

# Scaling as a key conceptual tool for the interpretation of empirical and experimental patterns in ecology and biology

THÈSE N° 8719 (2018)

PRÉSENTÉE LE 18 JUILLET 2018

À LA FACULTÉ DE L'ENVIRONNEMENT NATUREL, ARCHITECTURAL ET CONSTRUIT  
LABORATOIRE D'ÉCOHYDROLOGIE  
PROGRAMME DOCTORAL EN GÉNIE CIVIL ET ENVIRONNEMENT

ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

POUR L'OBTENTION DU GRADE DE DOCTEUR ÈS SCIENCES

PAR

Silvia ZAOLI

acceptée sur proposition du jury:

Prof. A. Meibom, président du jury  
Prof. A. Rinaldo, Dr A. Giometto, directeurs de thèse  
Prof. R. Stocker, rapporteur  
Prof. F. Simini, rapporteur  
Prof. P. De Los Rios, rapporteur



ÉCOLE POLYTECHNIQUE  
FÉDÉRALE DE LAUSANNE

Suisse  
2018



It was six men of Indostan  
To learning much inclined,  
Who went to see the Elephant  
(Though all of them were blind),  
That each by observation  
Might satisfy his mind.  
The First approached the Elephant,  
And happening to fall  
Against his broad and sturdy side,  
At once began to bawl:  
"God bless me! but the Elephant  
Is very like a WALL!"  
The Second, feeling of the tusk,  
Cried, "Ho, what have we here,  
So very round and smooth and sharp?  
To me 'tis mighty clear  
This wonder of an Elephant  
Is very like a SPEAR!"  
The Third approached the animal,  
And happening to take  
The squirming trunk within his hands,  
Thus boldly up and spake:  
"I see," quoth he, "the Elephant  
Is very like a SNAKE!"  
The Fourth reached out an eager hand,  
And felt about the knee  
"What most this wondrous beast is like  
Is mighty plain," quoth he:  
"'Tis clear enough the Elephant  
Is very like a TREE!"  
The Fifth, who chanced to touch the ear,  
Said: "E'en the blindest man  
Can tell what this resembles most;  
Deny the fact who can,  
This marvel of an Elephant  
Is very like a FAN!"  
The Sixth no sooner had begun  
About the beast to grope,  
Than seizing on the swinging tail  
That fell within his scope,  
"I see," quoth he, "the Elephant  
Is very like a ROPE!"  
And so these men of Indostan  
Disputed loud and long,  
Each in his own opinion  
Exceeding stiff and strong,  
Though each was partly in the right,  
And all were in the wrong.

*The blind men and an elephant* by John Godfrey Saxe



# Acknowledgements

First, I wish to thank my supervisor Andrea Rinaldo for his steady support and for the encouragement he gave me, which helped me get through any difficulty I encountered. The enthusiasm he showed for my work motivated me throughout my thesis and I learned much from his experience. I also wish to thank him for building the better working environment I could have hoped for: stimulating, welcoming and fun. Secondly, I wish to thank my co-supervisor, colleague and friend Andrea Giometto. He was a guide and an example for me, stimulating me to do always better and to aim high. I also wish to thank Amos Maritan and Emilio Marañón, whose help and suggestions were precious for my thesis.

I thank all my colleagues (and former colleagues) for making the ECHO lab the most entertaining and enjoyable work place, for their friendship and support. Thank you Damiano, Joseph, Luca, Jonathan, Javier, Paolo, Mitra, Laura, Enrico, Flavio, Ana-Clara, Pierre, Théo, Anna, Bernard, Jean-Marc. I cannot thank enough the friends I met in Lausanne for making me feel at home in Switzerland and for sharing with me the pic-nics at the lake, the beers, the games and all the fun moments we spent together. Thank you Linda, Giulia, Chiara, Zimmer, Valentina, Valentina, Daniele, Mario, Giulia, Gözde.

I thank my parents for their support and encouragement and for the long trips to Lausanne they endured to bring me food and love.

Finally, I thank Pierre for being there for me, always.

*Lausanne, 20 June 2018*

S. Z.



# Abstract

This thesis is concerned with regular patterns found in ecology and biology, their linkages and the statistical description of their fluctuations around average trends. Among such patterns, often conforming to power laws, well-known examples include the Species-Area relationship (SAR), quantifying the increase of the number of species  $S$  inhabiting an ecosystem with ecosystem area, and the scale-invariant body size spectrum, routinely observed, e.g., in aquatic ecosystems. In biology, Kleiber's law is an allometric relationship describing how metabolic rates scale with an organism's body size.

While ecological laws have often been studied independently, simple heuristic reasonings show that they are linked. The need for a unifying effort in ecology, coherently synthesizing the vast and diverse set of empirical observations across scales, has been often voiced. However, a theoretical framework answering this need was still lacking. Furthermore, ecological variables are the result of the interplay between several stochastic ecological processes, and are therefore stochastic variables fluctuating around average values. Ecological and biological scaling laws typically make predictions for such averages, but the issue of fluctuations received scarce attention in the literature. Similarly, biological fluctuations have been typically neglected in the study of the size-scaling of metabolic rates, even though body sizes and metabolic rates may have a significant variability within a species. Fluctuations may be relevant to interpret empirical observations, judge the reliability of predictions and understand ecosystem dynamics.

An hypothesis for the distribution of abundances and body sizes of species inhabiting an ecosystem of finite area is proposed here. The hypothesis is inspired by finite-size scaling and is used to derive macroecological patterns and their linkages within a coherent theoretical framework. Stochastic models of community dynamics are used to support the hypothesis, and the derived linkages are tested on empirical datasets. Several stochastic models of community dynamics are also used here to study the fluctuations of  $S$  and how they scale with the average  $S$ . The intra-specific variability of metabolic rates and body sizes is investigated experimentally using freshwater phytoplankton species by nanoscale secondary ion mass spectrometry (NanoSIMS).

The linkages among ecological scaling laws predicted by the theoretical framework are verified in several empirical datasets. Theoretical generalizations including deviations from pure power-law behavior and heavy-tailed intra-specific size distributions are also addressed. The theoretical study of the relative scaling of the fluctuations of  $S$  with the mean  $\langle S \rangle$  in various community dynamics models shows that different ecological processes predict radically

## Acknowledgements

---

different fluctuations scalings, highlighting the need of empirical investigations to sort out which scenario applies to real ecosystems. Experiments on phytoplankton metabolic rate scaling with body size suggest that intra-specific metabolic rate distributions are described by a universal scaling form across different taxa and over three orders of magnitude in body size. This thesis, along with previous works, suggests that scaling concepts derived for inanimate systems can provide new insights into the dynamics of ecosystems and help unraveling regularities across scales of biological complexity.

Key words: Ecology, scaling laws, allometry, Species-Area relationship, community size-spectrum, macroecology, biodiversity, community dynamics, finite-size scaling, Kleiber's law, metabolic rate, body mass, intraspecific trait diversity



# Sommario

Questa tesi si occupa di pattern regolari osservati in ecologia e biologia, dei loro legami e della descrizione statistica delle loro fluttuazioni attorno a un trend medio. Tra tali pattern, che spesso seguono una legge di potenza, esempi noti includono la relazione Specie-Area (SAR), che quantifica l'aumento del numero di species  $S$  con l'area dell'ecosistema che occupano, e lo spettro di taglie corporee osservato regolarmente, ad esempio, in ecosistemi acquatici. In biologia, la legge di Kleiber è una relazione allometrica tra la taglia di un organismo e il suo tasso metabolico.

Nonostante le leggi ecologiche siano state spesso studiate indipendentemente, ragionamenti euristici dimostrano che sono legate. Il bisogno di un tentativo di unificazione in ecologia, che sintetizzi coerentemente le vaste e variegata osservazioni empiriche attraverso scale diverse, è stato spesso espresso, ma un framework teorico rispondente a questo bisogno mancava. Inoltre, le variabili ecologiche risultano dell'interazione di processi ecologici stocastici, e sono quindi variabili stocastiche che fluttuano attorno a un valor medio. Le leggi di scaling, generalmente, fanno previsioni per questo valor medio, ma le fluttuazioni hanno ricevuto scarsa attenzione in letteratura. Allo stesso modo, le fluttuazioni biologiche sono solitamente trascurate nello studio dello scaling del tasso metabolico con la taglia, nonostante entrambe le variabili abbiano una significativa variabilità all'interno di una specie. Le fluttuazioni possono essere rilevanti per interpretare le osservazioni empiriche, giudicare l'affidabilità delle previsioni e comprendere la dinamica dell'ecosistema.

Qui viene proposta un'ipotesi per la distribuzione di abbondanze e taglie di specie che vivono in un ecosistema di area finita. Tale ipotesi si ispira al finite-size scaling ed è usata per derivare i pattern macroecologici e il loro legame. Modelli stocastici di dinamica di comunità sono usati per dare supporto all'ipotesi. Diversi modelli di dinamica di comunità sono usati anche per studiare le fluttuazioni di  $S$  e il loro scaling con la media  $\langle S \rangle$ . La variabilità intra-specifica dei rate metabolici e delle taglie è investigata sperimentalmente usando specie di fitoplancton di acqua dolce tramite spettrometria di massa degli ioni secondari alla nanoscala (NanoSIMS).

I legami tra leggi di scala ecologiche predetti dal framework teorico sono verificati in diversi dataset empirici. Vengono affrontate generalizzazioni teoriche tra cui deviazioni da una pura legge di potenza e distribuzioni di taglie intraspecifiche 'heavy-tailed'. Lo studio teorico dello scaling delle fluttuazioni relative di  $S$  con la media  $\langle S \rangle$  in vari modelli di dinamica di comunità mostrano che diversi processi ecologici prevedono scaling delle fluttuazioni radicalmente diversi, sottolineando il bisogno di investigazioni empiriche per individuare quale scenario si applichi a ecosistemi reali. Gli esperimenti sullo scaling del tasso metabolico del fitoplancton

## Acknowledgements

---

con la taglia corporea suggeriscono che le distribuzioni intraspecifiche del tasso metabolico siano descritte da una forma di scaling universale attraverso taxa differenti e tre ordini di grandezza di taglia.

Questa tesi, insieme a lavori precedenti, suggerisce che i concetti di scaling derivati per sistemi inanimati possano fornire una nuova visione della dinamica degli ecosistemi e aiutare a scoprire regolarità attraverso scale di complessità biologica.

