag{6.10}$$

(where $z = \tau - 1$ and $\phi = 1 + H$, defined for reasons to appear in the latter) which poses a nontrivial scaling result that we shall use in the following. The basic result is that when $L \to \infty$ and $z > 1$ one has the asymptotic behavior $\langle a \rangle \propto a^{1-z}$, whereas if $z < 1$ one has $\langle a \rangle \propto L^{\phi(1-z)}$ (see Equation 6.18). In the case at hand, regardless of boundary conditions in order to average sizes of tributaries larger than a we sample anywhere in the basin – thus $p(a|L)$ is the appropriate distribution and τ the relevant scaling exponent – one has:

$$\langle a \rangle_a \;=\; \frac{L^{(1+H)(2-\tau)}}{a^{1-\tau}} \;\propto\; a^{\tau-1} \;=\; a^{1-\frac{d_f}{1+H}}, \tag{6.11}$$

a result that we shall compare with data in the following. Note that for scaling with τ_b as in Eq. 6.9 one would have had $\langle a \rangle_a \;\sim\; a^{H/(1+H)}$ and for τ_{ms} as in Eq. 6.6 one would have obtained $\langle a \rangle_a \;\sim\; a^{d_f/(1+H)}$.

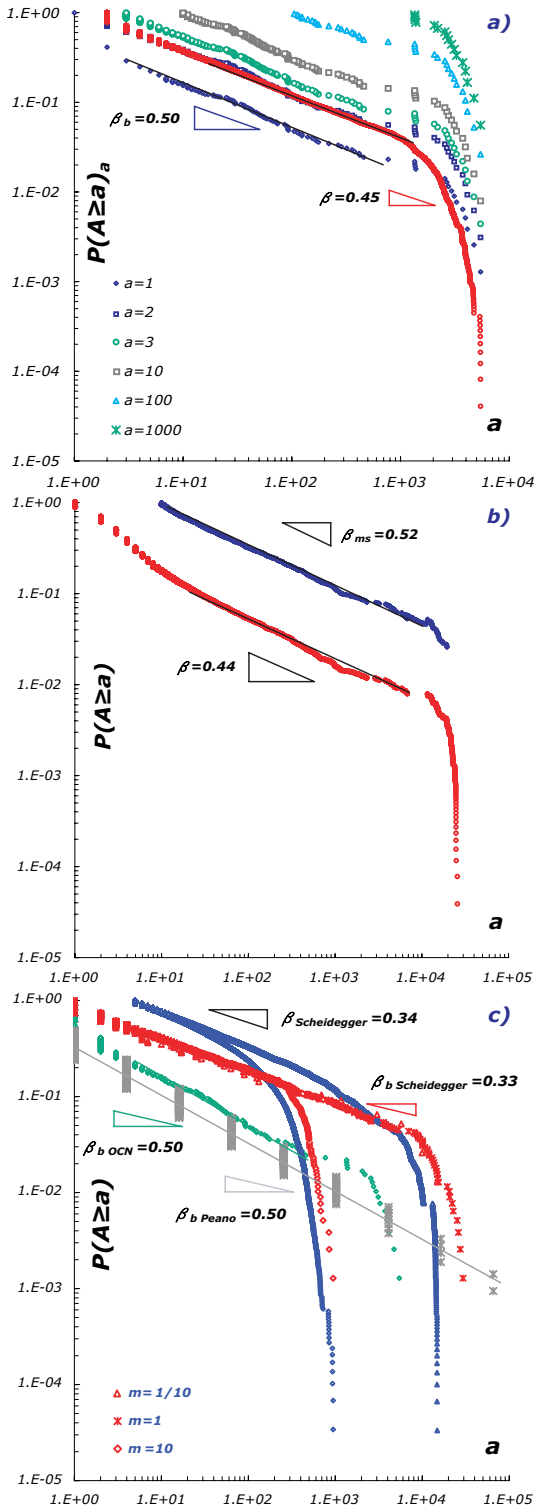

Figure 6.2: Cumulative probability distributions of areas $P(A \geq a|L)$ sampled on the whole basin and $P_{ms}(A \geq a|L)$ sampled along a mainstream. *a)*: Statistics on OCNs for different threshold values a (see Legend).

We shall present now the derivation of the distribution of distances of tributaries characterized by their area larger than a threshold a (and thus by related characteristic flowrates, riparian proper, cross-section width, depth and velocity among other features). Notationwise, we shall define the random variable drawn from the random population of possible distances as D, and as $P(D \geq d)_a$ the probability of exceeding a given distance d conditional on the confluence of tributaries of area (at least) a.

We approach this derivation initially for the multiple outlet boundary leading to a distribution of areas with scaling exponent τ_b. We assume that the distribution of distances (conditional on a given tributary area a) is defined by the probability of finding $d-1$ consecutive tributaries with area smaller than a, and finding two tributaries of area equal or greater than a at both ends. Distances are walks in terms of steps between tributaries, so d is in pixel units. If statistical independence holds for tributaries area, for $a \gg 1$ so $d \gg 1$, such probability would be given by $P(A < a)^{d-1} = (1 - P(A \geq a))^{d-1} \propto \exp\left[-(a^{1-\tau_b})\ d\right]$. One doubts, however, that the statistical independence might indeed hold for small distances. At small areas, in fact, the determination of the probability is most uncertain owing to the issue of the proper drainage density for network extraction – hence probabilities become reliable only for sizable areas. For large areas, however, the occurrence of consecutive tributaries cannot be independent at small distances, as one cannot possibly have two large consecutive basins for distances smaller than $L_\perp \sim L^H$. This assumption will also be tested in the following.

For two tributaries of size $\approx a$, draining in a quasi-parallel arrangement along a boundary, one may speculate that the minimum distance is roughly proportional to their width $L_\perp$. This implies that the mainstream is somewhat straight and that for smaller distances the receiving boundary can only collect smaller basins. The above relationships imply:

$$d(a) \ \sim \ a^{\frac{H}{1+H}}. \tag{6.12}$$

Note that in case the boundary is draining tributaries from both sides, as in the case of a mainstream, we expect the distribution to be a fraction of that given by the ansatz of Equation (6.12). Note also that the unconditional probability $p(d)$ would be given by the relationship $p(d)dd = p(a)da$ and $da/dd \propto d^{1/H}$, from which one obtains, after some calculations, using Equation 6.6 with $d_f = 1$:

$$p(d) \ \sim \ d^{-\frac{1+H}{H}} f_d\left(\frac{d}{L^H}\right), \tag{6.13}$$

more generally the power law is $d^{-\frac{\tau_{ms}(1+H)-1}{H}}$, from which the unconditional probability of exceedence is given by $P(D \geq d) \propto d^{-(\tau_{ms}-1)\frac{1+H}{H}}$, with the scaling argument $d^{-1/H}$. From the computational viewpoint, one finds easier to compute the probability $P(D \geq d)_a$ of finding a distance conditional on a given size a of the tributary. The basic relationship would be given by

$$p(d) \ = \ \int da\ p(d|a)\ p(a), \tag{6.14}$$

from which, after some calculations and power counting by assuming $p(d|a) \sim d^{-\psi}\hat{f}_d(d/a^{H/(1+H)})$,

one obtains $\psi + 1/H = (1+H)/H$ and hence $\psi = 1$ independently of τ, or:

$$p(d|a) \;=\; d^{-1}\; \hat{f}_d\left(\frac{d}{a^{H/(1+H)}}\right), \tag{6.15}$$

which, interestingly, leads to the testable scaling form

$$P(\geq d) \;=\; P(D \geq d)_a \;=\; \hat{F}_d\left(\frac{d}{a^{H/(1+H)}}\right), \tag{6.16}$$

(where both notations are freely interchanged) that follows from Equation 6.15 with:

$$\hat{F}_d(x) \;=\; \int_x^\infty dy\; y^{-1}\; \hat{f}_d(y), \tag{6.17}$$

whose collapse for different values of a can be directly compared with observations from real or artificial networks. Notationwise (in analogy with previous choices), we shall use the notation $F_{ms}(x)$ and $F_b(x)$ for mainstream or multiple-outlet boundary estimates of the scaling function in Equation 6.16, respectively. The theoretical prediction, in fact, suggests that the larger the threshold area a for tributaries of equal size, the larger, in the average, the random distance D among them – hence different values of a should produce broader distributions with larger mean and variance. However, when probabilities $P(D \geq d)_a$ are plotted against $d/a^{H/(1+H)}$ a collapse onto a single curve is expected. Notice that the scaling function $F_d(x) \to 1$ when $x \to 0$ and $F_d(x) \to 0$ when $x \to \infty$, and it depends on the boundary conditions adopted (e.g. if drainage occurs along a mainstream, areas flow in from both sides, and distances are shorter).

The expected distance between two rivers of area $\geq a$ would therefore be given in terms of the probability of finding such area a, i.e. (for a mainstream)

$$\langle d \rangle_a \;\sim\; \langle a^{\frac{Hd_f}{1+H}} \rangle \;\propto\; \frac{\int_{\mathrm{a}}^\infty p_{ms}(a|L)\; a^{Hd_f/(1+H)}\; da}{\int_{\mathrm{a}}^\infty p_{ms}(a|L)\; da}\,. \tag{6.18}$$

Standard finite-size arguments require:

$$\int_{\mathrm{a}}^\infty a^{-z}\; f\left(\frac{a}{L^{1+H}}\right)\; da \;\sim\; \begin{cases} L^{\,(1+H)(1-z)} & \text{for } z < 1 \\ \mathrm{a}^{1-\mathrm{z}} & \text{for } z > 1 \end{cases}, \tag{6.19}$$

where $z = \tau_{ms} - \frac{Hd_f}{1+H}$, and the integral is dominated by the upper/lower cutoff when $z < 1$ and $z > 1$ respectively. The coefficient $z = \tau_{ms} - \frac{Hd_f}{1+H} \geqslant 1$ because $H \leqslant 1 \leqslant d_f$. Since z, $\tau > 1$, one finds the mean distance between two subbasins of area $\geqslant a$ as:

$$\langle d \rangle_a \;\sim\; \frac{a^{1-\left(\tau_{ms} - \frac{Hd_f}{1+H}\right)}}{a^{1-\tau_{ms}}} \;\sim\; a^{\frac{Hd_f}{1+H}} \;\sim\; \langle a \rangle_a^H, \tag{6.20}$$

which can be compared with field observations directly. On average for real basins $d_f \simeq 1$ so simplifying the scaling relation is $\langle d \rangle_a \;\sim\; a^{\frac{H}{1+H}}$. In Equation 6.19 we have used the result $\langle a \rangle_a \sim a^{\tau_{ms}-1}$ which can be derived with the same above procedure. Analogously one may compute the second moment $\langle d^2 \rangle_a = \langle a^{2H/(1+H)} \rangle$ which poses a few additional finite-size problems. In fact, through scaling arguments similar to the

ones seen before, with $d_f \simeq 1$, we obtain

$$\langle d^2 \rangle_a \sim a^{\tau_{ms}-1} L^{2H-1} = a^{\frac{1}{1+H}} L^{2H-1}, \tag{6.21}$$

and the variance $\sigma^2(d)_a = \langle d^2 \rangle - \langle d \rangle^2 = \langle (d - \langle d \rangle)^2 \rangle$, using Equation 6.20 and the square of Equation 6.18, is given by

$$\sigma^2(d)_a \sim a^{\frac{2H}{1+H}} \left(-1 + const \frac{L^{1+H}}{a} \right)^{\frac{2H-1}{1+H}}, \tag{6.22}$$

Both statistics foresee notable corrections dependent on size L, which will also be tested against data (see inset in Figure 6.3).

We finally note that in the case of straight, long enough (i.e longer than L^H) multiple-outlet boundary one has to carry out the same operations via the distribution $p_b(a|L)$ with $\tau_b = 2 - 1/(1+H)$ (see Equation 6.9), thus obtaining:

$$\langle d \rangle_a \sim a^{\frac{H d_f}{1+H}} \sim \langle a \rangle_a^{d_f} / L, \tag{6.23}$$

which indicates that Scheidegger's network draining along a straight boundary should scale approximately like real rivers along mainstreams (for both $d_f = 1$, exactly in the case of Scheidegger and approximately for real rivers).

All the above relationships will be compared against data in the next Section. Note, finally, that for Peano, OCNs and real basins one has $H \simeq 1 \simeq d_f$, whereas for Scheidegger river networks $d_f = 1$ while $H = 1/2$ derive from the exact properties of the random walk (*Hughes*, 1995) that defines the outer boundaries. Thus we value comparative studies among them, as notable differences in scaling structures should arise in the observational properties.

6.3 Comparative analysis: Tributaries draining onto multiple-outlet boundaries or mainstreams

The dataset specifically used for the computation of the distributions of distances for real rivers pertains three sample river basin in Northern Italy (see caption of Figure 2.1, Chapter 2), the Tanaro river basin (Piemonte, Italy), the Cordevole and Mis river basins (Veneto, Italy), part of our relatively large geomorphologic database where we have chosen the actual basins for their differing size, relief and geologic and vegetational context. Table 6.1 illustrates the rather narrow range of variation for the possible values of H, d_f, τ of real networks, which follows usual patterns (see e.g., *Maritan et al.*, 1996; *Rodriguez-Iturbe and Rinaldo*, 1997; *Rinaldo et al.*, 1998). Conversely, conceptual constructs like Peano or Scheidegger have rather different features (*Rinaldo et al.*, 1998), some exactly determined (*Huber*, 1991; *Marani et al.*, 1991; *Takayasu et al.*, 1991), thus allowing us to check theoretical results against the scaling exhibited by computational entities, such as real river networks, and OCNs (single and multiple outlet). All geomorphological quantities, including distance d and area a, are expressed in pixel units.

Figure 6.3 illustrates a preliminary test of the theoretical scaling results that are derived solely from the the probability of areas $p(a|L)$. Specifically, we discuss the

scaling relationship of $\langle a \rangle_a$ (the mean area of all basins larger than a threshold area a encountered along a boundary or a mainstream) with the minimum tributary area a. From theory we predict (Equation 6.11) that such scaling is of the type $\langle a \rangle_a \propto a^{\tau-1}$ where τ is the scaling exponent of the distribution of area anywhere within the network, and thus independent of boundary conditions (*Rodríguez-Iturbe et al.*, 1992; *Maritan et al.*, 1996). The computed values match quite accurately Peano's exact value $\tau - 1 = 1/2$ (*Marani et al.*, 1991; *Colaiori et al.*, 1997; *Rodriguez-Iturbe and Rinaldo*, 1997), the), Scheidegger's exact value $\tau - 1 = 1/3$ (*Huber*, 1991; *Takayasu et al.*, 1991) and the expected value from real networks ($\tau - 1 = 0.42 - 0.46$). The implied validation of the scaling argument leading to Equation 6.11 is deemed particulary significant.

Figure 6.4 shows the empirical distributions of distance conditional on area $P(D \geq d)_a$, computed for the sample networks shown in Figure 2.1, Chapter 2. Notice that at larger values of a, the broader become the distributions and the larger the random distance D in the average, as expected. The double logarithmic plot chosen emphasizes the lack of a power-law character (predicted by Equation 6.16) strictly conditional either on the threshold area a or on the drainage sites chosen as boundary conditions (a line or a mainstream). In fact: the top graph *a)* shows multiple-outlet OCNs draining on a line; *b)* shows the results for mainstreams of Tanaro basin; and *c)* illustrates Scheidegger's distribution of distances draining along a line orthogonal to the mainstream. Figure 6.4 *a)* shows also an approximation to the unconditional probability $P(D \geq d)$ obtained by ensemble averaging. This is operationally mimicked by the average of the exceedence probabilities calculated by putting together all distances computed for each value of the parameter a (note that such distances span the entire range of possible values of D). Although the sampling of a was not designed to produce a proper ensemble average, we find that, as theory predicts and regardless of details on how this choice approximates the unconditional probability, the computational slope is rather close to the theoretical value of -1 for the exceedence probability (which of course implies a scaling of -2 for the unconditional pdf). The effects of the different aggregation structure are clearly reflected by the observational conditional exceedence probabilities of D. In all cases, in fact, $P(D \geq d)_a$ proves to be a function of the threshold value a for the confluent tributary area, as larger thresholds naturally imply larger distances among equally-sized tributaries.