Parole chiave: Ecologia, leggi di scala, allometria, Species-Area Relationship, community size-spectrum, biodiversità, dinamica di comunità, finite-size scaling, legge di Kleiber, tasso metabolico, massa corporea, diversità dei tratti intraspecifica

# Contents

<b>Acknowledgements</b>	<b>i</b>
<b>Abstract (English/Italian)</b>	<b>iii</b>
<b>List of figures</b>	<b>xi</b>
<b>List of tables</b>	<b>xiii</b>
<b>Introduction</b>	<b>1</b>
<b>1 Covariations in ecological scaling laws</b>	<b>7</b>
1.1 Introduction . . . . .	7
1.2 Finite-size scaling . . . . .	12
1.2.1 Properties of Eq. (1.8) . . . . .	12
1.2.2 Data collapse . . . . .	14
1.3 Scaling Framework . . . . .	14
1.4 Derivation of ecological scaling laws . . . . .	16
1.4.1 The SAR and other scaling relationships obtained from moments of $p(n, m A)$ . . . . .	17
1.4.2 The community size spectrum . . . . .	19
1.4.3 The scaling of the maximum body mass . . . . .	19
1.4.4 Scaling of the largest sampled value with sample size for power-law distributions . . . . .	20
1.4.5 The RSA . . . . .	21
1.4.6 Some observations on the results . . . . .	22
1.4.7 Compatibility with previous works . . . . .	23
1.5 Stochastic models of community dynamics . . . . .	25
1.5.1 Basic community dynamics model . . . . .	27
1.5.2 Bounded multiplicative process . . . . .	30
1.5.3 Variation on the speciation dynamics . . . . .	31
1.5.4 Fluctuating number of species . . . . .	31
1.5.5 Value of $\eta$ in the community dynamics models . . . . .	32
1.6 Generalizations of the scaling framework . . . . .	33
1.6.1 Intra-specific size distribution . . . . .	33

## Contents

---

1.6.2	Kleiber's law with curvature . . . . .	37
1.6.3	Alternative hypotheses on $p(m A)$ . . . . .	39
1.6.4	Multiple trophic levels . . . . .	42
1.7	Discussion . . . . .	43
1.8	Appendix . . . . .	46
1.8.1	Relative species abundance (RSA) . . . . .	46
1.8.2	Community size-spectrum with intra-specific size-distribution . . . . .	47
1.8.3	Supplementary figures and tables . . . . .	49
<b>2</b>	<b>Ecological scaling laws in macroecological data</b>	<b>57</b>
2.1	Introduction . . . . .	57
2.2	Power-laws in empirical data . . . . .	60
2.2.1	Plotting a power-law . . . . .	60
2.2.2	Fitting a power-law . . . . .	62
2.2.3	Power-laws with cut-off at large value . . . . .	64
2.3	Test of the relationship $\eta = \gamma + \delta$ in tropical forests . . . . .	64
2.3.1	The datasets . . . . .	64
2.3.2	Estimating the scaling exponents $\delta$ and $\eta$ . . . . .	65
2.3.3	Estimating the scaling exponents $\gamma$ . . . . .	65
2.3.4	Intra-specific body mass distribution for trees . . . . .	68
2.3.5	Test of the linking relationship . . . . .	68
2.4	Test of equation $z = 1 - \Phi - \max\{0, \lambda(1 + \alpha - \eta)\}$ on a dataset of lizard species . . . . .	70
2.5	Test of the relationship $z = \xi(\delta - 1)$ on a dataset of mammal species . . . . .	71
2.6	Discussion . . . . .	71
<b>3</b>	<b>Fluctuations scaling of the species-area relation</b>	<b>77</b>
3.1	Introduction . . . . .	78
3.2	Species richness from minimalist models of community dynamics . . . . .	80
3.3	Neutral models with competition . . . . .	83
3.3.1	Multi-species Voter model (MSV) . . . . .	83
3.3.2	Neutral-Spatially Implicit model (N-SI) . . . . .	84
3.3.3	Comparison with the neutral model without competition . . . . .	84
3.4	Non-neutral models . . . . .	85
3.5	The role of perturbations . . . . .	89
3.6	Effect of habitat stochasticity in a metapopulation dynamics model . . . . .	93
3.7	Discussion . . . . .	96
3.8	Appendix . . . . .	99
3.8.1	Persistence time distribution . . . . .	99
3.8.2	Deterministic model of a system subject to perturbations . . . . .	99

<b>4 Single-cell metabolic rates size-scaling in freshwater phytoplankton</b>	<b>105</b>
4.1 Introduction . . . . .	105
4.2 Methods . . . . .	108
4.2.1 Generalized Kleiber's law . . . . .	108
4.2.2 Maintenance of phytoplanktonic cultures . . . . .	109
4.2.3 Labeling experiment . . . . .	110
4.2.4 NanoSIMS measurements . . . . .	110
4.2.5 C-N analysis . . . . .	111
4.2.6 Computation of volumes and uptake rates . . . . .	111
4.2.7 Hypothesis testing . . . . .	112
4.2.8 Principles of NanoSIMS . . . . .	114
4.3 Results . . . . .	116
4.3.1 Scaling of uptake rates with cell volume . . . . .	116
4.3.2 Correlation of C and N uptake rates . . . . .	116
4.3.3 Marginal probability distributions of body size . . . . .	118
4.3.4 Marginal probability distributions of uptake rates . . . . .	118
4.3.5 Joint distribution of body size and uptake rates . . . . .	119
4.3.6 Contribution of intrinsic variability of $b$ to the total observed variability	120
4.3.7 Diffusion-limited and transport-limited uptake . . . . .	120
4.4 Discussion . . . . .	122
4.5 Appendix . . . . .	124
4.5.1 Properties of $p(v \langle v \rangle)$ and $p(b \langle v \rangle)$ . . . . .	124
4.5.2 Single-cell content of $C$ and $N$ . . . . .	125
4.5.3 Appendix: Possible forms of $p(v, b \langle v \rangle)$ . . . . .	126
<b>Conclusions</b>	<b>129</b>
<b>Bibliography</b>	<b>141</b>
<b>Curriculum Vitae</b>	<b>143</b>



# List of Figures

1.1	Types of scaling functions. . . . .	12
1.2	Relative species abundance $p(n A)$ computed numerically via (1.42) and collapse of the curves for different values of $A$ . . . . .	21
1.3	Scheme of predictions on the values or bounds of scaling exponents based on the linking relationships (1.46a-e). . . . .	24
1.4	Density-scatter plot of $\delta + \gamma$ and $\delta$ versus $\eta$ in simulations of the stochastic community dynamics model. . . . .	26
1.5	Scaling patterns from the basic community dynamics model. . . . .	27
1.6	Basic model statistics with parameters $z = 1/4, w = 10^{-3}, \alpha = 3/4, \theta = 1/4$ . . . . .	49
1.7	Basic model statistics with parameters $z = 1/4, w = 10^{-4}, \alpha = 3/4, \theta = 1/4$ . . . . .	49
1.8	Basic model statistics with parameters $z = 1/4, w = 10^{-5}, \alpha = 3/4, \theta = 1/4$ . . . . .	50
1.9	Basic model statistics with parameters $z = 1/4, w = 10^{-3}, \alpha = 1/2, \theta = 1/4$ . . . . .	50
1.10	Basic model statistics with parameters $z = 1/4, w = 10^{-3}, \alpha = 1/4, \theta = 1/4$ . . . . .	51
1.11	Basic model statistics with parameters $z = 1/4, w = 10^{-3}, \alpha = 3/4, \theta = 1/2$ . . . . .	51
1.12	Basic model statistics with parameters $z = 1/2, w = 10^{-3}, \alpha = 3/4, \theta = 1/4$ . . . . .	52
1.13	Variation of the basic model, described in section 1.5.3. Statistics computed with parameters $z = 1/4, w = 10^{-6}, \alpha = 3/4, \theta = 1/2$ . . . . .	52
1.14	Model <i>a</i> statistics computed with the parameter set: $\alpha = 3/4, \theta = 1/4, w = 10^{-2}, v_0 = 1/2$ and $c = 10^{-5}$ . . . . .	53
1.15	Model <i>b</i> statistics computed with the parameter set: $\alpha = 3/4, \theta = 1/4, w = 10^{-2}, v_0 = 1/2$ and $c = 10^{-5}$ . . . . .	53
1.16	Model <i>a</i> statistics computed with the parameter set: $\alpha = 3/4, \theta = 1/4, w = 10^{-2}, v_0 = 1/2, c = 10^{-5}$ and $\bar{q} = 1.2$ . . . . .	54
1.17	Model <i>b</i> statistics computed with the parameter set: $\alpha = 3/4, \theta = 1/4, w = 10^{-2}, v_0 = 1/2, c = 10^{-5}$ and $\bar{q} = 1.2$ . . . . .	54
2.1	Time line associated with the description of different macroecological patterns and their integration through statistical mechanic approaches. . . . .	58
2.2	Empirical evidence of scaling ecological patterns in different ecosystems. . . . .	61
2.3	Empirical evidence of scaling patterns in Barro Colorado Island. . . . .	66
2.4	Empirical evidence of scaling patterns in the Luquillo forest. . . . .	66
2.5	Estimation of the size spectrum exponent in the Luquillo forest. . . . .	67
2.6	Estimation of the exponent $\gamma$ . . . . .	68

## List of Figures

---

2.7	Intra-specific tree size-distribution in the BCI and Luquillo forests. . . . .	69
2.8	Empirical evidence of scaling patterns in LIZ dataset. . . . .	72
3.1	A conceptual scheme highlighting the three possible scenarios for the probabilistic structure of SARs. . . . .	81
3.2	$p(S A)$ for the MSV model. . . . .	85
3.3	$p(S A)$ for the N-SI model. . . . .	86
3.4	Dynamics of the number of species $S$ in the models presented in the text. . . .	87
3.5	Statistics of $S$ for model M1. . . . .	89
3.6	Statistics of $S$ for model M2. . . . .	90
3.7	Failure of the collapse in Eq. (3.8) for the N-SI model. . . . .	90
3.8	Comparison of Poisson distributions and lognormal distributions with the same mean. . . . .	91
3.9	Scaling of average number of species $\langle S \rangle$ vs the ecosystem size $A$ for models N-SI, M1 and M2. . . . .	91
3.10	Scaling of average number of species $\langle S \rangle$ versus ecosystem size $A$ in models N-SI, N-SI-P1 and N-SI-P2. . . . .	94
3.11	$p(S A)$ for the metapopulation model. . . . .	95
3.12	Persistence time distributions for the N-SI model. . . . .	100
3.13	Persistence-time distribution for the MSV model. . . . .	101
3.14	Persistence time distributions for model M1. . . . .	101
3.15	Persistence time distribution for model M2. . . . .	102
3.16	Dynamics of the deterministic model of a system subject to perturbations described in section 3.5 . . . . .	104
4.1	Examples of ratio images obtained with the NanoSIMS. . . . .	113
4.2	The physical principle of SIMS. . . . .	115
4.3	Results of the joint measurements of volume and metabolic rates. . . . .	117
4.4	Body size distributions for the three strains and their collapse. . . . .	118
4.5	Nutrient uptake rate distributions for the three strains and their collapse. . . .	119