In Figure 6.4: *a)*, distribution of distances for the multiple-outlet OCN draining along a straight line of Figure 2.1 *a)* in Chapter 2 (see also Figure 6.1 *a.1)* and *a.2)*), for different values of a namely reported in Legend on right of the graphic; darker line (empty circles) is obtained by ensemble averaging of the exceedence probabilities putting together all distances for each value of a (note that such distances span the entire range of possible values of D); *b)*, statistics of distances made along the mainstream of Tanaro basin (Figure 6.1 *b*) and Figure 2.1 *b)*, Chapter 2). The chosen values of a are on Legend attached on the right; *c)*, statistics of Scheidegger network for $m = L_\perp / L_{||} = 1$ draining onto a straight boundary. Values of the parameter a related to each distribution are on Legend on right of the plot; observe how for higher values of a distributions are quite detached from the other ones, confirming the fast deviation of scaling exponents noted computationally for the model (Figure B.2). Note also that the larger the size that defines tributary confluent basins, the longer it takes to find them (in the average) walking along

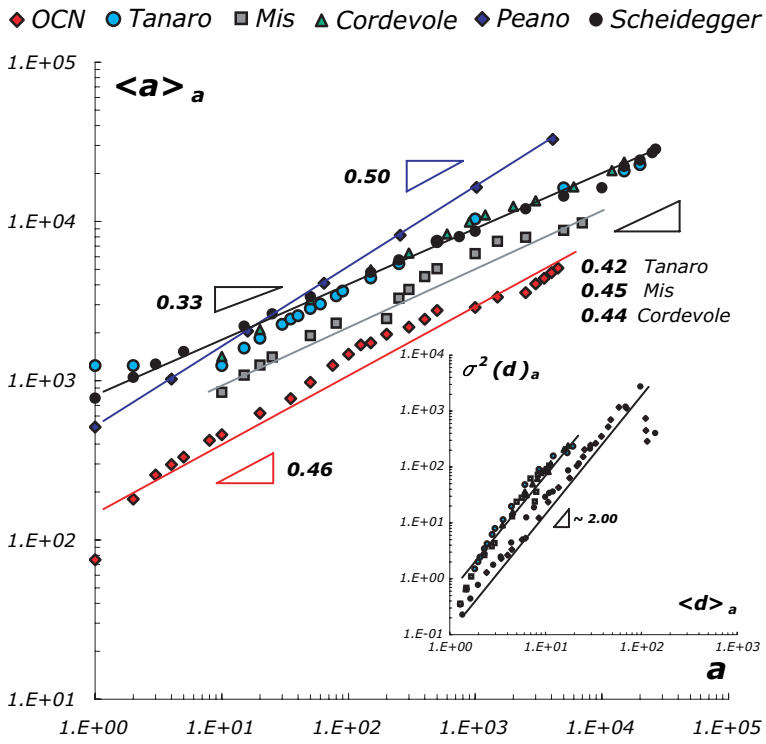

Figure 6.3: When we average the size of basins larger than a encountered along a mainstream or a multiple-outlet boundary, we observe a scaling relationship between the average size $\langle a \rangle_a$ and the threshold area a that defines the tributary, i.e. $\langle a \rangle_a \sim a^{\tau-1}$, where τ is the scaling exponent of the distribution of areas anywhere in the basin. Here we document an excellent agreement of theory and data, in all cases, specifically: (diamonds) Peano's construction; (solid dots) Scheidegger's network; (circles, squares and triangles) real river basins. The estimated scaling exponents are shown in the Figure. Inset: the logarithmic corrections to the relation between $\sigma^2(\langle d \rangle_a)$ *vs* $\langle d \rangle_a$ show here, a function of a (see Figure 6.7 and 6.8). An average exponent for all basins is roughly equal to 2, confirming the variance-mean theoretical expectation $\sigma^2(\langle d \rangle_a) \sim \langle d \rangle_a^{2\alpha}$ with $\alpha \simeq 1$.

the chosen draining boundary. See for whichever type of networks that distributions of distances (*a)*, *b)*, *c)*), for each value of the parameter a, lead to a characteristic length $d(a)$ in function of the parameter a, so they can be considered as exponential distribution, like $P[D \geqslant d]_a = const \exp(-\beta \mathrm{d})$ with $\beta \to 0$.

Figure 6.5 shows the collapse plot attempted because of the theoretical prediction (Equation 6.16), with estimates of the best collapse that indeed conforms well to the theoretical predictions (see Caption for details). Two theoretical predictions are validated, namely the scaling form where d is rescaled with respect to the mean area $\langle d \rangle_a \sim \langle a \rangle_a^{H,\, d_f}$ (with the exponent H for real basins, d_f for straight mainstream) of all tributaries with area $\geq a$, and the scaling with $a^{H/(1+H)}$. Values of H for the best collapse match the viable values from direct observation. It is safe to assume, in this context, that the estimate of $d_f \approx 1$. In all cases the theoretical prediction is verified quite well, and differences arising in the scalings are explained by the different boundary conditions or aggregation structures, like e.g. in the case of Scheidegger's construction which exhibits by construction $H = 1/2$, $d_f = 1$ and $\tau = 4/3$ (*Huber*, 1991). Overall, the theoretical framework seems quite robustly verified.

In Figure 6.5, on the left side, the collapse of $P(D \geq d)_a$ is tested as a function of $d/a^{\frac{Hd_f}{1+H}}$, as predicted by theory for mainstreams (Equation 6.16), and on the right, as a function of $d/\langle a \rangle_a^H$ or $d/\langle a \rangle_a^{d_f}$ depending on boundary conditions (Eqs. 6.20 and 6.23 into Eq. 6.16). Here we show *a.1)* and *a.2)* for OCNs, *b.1)* and *b.2)* for the Tanaro river network, *c.1)* and *c.2)* for Scheidegger model with side ratio $m = 1$. Insets in *a)* and *b)*: sensitivity tests are shown on the estimate of the best exponents, where $R^2 = \sum_{i=1}^{N} \sigma_i^2$, i.e. the sum of variations with respect to the mean for each i-th logarithmic bin considered. Larger red dots represent mean values for each bin and error logarithmic bars are standard deviations. The fitting shown with the red continuous line is carried out by an exponential function. For Scheidegger networks draining on a line, the collapse for higher tributaries' areas has to be done with the deviation exponents from $H/1+H$ and d_f (Insets in *c)*), as observed before and as supported by our calculations (Figure B.2).

Note also that it is of great practical interest the fitting of the collapse functions $F_{ms}(x)$ and $F_b(x)$, in this case simply addressed via exponential functions (Figure 6.6). Although such choice clearly oversimplifies the fitting problem, the single parameter serves well our purposes and clearly represents the inverse of the characteristic value (of $d/a^{Hd_f/(1+H)}$, or $d/\langle a \rangle_a$ respectively) for cutoff to occur. Here we obtain $F_{ms}(x) \sim \exp(-2.58\mathrm{x})$ and $F_b(x) \sim \exp(-1.10\mathrm{x})$ respectively. These may indeed be useful for practical applications.

In Figure 6.6, we choose an exponential which would be the case should the probability obey $P(D \geq d) \propto P(A < a)^{d-1}$. We note that, as expected, the mainstream collapse function is generally shorter than in the boundary case, with higher exponents in semilog plot, because tributaries drain independently on both sides, whereas for line boundaries – all other conditions being equal – the distances are larger in the average. For Scheidegger's construct (*c)*) the exponent β on semilog plot is higher than the OCN one (*a)*) and lower than the Tanaro one (*b)*). The analytic form of the exponential function is of the type $F(x) = \exp(-\mathrm{Cx})$ where C is a suitable constant (reported in

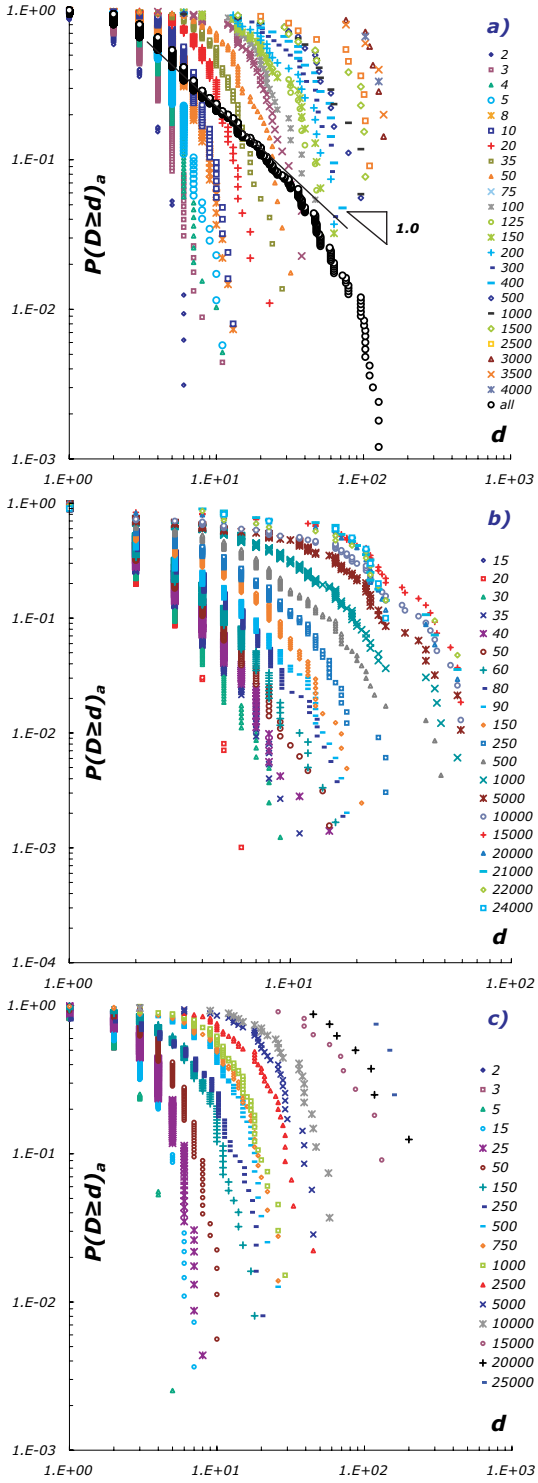

Figure 6.4: Exceedence probability distributions $P(D \geq d)_a$ of the random distance D between tributaries versus the current value d of distance, clearly a function of the threshold size a that defines a meaningful confluent tributary.

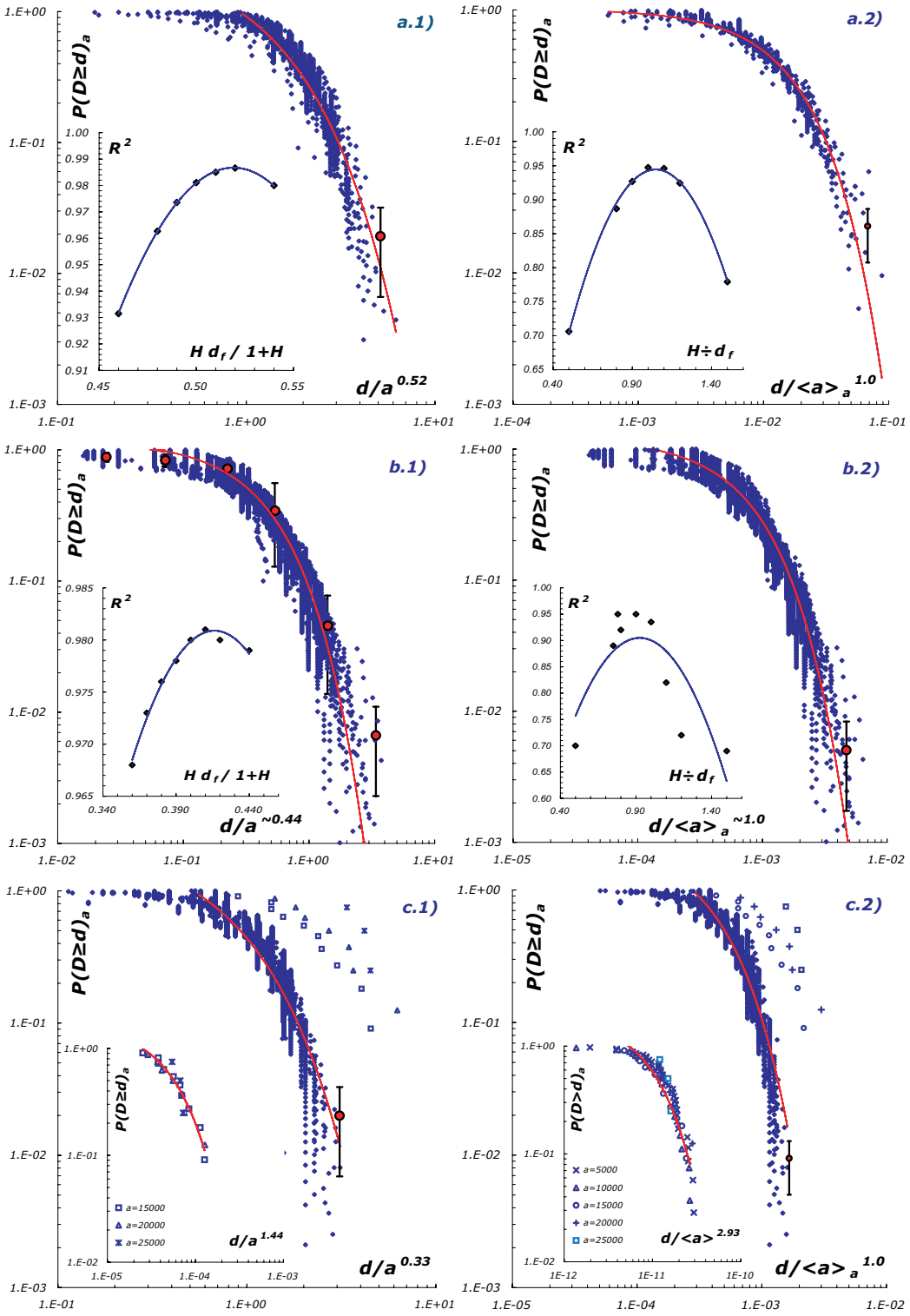

Figure 6.5: Collapse tests for $P(D \geq d)_a$.

each inset) where $x_c = 1/C$ is a cutoff value of the distribution of $d/a^{\frac{H}{1+H}}$. Note the finite-size effects for Scheidegger's networks (see also Figure B.2).

Figures 6.7 and 6.8 show the results of an articulate set of calculations required to describe the scaling of mean and variance of d given a, $\langle a \rangle$ for all networks shown in Figure 2.1, Chapter 2. Here we illustrate both the scaling of the second moment $\langle d^2 \rangle$ and of the variance $\sigma^2(d)_a = \langle d^2 \rangle_a - \langle d \rangle_a^2$ as a function of a. Note that the theoretical predictions (Equations 6.19 to 6.22) are reasonably satisfied regardless of the fractal properties of the network – noting that in the case of Scheidegger networks this represents quite a demanding test owing to its exact scaling coefficients ($H = 1/2$, $d_f = 1$, $\tau = 4/3$, $\tau_{ms} = 5/3$).

The validity of the finite-size scaling argument is strengthened by the analysis of the ratio of consecutive moments $\langle d^n \rangle_a / \langle d^{n-1} \rangle_a \propto a^{H/(1+H)}$ for different (integer) values of n, which probes the structure of the cutoff function $F_{ms}(x)$. Once such ratio is plotted against a in a double logarithmic graph (as in Figure 6.9, see *Rigon et al.* (1996) for computational details on the binning requirements for sampling the moments for several samples of area a), one might estimate the slope, which indeed should be equal to $H/(1+H)$ from Equation 6.16. The consistency of the ratio of several consecutive moments (here arbitrarily translated vertically for convenience) is a solid test for the scaling argument adopted.

Figure 6.10 illustrates an a posteriori test of the validity of the statistical independence assumption, i.e. $P(D \geq d)_a \propto P(A < a)^{d-1}$, and is carried out by plotting directly the probability of exceeding a distance d given the threshold area a with the theoretical value $(1 - P(A \geq a))^{d-1}$ valid under strict statistical independence. We note, that, as we expected, deviations appear both at large and small areas, thus denying generality to the position. Notice, however, that although the independence argument would have called for a scaling structure for $P(D \geq d)_a$ which is actually unsupported by our data, differences do not seem overwhelming. Note that all relationships between tributaries' areas and distances are then rightly derived through the finite-size scaling assumption rather than by the expression postulating statistical independence.

6.4 Discussion

The collapse of all conditional distributions of distance $P(D \geq d)_a$ when plotted against $d/a^{H/(1+H)}$ or $d/\langle a \rangle^{d_f}$ (depending on whether drainage is allowed along a mainstream or a multiple-outlet line boundary) defines a general and interesting ecological measure. As predicted by Equation 6.16, the distribution is shown to exhibit a finite-size scaling form that derives from the probability distribution of contributing area at any point along a given boundary or mainstream, and its related scaling exponents τ_{ms}, τ_b. The major effect of the constraints that can be imposed by the draining boundary is not surprising because areas tributary to mainstreams or boundaries define a different population of random variables.

Most important is deemed the "universal" behavior of $P(D \geq d)_a$ against d/a^{ψ} or, equivalently, $d/\langle a \rangle^{\phi}$ (with appropriate ψ, ϕ depending on B.C., see Table 6.1 for a complete comparison and a detailed summary of scaling relationships), where a, $\langle a \rangle$ indicate,

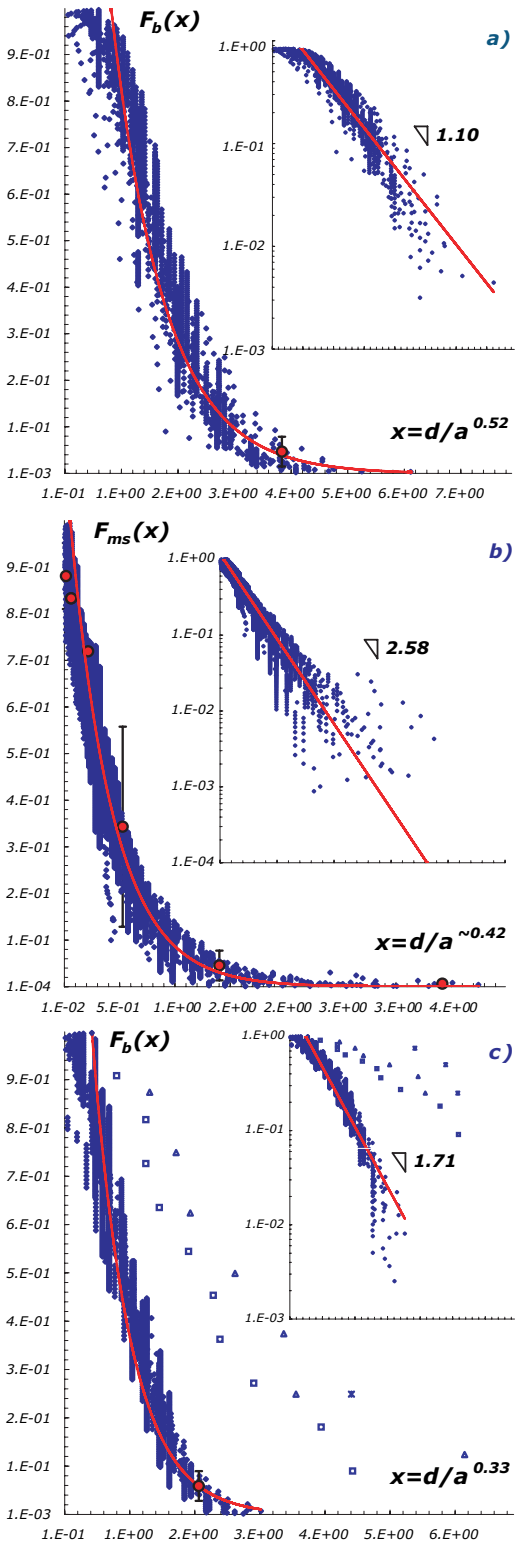

Figure 6.6: The detailed structure of the collapse function (i.e. homogeneity or scale function) $F_b(x)$ and $F_{ms}(x)$ plotted versus x, where x is the argument $d/a^{\frac{H}{1+H}}$.

respectively, the threshold area for a tributary and the mean area of all tributaries whose area is larger than a (see Figure 6.1). The collapse of all distributions that are clearly distinct for different values of a, onto a single universal curve suggests that the relationship $P(D \geq d)_a \sim F(d/a^{\psi})$ is valid for all areas a and thus seemingly of general nature (Figures 6.4 and 6.5). This is a distinctive statistical tool that has notable implications. In fact, the traditional geomorphological tenet that assumes link lengths as uniformly distributed (see e.g., *Leopold et al.*, 1964) is incompatible with our findings that pose a few conceptual and practical problems. Uniformly distributed link lengths do not seem to hold, in fact, on one hand because given a threshold area for the definition of a meaningful tributary, no observational dataset indeed conforms to some kind of uniform distribution – the cutoff is inevitably far from a step function. On the other hand, the distribution cannot escape the problem of the choice of a, which makes the very definition of link quite arbitrary otherwise. On this basis alone we suggest that the concept needs be readdressed in the present framework. Obviously this has implications for biodiversity and ecosystem services provided by the river basin, as the confluence of tributary areas is crucial to flow, width and riparian area at any basin site.

The analytical specification of the cutoff function $F(d/\langle a \rangle^{\phi})$ differs from multiple-outlet boundaries of mainstreams for the fact that allowing drainage from either sides naturally shortens the distance between comparable (or larger) confluences. We notice, however, that the scaling ansatz $d \propto a^{H/(1+H)}$ seems to work properly also in this case, thereby suggesting that the distribution of distances along a support draining from either side is a obtained by rescaling the analog for an open boundary by some constant (Figure 6.9). Our guess of analytic form of the scaling function is $F(x) = C \exp(-\beta \, \mathrm{x})$ where $\beta = 1/x_c$ is the exponent found via linear estimates on semi-log plot between $F(x)$ and x. In such case x_c acts as a cutoff value of the distribution of $d/\langle a \rangle^{\phi}$. The numerical estimates produced for C, x_c are described in Section 3 and are possibly general. Further confirmations from data will be sought elsewhere.

The computation of mean and variance of the empirical distances matches nicely our theoretical predictions. The mean distance between tributaries of area larger than a is definitely proportional to $\langle a \rangle^H$ (hence often $\langle d \rangle_a \approx \langle a \rangle$), whereas the mean distance plotted against a obeys the predicted scaling, resulting in $\langle d \rangle_a \sim a^{H/(1+H)} \approx a^{0.52 \pm 0.04}$ except for singular cases like Scheidegger's (where the exponent is close to the predicted value of 1/3 – see Appendix B). The theoretical predictions, based on two different arguments, capture well the structure of the distribution.

Overall the probabilistic structure of tributaries thus seems fully captured.

6.5 Conclusions

The following conclusions are worth mentioning:

- the probabilistic structure of the alongstream distance between tributaries of (any) comparable size a drawn from a river network has been defined theoretically and verified observationally. Its importance is linked to the ecological implications of the availability of flow, width and riparian area available at any stream site, all

bearing implications for the type of biodiversity and ecosystem services produced by the river basin;

- the pdf of distance depends on the threshold size needed to define a tributary – obviously if a unit area is employed, the distance turns out to be unit (i.e. the link length is also unit), and grows, in a statistically well-defined sense, as basins grow larger. Interestingly, however, when the distribution is plotted against d/a or $d/\langle a \rangle$ ($\langle a \rangle$ being the mean area of all basins larger than a) collapses onto a single curve. Notice that the dependence of link length on a proves that the traditional tenet that assumes uniform link length distributions is a spinoff of the rather arbitrary threshold for channelization used to extract fluvial patterns, and cannot hold conceptually as it depends on the choice of cumulated area that defines a tributary;

- the distribution of distances that separate a given average injection of flow, width or riparian area (i.e. confluent area) seems to represent more general a concept than that of link length. Our empirical findings support well the theoretical arguments. In particular, exactly solved network constructions like Peano or Scheidegger networks (whose merits and limits in describing the geometry of river networks are described elsewhere but are deemed noteworthy) have been extensively used to validate our findings;

- overall we have proposed an explicit form of the probabilistic structure of tributaries in the form $P(\geq d) = P(D \geq d)_a = \exp(-Cd/a^{H/(1+H)})$ where the constant C is dependent on the choice of boundary conditions (i.e. whether mainstream areas of multiple-outlet line sources are considered).

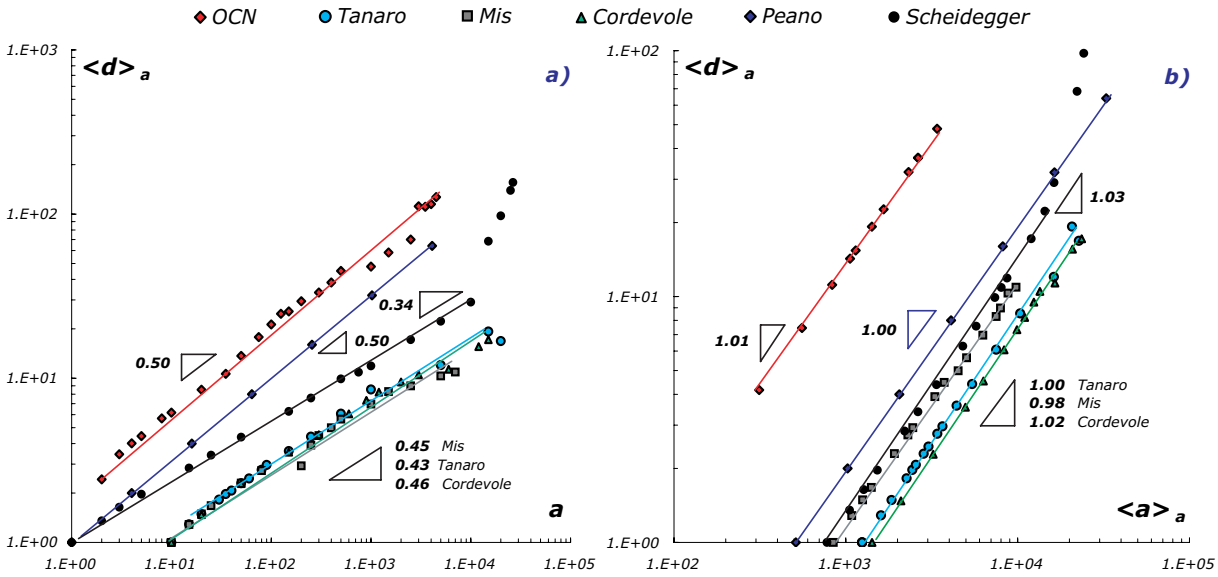

Figure 6.7: Scaling of the mean distance $\langle d \rangle_a$ with a and $\langle a \rangle_a$: left (a)), the scaling exponent $\frac{Hd_f}{1+H}$ is derived from the relation between $\langle d \rangle_a$ *vs* a; right (b)), we recover the exponents: d_f, in the case of straight draining boundary; H, in the case of a fractal mainstream.

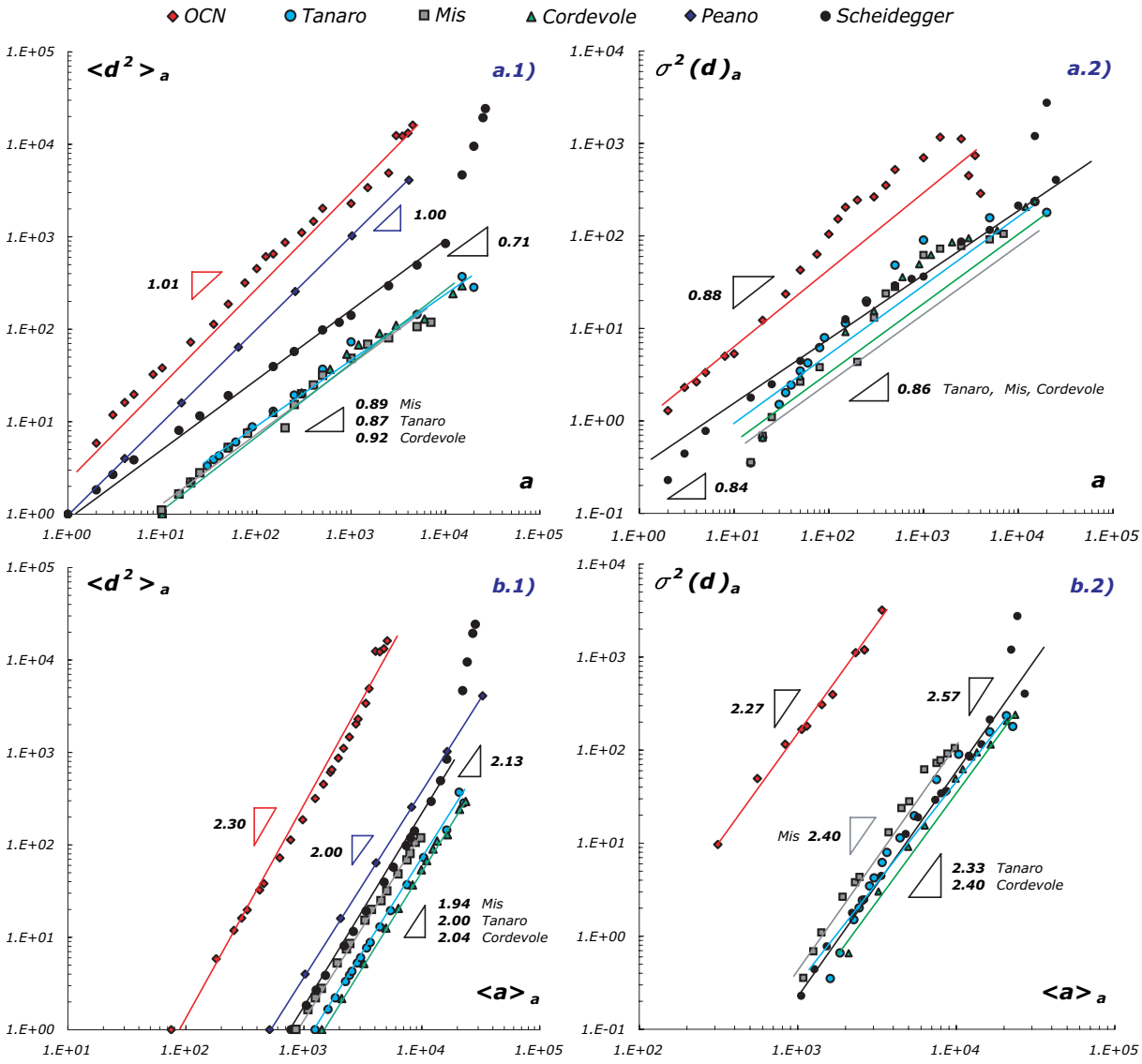

Figure 6.8: Scaling of the second moment, $\langle d^2 \rangle$, and variances, $\sigma^2(d)_a$, with a (*a.1)* and *a.2)*) and $\langle a \rangle_a$ (*b.1)* and *b.2)*), respectively. Note that the scaling exponents are affected by the cutoff dependence on L foreseen by Equations 6.20 and 6.21.

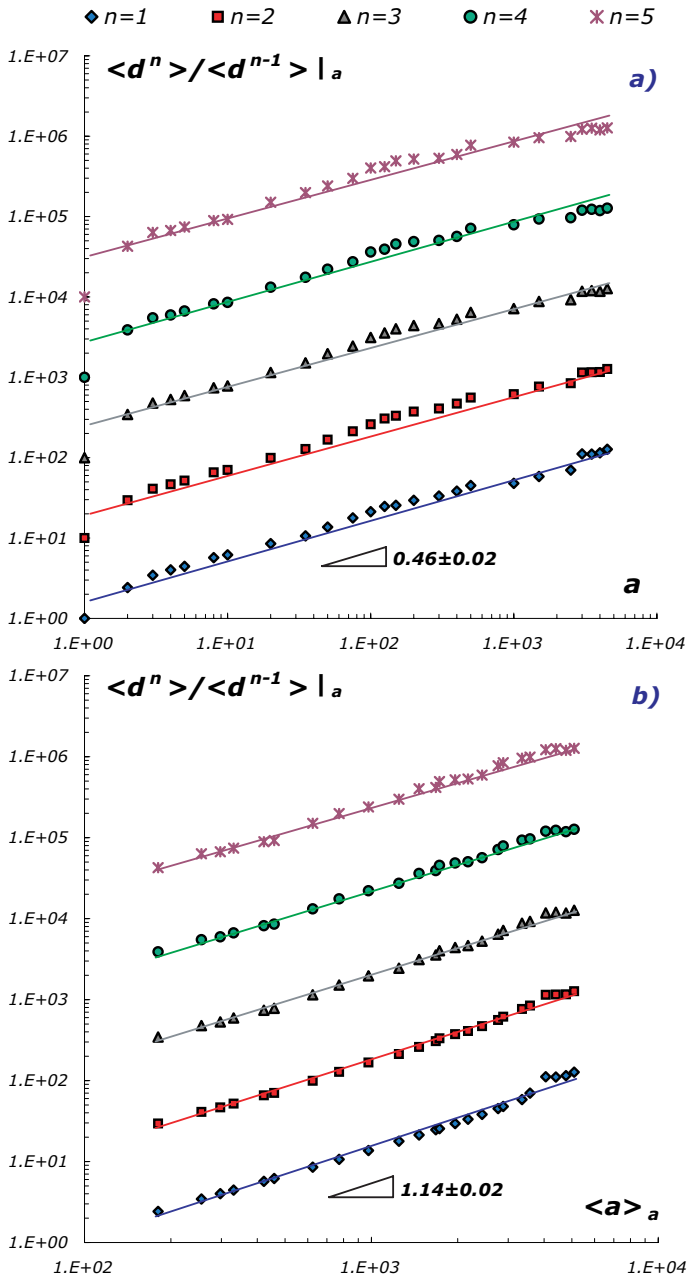

Figure 6.9: Scaling of the moments of the tributaries' distances for the OCN of Figure 6.1 *a.1)*, *a.2)*, and Figure 2.1 *a)* in Chapter 2. The plot shows the ratio $\langle d^n \rangle_a / \langle d^{n-1} \rangle_a$ where n is the moment order. The lowest curve is the moment of order one, the others show higher-order moments (up to 5-th order) which exhibit lower exponents than the meaningful but the stability results excellent. Note that the curves are conveniently offset vertically by powers of ten; *a)* shows the relation $\langle d^n \rangle_a / \langle d^{n-1} \rangle_a \propto a^{\frac{H d_f}{1+H}}$ with mean exponent of 0.46 and standard deviation of 0.02; *b)* shows the relation $\langle d^n \rangle_a / \langle d^{n-1} \rangle_a \propto \langle a \rangle_a^{d_f}$ with mean exponent equal to 1.14 ± 0.02.