# List of Tables

1.1	Scaling exponents measured in the basic model for different sets of parameters' values. . . . .	55
1.2	Scaling exponents measured in the variation of the basic model described in section 1.5.3. . . . .	56
1.3	Scaling exponents measured in the community dynamics models with fluctuating numbers of species. . . . .	56
1.4	Scaling exponents measured in the models with fluctuating numbers of species and $\bar{q} > 1$ . . . . .	56
2.1	References to empirical measurements of the ecological patterns discussed in the main text. . . . .	74
2.2	Estimates for the scaling exponents of the ecological patterns depicted in Fig. 2.2. . . . .	75
2.3	Estimates of scaling exponents $\eta$ , $\delta$ and $\gamma$ in the BCI forest. . . . .	75
2.4	Estimates of scaling exponents $\eta$ , $\delta$ and $\gamma$ in the Luquillo forest . . . . .	75
2.5	Estimates of scaling exponents $z$ , $\delta$ , $\Phi$ and $\gamma$ for the LIZ dataset. . . . .	75
2.6	Estimates of scaling exponents $z$ , $\xi$ and $\delta$ for the SSI dataset. . . . .	75
2.7	Summary of the empirical tests performed. References are to equations in the text. . . . .	75



# Introduction

Ecological systems are exceedingly complex, encompassing an extraordinary diversity of organisms unique in their forms, functions and interactions among themselves and with the environment, itself variable in space and time. Such overwhelming complexity has occasionally discouraged ecologists on whether a general and unifying description of ecological systems is even possible [Simberloff, 2004]. Nevertheless, the large amount of disparate empirical evidence collected during the last century highlighted the existence of some simple and regular patterns characterizing ecosystems, communities and organisms [Marquet et al., 2005]. Patterns may describe the mutual dependence of two ecological variables or the probability distribution describing their occurrence. As an example, the probability distribution of individual sizes in a community (the community size-spectrum) is recurrently observed to follow a power-law behavior, with larger organisms being rarer than smaller ones according to a precise rule [Sheldon et al., 1972, Cavender-Bares et al., 2001, Marañón, 2015, Muller-Landau et al., 2006, Condit et al., 2012]. This pattern has been observed for different types of communities (e.g. marine or terrestrial), featuring a shape invariant across time, space and environmental conditions. Similarly, many other regular patterns, remarkably constant and often conforming to power-law shapes, have been found to describe attributes of individual organisms (e.g. their metabolic requirements and vital rates in relation to their size), of communities (e.g. the distribution of species' abundance and body-size) and of entire ecosystems (e.g. spatial patterns of biodiversity). In contrast with the view of Simberloff [2004], such regular patterns, or 'ecological laws', seem to imply some degree of generality in the working principles of organisms and ecological communities, suggesting that similar, overarching processes associating them dominate over local or specific mechanisms. Additionally, the ubiquitous power-law shape hints that the processes underlying them do not have a preferred scale and apply through different categories of organisms and across scales of organismic size and ecosystem size and type. Therefore, these simple patterns have been considered as a way to tackle the complexity of nature, as they point out that among the plethora of processes taking place in ecosystems and of the immense heterogeneity of characteristics of organisms there are some very general ones which determine the emergent properties, and that these emergent properties are quite simply described. To make a concrete example, power-law community size-spectra with exponents in a quite narrow range have been observed in marine communities for a wide range of environmental conditions and for organisms ranging from bacteria and unicellular autotrophs to large fish [Rodríguez and Mullin, 1986, Quinones

## Introduction

---

et al., 2003, Sprules and Barth, 2016], suggesting that such distributions emerge as ubiquitous properties of complex marine food webs. The different categories of organisms following this pattern differ in many ways, e.g. in feeding and reproduction mechanisms, and changing environmental conditions mean changes in the abiotic factors affecting communities, e.g. primary resources supply or temperature. The observation that the functional shape of the community size-spectrum is robust to these changes implies that the dominant processes determining it are also robust (although the exponent has been seen to vary with ecosystem type and resource supply rate [Cavender-Bares et al., 2001, Marañón, 2015]).

Power-laws are a simple and powerful tool to describe scaling patterns, as they describe their behavior with only two parameters, their exponent and proportionality constant. While in some cases they might be just an approximation of the real scaling of a variable and deviations might arise at small or large scale due to scale-dependent processes, they have proved successful in describing the order of magnitude of change of several ecological and biological variables. For example, at the organismic scale, the  $3/4$  power-law scaling of the metabolic rate of an individual with its body mass, called Kleiber's law, is supported by many physiological studies [Kleiber, 1932, 1947, Calder, 1984, Peters, 1986, West and Brown, 2004, Brown et al., 2004, Savage et al., 2004], showing that mass explains much of the metabolic rate variability of organisms across more than 27 orders of magnitude in body mass [West and Brown, 2004]. At the macro scale, a power-law seems to be the best description of most instances of the island species-area relationship (SAR) [Triantis et al., 2012], quantifying the increase in the number of species  $S$  coexisting in an isolated ecosystem with the ecosystem area.

Following a very large body of empirical observations, explanations have been proposed for how single patterns emerge from microscopic processes and dynamics at the level of communities [MacArthur and Wilson, 1967, Hubbell, 2001, Azaele et al., 2016, Blanchard et al., 2017] and of single organisms [West et al., 1997, West, 1999, West and Brown, 2004, Banavar et al., 1999b, 2010, 2014]. Such mechanistic explanations predicting the functional forms of observed patterns in terms of microscopic dynamics are precious, as they provide an understanding of the working principles of organisms and communities and highlight the dominant processes which, among all those which are present, determine the observed patterns. This knowledge is an important step for predicting how organisms and communities would respond to changing conditions. For example, once the processes determining the size-spectrum of marine community are identified and a mechanistic understanding of their interplay is reached, predicting the effect of a change in environmental factors on the size structure of the community would amount to studying its effect on the single processes.

Nevertheless, what is still lacking in ecology is a synthetic view, bridging patterns at the organismic scale with those at the scale of communities and at the macroscale [Marquet, 2017]. In fact, the different scales and patterns are tightly linked. The properties of an individual, in particular its body mass, determine key organismic characteristics like its metabolic rate, growth rate and life span [Brown et al., 2004] which, in turn, affect its role in the community. Therefore,

patterns at the organismic level certainly have an influence on patterns at the community level. Additionally, patterns describing different aspects of the same community will affect each other due to consistency constraints. Finally, patterns observed at the macroscale, e.g. in the comparison of ecosystems of different sizes or different geographical location, will emerge from the properties of the communities inhabiting those ecosystem. The clarification of the nature of these linkages within and across scales requires a unifying framework, providing a coherent view of the widest possible set of ecological patterns.

A powerful approach which has emerged recently is that of considering macroecological observables as macroscopic variables emerging from the microscopic behavior of the system, an idea that stems from statistical mechanics. Statistical mechanics studies systems of many interacting particles and predicts macroscopic properties of the system by probability arguments. In ecological systems, interacting particles are substituted by interacting individuals or species. While the fundamental units of a physical system, particles, differ in few ways (velocity, mass, charge...), those of an ecological system differ in many. The first difficulty in applying the ideas of statistical physics to ecology is, therefore, to identify which properties of individuals (or species) are important, at least for the phenomena of interest. For example, one could hypothesize that birth rate, death rate and resource use are fundamental characteristics of an individual, while the particular taxa to which it belongs is an information that can be discarded. Macroscopic observables (e.g., the various ecological and biological patterns) then follow from the probability of microscopic states of the system and from state variables, that is, some fixed variables constraining the behavior of the system (e.g., ecosystem area, the total number of individuals and species inhabiting the ecosystem). The second challenge is therefore that of identifying the correct form of the microscopic probability distribution and the appropriate state variables. The predictions and success of this approach depend on the identification of all these elements.

Following this line of thought, Chapters 1 and 2 of this thesis concern a novel theoretical framework in which many of the observed empirical patterns related to abundance, size and biodiversity are derived in a consistent way from which their linkages emerge naturally. The approach is based on the proposal of a finite-size scaling hypothesis describing the joint probability distribution of abundance and typical body mass for species inhabiting an ecosystem of finite area  $A$ , i.e. the probability that one species, picked at random, has abundance  $n$  and the typical body mass of its individuals is  $m$ , termed the fundamental distribution. Body mass is considered as a 'master trait' summarizing species differences, on account of it being a strong predictor of resource consumption rates, vital rates and other organismic properties [Brown et al., 2004]. Ecosystem area is assumed as the only fixed variable in the system, imposing not only a space constraint but also a constraint in the total resource supply rate. Such constraint operates through finite-size cut-offs in observed laws, which are reflected in its appearance in the fundamental distribution. As many of the empirically observed patterns descend from the fundamental distribution, the central idea of the proposed framework is that such distribution is constrained by the shapes of patterns which are observed empirically, which must be reproduced. The aim of this framework, therefore, is not that of predicting the shape of the single

patterns but rather that of identifying the conditions under which the patterns are consistent with each other and with the resource constraint. The value of this approach lies in the explicit identification of the linkages among patterns, allowing to interpret empirical measurements of scaling exponents within a coherent framework. **Chapter 1** is devoted to the formulation of the theoretical framework and to the derivation of the linkages among ecological laws. A class of stochastic models of community dynamics, accounting for competition for finite resources and including only the key processes in determining ecological patterns, is shown to support the hypotheses made in the framework regarding the form of the joint probability distribution of species abundance and typical body mass. **Chapter 2** presents empirical tests of the framework predictions on the available macroecological data. The linking relationships constraining the covariations of scaling exponents are successfully verified on tropical forests and island communities of mammals and lizards. The chapter also deals with challenges posed by sampling biases and deviations from power-law behavior at small and large scale emerging in the data analysis.

Within the first two chapters, ecological patterns are treated as ‘deterministic’: the area of the ecosystem determines univocally the number of species, the mass of an individual determines univocally its metabolic rate and so on. While such relationships are successful at describing average trends of empirical data, fluctuations will be present around such trends. Their causes, effects and importance depend on the particular pattern considered. Fluctuations may limit the predictive power of ecological laws and omitting to consider their effect on empirical measurements may lead to their misinterpretation. To include fluctuations in the theoretical framework proposed in the first two chapters of the thesis or in any other ecological model, in order to assess their potential role and effect, one needs to characterize their statistical properties. Knowledge of the behavior and amplitude of the variability of ecological and biological variables is necessary also to interpret correctly empirical data and to make appropriate use of ecological laws. This thesis deals with this issue in chapters 3 and 4, respectively in the case of the SAR and of Kleiber’s law. In the first case, fluctuations are studied through numerical modeling, while in the second through laboratory experiments at the single-cell level for freshwater phytoplankton using state-of-the-art technology.

As an example of paramount importance in ecology, the island SAR, describing species increase in isolated ecosystem of increasing area, has captured broad interest in ecology, in particular relatively to its use to predict the numbers of species going extinct due to habitat loss or fragmentation [Durrett and Levin, 1996, Hanski, 2013, Borile et al., 2013]. Typically, SARs are meant as trends linking the average number of species to the ecosystem area. However, fluctuations in the number of species  $S$  are expected due to the intrinsic stochasticity of ecosystem dynamics and to external perturbations. Therefore, here SARs are interpreted as portraying only average trends in ecosystems, whose power in predicting the results of empirical biodiversity measurements depends on the amplitude of fluctuations. In fact, if relative fluctuations are large, a point empirical measurement of  $S$  may differ significantly from the expected value. Such scenario poses problems both in the interpretation of field

measurements and in the use of SARs for predictions, demanding for an assessment of the behaviour of the variability of  $S$ . **Chapter 3** examines the statistics of  $S$  in a number of community dynamics models which implement distinct ecological processes possibly driving the  $S$  dynamics in real ecosystems. The different models are studied numerically and compared in terms of the scaling of the relative fluctuations of  $S$  with the mean. The model exploration starts from a minimal neutral model, where species have the same vital rates and resource consumption rate and  $S$  variability is only driven by the stochasticity of diversification and extinction events, and moves to non-neutral models with resource competition, models with habitat stochasticity, and neutral models with external perturbations. The aim of this theoretical study is to determine how the scaling of fluctuations changes in response to the introduction of different ecological processes. Predictions of different models differ remarkably, either foreseeing relative fluctuations which decay with increasing mean or instead ones whose amplitude remains constant as the mean increases. In the former case, a deterministic SAR makes reliable predictions for ecosystems with large enough  $S$ , while in the latter the stochasticity of  $S$  hinders deterministic predictions in ecosystems of all sizes. Additionally, perturbations might (depending on their characteristics, varied in the models) prevent the system from reaching its stationary state, making stationary predictions inappropriate to describe empirical biodiversity measurements. The model survey shows that the variability of the number of species  $S$  underpinning the applicability of a deterministic SAR does not have a universal behaviour but rather depend on the different ecological processes reflected in various model assumptions. This conclusion suggests that empirical investigations should determine which fluctuation scaling scenario applies in a specific context to guide future modeling efforts and identify the process dominating community dynamics.

Chapter 4 deals again with fluctuations but concerning a different biological pattern, the scaling of metabolic rate with organismic size. Metabolic rate is the rate at which organisms use energy, or equivalently the resources from which energy is obtained, for life-sustaining processes, growth and reproduction. Abundant evidence from physiological studies suggests that this rate scales as a power of the organism's body mass, a large part which yielding an exponent of this scaling of  $3/4$  [Kleiber, 1932, Peters, 1986, Savage et al., 2004]. However, a good number of studies showed deviations from this rule (see e.g. Dodds et al. [2001], Glazier [2005]), with phytoplankton being a remarkable exception [Marañón, 2015]. Size-scaling of metabolic rate is also linked to the the size-scaling of other important organismic properties [Brown et al., 2004], especially growth rates, therefore it has a central role in structuring communities, for example freshwater and marine microbial ones [Marañón, 2015] which are the object of the experiments designed and carried out in the context of the present thesis and presented in the chapter. Understanding how these communities are assembled and how they respond to changes in environmental conditions, e.g. resource supply, is an issue of great importance given the central role of phytoplankton in the regulation of global climate [Falkowski, 1998]. Typically, size-scaling of metabolic rates is studied for species averages. However, at the intra-specific level body mass and metabolic rates are heterogeneous, and their fluctuations around species averages are arguably correlated. A more appropriate description of

the scaling relationship should also acknowledge such fluctuations, providing their statistical characterization. Intra-specific mass variability has already been studied in Giometto et al. [2013] for several different protist species and shown to be appropriately described by a scaling form incorporating few species-specific details. Not much is known, instead, on the metabolic rate variability and on its covariation with body mass at the intra-specific level, although phenotypic variability has been recognized as an important factor for the functionality of microbial populations and communities [Ackermann, 2015]. **Chapter 4** presents the results of experiments on three phytoplanktonic species (*Synechococcus sp*, *Scenedesmus obliquus* and *Cryptomonas ovata*) aimed at measuring jointly their volume and nutrient uptake rates, which are proxies of metabolic rate, at the single cell level. The scope of such measurements is to test a scaling hypothesis for the intra-specific distribution of metabolic rates proposing that, similarly to the body mass distribution, it is described by a general scaling form where average species body mass acts as the only species-dependent constraint. The statistical characterization of the two marginal distributions is the first step in the direction of the formulation of a generalized scaling relationship of metabolic rate with size. Correlations between the two variables are also investigated by the experiments, although with limitations related to the methodological technique.

The importance of such a general description of correlated fluctuations of size and metabolic rate on the basis of minimum biological detail is two-fold. At a fundamental level, its feasibility would prove the generality of mechanisms underlying the heterogeneity of metabolic requirements of organisms. For ecology, it would provide a theoretical framework synthesizing correlated fluctuations of size and metabolic rates of different origins at the intra- and inter-specific levels, necessary to test their potential role in shaping size-related patterns and their response to different environmental conditions, such as resource supply.

To summarize, this thesis combines methods and ideas coming from statistical physics, primarily the use of finite-size scaling to describe joint probability distributions of correlated variables constrained by the system size (in Chapter 1) or by a characteristic size (in Chapter 4), mathematical modeling and experiments to study quantitative laws in ecology and biology, specifically to gain insight on their linkages and to assess the features and role of fluctuations.



# 1 Covariations in ecological scaling laws

*Scaling laws in ecology, intended both as functional relationships among ecologically-relevant quantities and the probability distributions that characterize their occurrence, have long attracted the interest of empiricists and theoreticians. Empirical evidence exists of power laws associated with the number of species inhabiting an ecosystem, their abundances and traits. Although their functional form appears to be ubiquitous, empirical scaling exponents vary with ecosystem type and resource supply rate. The idea that ecological scaling laws are linked had been entertained before, but the full extent of macroecological pattern covariations, the role of the constraints imposed by finite resource supply and a comprehensive empirical verification were still lacking. In this Chapter, such gap is addressed by presenting a theoretical scaling framework that predicts the linkages of several macroecological patterns related to species' abundances and body sizes. Such framework, based on a finite-size scaling approach, accounts for deviations from pure power-law patterns entailed by the finite size of the ecosystem. The proposed framework is consistent with the stationary state statistics of a broad class of resource-limited community dynamics models, regardless of parametrization and model assumptions. Testable predictions are provided in the form of algebraic relationship among scaling exponents. The approach presented in this Chapter places the observed variability of ecological scaling exponents into a coherent statistical framework where patterns in ecology embed constrained fluctuations. Moreover, the framework can be generalized to assimilate empirical evidence that is specific to certain ecosystems. The implications of such generalizations possible deviations from pure power-law patterns and multiple trophic level ecosystems, are explored.<sup>1</sup>*

## 1.1 Introduction

Over the span of decades, the study of ecosystems uncovered remarkable regularities recurring in nature in the form of patterns of diversity, size and abundance which have often been elevated to the rank of laws owing to their ubiquity [Levin, 1992, Marquet, 2000]. Among the most relevant (and studied) patterns we count, for example: the Species-Area Relationship

---

<sup>1</sup>This chapter features contents from Zaoli et al. [2017].

## Chapter 1. Covariations in ecological scaling laws

---

(SAR) [MacArthur and Wilson, 1963, 1967], quantifying the increase in the number of species  $S$  coexisting in disjoint habitats of increasing area  $A$  [Rybicki and Hanski, 2013, Hanski, 2016]; the community size-spectrum [Sheldon et al., 1972, Cavender-Bares et al., 2001, Rinaldo et al., 2002, White et al., 2007], describing the probability that an individual in a community has size  $m$ , regardless of species; Kleiber's law [Kleiber, 1932, 1947], that describes metabolic rate-organismic body size relations operating across more than 27 orders of magnitude of body mass [West et al., 2002]; Damuth's law [Damuth, 1981, 1987, 1993], the scaling of a species' abundance (i.e. its population within an ecosystem of area  $A$ ) with its characteristic mass.

Patterns may equally refer to functional relationships between ecologically-relevant quantities  $X, Y$ , like e.g. for SARs, or to the probability distributions  $p(X)$  characterizing their occurrence, like e.g. for community size-spectra. Much field work has been devoted to measuring such patterns in diverse types of ecosystems (e.g. terrestrial, aquatic) and organisms (e.g. mammals, birds, phytoplankton and trees) and for many of them strong evidence suggests a power-function shape [Levin, 1992, Marquet, 2000, Marquet et al., 2005], that is, respectively,  $X \propto Y^a$  or  $p(X) \propto X^{-a}$ . The empirically verified ubiquity of power-laws pushed theoreticians to formulate mechanistic models to explain their emergence and the value of their scaling exponents  $a$ , e.g. Brown [1995], West [1999], Hubbell [2001], West et al. [2002], West, G.B. et al. [2003], Gherardi et al. [2013], Giometto et al. [2015], Azaele et al. [2016], Cuesta et al. [2017], Blanchard et al. [2017], MacArthur and Wilson [1967]. Empirical evidence, however, shows that such values vary considerably across ecosystems [Cavender-Bares et al., 2001, Finkel et al., 2004, Marañón, 2015], suggesting that exponents of scaling ecological laws are far from universal, although the power-law form proves remarkably robust (see Chapter 2).

Typically, studies focused on single patterns, often overlooking their mutual linkages (but see Southwood et al. [2006], Banavar et al. [2007], Marquet [2017]), and therefore leading to canonical estimates of scaling exponents which may not be simultaneously achievable in a single ecosystem due to extant and consistency constraints. However, the existence of linkages among laws characterizing an ecosystem, and therefore of relationships constraining the co-variation of their scaling exponents (when applicable), is a fact. Simple heuristic arguments can be used to show such linkages. Consider, for example, a community of  $N$  individuals and  $S$  species hosted within an ecosystem of area  $A$ . Let the  $i$ -th species have abundance  $n_i$  and typical body mass  $m_i$  and let  $s(m)$  be the community size-spectrum in such an ecosystem and  $p(m)$  the probability distribution of species' typical body masses. The total number of individuals with size in  $[m, m + dm]$ , which is given by  $Ns(m)dm$ , is the product of the number of species with typical size falling within the same interval,  $Sp(m)dm$ , and the average abundance of a species having that typical size  $m$ , here labeled as  $\langle n|m \rangle$ :

$$N s(m) = S p(m) \langle n|m \rangle. \quad (1.1)$$

Empirical evidence suggests that the three mass-related patterns involved in Eq. (1.1) are described, at least approximately, by power-laws, i.e.,  $s(m) \propto m^{-\eta}$  [Sheldon et al., 1972, Cavender-Bares et al., 2001, Rinaldo et al., 2002, White et al., 2007],  $p(m) \propto m^{-\delta}$  [Marquet and

Taper, 1998, Marquet et al., 2005] and Damuth' law  $\langle n|m \rangle \propto m^{-\gamma}$  [Damuth, 1981, Marquet et al., 1990, Damuth, 1993, Marquet, 2000]. Eq. (1.1) implies a consistency relationship among the three scaling exponents:

$$\eta = \delta + \gamma. \quad (1.2)$$

Eq. (1.1) suggests also a second relationship, stemming from the dependence of the empirical patterns involved on ecosystem area  $A$ . In fact, the dependence of the number of species  $S$  on  $A$  (SAR) may be considered to assume a power-law form (at least in a range of scales), say  $S \propto A^z$ , one of the most widespread 'laws' of macroecology since the equilibrium theory of island biogeography [MacArthur and Wilson, 1967]. Evidence exists for typical values  $z \approx 1/4$  [MacArthur and Wilson, 1963, Okie and Brown, 2009, Lomolino, 1982]. If one assumes  $N \propto A^v$ , where the most probable value of  $v$  is 1, an additional power dependency on  $A$  for Eq. (1.1) is needed for consistency. Neglecting, for the moment being, possible  $A$ -dependent cutoffs in Eq. (1.1), the only pattern that can depend on powers of  $A$  is  $\langle n|m \rangle$ , i.e.  $\langle n|m, A \rangle \propto A^\Phi m^{-\gamma}$ , implying

$$v = z + \Phi. \quad (1.3)$$

This example and a few others identified earlier [Rinaldo et al., 2002, Southwood et al., 2006, Banavar et al., 2007] highlight the need for a framework that comprehensively accounts for linking relationships among macroecological scaling laws.