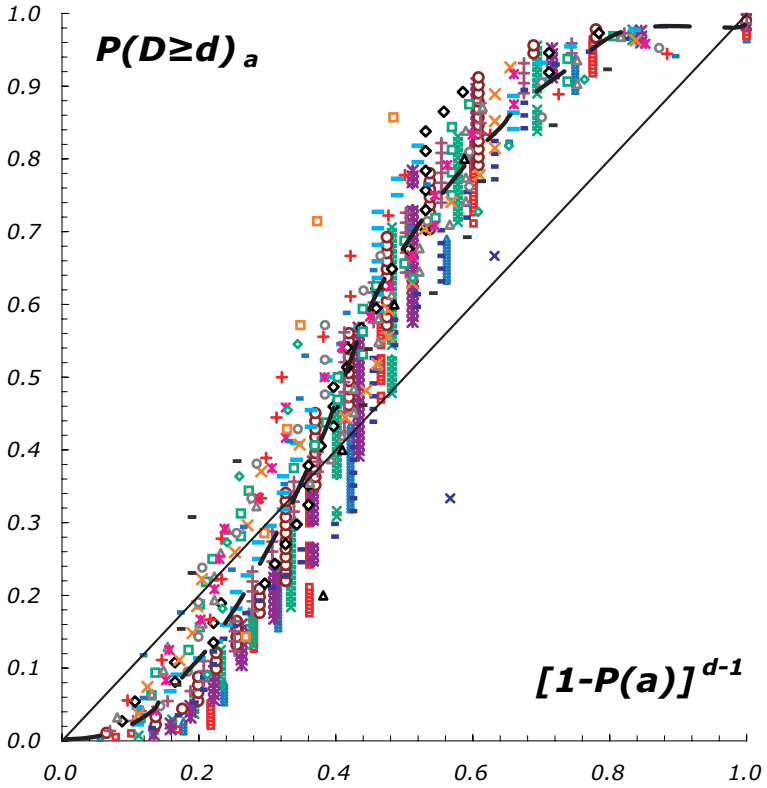

Figure 6.10: Independence test between probabilities of areas and distances. We have assumed the possibility to simplify the relation $P(D \geqslant d)_a \backsimeq (1 - P(A \geq a))^{(d-d_c)}\ \theta(d - d_c(a))$ assuming $d_c = 1$ and neglecting the nonlinear term $\theta(d, a)$ with d_c that can be approximated by the distance $d \propto a^{\frac{Hd_f}{1+H}}$. Based on the obtained curves we can infer the independence between areas along the mainstream at distances longer than $a^{\frac{Hd_f}{1+H}}$, so we have great areas with great distances only. In this case $P(A \geqslant a)$ is the exact value found on the complete exceedence probability distribution of areas (for $a = 1$), relative to each exceedence probability distribution of distances $P(D \geqslant d)_a$.

Chapter 7

Final Remarks

" I do not know what I may appear to the world; but to myself I seem to have been only like a boy playing on the seashore, and diverting myself in now and then finding a smoother pebble or a prettier shell than ordinary, whilst the great ocean of truth lay all undiscovered before me"
Isaac Newton

In the present thesis it has been tried to give a deep probabilistic description of how some biodiversity and geomorphological patterns change with scale and resolution. Some assumptions made can result too much simple but in order to give a test in their support empirical data have been used. For example while it is true that interactions between species exist, at the macroecological scale, that is not really a physical scale but rather a way of analysis, the neutral theory that assumed individuals competitively equivalent works and is able to reproduce the real biodiversity patterns. Fishes and trees patterns of the Mississippi-Missouri River System are correctly reproduced. This does not mean that interactions or other real processes does not matter, rather the study of ecosystems and its management can be performed in a simple but efficient way using the neutral theory that is based only on a stochastic birth-death and dispersal dynamics.

Further investigations will explore more in details the role of the prevalent climate on the clustering patterns of the species selecting different ecosystem throughout the world or regions with homogeneous climate in an heterogeneous region like the MMRS. For example it is unclear how to explain that in subtropical forests the aggregation of trees is weaker for large diameter classes (*Li et al.*, 2009) while it is the contrary in a "globally temperate" region like the MMRS. Therefore the comparison should be done as much as possible at the same relevant scale of analysis since scale issues due to the different extension of the studied area can arise. E.g. the extension of the MMRS and its degree of heterogeneity vs the regions of investigation found in literature is extremely large.

Other studies will explore in more detail the role of the neutral model in the clumping of individuals. For example how do species aggregate in a hierarchical competition-colonization trade-off model that explicitly take in account the differences between them? While it is true that here the species have been analysed under the neutral assumption that captures the actual biodiversity patterns (among which also the spatial aggregation) it is proofed that in some cases neutral models are not able to reproduce the actual

observed biodiversity patterns especially at small scales or where the environmental fluctuations are big and fast.

A possible further direction of research can be the characterization of the fractal dimension of the boundaries of the patches formed by the aggregation of species. Since the distribution of the cluster size is a power-law we expect fractality for the boundaries of the patches. Therefore due to the resolution dependence of the cluster-size distribution we expect also a different fractal dimension at different resolutions. Further study will also be in the direction of characterizing the intra and inter-species distances between clusters as well the characterization of the individuals in the ecosystem in function of their ecological traits (e.g. the diameter) and in function of some geomorphological variables (e.g. the slope and the elevation). As said the neutral model is not a tool that neglect the existence of other factors such as geomorphological influences, biological interactions and environmental heterogeneities, rather it is a null model testing their respective importance without using them a priori, as niche models do, as determinants of the spatial patterns of species in ecosystems.

Appendix A

Simulations for the Tanaro river basin with $\theta = 80$ and exponential-Cauchy dispersal

Here the results of the simulations performed on the Tanaro basin shape, in both the RN and 2-D topologies, for the exponential-Cauchy kernel (case *b* in table 3.2 chapter 3) have been reported. Figures A.1, A.2, A.3, A.4, and A.5, report the biodiversity patterns of the power-law of the cluster-size, the species-area relationship, the range-rank curve, the rank-abundance curve, the pdf of the LSR, JSI and RSR respectively. All the comments about these results are in chapter 5.

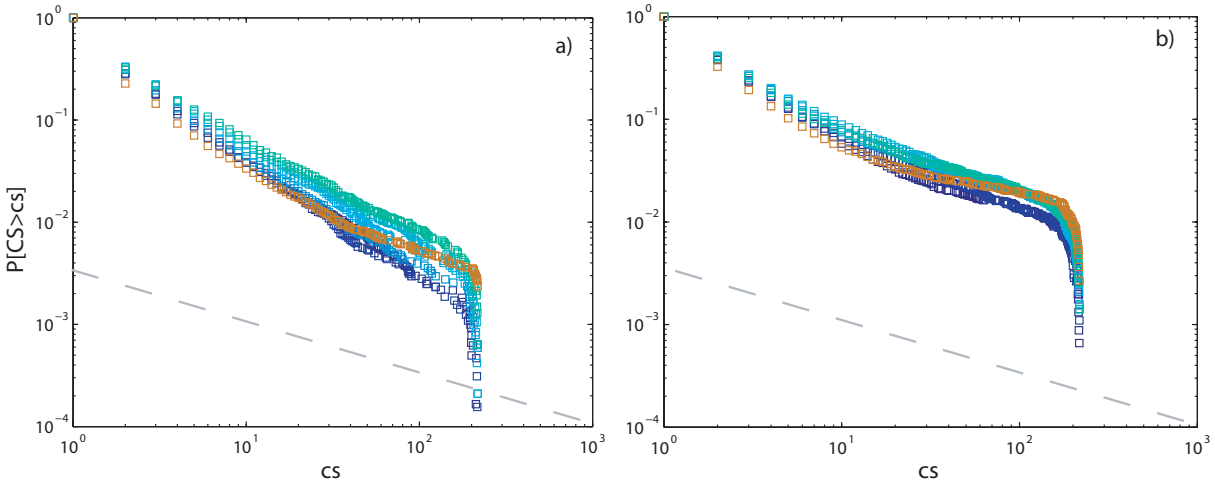

Figure A.1: Numerical results of $P(CS \geq c)$ for the Tanaro river basin with exponential-Cauchy kernel dispersal increasing d_l (see legend) and $\theta = 80$: a) RN, b) 2-D. The slope of the grey line is -0.5.

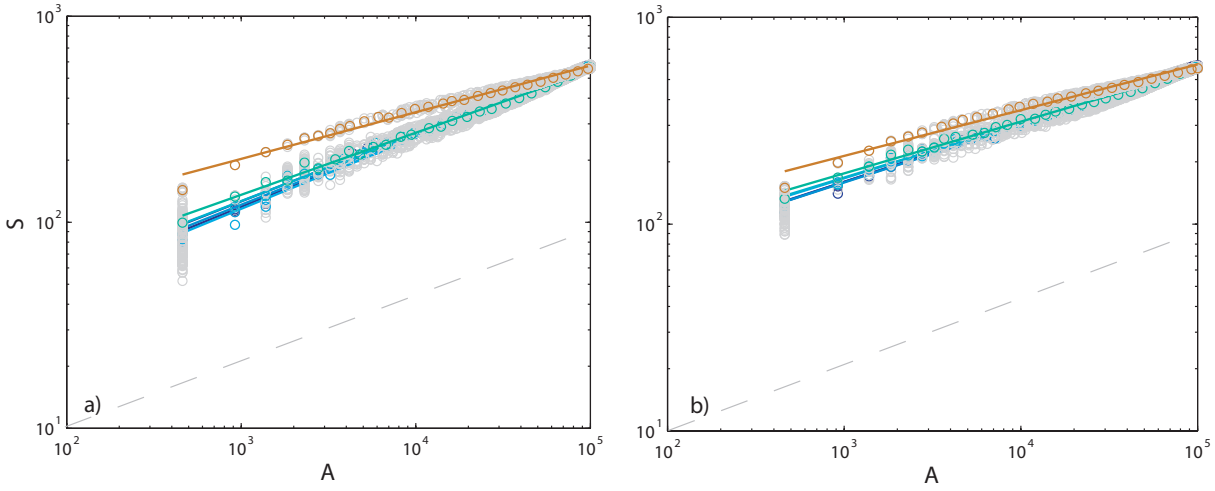

Figure A.2: Numerical results of the SAR for the Tanaro river basin with exponential-Cauchy kernel dispersal increasing d_l (see legend) and $\theta = 80$: a) RN, b) 2-D. The slope of the grey line is -0.5.

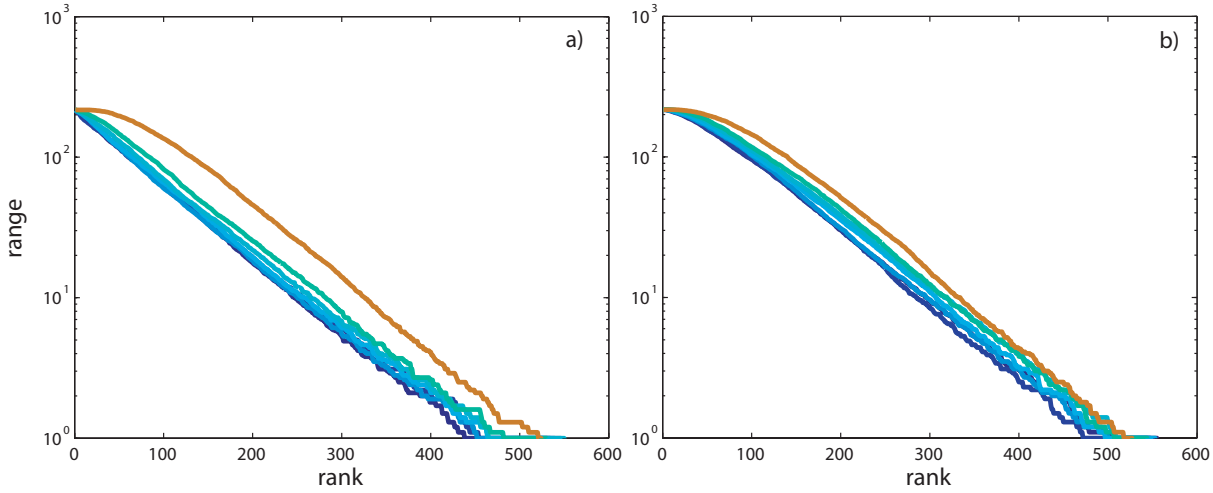

Figure A.3: Numerical results of the Range-Rank for the Tanaro river basin with exponential-Cauchy kernel dispersal increasing d_l (see legend) and $\theta = 80$: a) RN, b) 2-D.

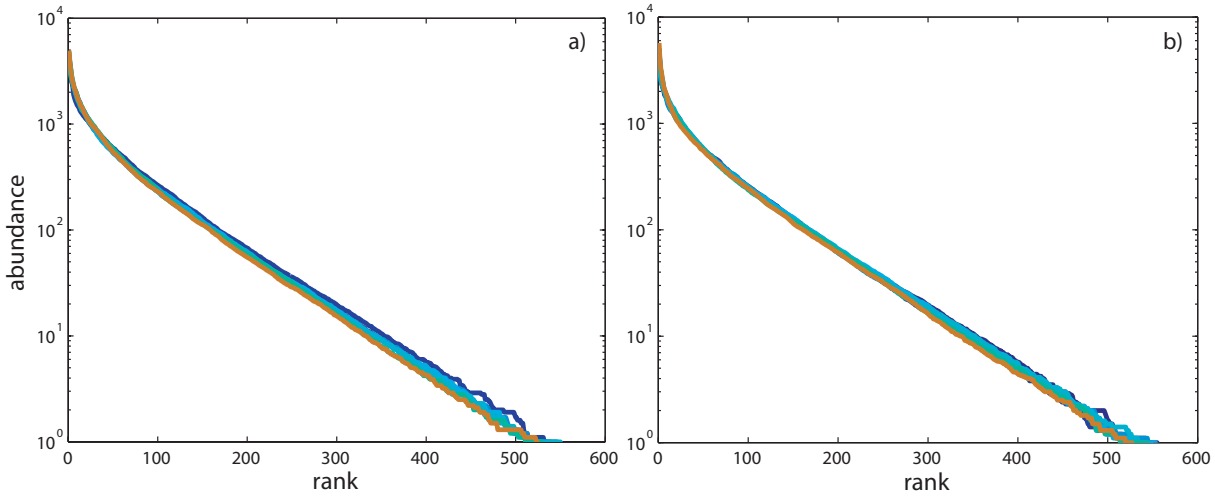

Figure A.4: Numerical results of the Rank-Abundance for the Tanaro river basin with exponential-Cauchy kernel dispersal increasing d_l (see legend) and $\theta = 80$: a) RN, b) 2-D.

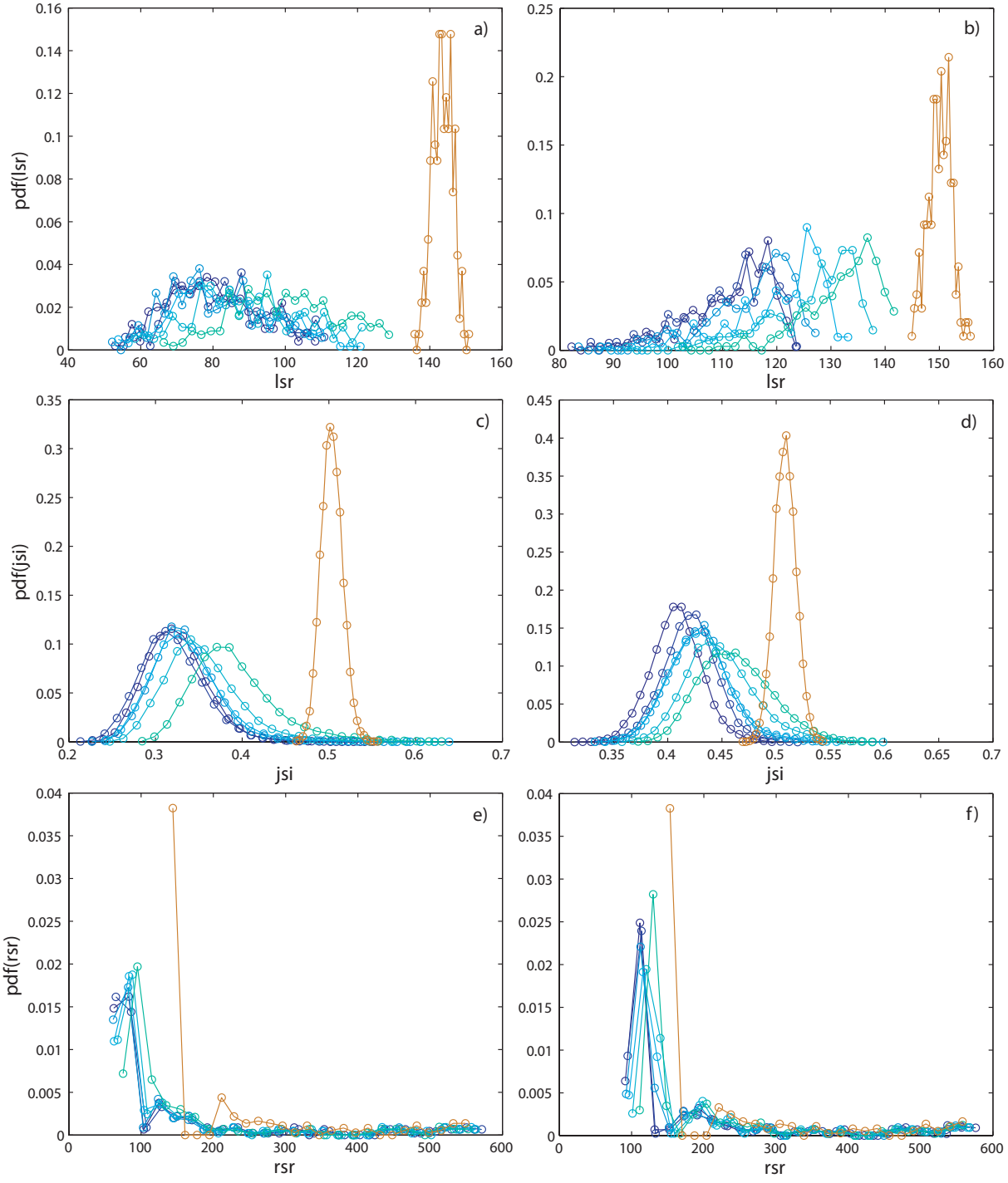

Figure A.5: Numerical results of the pdf of lsr (a, b), jsi (c, d) and rsr (e, f) for the Tanaro river basin with exponential-Cauchy kernel dispersal increasing d_l (see legend) and $\theta = 80$: RN (left plots); 2-D (right plots).

Appendix B

Scheidegger's construction and the distance between tributaries

Testing the theory against Scheidegger's construct proves particularly demanding and interesting, owing to the strong anisotropy that such structure produces, and to the slow convergence to exact power-law statistics (*Huber*, 1991; *Takayasu et al.*, 1991; *Rodriguez-Iturbe and Rinaldo*, 1997). It also relates to subtrees rooted along the primary path of a binary tree (*Troutman and Karlinger*, 1993). In this context, we have performed our analysis for different networks varying the ratio, say m, between the perpendicular and the parallel main axes. Specifically, we have carried out the same analysis on $\langle d \rangle_a$ as shown in Figure 6.7 *a)*, *b)*, and on $\sigma(d)$ in Figure 6.8 *a.2)*, *b.2)*. Figure B.1 shows, clockwise, Scheidegger's networks with $m = 1/10, 1, 10$.

It is clear that statistical measures of distances for the Scheidegger model are strongly dependent on the size of the lattice employed. It is relevant to note the deviation from the predicted value of the exponent $\frac{H}{1+H} = 1/3$ (Equation 6.22 and Figure B.2 *a.1)*) for $\langle d \rangle$ in function of the parameter a, and from the exponent $d_f = 1.00$ (Equation 6.22 and Figure B.2 *b.1)* for $\langle d \rangle$ in function of the mean area $\langle a \rangle_a$. Same deviations are found in the variance $\sigma(d)^2$ in function of a and $\langle a \rangle$ (Figure B.2 *a.2)* and *b.2)*). The slow logarithmic growth of Scheidegger's construct (*Huber*, 1991) is, in fact, equivalent to the use of a model with a high ratio m, which leads to incomplete aggregation. Several analysis have been performed on Scheidegger basins with very low value of m (for example using models with global area 1000×100 and 5000×100), and the statistics of areas at the boundary match the theoretical results. Otherwise exponents gradually deviate from the exact values (Figure B.2). The explanation for these deviation lies on the low value of the Hurst exponent characterizing Scheidegger's construct, $H = 1/2$. Only considering low values of m, in fact, or equivalently considering a very long network, it is possible to tend to reproduce the correct statistics between tributaries distances and areas, a phenomenon fully understood theoretically by *Huber* (1991). Therefore we note, also from Figure 6.2 *c)*, that boundary areas (whose distribution is a power law with exponent τ_b) and global areas (whose scaling exponent is τ) exhibit both the typical behavior as power laws, and are not susceptible to the level of growth of the stochastic construct. In this way only finite-size effects alter the scaling, and this happens is particularly significant for high values of m.

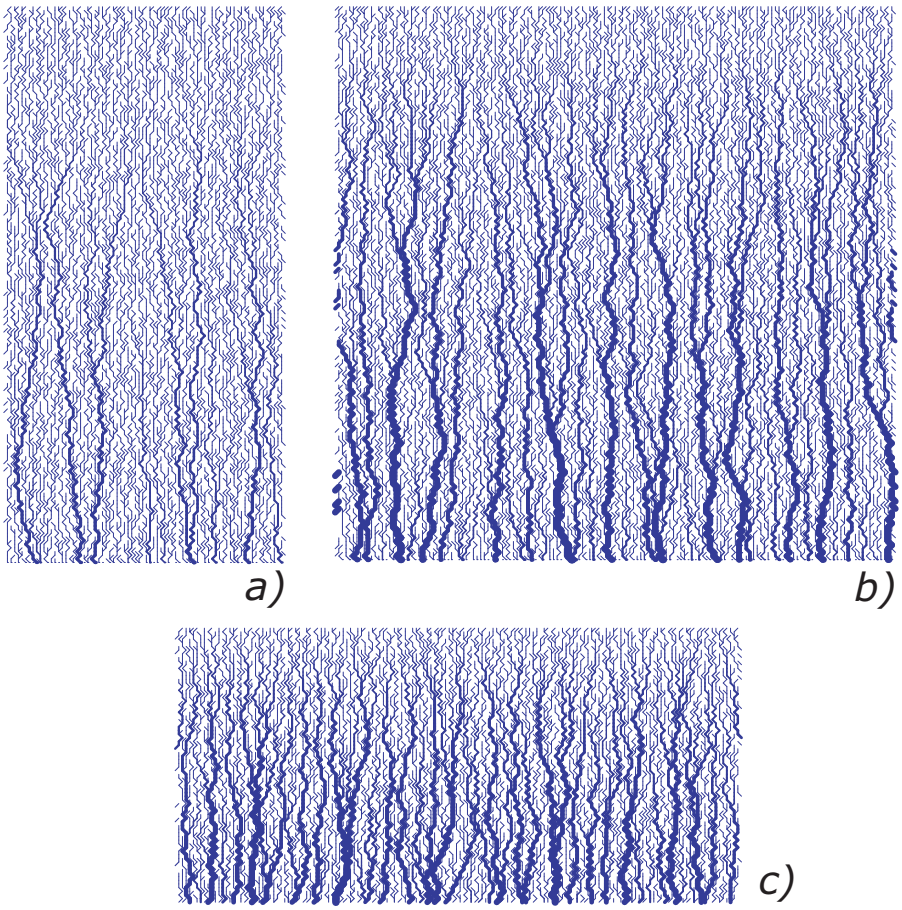

Figure B.1: Scheidegger networks arranged, clockwise, with increasing aspect ratios, that is, of perpendicular *vs* lateral dimension of the tributaries, specifically it is in growing order $m = L_{\perp}/L_{||} = 1/10,\ 1,\ 10$ *(a), b)* and *c)* respectively). Note that this geometric and topological arrangement is seen as a demanding test for the theoretical results owing to the pronounced anisotropy of the network forms (see Figure B.2).

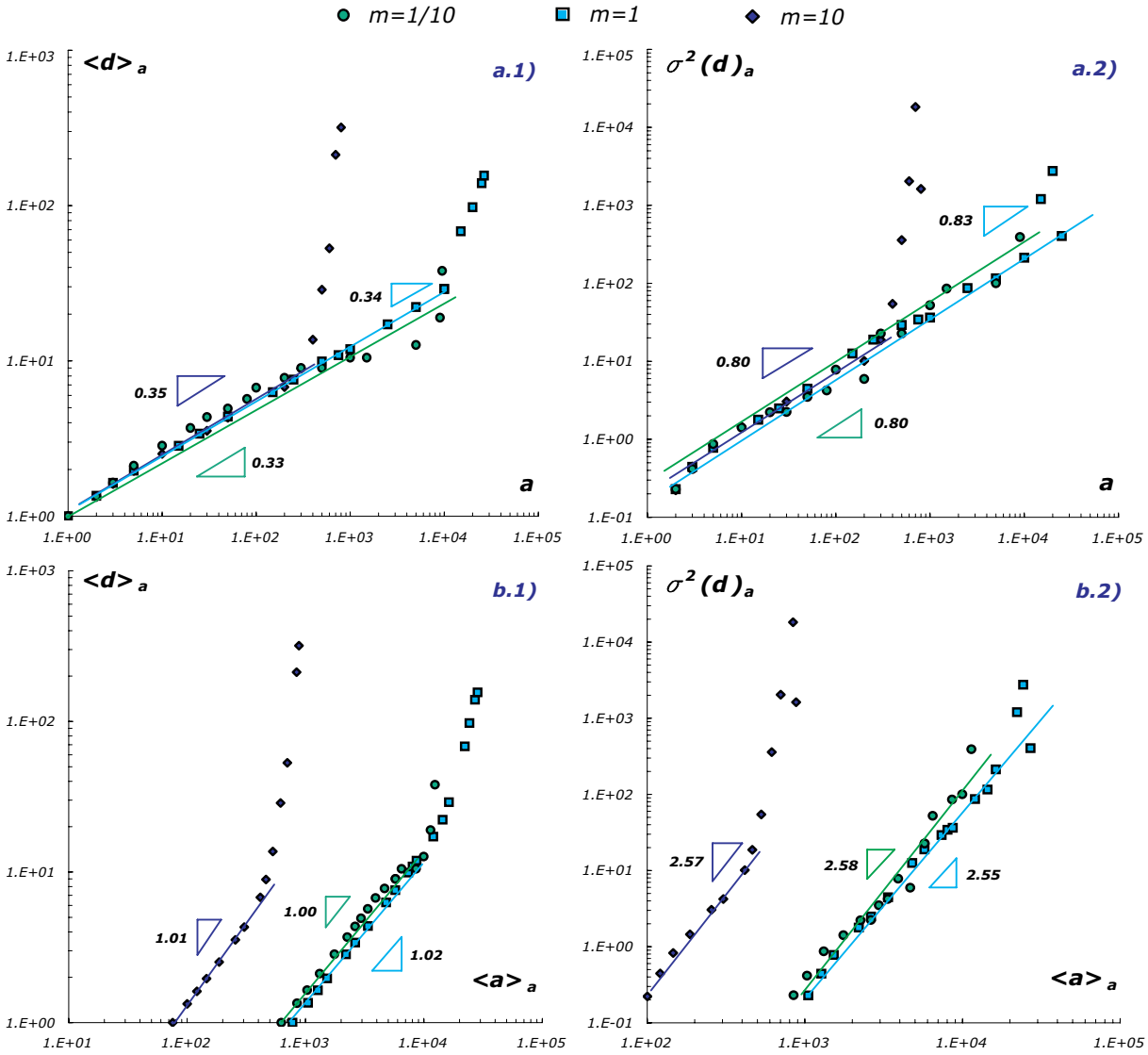

Figure B.2: Scaling relations between $\langle d \rangle_a$ and a, and $\langle a \rangle_a$, for the Scheidegger networks shown in Figure B.1. Here various aspect ratios are used, specifically in growing order with $m = L_\perp / L_{||} = 1/10,\ 1,\ 10$. The first case is perhaps the most appropriate to describe the exact results of the exponent related to the tributaries distribution, i.e. $\frac{Hd_f}{1+H} = 1/3$, because it eases the formation of subbasins with the exact value $H = 1/2$. Other lattice ratios lead to a deviation from the exact values of exponents and deviations appear as shown due to the slow logarithmic growth of the construct. The largest deviations relate to the highest value of m. *a.1)* and *b.1)* show the computational relations $\langle d \rangle_a$ *vs* a and $\langle d \rangle_a$ *vs* $\langle a \rangle_a$ respectively, *a.2)* and *b.2)* the scaling of their related variance.

Appendix C

List of Publications/Manuscripts/Funds

- Convertino, M., R. Rigon, A. Maritan, I. Rodriguez-Iturbe, and A. Rinaldo (2007), Probabilistic structure of the distance between tributaries of given size in river networks, *Water Resources Research*, 43, W11418, Nov., *10.1029/2007WR006176*;
- Convertino, M., R. Muneepeerakul, S. Azaele, E. Bertuzzo, A. Rinaldo, and I. Rodriguez-Iturbe (2009), On neutral metacommunity patterns of river basins at different scales of aggregation, *Water Resources Research*, 45, W08424, Aug., *10.1029/2009WR007799*;
- Convertino, M., R. Muneepeerakul, S. Azaele, M. Konar, A. Rinaldo, and I. Rodriguez-Iturbe (2009), Neutral Metacommunity Clustering and SAR: River Basin vs Landscape Species Patterns, *Water Resources Research*, in submission;
- Stark, C.P., E. Choi, and M. Convertino, Landslide rupture and length-depth scaling, *Earth and Planetary Science Letters*, submitted.

I wish to formally acknowledge the fundings that allowed my research all along the PhD program at the University of Padova and during my research periods at Princeton University. The following National Science Foundation projects, in which I was an official active collaborator, are kindly acknowledged: (i) "Co-Organization of River Basin Geomorphology and Vegetation (PI: Ignacio Rodriguez-Iturbe; Awardee: Princeton University; Award Number: 0642517) at Princeton University, Jan.-Dec. 2008; (ii) "An Exploration of the Role of Mountain River Sinuosity in Landscape Dynamics" (PI: Colin Stark; Awardee: Columbia University; Award Number: 0617557), June-Aug. 2009; (iii) "Collaborative Research: Geomorphic transport laws, landscape evolution, and fractional calculus" (PI: Colin Stark; Awardee: Columbia University; Award Number: 0823953), June-Aug. 2009. Additionally I would like to thank the European Union AQUATERRA fundings ("Integrated modeling of the river-sediment-soil-groundwater

system, advanced tools for the management of catchment areas and river basins in the context of global change" of the International Research Program Global Change and Ecosystems-EU n GOCE-CT-2003-505428; PIs: Andrea Rinaldo and Marco Marani; Awardee: Department IMAGE and "Dino Tonini" International Center for Hydrology - University of Padova, Italy) for my individual departmental projects: (i) "Analysis of the geomorphological properties relevant of the transport phenomena on large scale in river basins"; (ii) "Theoretical and modeling study of the spatial distribution and biodiversity of freshwater fishes in fluvial networks". Scholarships from INPDAP (Department of Education, Italy), years 2007, 2008 and 2009, from Regione Veneto/University of Padova, academic years 2006/2007, 2007/2008, 2008/2009, and from "Ing. Aldo Gini" Foundation at the University of Padova for study and research abroad (year 2008 at the Civil and Environmental Engineering Department at Princeton University) are gratefully acknowledged.

Bibliography

Alerstam, T., A. Hedenstrom, and S. Akesson (2003), Long-distance migration: evolution and determinants, *Oikos*, *103*.