The drawback of of such back-of-the-envelope computations is that they neglect possible finite-size corrections to power-law behavior, which arise due to physical or biological limits of the available range of sizes, and they do not allow accounting for patterns which are not pure power-laws. For example, empirical evidence suggests that the body size of the largest organism found in an ecosystem scales with its area [Burness et al., 2001] and/or is limited by mechanical properties of biological material [Thompson, D'Arcy W., 2001], therefore supporting the existence of an upper limit beyond which the conditions or resources needed to make organismic bodies sustainable fail (be they mechanical or altogether ecological). Power-law dependencies cannot, consequently, extend over an infinite domain of body masses, but they necessarily exhibit a viable finite range. When the ecosystem resource supply is limited, and assuming such supply scales with the ecosystem area  $A$ , an  $A$ -dependent cut-off for the maximum size is expected to apply. Additionally, in some cases back-of-the-envelope calculations are not possible, therefore a self-consistent framework is needed to obtain all linking relationships among patterns.

The ubiquity of power-laws and the presence of finite-size constraints suggested a parallel with finite-size scaling theory for critical phenomena [Kadanoff et al., 1967, Fisher and Barber, 1972, Stanley, 1999, Stanley and Amaral, 2000], which proved a valid method to treat scaling in ecosystems in a model-independent fashion [Banavar et al., 1999a, 2007, Zillio et al., 2008, Simini et al., 2010]. To illustrate the central concept of finite-size scaling theory, let us consider

## Chapter 1. Covariations in ecological scaling laws

---

the body-mass of an organism,  $m$ , as a reference variable. Let  $m$  be defined in units of the minimum viable mass  $m_0$ , so that  $m$  is dimensionless. In a system of infinite size, and assuming no physiological constraints to body size, there is no upper bound on  $m$  and we can conceive that, over evolutionary times, the upper limit of the spectrum of masses present in the system diverges. In such case, the quantities depending on  $m$ , e.g. the community size spectrum  $s(m)$  or Damuth's abundance  $\langle n|m \rangle$ , would in principle be allowed to attain a pure power-law behavior across the whole  $[1, \infty)$  domain. Nevertheless, if the system is endowed with a finite size  $A^2$ ,  $m$  will hold an upper limit that depends on  $A$ , say  $A^\lambda$  with  $\lambda > 0$ . When  $m \sim A^\lambda$ , quantities depending on  $m$  will show finite-size effects, exhibiting a fast decay to zero (e.g. exponential). The finite-size scaling hypothesis posits that such finite size-effects will have the shape of a cut-off function depending only on the dimensionless variable  $m/A^\lambda$ , e.g.

$$s(m) = m^{-\eta} h\left(\frac{m}{A^\lambda}\right), \quad (1.4)$$

where  $h$  is a function such that  $h(x) \sim \text{const}$  when  $x \ll 1$  and  $\lim_{x \rightarrow \infty} h(x) = 0$ . When  $m$  is far from the upper limit, power-law behavior is conserved, but when  $m/A^\lambda \sim 1$  finite-size effects pitch in and cause a rapid decay to near-zero values of the fraction of organisms in that size range. The validity of the scaling hypothesis can be verified by data collapse [Kadanoff et al., 1967, Fisher and Barber, 1972, Stanley, 1999, Stanley and Amaral, 2000]. See Box 1.2 for an explanation of how data collapse is performed and for the derivation of some consequences of the finite-size scaling hypothesis in Eq. (1.4). Approaches based on a finite-size scaling description of patterns are model-free, as they do not make any assumptions on the processes that generate such patterns, but can give information on the value of exponents (through data collapse) and on their covariations (see 1.4).

The use of finite-size scaling in ecology is not new. For instance, in Zillio et al. [2008], a scaling argument has been proposed for the Relative Species Abundance (RSA)  $p(n|A)dn$ , which is the probability that a species picked at random from a community inhabiting an ecosystem of area  $A$  has abundance in  $[n, n + dn]$ . It was suggested therein that the  $n$  dependence of the RSA might be captured by a scaling variable  $n/f(A)$ , where  $f(A)$  is the typical value of  $n$  in an ecosystem of size  $A$ . Such hypothesis was successful in describing the RSA of several forest plots [Zillio et al., 2008]. The authors identified its relationship with the SAR by noting that  $\langle n|A \rangle = N(A)/S(A)$ , where  $N(A)$  is the total number of individuals in the ecosystem.

Banavar et al. [2007] first proposed to use a finite-size scaling framework to investigate the linkage of ecological laws. Therein, a number of laws are linked by a two-variables scaling *ansatz* for  $p(n, m|A)$ , the joint probability density function of abundance and body-mass, i.e.,

$$p(n, m|A) = n^{-\sigma_1} m^{-\sigma_2} F\left(\frac{n}{A^{\Phi_1}}, \frac{m}{A^{\Phi_2}}\right) \quad (1.5)$$

---

<sup>2</sup>  $A$  is measured in units of a reference area  $a_0$ , therefore it is dimensionless. The reference area  $a_0$  is defined as the smallest ecosystem area that can sustain an organism of mass  $m_0$ .

where  $\sigma_1, \sigma_2, \Phi_1, \Phi_2 > 0$ . Note that the properties of  $F$  proposed therein are:

$$\lim_{x \rightarrow \infty} F(x, y) = 0 = \lim_{y \rightarrow \infty} F(x, y) \quad (1.6a)$$

$$\lim_{x \rightarrow 0} F(x, y) = f(y), \quad \lim_{y \rightarrow 0} F(x, y) = g(x) \quad (1.6b)$$

where  $f(y)$  is such that  $\lim_{y \rightarrow 0} f(y) = \text{const}$ ,  $\lim_{y \rightarrow \infty} f(y) = 0$  and  $g(x)$  has properties analogous to those of  $f(y)$ . In this framework, observable laws can be obtained by computing, for example, the marginals,  $p(n|A)$  and  $p(m|A)$ , and the moments of the joint distribution:

$$I_{i,j} = \langle n^i m^j \rangle = \int \int n^i m^j p(n, m|A) dn dm \propto A^{\psi(i,j)} \quad (1.7)$$

where  $\psi(i, j)$  is a function of the exponents  $\sigma_1, \sigma_2, \Phi_1, \Phi_2$ . The normalization of (1.5) is imposed as  $I_{0,0} = 1$ , which provides conditions on the exponents appearing therein. By suitably tuning  $i$  and  $j$ , one obtains several relevant quantities. The average species' abundance, for example, is obtained from the moment of order (1,0),  $I_{1,0} = \langle n \rangle$ . Linking relationships between their scaling exponents are obtained, as will be detailed better in section 1.4.

Although the above framework recovered many empirically observed laws connected to abundance and size, it fails to obtain Damuth's law [Damuth, 1981] which is solidly supported by empirical observations. Consequently, instead of Eq. (1.2) it predicts  $\eta = \delta$ . In this Chapter, an alternative scaling hypothesis for  $p(n, m|A)$  is proposed which allows to link a larger set of ecological laws, including Damuth', and which accounts for resource limitation and finite-size effects. Such framework predicts a set of linking relationships among scaling exponents which clarify exactly how ecological patterns influence each other. The predicted relationships are compatible with the available field data (see Chapter 2). Moreover, the scaling hypothesis at the base of the framework is supported by a large class of stochastic models of community dynamics describing a set of species competing for finite resources. The proposed framework is suited to generalizations that account for several empirically-inspired facts and assumptions that possibly encompass different conditions actually realized in nature. The framework is thus a tool to investigate the effects of ecosystem-specific assumptions on pattern covariations. In particular, the following generalizations are examined: intra-specific size-distributions, i.e. a spread in the body mass of individuals belonging to the same species; a curvature in Kleiber's law, which has been suggested by some experimental and theoretical works [Kolokotronis et al., 2010, Ballesteros et al., 2018, Banavar et al., 2014, Dodds et al., 2001, Marañón et al., 2013, Mori et al., 2010]; a non-pure power-law forms of  $p(m|A)$  and an area-independent cut-off on the maximum species' typical body mass, which may arise due to physiological limits or competition-based constraints and, finally, a two-trophic-levels community. We highlight if, and how, the considered generalizations change the predictions for the linking relationships.

The chapter is organized as follows. First, Box 1.2 presents some facts about finite-size scaling which will be useful in the following. Then, in section 1.3, the scaling hypothesis is introduced and justified. In section 1.4 the scaling hypothesis is used to obtain empirically observed

scaling laws and the linking relationships among their exponents. Section 1.5 is devoted to showing that models of community dynamics support the framework. Finally, section 1.6 illustrates several generalization which can be implemented in the framework.

## 1.2 Finite-size scaling

Consider a probability distribution of the form

$$p(x|C) = \frac{1}{x^a} f\left(\frac{x}{C^b}\right), \quad (1.8)$$

where  $C$  is a constant and  $x \geq x_0$  a continuous variable. To allow for the convergence of the moment of order  $k$  (order 0 being the normalization constant), one must ask  $f(x) = o(x^{a-k-1})$  when  $x \rightarrow \infty$ . The behavior in  $x \sim 0$  may instead vary.  $C$  represents a characteristic scale for the variable  $x$ , whose meaning depends on the shape of the scaling function  $f$ . If  $f(y)$  is a function approximately constant for  $y \ll 1$  and which decays for  $y \rightarrow \infty$ ,  $C^b$  represents a characteristic scale for the maximum value of  $x$ , in fact  $p(x|C)$  decays when  $x > C^b$  (see Fig. 1.1A). An example of function with these properties is  $f(y) = e^{-y}$ . Instead, when  $f(y)$  is a function with an internal mode, decaying both when its argument is small and large (e.g.,  $f(y) = e^{-(\log y)^2}$ ),  $C^b$  represents a characteristic scale for the typical value of  $x$  (see Fig. 1.1B). Both types of scaling function will be used in this chapter and in the following ones.

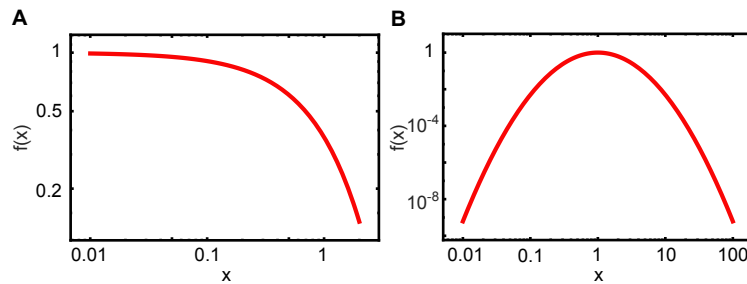


Figure 1.1 – The two types of scaling functions described in the text, plotted in double logarithmic scales. A)  $f(x) = e^{-x}$ ; B)  $f(x) = e^{-\log(x)^2}$ .

### 1.2.1 Properties of Eq. (1.8)

The interest in the scaling form of Eq. (1.8) is that it implies a particular scaling of the moments of  $x$  with  $C$ , regardless of the specific function  $f$  (but depending on its general shape). In order to derive the moments' scaling, the two cases distinguished above must be treated separately. Let us call case A the one where  $f(y) \approx c$  when  $y \ll 1$  and decays at large  $y$  values, and case B the one where  $f(y)$  has an internal mode.

In case A, the scaling of moments of  $x$  can be obtained as follows:

$$\begin{aligned}
 I_k = \langle x^k \rangle &= \int_{x_0}^{\infty} x^k p(x|C) dx = \int_{x_0}^{\infty} x^{k-a} f\left(\frac{x}{C^b}\right) dx = C^{b(1+k-a)} \int_{x_0/C^b}^{\infty} y^{k-a} f(y) dy \\
 &\simeq C^{b(1+k-a)} \left[ c \int_{x_0/C^b}^{\epsilon} y^{k-a} dy + \int_{\epsilon}^{\infty} y^{k-a} f(y) dy \right] = C^{b(1+k-a)} [c' C^{-b(1+k-a)} + c''],
 \end{aligned} \tag{1.9}$$

where the change of variable  $y = x/C^b$  was performed,  $\epsilon \ll 1$  and  $c'$  is a constant that does not depend on  $C$ . The integral  $c'' = \int_{\epsilon}^{\infty} y^{k-a} f(y) dy$  converges owing to the properties of  $f$  and does not depend on  $C$ . In the limit of large  $C$ , the moments display the scaling

$$I_k \propto C^{\max\{0, b(1+k-a)\}}. \tag{1.10}$$

In this chapter, the role of  $C$  is played by the area of the ecosystem, therefore the large  $C$  limit applies. Normalization poses a constraint on  $a$ , in fact it requires

$$1 = I_0 \propto C^{\max\{0, b(1-a)\}}. \tag{1.11}$$

Because the normalization constant cannot depend on  $C$ , one has  $a > 1$ .

In case B, the scaling of the moment is computed as

$$\begin{aligned}
 I_k = \langle x^k \rangle &= \int_{x_0}^{\infty} x^k p(x|C) dx = \int_{x_0}^{\infty} x^{k-a} f\left(\frac{x}{C^b}\right) dx = C^{b(1+k-a)} \int_{x_0/C^b}^{\infty} y^{k-a} f(y) dy \\
 &\simeq C^{b(1+k-a)} \int_0^{\infty} y^{k-a} f(y) dy \propto C^{b(1+k-a)},
 \end{aligned} \tag{1.12}$$

where the prolongation of the integral to 0 is possible when  $x_0/C^b \ll 1$  thanks to the properties of  $f(y)$ , decaying fast in  $y \sim 0$ . In this case, normalization imposes  $a = 1$ , therefore  $I_k \propto C^{bk}$ .

Note that, since  $\langle x \rangle = I_1 \propto C^b$ , in case B Eq. (1.8) can be written as

$$p(x|\langle x \rangle) = \frac{1}{x} \tilde{F}\left(\frac{x}{\langle x \rangle}\right). \tag{1.13}$$

Additionally, subsequent moments have the property

$$\frac{\langle x^{k+1} \rangle}{\langle x^k \rangle} \propto \langle x \rangle, \tag{1.14}$$

i.e. all moments can be expressed in terms of the mean.

### 1.2.2 Data collapse

Data collapse is a method to verify that probability distributions corresponding to different values of  $C$ , obtained from empirical measurements or from numerical simulations, are described by the scaling form in Eq. (1.8). It consists in plotting  $x^a p(x|C)$  against  $\frac{x}{C^b}$ . If the curves satisfy Eq. (1.8) they will all collapse on the same curve,  $f$ , for the appropriate value of  $a$  and  $b$ . Bhattacharjee and Seno [2001] have proposed an algorithm to estimate the value of  $a$  and  $b$  and their uncertainty by optimizing the collapse. The algorithm computes the value of a functional  $P_b(a, b)$  measuring the area comprised between each set of curves after their rescaling according to the exponents  $a, b$ . The optimal exponents  $a_{opt}$  and  $b_{opt}$  are then estimated as the minimum of this functional and a confidence level is obtained by finding, for each exponent, the value at which the functional is 1% larger than its value at the minimum.

## 1.3 Scaling Framework

This section is devoted to introducing the scaling framework, with particular attention to the assumptions made. Specifically, the framework applies to a community inhabiting an ecosystem of area  $A$  and feeding on a common resource, supplied at a finite rate  $\mathcal{R}$ . The finite resource supply rate constrains the total community consumption, and therefore its size and composition, as will be detailed below. At the core of the framework is a hypothesis for the joint probability  $p(n, m|A) dn dm$  that a species chosen at random holds abundance and body size (mass) in the ranges  $[n, n + dn]$  and  $[m, m + dm]$ . The probability distribution  $p$  is termed “fundamental distribution”, as community structure and resource use follow from it. The choice of the relevant variables, here  $n, m$  and  $A$ , is the minimum one that allows obtaining scaling laws related to mass and abundance consistently with a finite resource constraint. A general theoretical argument must in fact incorporate a minimum of biological detail to be applicable to the widest range of organisms. Other quantities, such as the number of species  $S$  or the total abundance  $N$ , emerge as predictions of the framework once the constraint has been imposed. Considering body mass is deemed necessary to implement the resource constraint, because the resource consumption of an individual, epitomized by its metabolic rate, is determined by its mass through an allometric relation known as Kleiber’s law [Kleiber, 1932],  $b = cm^\alpha$  where  $c$  is a constant and the typical value for the exponent is  $\alpha = 3/4$  [West et al., 1997, Savage et al., 2007, Banavar et al., 1999b]. Assuming that the resource supply rate is proportional to the ecosystem area, i.e.  $\mathcal{R} = rA$ , where  $r$  is the resource supply per unit area (assumed constant), the total community metabolic rate  $B \propto \sum_{j=1}^S n_j \bar{m}_j^\alpha$  ( $\bar{m}_j$  is the typical mass of species  $j$ ) is constrained to be

$$B = \mathcal{R} = rA \propto A. \quad (1.15)$$

In fact, if the scaling of  $B$  with  $A$  was super-linear, the metabolic rate would overgrow the resource supply rate in the limit of large ecosystem area, resulting in a non sustainable commu-



nity. Conversely, in the case of sub-linear scaling, the fraction of consumed resources would tend to zero in the limit of large ecosystem area. For an ecosystem at stationarity, the total metabolism of the community is expected to be the maximum sustainable one, corresponding to a direct proportionality of  $B$  and  $A$ . Note that the constant  $r$  appearing in equation (1.15) determines the type of ecosystem (with abundant or scarce resource supply rate, which could be interpreted, e.g., as a tropical forest or a desert). Therefore, all scaling with ecosystem area obtained via this framework are valid for a fixed value of  $r$ , i.e. for a set of islands, lakes or forests sharing a similar climate or environmental conditions.

The scaling hypothesis formulated for the fundamental distribution must be such that the scaling laws derived from it reproduce the empirical scaling observed in the field. The minimum viable mass for an organism is assumed to be  $m_0 > 0$  independent of  $A$ , so that  $p(n, m|A)$  is zero for  $m < m_0$ . Mass and area are measured in units of  $m_0$  and of a reference unit area  $a_0$ , so that  $m$  and  $A$  are dimensionless. To facilitate the formulation of an empirically driven scaling hypothesis, one can factorize the fundamental distribution in two parts:

$$p(n, m|A) = p(m|A)p(n|m, A). \quad (1.16)$$

The first factor is the probability density of finding a species of typical mass  $m$  in an ecosystem of area  $A$ . Empirical evidence [Marquet and Taper, 1998, Marquet et al., 2005] points to a power-law dependence of  $p(m|A)$  on  $m$ , therefore the hypothesis made here is

$$p(m|A) = (\delta - 1)m^{-\delta} \quad (1.17)$$

with  $\delta > 1$  for normalization. The scaling hypothesis for  $p(n|m, A)$  is dictated by the need to reproduce the empirically observed Damuth's law,

$$\langle n|m, A \rangle = m^{-\gamma} A^\Phi h\left(\frac{m}{A^\lambda}\right). \quad (1.18)$$

where the function  $h$ , providing a cut-off to the power-law behavior in  $m$ , has the properties

$$h(x) = o\left(x^{-2+\delta+\gamma}\right) \quad \text{as } x \rightarrow \infty \quad (1.19a)$$

$$h(x) \simeq h_0 \quad \text{when } x \ll 1. \quad (1.19b)$$

The presence of a cut-off function in  $\langle n|m, A \rangle$ , imposing an upper limit on viable body masses, is supported by the results of a stochastic model of community dynamics, as shown in section 1.5. Note that Damuth's law typically refers only to the mass dependence of  $\langle n \rangle$ , while here also the  $A$  dependence is considered. The simplest hypothesis for the scaling form of  $p(n|m, A)$  allowing to obtain Eq. (1.18) is

$$p(n|m, A) = g(m, A) \hat{G}\left(\frac{n}{\langle n|m, A \rangle}\right) \quad (1.20)$$

where the term  $g(m, A)$  allows a further dependence on  $m$  and  $A$  required for normalization.

## Chapter 1. Covariations in ecological scaling laws

---

The function  $\hat{G}(x)$  has the properties:

$$\int_0^{\infty} x^j \hat{G}(x) dx < \infty \quad j = 0, 1, 2. \quad (1.21)$$

An example of a function that satisfies the above requests is  $\hat{G}(x) = e^{-x}$ . The properties of  $h$  and  $\hat{G}$  are such as to allow convergence of the  $(j, k)$ -th moment of  $p(n, m|A)$  for  $j = 0, 1, 2$  and  $k \in [0, 1]$ . These are all the moments needed to derive ecological scaling laws (see section 1.4). Furthermore, the properties of  $\hat{G}$  ensure that  $\langle n|m, A \rangle = \int_0^{\infty} dn n p(n|m, A) = m^{-\gamma} A^{\Phi} h\left(\frac{m}{A^{\lambda}}\right)$ , that is, Damuth's law is reproduced.