Alexander, R. B., R. A. Smith, and G. E. Schwarz (2000), Effect of stream channel size on the delivery of nitrogen to the Gulf of Mexico, *Nature*, *403*, 758–761, doi: 10.1038/35001562.

Ammerman, A., and L. Cavalli-Sforza (1984a), *Neolithic Transition and the Genetic of Population in Europe*, Princeton University Press, USA.

Ammerman, A., and L. Cavalli-Sforza (1984b), *Neolithic Transition and the Genetic of Population in Europe*, Princeton University Press, USA.

Angermeier, P. L., and M. R. Winston (1998), Local vs. regional influences on local diversity in stream fish communities of Virginia, *Ecology*, *79*(3), 911–927.

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

Azaele, S. (2006), Stochastic equations for the evolution of ecosystems, Ph.D. thesis, Università di Padova, Advisor: Prof. A. Maritan.

Azaele, S., S. Pigolotti, J. R. Banavar, and A. Maritan (2006), Dynamical evolution of ecosystems, *Nature*, *444*, 926–928, doi:10.1038/nature05320.

Azaele, S., R. Muneepeerakul, A. Maritan, A. Rinaldo, and I. Rodriguez-Iturbe (2009), Predicting spatial similarity of freshwater fish biodiversity, *Proceedings of the National Academy of Science*, *106*, 7058–7062, doi:10.1073/pnas.0805845106.

Bak, P., K. Chen, and C. Tang (1990), A forest-fire model and some thoughts on turbulence, *Physics Letters A*, *147*, 297–300, doi:10.1016/0375-9601(90)90451-S.

Ball, P. (1999), *The Self-Made Tapestry: Pattern Formation in Nature*, Oxford University Press.

Banavar, J. R., J. L. Green, J. Harte, and A. Maritan (1999), Finite size scaling in ecology, *Phys. Rev. Lett.*, *83*(20), 4212–4214.

Banavar, J. R., F. Colaiori, A. Flammini, A. Maritan, and A. Rinaldo (2000), Topology of the Fittest Transportation Network, *Physical Review Letters*, *84*, 4745–4748, doi: 10.1103/PhysRevLett.84.4745.

Banavar, J. R., F. Colaiori, A. Flammini, A. Maritan, and A. Rinaldo (2001), Scaling, optimality and landscape evolution, *J. Stat.Phys.*, *104*, 1–33.

Banavar, J. R., J. Damuth, A. Maritan, and A. Rinaldo (2007), Scaling in Ecosystems and the Linkage of Macroecological Laws, *Physical Review Letters*, *98*(6), doi:10.1103/PhysRevLett.98.068104.

Bangert, R. K., E. V. Lonsdorf, S. M. Shuster, D. Fisher, J. A. Schweitzer, J. K. Bailey, and T. G. Whitman (2008), Genetic structure of a foundation species: scaling community phenotypes from the individual to the region, *Heredity*, *100*, 121–131, doi:10.1038/sj.hdy.6800914.

Bartumeus, F., and S. A. Levin (2008), Movement Ecology Special Feature: Fractal reorientation clocks: Linking animal behavior to statistical patterns of search, *Proceedings of the National Academy of Science*, *105*, 19,072–19,077, doi:10.1073/pnas.0801926105.

Battin, T. J., L. A. Kaplan, J. Denis Newbold, and C. M. E. Hansen (2003), Contributions of microbial biofilms to ecosystem processes in stream mesocosms, *Nature*, *426*, 439–442.

Battin, T. J., L. A. Kaplan, S. Findlay, C. Hopkinson, E. Marti, A. Packman, J. Newbold, and F. Sabater (2008), Biophysical controls on organic carbon fluxes in fluvial networks, *Nat. Geosci.*, *1*, 95–100.

Bechtold, W. (2003), FIA-sampling and plot design fact sheet, *http://fia.fs.fed.us/library/fact-sheets/p3-factsheets/tree-growth.pdf*.

Bell, G. (2001), Neutral Macroecology, *Science*, *293*, 2413–2418.

Benda, L., L. Poff, D. Miller, T. Dunne, G. Reeves, G. Pess, and M. Pollock (2004), The Network Dynamics Hypothesis: How Channel Networks Structure Riverine Habitats, *BioScience*, *54*, 413–427.

Bertuzzo, E., A. Maritan, M. Gatto, I. Rodriguez-Iturbe, and A. Rinaldo (2007), River networks and ecological corridors: reactive transport on fractals, migration fronts, hydrochory, *Water Resources Research*, *41*(W04419), doi:10.1029/2006WR005533.

Bertuzzo, E., S. Azaele, A. Maritan, M. Gatto, I. Rodriguez-Iturbe, and A. Rinaldo (2008), On the space-time evolution of a cholera epidemic, *Water Resources Research*, *44*(W01424), doi:10.1029/2007WR006211.

Bertuzzo, E., R. Muneepeerakul, H. Lynch, I. Rodriguez-Iturbe, and A. Rinaldo (2009), On the geographic range of freshwater fish in river basins, *Water Resources Research*, doi:10.1029/2009WR007997, in press.

Bissonette, J. (1997), *Wildlife and Landscape Ecology: Effects of Pattern and Scale*, Springer.

Blake, S., S. L. L. Deem, E. Mossimbo, F. Maisels, and P. Walsh (2009), Forest elephants: Tree planters of the congo, *Biotropica*, *41*(4), 459–468, doi:10.1111/j.1744-7429.2009.00512.x.

Boer, M., R. Sadler, R. Bradstock, A. Gill, and P. Grierson (2008), Spatial scale invariance of southern australian forest fires mirrors the scaling behavior of fire-driving weather events, *Landscape Ecology*, *23*.

Bohrer, G., R. Nathan, and S. Volis (2005), Effects of long-distance dispersal for metapopulation survival and genetic structure at ecological time and spatial scales, *Journal of Ecology*, *93*, 1029–1040, doi:10.1111/j.1365-2745.2005.01048.x.

Borgono, F., P. D'Odorico, F. Laio, and L. Ridolfi (2009), Mathematical models of vegetation pattern formation in ecohydrology, *Rev. Geophys.*, in press.

Botter, G., E. Bertuzzo, A. Bellin, and A. Rinaldo (2005), On the lagrangian formulations of reactive solute transport in the hydrologic response, *Water Resour. Res.*, *41*(W04008), doi:10.1029/2004WR003544.

Bowman, J., A. Jaeger, and L. Fahrig (2002), Dispersal distance of mammals is proportional to home range size, *Ecology*, (83(7)), 2049–2055.

Boyer, D., G. Ramos-Fernández, O. Miramontes, J. L. Mateos, G. Cocho, H. Larralde, H. Ramos, and F. Rojas (2006), Scale-free foraging by primates emerges from their interaction with a complex environment., *Proceedings. Biological sciences / The Royal Society*, *273*(1595), 1743–1750, doi:10.1098/rspb.2005.3462.

Broquet, T., and E. J. Petit (2009), Molecular estimation of dispersal for ecology and population genetics, *Annual Review of Ecology, Evolution, and Systematics*, *40*(1), doi:10.1146/annurev.ecolsys.110308.120324.

Brown, K. A., and J. Gurevitch (2004), Long-term impacts of logging on forest diversity in Madagascar, *Proceedings of the National Academy of Science*, *101*, 6045–6049.

Bullock, J., R. Kenward, and R. Hails (2002), *Dispersal Ecology: 42nd Symposium of the British Ecological Society*, Cambridge University Press.

Burkman, B. (2005), FIA-sampling and plot design fact sheet, *http://fia.fs.fed.us/library/fact-sheets/data-collections/SamplingandPlotDesign.pdf*.

Burness, G. P., J. Diamond, and T. Flannery (2001), Dinosaurs, dragons, and dwarfs: The evolution of maximal body size, *Proceedings of the National Academy of Science*, *98*, 14,518–14,523.

Burns, J., and J. Thomson (2006), A test of spatial memory and movement patterns of bumblebees at multiple spatial and temporal scales, *Behavioral Ecology*, *17*, 48–55, doi:10.1093/beheco/arj002.

Burt, W. (1943), Territoriality and home range concepts as applied to mammal, *Journal of Mammology*, (24), 346–352.

Cain, M., B. Milligan, and A. Strand (2000), Long-distance seed dispersal in plant populations, *American Journal of Botany*, *87*, 1217–1227.

Camporeale, C., and L. Ridolfi (2006), Riparian vegetation distribution induced by river flow variability: a stochastic approach, *Water Resources Research*, *42*(W10415), doi: 10.1029/2006WR004933.

Campos, D., J. Fort, and V. Mendez (2006), Transport on fractal river networks: Application to migration fronts, *Theoretical Population Biology*, *69*, 88–93, doi: 10.1016/j.tpb.2005.09.001.

Caylor, K. K., and H. H. Shugart (2006), Pattern and process in savanna ecosystems, in *Dryland Ecohydrology*, edited by P. D'Odorico and A. Porporato, pp. 259–282, Springer-Verlag.

Caylor, K. K., T. M. Scanlon, and I. Rodriguez-Iturbe (2004), Feasible optimality of vegetation patterns in river basins, *Geophysical Research Letter*, *31*, 13,502–+, doi: 10.1029/2004GL020260.

Caylor, K. K., S. Manfreda, and I. Rodriguez-Iturbe (2005), On the coupled geomorphological and ecohydrological organization of river basins, *Advances in Water Resources*, *28*, 69–86, doi:10.1016/j.advwatres.2004.08.013.

Chave, J., H. C. Muller-Landau, and S. A. Levin (2002), Comparing classical community models: Theoretical consequences for patterns of diversity, *American Naturalist*, *159*, 1–23.

Clark, J. S., M. Silman, R. Kern, E. Macklin, and J. Hillerislambers (1999), Seed dispersal near and far: Patterns across temperate and tropical forests, *Ecology*, *80*(5), 1475–1494, doi:10.2307/176541.

Clauset, A., C. Rohilla Shalizi, and M. E. J. Newman (2007), Power-law distributions in empirical data, *ArXiv e-prints*.

Colaiori, F., A. Flammini, A. Maritan, and J. R. Banavar (1997), Analytical and numerical study of optimal channel networks, *Phys. Rev. E*, *55*(2), 1298–1310, doi: 10.1103/PhysRevE.55.1298.

Colwell, R. K. (2000), A barrier runs through it ... or maybe just a river, *Proceedings of the National Academy of Science*, *97*, 13,470–13,472, doi:10.1073/pnas.250497697.

Condit, R., S. P. Hubbell, J. V. Lafrankie, R. Sukumar, N. Manokaran, R. B. Foster, and P. S. Ashton (1996), Species-Area and Species-Individual Relationships for Tropical Trees: A Comparison of Three 50-ha Plots, *Journal of Ecology*, *84*, 549–562.

Condit, R., P. S. Ashton, P. Baker, S. Bunyavejchewin, S. Gunatilleke, N. Gunatilleke, S. P. Hubbell, R. B. Foster, A. Itoh, J. V. Lafrankie, H. S. Lee, E. Losos, N. Manokaran, R. Sukumar, and T. Yamakura (2000), Spatial patterns in the distribution of tropical tree species, *Science*, *288*(5470), 1414–1418.

Connell, J. (1971), On the role of natural enemies in preventing competitive exclusion in some marine animals and in rain forest trees, in *Dynamic of Population*, edited

by P. den Boer and G. Gradwell, Proc. Advanced Study Institute on Dymamics of Numbers in Populations, Oosterbeek. Centre for Agricultural Publishing, and Documentation.

Connor, E., A. Courtney, and J. Yoder (2000), Individuals-area relationships: the relationship between animal population density and area, *Ecology*, *81*(3), 734–748.

Convertino, M., R. Rigon, A. Maritan, I. Rodriguez-Iturbe, and A. Rinaldo (2007), Probabilistic structure of the distance between tributaries of given size in river networks, *Water Resour. Res.*, *43*, W11,418, doi:10.1029/2007WR006176.

Convertino, M., R. Muneepeerakul, S. Azaele, E. Bertuzzo, A. Rinaldo, and I. Rodriguez-Iturbe (2009), On neutral metacommunity patterns of river basins at different scales of aggregation, *Water Resour. Res.*, *45*, W08,424, doi:10.1029/2009WR007799.

Crawley, M. J., and J. E. Harral (2001), Scale dependence in plant biodiversity, *Science*, *291*(5505), 864–868, doi:10.1126/science.291.5505.864.

Crocker, W. (1938), Life-span of seeds, *The Botanical Review*, *4*, 235–274.

Damschen, E. I., L. A. Brudvig, N. M. Haddad, D. J. Levey, J. L. Orrock, and J. J. Tewksbury (2008), Movement Ecology Special Feature: The movement ecology and dynamics of plant communities in fragmented landscapes, *Proceedings of the National Academy of Science*, *105*, 19,078–19,083, doi:10.1073/pnas.0802037105.

Davies, S., A. White, and A. Lowe (2004), Long-distance dispersal of plants, *Heredity*, *93*, 566–576, doi:doi:10.1038/sj.hdy.6800555.

Dieckmann, U., B. O'Hara, and W. Weisser (1998), The evolutionary ecology of dispersal, *Tech. rep.*, Adaptive Dynamics Netwok Project.

Dornelas, M., S. R. Connolly, and T. P. Hughes (2006), Coral reef diversity refutes the neutral theory of biodiversity, *Nature*, *440*, 80–82, doi:10.1038/nature04534.

Duncan, R. (1993), Testing for life historical changes in spatial patterns of four tropical trees species in Westland, New Zealand, *Journal of Ecology*, *81*, 403–416.

Economo, E., and K. T.H. (2007), Species diversity in neutral metacommunities: a network approach, *Ecology Letters*, *11*, 52–62, doi:10.1111/j.1461-0248.2007.01126.x.

Edmunds, P., and J. Bruno (1996), The importance of sampling scale in ecology: kilometer-wide variation in coral reef communities, *Marine Ecology Progress Series*, *143*, 165–171.

Edwards, C., and P. Bohlen (1977), *Biology and Ecology of Earthworms*, Chapman and Hall, London.

Etienne, R. S., and D. Alonso (2007), Neutral Community Theory: How Stochasticity and Dispersal-Limitation Can Explain Species Coexistence, *Journal of Statistical Physics*, *128*, 485–510, doi:10.1007/s10955-006-9163-2.

Fagan, W. (2002), Connectivity, fragmentation, and extinction risk in dendritic metapopulations, *Ecology*, *83*.

Fargione, J., C. S. Brown, and D. Tilman (2003), Community assembly and invasion: An experimental test of neutral versus niche processes, *Proceedings of the National Academy of Science*, *100*, 8916–8920.

Fernandes, C. C., J. Podos, and J. G. Lundberg (2004), Amazonian Ecology: Tributaries Enhance the Diversity of Electric Fishes, *Science*, *305*, 1960–1962, doi:10.1126/science.1101240.

Fisher, S. G., J. B. Heffernan, R. A. Sponseller, and J. R. Welter (2007), Functional ecomorphology: Feedbacks between form and function in fluvial landscape ecosystems, *Geomorphology*, *89*, 84–96, doi:10.1016/j.geomorph.2006.07.013.

Fort, J., and V. Méndez (1999), Time-Delayed Theory of the Neolithic Transition in Europe, *Physical Review Letters*, *82*, 867–870, doi:10.1103/PhysRevLett.82.867.

Garber, P. (1989), Role of spatial memory in primate foraging patterns: Saguinus mystax and Saguinus fuscicollis, *American Journal of Primatology*, *19*, 203–216, doi:10.1002/ajp.1350190403.

Garcia-Martin, H., and N. Goldenfeld (2006), On the origin and robustness of power-law species-area relationships in ecology, *PNAS*, *103*(27), 10,310–10,315, doi:10.1073/pnas.0510605103.

Gascon, C., J. R. Malcolm, J. L. Patton, M. N. F. da Silva, J. P. Bogart, S. C. Lougheed, C. A. Peres, S. Neckel, and P. T. Boag (2000), Riverine barriers and the geographic distribution of Amazonian species, *Proceedings of the National Academy of Science*, *97*, 13,672–13,677.

Gastner, M., B. Oborny, D. Zimmermann, and G. Pruessner (2009), Transition from connected to fragmented vegetation across an environmental gradient: scaling laws in ecotone geometry, *The American Naturalist*, *174*.

Gaston, K. J. (2000), Global patterns in biodiversity, *Nature*, *405*, 220–227, doi:10.1038/35012228.

Gebert, W., A. Graczyk, and D. Krug (1987), *Runoff in the United States, 195180*, http://aa179.cr.usgs.gov/metadata/wrdmeta/runoff.htm.