The normalization condition for  $p(n|m, A)$  reads:

$$\begin{aligned} 1 &= \int_0^{\infty} p(n|m, A) dn = g(m, A) \int_0^{\infty} \hat{G}\left(\frac{n}{\langle n|m, A \rangle}\right) dn \\ &= g(m, A) \langle n|m, A \rangle \int_0^{\infty} \hat{G}(x) dx = g(m, A) \langle n|m, A \rangle C, \end{aligned} \quad (1.22)$$

where  $C = \int_0^{\infty} \hat{G}(x) dx$ . Therefore,  $g(m, A) = (C \langle n|m, A \rangle)^{-1}$ , which substituted in Eq. (1.20) gives

$$p(n|m, A) = \frac{1}{C \langle n|m, A \rangle} \hat{G}\left(\frac{n}{\langle n|m, A \rangle}\right) = n^{-1} G\left(\frac{n}{\langle n|m, A \rangle}\right), \quad (1.23)$$

where  $G(x) = \frac{1}{C} x \hat{G}(x)$ , implying the condition  $\int_0^{\infty} \frac{G(x)}{x} dx = 1$ . In conclusion, joining the two hypothesis in Eqs (1.17) and (1.23), the normalized fundamental distribution reads:

$$p(n, m|A) = (\delta - 1) n^{-1} m^{-\delta} G\left[\frac{nm^{\gamma}}{A^{\Phi}} \frac{1}{h(m/A^{\lambda})}\right]. \quad (1.24)$$

Note that Eq. (1.24) can be expressed as  $p(n, m|A) = n^{-1} m^{-\delta} F\left(\frac{n}{A^{\Phi}}, \frac{m}{A^{\lambda}}\right)$ , with  $F(x, y) = (\delta - 1) G\left[\frac{xy^{\gamma}}{h(y)}\right]$  and  $\tilde{\Phi} = \Phi - \gamma\lambda$ . In such form, at first glance Eq. (1.24) may look like a particular case of the scaling framework of Banavar et al. [2007] (Eq. (1.5)). However, the assumptions on the scaling properties of  $F$  made therein (Eqs (1.6)) are different than here, making the two frameworks and the observable scaling laws and linking relationships distinct. Most importantly, the hypotheses of Banavar et al. [2007] are incompatible with Damuth's law.

### 1.4 Derivation of ecological scaling laws

The following empirically-observed ecological scaling laws can be obtained from the fundamental distribution:

- the island SAR<sup>3</sup>,  $S \propto A^z$ ;

---

<sup>3</sup>The so-called island SAR [Preston, 1962], obtained by counting species inhabiting disjoint patches of land (e.g. islands, lakes or, in general, areas separated by environmental barriers from the surroundings which we can think

- the scaling of the total number of individuals  $N$  with  $A$  [Hubbell, 2001],  $N \propto A^\nu$ ;
- the scaling of the total body mass  $M$  with  $A$ ,  $M \propto A^\mu$ ;
- Damuth's law,  $\langle n|m, A \rangle \propto m^{-\gamma} A^\Phi h(m/A^\lambda)$ ;
- the community size-spectrum,  $s(m) \propto m^{-\eta}$ ;
- the species' mass distribution  $p(m|A) \propto m^{-\delta}$  (which is a marginal of the fundamental distribution explicitly given as an assumption of the framework in Eq. (1.17));
- the scaling of the largest organism's mass with  $A$  [Burness et al., 2001, Okie and Brown, 2009],  $m_{\max} \propto A^\xi$ ;
- Taylor's law [Cohen et al., 2012, Giometto et al., 2015], describing the scaling of the variance of population abundance with its mean,  $\text{var}(n) \propto \langle n \rangle^\beta$  (here, reference is made to the 'spatial Taylor's law', dealing with the variability of species abundance in space, and not to its temporal counterpart);
- the relative species' abundance (RSA) [Preston, 1948], i.e., the probability of finding a species with abundance  $n$ .

First, one can verify that equation (1.20) implies equation (1.18), i.e. Damuth's law:

$$\begin{aligned} \langle n|m, A \rangle &= \int_0^\infty n p(n|m, A) dn = \int_0^\infty G \left[ \frac{nm^\gamma}{A^\Phi} \frac{1}{h(m/A^\lambda)} \right] dn \\ &= A^\Phi m^{-\gamma} h \left( \frac{m}{A^\lambda} \right) \int_0^\infty G(x) dx = A^\Phi m^{-\gamma} h \left( \frac{m}{A^\lambda} \right), \end{aligned} \quad (1.25)$$

where  $\int_0^\infty G(x) dx = 1$  was set without loss of generality.

The following sections 1.4.1–1.4.5 are devoted to the derivation of the other laws and of their linking relationships.

### 1.4.1 The SAR and other scaling relationships obtained from moments of $p(n, m|A)$

The SAR, the scaling of the total biomass  $M$  and of the total number of individuals  $N$  with  $A$  and Taylor's law are obtained from the moments of the fundamental distribution. The scaling

---

of as closed ecosystems) must be distinguished from nested SARs, where areas are sub-patches of a single larger domain [Harte et al., 2009, Azaele et al., 2016]. The two SARs are quite different, as the nested SAR is related to the spatial distribution of individuals, while the island SAR stems from complex eco-evolutionary dynamics shaping the community.

## Chapter 1. Covariations in ecological scaling laws

of the  $(j, k)$ th-moment ( $j = 1, 2, k \in [0, 1]$ ) with  $A$ , for large  $A$ , can be computed as:

$$\begin{aligned}
 I_{j,k} &= \int_1^\infty \int_0^\infty n^j m^k p(n, m|A) dn dm \\
 &= (\delta - 1) \int_1^\infty m^{k-\delta} \int_0^\infty n^{j-1} G \left[ \frac{nm^\gamma}{A^\Phi} \frac{1}{h(m/A^\lambda)} \right] dn dm \\
 &= (\delta - 1) A^{j\Phi} \int_1^\infty m^{k-\delta-j\gamma} h^j \left( \frac{m}{A^\lambda} \right) dm \int_0^\infty x^{j-1} G(x) dx \\
 &\propto A^{j\Phi} \int_1^\infty m^{k-\delta-j\gamma} h^j \left( \frac{m}{A^\lambda} \right) dm \\
 &\propto A^{j\Phi+\lambda(1+k-\delta-j\gamma)} \int_{1/A^\lambda}^\infty y^{k-\delta-j\gamma} h^j(y) dy \\
 &\propto A^{j\Phi+\lambda(1+k-\delta-j\gamma)} \left[ h_0^j \int_{1/A^\lambda}^\epsilon y^{k-\delta-j\gamma} dy + \int_\epsilon^\infty y^{k-\delta-j\gamma} h^j(y) dy \right] \\
 &\propto A^{j\Phi+\lambda(1+k-\delta-j\gamma)} \left[ c_1 A^{-\lambda(1+k-\delta-j\gamma)} + c_2 \right] \\
 &\propto A^{j\Phi+\max\{0, \lambda(1+k-\delta-j\gamma)\}},
 \end{aligned} \tag{1.26}$$

with  $\epsilon \ll 1$ ,  $c_1$  and  $c_2$  constants. We used the property (1.21) to ensure that the integral  $\int_0^\infty x^{j-1} G(x) dx$  converges for  $j = 1, 2$ , and properties (1.19a-1.19b) to evaluate the integral in  $m$  (in particular,  $h(x) \simeq h_0$  constant for  $x \in (0, \epsilon]$ ).

The SAR is obtained by imposing the metabolic constraint (Eq. 1.15). The total metabolic rate of the community can be computed as:

$$B \propto S \langle nm^\alpha \rangle = S I_{1,\alpha}, \tag{1.27}$$

where Kleiber's law has been used to say that the metabolic rate of an organism of body mass  $m$  is  $b \propto m^\alpha$  and  $\langle nm^\alpha \rangle$  is the average metabolic rate of a species (safe for Kleiber's law proportionality constant). The constraint in Eq. (1.15) implies then that the species number  $S$  must scale as

$$S \propto A(I_{1,\alpha})^{-1} \propto A^z, \text{ with } z = 1 - \Phi - \max\{0, \lambda(1 + \alpha - \delta - \gamma)\}. \tag{1.28}$$

The scaling of the total biomass  $M$  and of the total number of individuals  $N$  are obtained as:

$$M = S \langle nm \rangle = S I_{11} \propto A^{1+\max\{0, \lambda(2-\delta-\gamma)\} - \max\{0, \lambda(1+\alpha-\delta-\gamma)\}}, \tag{1.29}$$

$$N = S \langle n \rangle = S I_{10} \propto A^{1-\max\{0, \lambda(1+\alpha-\delta-\gamma)\}}. \tag{1.30}$$

The exponent of Taylor's law is obtained by noting that

$$\begin{aligned}
 \text{var}(n) &= \langle n^2 \rangle - \langle n \rangle^2 = I_{2,0} - I_{1,0}^2 \\
 &= k_1 A^{2\Phi+\max\{0, \lambda(1-\delta-2\gamma)\}} - k_2 A^{2\Phi+2\max\{0, \lambda(1-\delta-\gamma)\}} \propto A^{2\Phi} \propto \langle n \rangle^2
 \end{aligned} \tag{1.31}$$

where  $k_1, k_2$  are constants and we used the fact that  $\delta > 1$ . Therefore,  $\beta = 2$  regardless of the other exponents' values. An analogous computation can be performed for the abundance of species with fixed typical size, yielding  $\langle n^2 | m \rangle \propto \langle n | m \rangle^2$ . Thus, also the mean abundance in a fixed size range satisfies Taylor's law with  $\beta = 2$ .

The following linking relationships are therefore obtained:

$$z = 1 - \Phi - \max\{0, \lambda(1 + \alpha - \delta - \gamma)\} \quad (1.32a)$$

$$\mu = 1 + \max\{0, \lambda(2 - \delta - \gamma)\} - \max\{0, \lambda(1 + \alpha - \delta - \gamma)\} \quad (1.32b)$$

$$\nu = 1 - \max\{0, \lambda(1 + \alpha - \delta - \gamma)\} \quad (1.32c)$$

Note that Eq. (1.32a) and Eq. (1.32c) imply Eq. (1.3), which was obtained with a heuristic argument in the introduction.

### 1.4.2 The community size spectrum

The community size spectrum  $s(m|A)$  is equal to:

$$\begin{aligned} s(m|A) &= \frac{S}{N} \int_0^\infty n p(n, m|A) dn = \frac{\delta - 1}{A^\Phi} m^{-\delta} \int_0^\infty G\left(\frac{nm^\gamma}{A^\Phi h(m/A^\lambda)}\right) dn \\ &= \frac{\delta - 1}{A^\Phi} m^{-\delta} \left(\frac{A^\Phi}{m^\gamma} h\left(\frac{m}{A^\lambda}\right)\right) \int_0^\infty G(x) dx \propto m^{-\delta - \gamma} h\left(\frac{m}{A^\lambda}\right) \end{aligned} \quad (1.33)$$

where in the first line we have used (1.28), (1.30) and  $\delta > 1$ , and in the second line we performed the change of variable  $x = nm^\gamma / (A^\Phi h(m/A^\lambda))$  and the properties of  $G$  ensure the convergence of the integral. The size spectrum has a power-law dependence on  $m$  and one can identify  $\eta = \gamma + \delta$ , corresponding to Eq. (1.2). Note that both  $s(m|A)$  and  $\langle n | m, A \rangle$  display a cutoff at  $m \propto A^\lambda$ . Nevertheless, their scaling exponents, determining the behavior before the cut-off, satisfy the same relationship that was found with a heuristic calculation which neglected cut-offs.