Getz, W. M., and D. Saltz (2008), Movement Ecology Special Feature: A framework for generating and analyzing movement paths on ecological landscapes, *Proceedings of the National Academy of Science*, *105*, 19,066–19,071, doi:10.1073/pnas.0801732105.

Gewin, V. (2006), Beyond Neutrality–Ecology Finds Its Niche, *PLoS Biology*, *4*, 1306–1310.

Gilbert, E., and M. Singer (1975), Butterfly ecology, *Annual Review Ecological System*, *6*, 365–395.

Gillooly, J. F., and A. Allen (2007), Linking global patterns in biodiversity to evolutionary dynamics using metabolic theory, *Ecology*, *88*, 1890–1894.

Giplin, M., and I. Hanski (1991), *Metapopulation Dynamics: Empirical and Theoretical Investigations*, Academic Press, London.

Girdler, E. B., and B. Barrie (2008), The scale-dependent importance of habitat factors and dispersal limitation in structuring great lakes shoreline plant communities, *Plant Ecology*, *198*(2), 211–223, doi:10.1007/s11258-008-9396-z.

Gotelli, N. (2006), Null versus neutral models: what's the difference?, *Ecography*, *29*.

Gotelli, N. J. (2002), Ecology: Biodiversity in the scales, *Nature*, *419*, 575–576.

Grant, E., W. Lowe, and W. Fagan (2007), Living in the branches: population dynamics and ecological processes in dendritic networks, *Ecology Letters*, *10*.

Graves, G. R., and C. Rahbek (2005), Source pool geometry and the assembly of continental avifaunas, *Proceedings of the National Academy of Science*, *102*, 7871–7876.

Green, J. L., and B. J. M. Bohannan (2006), Spatial scaling of microbial biodiversity, *Microbial Ecology*, *21*(9), doi:10.1016/j.tree.2006.06.012.

Green, J. L., A. J. Holmes, M. Westoby, I. Oliver, D. Briscoe, M. Dangerfield, M. Gillings, and A. J. Beattie (2004), Spatial scaling of microbial eukaryote diversity, *Nature*, *432*, 747–750, doi:10.1038/nature03034.

Gregory, S., F. Swanson, W. McKee, and K. Cummins (1991), An ecosystem perspective of riparian zones, *Bioscience*, *41*(8), 540–551.

Guégan, J.-F., S. Lek, and T. Oberdorff (1998), Energy availability and habitat heterogeneity predict global riverine fish diversity, *Nature*, *39*, 382–384.

Guilhaumon, F., O. Gimenez, K. J. Gaston, and D. Mouillot (2008), Taxonomic and regional uncertainty in species-area relationships and the identification of richness hotspots, *Proceedings of the National Academy of Science*, *105*, 15,458–15,463, doi:10.1073/pnas.0803610105.

Hanski, I. (1999), *Metapopulation Ecology*, Oxford University Press, Oxford.

Hanski, I., and M. Giplin (1997), *Metapopulation biology : ecology, genetics and evolution*, Academic Press, San Diego.

Harms, K. E., S. J. Wright, O. Calderón, A. Hernández, and E. A. Herre (2000), Pervasive density-dependent recruitment enhances seedling diversity in a tropical forest, *Nature*, *404*, 493–495.

Harper, J. L., and D. L. Hawksworth (1994), Biodiversity: measurement and estimation, *Royal Society of London Philosophical Transactions Series B*, *345*, 5–12.

Harte, J., A. Kinzig, and J. Green (1999), Self-Similarity in the Distribution and Abundance of Species, *Science*, *284*, 334–+, doi:10.1126/science.284.5412.334.

Harte, J., E. Conlisk, A. Ostling, J. L. Green, and A. B. Smith (2005), A theory of spatial structure in ecological communities at multiple spatial scales, *Ecology Monographs*, *75*(2), 179–197, doi:10.1890/04-1388.

Hausdorf, B., and C. Henning (2007), Null model tests of clustering of species, negative co-occurrence patterns and nestedness in meta-communities, *Oikos*, *116*, 818–828.

He, F., and K. Gaston (2007), Estimating abundance from occurrence: an undetermined problem, *The American Naturalist*, *170*, 655–659.

He, F., and S. P. Hubbell (2003), Percolation theory for the distribution and abundance of species, *Phys. Rev. Lett.*, *91*(19), 198,103.

He, F., P. Legendre, C. Bellehumeur, and J. LaFrankie (1994), Diversity pattern and spatial scale: a study of a tropical forest of malaysia, *Environmental and Ecological Statistics*, *1*, 265–286.

Hein, S., B. Pfenning, T. Hovestadt, and H. Poethke (2004), Patch density, movements patterns, and realised dispersal distances in a patch-matrix landscape – a simulation study, *Ecological Modelling*, *174*, 411–420.

Holmes, E. (1993), Are diffusion model too simple? A comparison with telegraph models of invasion, *Am. Nat.*, *142*, 779–795.

Honnay, O., W. Verhaeghe, and M. Hermy (2001), Plant community assembly along dendritic networks of small forest streams, *Ecology*, *82*(6), 1691–1702.

Horner-Devine, M. C., M. Lage, J. B. Hughes, and B. J. M. Bohannan (2004), A taxa-area relationship for bacteria, *Nature*, *432*, 750–753, doi:10.1038/nature03073.

Houchmandzadeh, B. (2002), Clustering of diffusing organisms, *Phys. Rev. E*, *66*(5), doi:10.1103/PhysRevE.66.052902.

Houchmandzadeh, B. (2008), Neutral clustering in a simple experimental ecological community, *Physical Review Letters*, *101*(7), 078103.

Houchmandzadeh, B., and M. Vallade (2003), Clustering in neutral ecology, *Phys. Rev. E*, *68*(6), 061,912, doi:10.1103/PhysRevE.68.061912.

Hubbell, S. (2006), Neutral theory and the evolution of ecological equivalence, *Ecology*, *87*.

Hubbell, S. P. (2001), *The Unified Neutral Theory of Biodiversity and Biogeography*, Princeton University Press.

Huber, G. (1991), Scheidegger's rivers, Takayasu's aggregates and continued fractions, *Physica A Statistical Mechanics and its Applications*, *170*, 463–470, doi:10.1016/0378-4371(91)90001-S.

Hughes, B. (1995), *Random Walks and Random Environments*, Oxford University Press, Oxford.

Hyatt, L. A., M. S. Rosenberg, T. G. Howard, G. Bole, W. Fang, J. Anastasia, K. Brown, R. Grella, K. Hinman, J. P. Kurdziel, and J. Gurevitch (2003), The distance dependence prediction of the janzen-connell hypothesis: a meta-analysis, *Oikos*, *103*(3), 590–602, doi:10.1034/j.1600-0706.2003.12235.x.

ICZN (1999), International code of zoological nomenclature-glossary, *http://www.iczn.org/*.

Ims, R., and H. Andreassen (2005), Density-dependent dispersal and spatial population dynamics, *Proceedings of the Royal Society B*, *272*, 913–918, doi:10.1098/rspb.2004.3025.

Janzen, D. H. (1970), Herbivores and the number of tree species in tropical forests, *The American Naturalist*, *104*(940), 501–528, doi:10.2307/2459010.

Kéfi, S., M. Rietkerk, C. L. Alados, Y. Pueyo, V. P. Papanastasis, A. Elaich, and P. C. de Ruiter (2007), Spatial vegetation patterns and imminent desertification in Mediterranean arid ecosystems, *Nature*, *449*, 213–217, doi:10.1038/nature06111.

Kimura, M., and J. Crow (1964), The number of alleles that can be maintained in a finite population, *Genetics*, (49), 725–738.

Kirley, M. (2005), Competition, cooperation and collective behaviour: resource utilization in non-stationary environments, *Intelligent Agent Technology, IEEE / WIC / ACM International Conference*, *0*, 572–578, doi: \href{http://doi.ieeecomputersociety.org/10.1109/IAT.2005.55}{http://doi.ieeecomputersociety.org/10.1109/IAT.2005.55}.

Konar, M., R. Muneepeerakul, S. Azaele, E. Bertuzzo, A. Rinaldo, and I. Rodriguez-Iturbe (2009), Climate change will impact large scale patterns of tree diversity in the mississippi-missouri river system, in press.

Leibold, M. e. a. (2004), The metacommunity concept: a framework for multi-scale community ecology, *Ecology Letters*, *7*, 601–613, doi:10.1111/j.1461-0248.2004.00608.x.

Lensink, R., and M. Neubert (2003), Demography And Dispersal: Life Table Response Experiments For Invasion Speed, *Ecology*, *84*, 1968–1978.

Leopold, L., M. Wolman, and J. Miller (1964), *Fluvial Processes in Geomorphology*, W.H. Freeman, San Francisco, California.

Levin, S. A. (1992), The Problem of Pattern and Scale in Ecology: The Robert H. MacArthur Award Lecture, *Ecology*, *73*(6), 1943–1967, doi:10.1029/2007WR006100.

Levine, J. M. (2000a), Complex interaction in a streamside plant community, *Ecology*, *81*(12), 3431–3444.

Levine, J. M. (2000b), Species diversity and biological invasions: Relating local process to community pattern, *Science*, *288*, 852–854.

Levine, J. M. (2003), A patch modeling approach to the community-level consequences of directional dispersal, *Ecology*, *84*(5), 1215–1224.

Lewis, T. (2006), Climate change, species-area curves and the extinction crisis, *Philosophical Transactions of the Royal Society B*, *361*(1), 163–171.

Li, L., S. Wei, Z. Huang, W. Ye, and H. Cao (2008), Spatial patterns and interspecific association of threes canopy species at different life stages in a subtropical forest, china, *Journal of Integrative Plant Biology*, *50*, 1140–1150.

Li, L., Z. Huang, W. Ye, H. Cao, S. Wei, Z. Wang, J. Lian, I. Sun, K. Ma, and F. He (2009), Spatial distribution of tree species in a subtropical forest of china, *Oikos*, *118*, 495–502, doi:10.1111/j.1600-0706.2009.16753.x.

Löbel, S. (2009), Metapopulations and metacommunity processes, dispersal strategies and life-history trade-offs in epiphytes, Ph.D. thesis, Uppsala University.

Löbel, S., and H. Rydin (2009), Dispersal and life history strategies in epiphyte metacommunities: alternative solutions to survival in patchy, dynamic landscapes, *Oecologia*, *161*(3), 569–579.

Lowe, W. (2008), Linking movement behavior to dispersal and divergence in plethodontid salamanders, *Molecular Ecology*, *17*.

Lowe, W., G. Likens, and B. Cosentino (2006), Self-organization in streams: the relationship between movement behaviour and body condition in a headwater salamander, *Freshwater Biology*, *51*.

Lowrance, R., R. Todd, J. Fail Jr., O. Hendrickson Jr., R. Leonard, and L. Asmussen (1984), Riparian forest as nutrient filters in agricultural watershed, *BioScience*, *34*, 374–377.

MacArthur, R., and E. Wilson (1963), An equilibrium theory of island biogeography, *Evolution*, *17*, 373–387.

MacArthur, R. H., and E. O. Wilson (2001), *The Theory of Island Biogeography*, Princeton University Press.

Malanson, G. (1993), *Riparian Landscapes*, Cambridge University Press.

Mandelbrot, B. (1983), *The Fractal Geomtery of Nature*, W.H. Freeman, San Francisco, California.

Marani, A., R. Rigon, and A. Rinaldo (1991), A Note on Fractal Channel Networks, *Water Resources Research*, *27*, 3041–3049, doi:10.1029/91WR02077.

Marani, M., S. Lanzoni, D. Zandolin, G. Seminara, and A. Rinaldo (2002), Tidal meanders, *Water Resources Research*, *38*(11), 110,000–1, doi:10.1029/2001WR000404.

Marani, M., E. Belluco, A. D'Alpaos, A. Defina, S. Lanzoni, and A. Rinaldo (2003), On the drainage density of tidal networks, *Water Resources Research*, *39*(2), 020,000–1, doi:10.1029/2001WR001051.

Marani, M., S. Lanzoni, S. Silvestri, and A. Rinaldo (2004), Tidal landforms, patterns of halophytic vegetation and the fate of the lagoon of Venice, *Journal of Marine Systems*, *51*, 191–210.

Marba, N., C. M. Duarte, and S. Agusti (2007), From the Cover: Allometric scaling of plant life history, *Proceedings of the National Academy of Science*, *104*, 15,777–15,780, doi:10.1073/pnas.0703476104.

Marco, D., S. Cannas, M. Montemurro, B. Hu, and S. Cheng (2009), Comparable ecological dynamics underlie earlyu cancer invasion and species dispersal, involving self-organizing processes, *Journal of Theoretical Biology*, (256), 65–75.

Maritan, A., A. Rinaldo, R. Rigon, A. Giacometti, and I. Rodríguez-Iturbe (1996), Scaling laws for river networks, *Phys. Rev. E*, *53*(2), 1510–1515, doi:10.1103/PhysRevE.53.1510.

Matthews, W. (1998), *Patterns in Freshwater Fish Ecology*, Kluver Academic, Norwell, MA.

McClain, M. e. a. (2003), Biogeochemical Hot Spots and Hot Moments at the Interface of Terrestrial and Aquatic Ecosystems, *Ecosystems*, *6*, 301–312, doi:10.1007/s10021-003-0161-9.

Méndez, V., D. Campos, and S. Fedotov (2004), Analysis of fronts in reaction-dispersal processes, *Pysical Review E*, *70*(6), 066,129–+, doi:10.1103/PhysRevE.70.066129.

Méndez, V. m. c., D. Campos, and S. Fedotov (2004), Front propagation in reaction-dispersal models with finite jump speed, *Phys. Rev. E*, *70*(3), 036,121, doi:10.1103/PhysRevE.70.036121.

Mennell, K., and R. Scholes (Eds.) (2008), *Assessment of South African Elephant Management*, Witwatersrand University Press, Johannesburg.

Morrissey, M., and D. Kerckhove (2009), The maintenance of genetic variation due to asymmetric gene flow in dendritic metapopulations, *The American Naturalist*, *174*.

Mouillot, D., and K. J. Gaston (2007), Geographical range size heritability: what do neutral models with different modes of speciation predict?, *Global Ecology and Biogeography*, *16*, 367–380, doi:10.1111/j.1466-8238.2007.00292.x.

Muneepeerakul, R. (2007), Biodiversity in river networks: Theoretical and empirical perspectives, Ph.D. thesis, Princeton University, Advisor: Prof. I. Rodriguez-Iturbe.

Muneepeerakul, R., S. A. Levin, A. Rinaldo, and I. Rodriguez-Iturbe (2007a), On biodiversity in river networks: a trade-off metapopulation model and comparative analysis, *Water Resources Research*, *43*(W07426), doi:10.1029/2006WR005857.

Muneepeerakul, R., J. S. Weitz, S. A. Levin, A. Rinaldo, and I. Rodriguez-Iturbe (2007b), A neutral metapopulation model of biodiversity in river networks, *Journal of Theoretical Biology*, *245*(2), 351–363, doi:10.1016/j.jtbi.2006.10.005.

Muneepeerakul, R., E. Bertuzzo, H. J. Lynch, W. F. Fagan, A. Rinaldo, and I. Rodriguez-Iturbe (2008), Neutral metacommunity models predict fish diversity patterns in Mississippi-Missouri basin, *Nature*, *453*, 220–222, doi:10.1038/nature06813.

Muneepeerakul, R., A. Rinaldo, and I. Rodriguez-Iturbe (2008b), Patterns of vegetation biodiversity: The roles of dispersal directionality and network structure, *Journal of Theoretical Biology*, *252*(2), 221–229, doi:10.1016/j.jtbi.2008.02.001.

Naiman, R. J., and H. Décamps (1997), The ecology of interfaces: Riparian zones, *Annual Review of Ecology and Systematics*, *28*(1), 621–658, doi:10.1146/annurev.ecolsys.28.1.621.

Nathan, R. (2005), Long-distance dispersal reserach: building a network of yellow brick roads, *Diversity and Distribution*, *11*, 125–130.

Nathan, R. (2006), Long-distance dispersal of plants, *Science*, *313*(5788), 786–788, doi:10.1126/science.1124975.

Nathan, R., U. Safriel, and I. Noy-Meir (2001), Field Validation and Sensitivity Analysis of a Mechanistic Model for Tree Seed Dispersal by Wind, *Ecology*, *82*(2), 374–388.

Nathan, R., G. G. Katul, H. S. Horn, S. M. Thomas, R. Oren, R. Avissar, S. W. Pacala, and S. A. Levin (2002a), Mechanisms of long-distance dispersal of seeds by wind, *Nature*, *418*, 409–413.

Nathan, R., G. G. Katul, H. S. Horn, S. M. Thomas, R. Oren, R. Avissar, S. W. Pacala, and S. A. Levin (2002b), Mechanisms of long-distance dispersal of seeds by wind, *Nature*, *418*, 409–413.

Nathan, R., W. M. Getz, E. Revilla, M. Holyoak, R. Kadmon, D. Saltz, and P. E. Smouse (2008a), Movement Ecology Special Feature: A movement ecology paradigm for unifying organismal movement research, *Proceedings of the National Academy of Science*, *105*, 19,052–19,059, doi:10.1073/pnas.0800375105.

Nathan, R., F. M. Shurr, O. Spiegel, O. Steinitz, A. Trakhtenbrot, and A. Tsoar (2008b), Mechanisms of long-distance seed dispersal, *Trends in Ecology and Evolution*, *23*, 638–647.

NatureServe (2004), Distribution of native U.S. fishes by watershed, *http://www.natureserve.org/getData/dataSets/watershedHucs/index.jsp*.

Nee, S., A. Read, J. Greenwood, and P. Harvey (1991), The relationship between abundance and body size in British birds, *Nature*, *351*, 312–313, doi:10.1038/351312a0.

Newman, M. E. J. (2003), The structure and function of complex networks, *SIAM Review*, *45*, 167–256.

Nilsson, C., G. Grelsson, M. Johansson, and U. Sperens (1989), Patterns of plant species richness along riverbanks, *Ecology*, *70*(1), 77–84.

Nilsson, C., A. Ekblad, M. Dynesius, S. Backe, M. Gardfjell, B. Carlberg, S. Hellqvist, and R. Jansson (1994), A comparison of species richness and traits of riparian plants between a main river channel and its tributaries, *Journal of Ecology*, *82*(2), 281–295.

Nogués-Bravo, D., M. B. Araújo, T. Romdal, and C. Rahbek (2008), Scale effects and human impact on the elevational species richness gradient, *Nature*, *453*, 216–219, doi: 10.1038/nature06812.

Oberdorff, T., J.-F. Guégan, and B. Hugueny (1995), Global scale patterns of fish species richness in rivers, *Ecography*, *18*, 345–352.

Ostling, A. (2005), Ecology: Neutral theory tested by birds, *Nature*, *436*, 635–636, doi: 10.1038/436635a.

Pascual, M., R. Manojit, F. Guichard, and G. Flierl (2002), Cluster-size distributions: signatures of self-organization in spatial ecologies, *Phil. Trans. R. Soc. Lond. B (2002) 357, 657666*, *357*, 657–666, doi:10.1098/rstb.2001.0983.

Pearson, R., and T. Dawson (2004), Long-distance plant dispersal and habitat fragmentation: identifying conservation targets for spatial landscape planning under climate change, *Biological Conservation*, *123*, 389–401, doi:10.1016/j.biocon.2004.12.006.

Perona, P., C. Camporeale, E. Perucca, M. Savina, P. Molnar, P. Burlando, and L. Ridolfi (2009), Modelling river and riparian vegetation interactions and related importance for sustainable ecosystem management, *Aquatic Sciences*, *71*, 266–278.

Peterjohn, W., and D. Correll (1984), Nutrient Dynamics in an Agricultural Watershed: Observations on the Role of A Riparian Forest, *Ecology*, *65*, 1466–1475.

Pigolotti, S., and M. Cencini (2009), Speciation-rate dependence in species-area relationships, *Journal of Theoretical Biology*, *260*(1), 83–89.

Plotkin, J., and H. Muller-Landau (2002), Sampling the species composition of a landscape, *Ecology*, *83*.

Plotkin, J. B., M. D. Potts, N. Leslie, Manokaran, J. LaFrankie, and P. S. Ashton (2000), Species-area curves, spatial aggregation, and habitat specialization in tropical forests, *J. Theor. Biol.*, *207*, 81–99, doi:10.1006/jtbi.2000.2158.

Plotkin, J. B., J. Chave, and P. S. Ashton (2002), Cluster analysis of spatial patterns in malaysian tree species, *The American Naturalist*, *160*, 629–644, doi:\href{http://dx.doi.org/10.1086/342823}{http://dx.doi.org/10.1086/342823}.

Poole, G. C. (2002), Fluvial landscape ecology: addressing uniqueness within the river discontinuum, *Freshwater Biology*, *47*(4), 641–660, doi:10.1046/j.1365-2427.2002.00922.x.

Porporato, A., and I. Rodriguez-Iturbe (2002), Ecohydrology: a challenging multidisciplinary research perspective, *Journal des Sciences Hydrologiques*, *47*.

Potts, M., J. Plotkin, S. Lee, N. Manokaran, P. Ashton, and W. Bossert (2001), Sampling biodiversity: effect of plot shape, *The Malaysian Forester*, *64*.

Preston, F. (1948), The commonness, and rarity, of species, *Ecology*, *29*.

Purves, D. W., and S. W. Pacala (2005), Ecological drift in niche-structured communities: Neutral pattern does not imply neutral process, in *Biotic Interactions in the Tropics*, edited by D. Burslem, M. Pinard, and S. Hartley, pp. 107–138, Cambridge University Press.

Rahbek, C. (2005), The role of spatial scale and the perception of large-scale scpecies-richness patterns, *Ecology Letters*, *8*, 224–239, doi:10.1111/j.1461-0248.2004.00701.x.

Real, L. (1994), *Behavioral Mechanisms in Evolutionary Ecology*, University Of Chicago Press, USA.

Revilla, E., and T. Wiegand (2008), Movement Ecology Special Feature: Individual movement behavior, matrix heterogeneity, and the dynamics of spatially structured populations, *Proceedings of the National Academy of Science*, *105*, 19,120–19,125, doi: 10.1073/pnas.0801725105.

Rigon, R., I. Rodriguez-Iturbe, A. Maritan, A. Giacometti, D. G. Tarboton, and A. Rinaldo (1996), On Hack's law, *Water Resources Research*, *32*, 3367–3374, doi: 10.1029/96WR02397.

Rigon, R., I. Rodriguez-Iturbe, and A. Rinaldo (1998), Feasible optimality implies Hack's law, *Water Resources Research*, *34*, 3181–3190, doi:10.1029/98WR02287.

Rinaldo, A., A. Marani, and R. Rigon (1991), Geomorphological dispersion, *Water Resources Research*, *27*, 513–525, doi:10.1029/90WR02501.

Rinaldo, A., I. Rodriguez-Iturbe, R. Rigon, R. Bras, E. Ijjasz-Vasquez, and A. Marani (1992), Minimum Energy and Fractal Structures of Drainage Networks, *Water Resources Research*, *28*(9), 2183–2195.

Rinaldo, A., I. Rodriguez-Iturbe, R. Rigon, E. Ijjasz-Vasquez, and R. L. Bras (1993), Self-organized fractal river networks, *Phys. Rev. Lett.*, *70*(6), 822–825, doi:10.1103/PhysRevLett.70.822.

Rinaldo, A., I. Rodriguez-Iturbe, and R. Rigon (1998), Channel networks, *Ann. Rev. Earth Planet. Sci.*, *26*, 289–306, doi:10.1029/98WR02287.

Rinaldo, A., J. R. Banavar, and A. Maritan (2006), Trees, networks, and hydrology, *Water Resources Research*, *42*, 6–+, doi:10.1029/2005WR004108.

Ritchie, M. E., and H. Olff (1999), Spatial scaling laws yield a synthetic thoery of biodiversity, *Nature*, *400*, 557–560, doi:10.1038/23010.

Rodríguez-Iturbe, I., and A. Porporato (2005), *Ecohydrology of Water-controlled Ecosystems*, Cambridge University Press.

Rodriguez-Iturbe, I., and A. Rinaldo (1997), *Fractal River Basins: Chance and Self-Organization*, Cambridge University Press.

Rodriguez-Iturbe, I., , E. Ijjasz-Vasquez, , R. Bras, and D. Tarboton (1992), Power-law distribution of mass and energy in river basins, *Water Resources Research*, *28*, 988–993.

Rodríguez-Iturbe, I., A. Rinaldo, R. Rigon, R. L. Bras, A. Marani, and E. Ijjász-Vásquez (1992), Energy Dissipation, Runoff Production, and the Three-Dimensional Structure of River Basins, *Water Resources Research*, *28*, 1095–1103, doi:10.1029/91WR03034.

Rodriguez-Iturbe, I., R. Muneepeerakul, E. Bertuzzo, S. A. Levin, and A. Rinaldo (2009), River networks as ecological corridors: a complex systems perspective for integrating hydrologic, geomorphologic and ecologic dynamics, *Water Resour. Res.*, *45*, W01,413, doi:10.1029/2008WR007124.

Rosindell, J., and S. Cornell (2009), Species-area curves, neutral models, and long-distance dispersal, *Ecology*, *7*(90), 1743–1750.

Russo, S., M. Potts, S. Davies, and S. Tan (2007), Determinants of tree species distributions: Comparing the roles of dispersal, seed size, and soil specialization in a bornean rain forest, in *Seed Dispersal: Theory and its Application in a Changing World*, CAB International.

Savage, V. M., E. J. Deeds, and W. Fontana (2008), Sizing up allometric scaling theory, *PLoS Comput Biol*, *4*(9), e1000,171+, doi:10.1371/journal.pcbi.1000171.

Scanlon, T. M., K. K. Caylor, S. A. Levin, and I. Rodriguez-Iturbe (2007a), Positive feedbacks promote power-law clustering of Kalahari vegetation, *Nature*, *449*, 209–212, doi:10.1038/nature06060.

Scanlon, T. M., K. K. Caylor, S. A. Levin, and I. Rodriguez-Iturbe (2007b), Positive feedbacks promote power-law clustering of Kalahari vegetation, *Nature*, *449*, 209–212, doi:10.1038/nature06060.

Scheidegger, A. (1967), A stochastic model for drainage patterns inro an intramontane trench, *Bull. Assoc. Hydrol.*, *12*, 15–20.

Schmid, P. E., M. Tokeshi, and J. M. Schmid-Araya (2000), Relation Between Population Density and Body Size in Stream Communities, *Science*, *289*, 1557–1560.

Schurr, F., O. Spiegel, O. Steinitz, A. Trakhtenbrot, A. Tsoar, and R. Nathan (2009), Long-Distance Seed Dispersal, *Annual Plant Reviews*, *38*(7), 204–237.

Seidler, T., and J. Plotkin (2006), Seed dispersal and spatial pattern in tropical trees, *PLoS Biology*, *4*.

SFPRTD-EU (2009), SCALES: Securing the conservation of biodiversity across administrative levels and spatial, temporal, and ecological scales, *http://www.scales-project.net/*.

Solé, R., and B. Goodwin (2002), *Signs of Life: How Complexity Pervades Biology*, Basic Book.

Soons, M., and W. Ozinga (2005), How important is long-distance seed dispersal for the regional survival of plant species?, *Journal of Biogeography*, *11*, 165–172, doi: 10.1111/j.1366-9516.2005.00148.x.

Stenzel, L., G. Page, W. J.C., J. Warriner, D. George, C. Eyster, B. Ramer, and K. Neuman (2007), Survival and natal dispersal of juvenile snowy plovers (charadrius alexandrinus) in central coastal california, *The Auk*, *124*(3), 1023–1036.

Stoeckel, J., D. W. Schneider, L. Soeken, K. Blodgett, and R. Sparks (1997), Larval Dynamics of a Riverine Metapopulation: Implication for Zebra Mussel Recruitment, Dispersal, and Control in a Large-River System, *Journal of the North America Benthiological Society*, *16*, 586–601.

Storch, D., P. Marquet, and J. Brown (2007), *Scaling Biodiversity*, Scaling Biodiversity, Edited by David Storch, Pablo Marquet, and James Brown. Series: Ecological Reviews. Cambridge: Cambridge University Press, 2007.

Storch, D., D. Sizling, L. Arnost, J. Reif, J. Polechova, E. Sizlingova, and K. Gaston (2008), The quest for a null model for macroecological patterns: geometry of species distributions at multiple spatial scales, *Ecology Letters*, *11*.

Sutherland, G., A. Harestad, P. K., and K. Lertzman (2000), Scaling of natal dispersal distances in terrestrial birds and mammals, *Conservation Ecology*, *1*(4).

Takayasu, H., M. Takayasu, A. Provata, and G. Huber (1991), Statistical models of river networks, *J. Stat. Phys.*, *65*, 725–740.

Thomas, C. D., E. J. Bodsworth, R. J. Wilson, A. D. Simmons, Z. G. Davies, M. Musche, and L. Conradt (2001), Ecological and evolutionary processes at expanding range margins, *Nature*, *411*, 577–581.

Thomas, C. D., A. Cameron, R. E. Green, M. Bakkenes, L. J. Beaumont, Y. C. Collingham, B. F. N. Erasmus, M. F. de Siqueira, A. Grainger, L. Hannah, L. Hughes, B. Huntley, A. S. van Jaarsveld, G. F. Midgley, L. Miles, M. A. Ortega-Huerta, A. Townsend Peterson, O. L. Phillips, and S. E. Williams (2004), Extinction risk from climate change, *Nature*, *427*, 145–148.

Tilman, D., and J. A. Downing (1994), Biodiversity and stability in grasslands, *Nature*, *367*, 363–365, doi:10.1038/367363a0.

Tilman, D., and P. Kareiva (1997), *Neolithic Transition and the Genetic of Population in Europe*, Princeton University Press, USA.

Troutman, B., and M. Karlinger (1993), A note on subtrees rooted along the primary path of a binary trees, *Discuss. Appl. Math.*, *42*, 87–93.

USDA (2009), *United States Average Annual Precipitation, 19611990*, http://rnp782.er.usgs.gov/atlas2/mld/prism0p.html, NRCS-USDA Natural Resources Conservation Service.

USEPA-USGS (2006), National hydrography dataset plus, *http://www.horizon-systems.com/*, Herndon, Va.

USFS (2009), Forest cover types, *http://www.nationalatlas.gov/mld/foresti.html*.

USFS-FIDO (2009), Forest inventory data online, *http://fiatools.fs.fed.us/fido/index.html*.

Vandermeer, J., I. Perfecto, and S. Philpott (2008), Clusters of ant colonies and robust criticality in a tropical agroecosystem, *Nature*, *451*, 457–460.

Vandewoestijne, S., T. Martin, S. Liegeois, and M. Baguette (2004), Dispersal, landscape occupancy and population structure in the butterfy *Melanargia galathea*, *Basic and Applied Ecology*, *5*(1574), 581–591.

Volkov, I., J. R. Banavar, S. P. Hubbell, and A. Maritan (2007), Patterns of relative species abundance in rainforests and coral reefs, *Nature*, *450*, 45–49, doi:10.1038/nature06197.

Wang, R., L. Shi, and Q. Zheng (2008), Trade-off between reciprocal mutualists: local resource availability-oriented interaction in fig/fig wasp mutualism, *Journal of Animal Ecology*, *77*.

Ward, J. V., K. Tockner, C. Arscott, and C. Claret (2002), Riverine landscape diversity, *Freshwater Biology*, *47*, 517–539, doi:10.1046/j.1365-2427.2002.00893.

Warton, D., I. Wright, D. Falster, and M. Westboy (2006), Bivarite line-fitting method for allometry, *Biol. Rev.*, *81*, 259–291.

West, G. B., J. H. Brown, and B. J. Enquist (1999), The Fourth Dimension of Life: Fractal Geometry and Allometric Scaling of Organisms, *Science*, *284*, 1677–+, doi:10.1126/science.284.5420.1677.

White, E., S. Ernest, A. Kerkhoff, and B. Enquist (2007), Relationships between body size and abundance in ecology, *Trends in Ecology and Evolution*, *22*, 312–313.

Wiens, J. A. (1989), Spatial scaling in ecology, *Functional Ecology*, *3*, 385–397.

Wiens, J. A. (2002), Riverine landscapes: taking landscape ecology into the water, *Freshwater Biology*, *47*, 501–515, doi:10.1046/j.1365-2427.2002.00893.

Wilson, D. (1992), Complex interactions in metacommunities, with implications for biodiversity and higher levels of selection, *Ecology*, *73*, 1984–2000.

Wu, J. (2004), Effect of changing scale on landscape pattern analysis: scaling relations, *Landscape Ecology*, *19*, 125–138.

Xenopoulos, M. A., and D. M. Lodge (2006), Going with the flow: using species-discharge relationships to forecast losses in fish biodiversity, *Ecology*, *87*(8), 1907–1914.

Young, W. R., A. J. Roberts, and G. Stuhne (2001), Reproductive pair correlations and the clustering of organisms, *Nature*, *412*, 328–331, doi:10.1038/35085561.

Zinck, R., and V. Grimm (2009), Unifying wildfire models from ecology and statistical physics, *The American Naturalist*, *174*.