### 1.4.3 The scaling of the maximum body mass

The last linking relationship stems from the observation that the maximum expected body mass observed in an ecosystem with  $S$  species is  $m_{\max}$  such that

$$P(m \geq m_{\max}) = \int_{m_{\max}}^\infty p(m|A) dm = 1/S. \quad (1.34)$$

In fact, the species present in the ecosystem can be interpreted as  $S$  samples drawn from  $p(m|A)$ , and when the above condition is satisfied there is, on average, only one sampled value among the  $S$  which is larger than  $m_{\max}$ , which is therefore the expected maximum mass

extracted. See box 1.4.4 for a more rigorous derivation of the scaling of the maximum sampled value with sample size. Substituting  $S \propto A^z$  we find  $\int_{m_{\max}}^{\infty} x^{-\delta} dx \propto A^{-z}$ , leading to:

$$m_{\max} \propto A^{\frac{z}{\delta-1}}, \quad (1.35)$$

which implies

$$z = \xi(\delta - 1). \quad (1.36)$$

Note that, if  $\lambda < \xi$ , the cut-off on the abundances might make the largest organism's mass  $m_{\max}$  not observable due to vanishing abundance. One would then have an effective scaling of the maximum observable body mass  $\tilde{m}_{\max} \propto A^\lambda$ .

#### **1.4.4 Scaling of the largest sampled value with sample size for power-law distributions**

This box presents a rigorous proof of the result used in section 1.4.3, stating how the expected largest value in a sample of size  $S$  extracted from a power-law distribution increases with  $S$ , adapted from Newman [2005]. Consider a power-law  $p(m) = (\delta - 1)m^{-\delta}$ , defined on  $[1, \infty]$ . The proof consists in computing the average of the probability distribution  $\pi(m)$  of the largest value in a sample of size  $S$ .  $\pi(m) dm$  is the probability that a particular element (out of  $S$ ) in the sample has value in  $[m, m + dm]$  and all the other have smaller values. Knowing that  $P(m) = \int_m^{\infty} p(m') dm'$  is the probability to find a value larger than  $m$  and therefore  $1 - P(m)$  the probability to find a smaller one,  $\pi(m)$  is expressed as

$$\pi(m) = Sp(m)(1 - P(m))^{S-1}, \quad (1.37)$$

where the factor  $S$  counts the possible choices for the element with the largest value. The average is obtained as

$$\langle m_{\max} \rangle = \int_1^{\infty} m \pi(m) dm = S \int_1^{\infty} m p(m) [1 - P(m)]^{S-1} dm. \quad (1.38)$$

$P(m)$  is obtained as:

$$P(m) = \int_m^{\infty} p(m') dm' = (\delta - 1) \int_m^{\infty} m'^{-\delta} dm' = m^{1-\delta}. \quad (1.39)$$

Substituting  $P(m)$  and  $p(m)$  in Eq. (1.38), one has

$$\begin{aligned} \langle m_{\max} \rangle &= S(\delta - 1) \int_1^{\infty} m^{-\delta+1} [1 - m^{1-\delta}]^{S-1} = S \int_0^1 \frac{y^{S-1}}{(y-1)^{1/(\delta-1)}} dy \\ &= SB(S, (\delta - 2)/(\delta - 1)), \end{aligned} \quad (1.40)$$

where the substitution  $y = 1 - m^{1-\delta}$  was made and  $B(a, b)$  is Legendre's beta function, defined by  $B(a, b) = \frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)}$ , with  $\Gamma(a)$  the complete  $\Gamma$ -function. By using Stirling's approximation, one can show that the function  $B(a, b)$  is approximated by a power function when either of its arguments are large. In particular, when  $a$  is large  $B(a, b) \propto a^{-b}$ . Therefore, when the sample size  $S$  is large

$$\langle m_{max} \rangle \sim SS^{-\frac{\delta-2}{\delta-1}} = S^{1/(\delta-1)}, \quad (1.41)$$

which confirms the result obtained with the heuristic method presented in section 1.4.3.

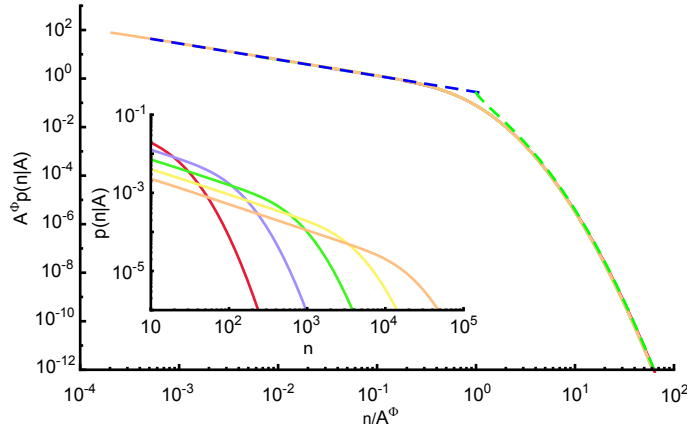


Figure 1.2 – Inset: Relative species abundance  $p(n|A)$  computed numerically via (1.42) with the choices of  $G(x)$  and  $h(x)$  specified in the text and parameter values  $h_0 = 1$ ,  $\sigma = 1/10$ ,  $\delta = 5/4$ ,  $\gamma = 3/4$ ,  $\lambda = 3/4$  and  $\Phi = 1$ . Colors refer to different value of  $A$ ,  $A = 10^2, 10^3, 10^4, 10^5, 10^6$ . Main plot: Collapse of  $p(n|A)$  curves for different values of  $A$  (see text). The collapsed curves (orange) are covered by the dashed lines (blue, green), which represent the approximation in Eq. (1.44).

### 1.4.5 The RSA

Finally, the RSA is the marginal of the fundamental distribution obtained integrating over  $m$ :

$$p(n|A) = \int_1^\infty p(n, m|A) dm. \quad (1.42)$$

The integral in Eq. (1.42) cannot be computed in the general case where the exact forms of  $h$  and  $G$  are unknown. Nevertheless, one can show via analytical approximations and numerical integration that for a particular choice of the two functions the scaling properties of  $p(n|A)$  agree with the scaling assumption put forward by Zillio et al. [2008]. Here, computations are performed for the particular choice  $G(x) = \frac{1}{\sqrt{\pi\sigma}} e^{-\frac{1}{\sigma}(\log x + \sigma/4)^2}$ , with  $\sigma > 0$  constant. Note that  $\int_0^\infty G(x)/x dx = 1$  as prescribed in section 1.3 and  $\int_0^\infty G(x) dx = 1$  as required in Eq. (1.25).

## Chapter 1. Covariations in ecological scaling laws

Regarding  $h(x)$ , it is sufficient to posit that it must be monotonically decreasing, but for ease of computation we consider  $h(x) = h_0 e^{-x}$ , with  $h_0$  a constant. With these choices, it is possible to derive an approximation (see Appendix 1.8.1 for details on the calculations) for the behavior of  $p(n|A)$  in two different  $n$  ranges:

$$p(n|A) = \begin{cases} \frac{1}{n} \frac{\delta-1}{\gamma} \left[ \frac{A^\Phi}{n} h_0 e^{-\frac{\sigma}{4}} \right]^{\frac{1-\delta}{\gamma}} & \text{if } n \simeq A^\Phi h_0 e^{-\sigma/4} \\ \frac{1}{n} \frac{\delta-1}{\sqrt{\pi\sigma}} \frac{e^{-\frac{1}{\sigma} \left[ \log\left(\frac{n}{h_0 A^\Phi}\right) + \frac{\sigma}{4} \right]^2}}{\frac{2\gamma}{\sigma} \left[ \log\left(\frac{n}{h_0 A^\Phi}\right) + \frac{\sigma}{4} \right]} & \text{if } n \gg A^\Phi h_0 e^{-\sigma/4} \end{cases} \quad (1.43)$$

Note that this result can be expressed as:

$$p(n|A) = \frac{1}{n} \tilde{Q}\left(\frac{n}{A^\Phi}\right) = \frac{1}{A^\Phi} Q\left(\frac{n}{A^\Phi}\right) \quad (1.44)$$

with

$$\tilde{Q}(x) = \begin{cases} \frac{\delta-1}{\gamma} \left[ \frac{1}{x} h_0 e^{-\frac{\sigma}{4}} \right]^{\frac{1-\delta}{\gamma}} & \text{if } x \simeq h_0 e^{-\sigma/4} \\ \frac{\delta-1}{\sqrt{\pi\sigma}} \frac{e^{-\frac{1}{\sigma} \left[ \log\left(\frac{x}{h_0}\right) + \frac{\sigma}{4} \right]^2}}{\frac{2\gamma}{\sigma} \left[ \log\left(\frac{x}{h_0}\right) + \frac{\sigma}{4} \right]} & \text{if } x \gg h_0 e^{-\sigma/4} \end{cases} \quad (1.45)$$

and  $Q(x) = \frac{1}{x} \tilde{Q}(x)$ . The validity of this approximation can be verified by integrating  $p(n|A)$  numerically for a particular choice of the exponents  $\delta$ ,  $\gamma$ ,  $\Phi$  and  $\lambda$  and of the parameters  $h_0$  and  $\sigma$ . Figure 1.2 (inset) shows the results of numerical integration for different values of  $A$ . Outside the inset, the curves are collapsed by plotting  $A^\Phi p(n|A)$  versus  $n/A^\Phi$ . The dotted lines represent the approximation in Eq. (1.44), which is shown to describe well the entire curve, excluded a small interval at the junction of the two approximations. In particular, the expression obtained for intermediate  $n$  values still holds for  $n \ll A^\Phi h_0 e^{-\sigma/4}$ . The result is valid regardless of parameters values. Note that  $\langle n \rangle \propto A^\Phi$ , therefore we can write  $p(n|A) = \frac{1}{\langle n \rangle} Q\left(\frac{n}{\langle n \rangle}\right)$ . This result coincides with the scaling assumption put forward by Zillio et al. [2008], with  $f(A) = \langle n \rangle$ .

### 1.4.6 Some observations on the results

In summary, the scaling framework predicts the following exact relationships among scaling exponents:

$$\eta = \gamma + \delta, \quad (1.46a)$$

$$z = 1 - \Phi - \max\{0, \lambda(1 + \alpha - \eta)\}, \quad (1.46b)$$

$$\mu = 1 + \max\{0, \lambda(2 - \eta)\} - \max\{0, \lambda(1 + \alpha - \eta)\}, \quad (1.46c)$$

$$\nu = 1 - \max\{0, \lambda(1 + \alpha - \eta)\}, \quad (1.46d)$$

$$\xi = \frac{z}{\delta - 1}. \quad (1.46e)$$



Eqs. (1.46b) and (1.46d) also imply

$$z + \Phi = \nu. \tag{1.47}$$

The exponent  $\beta$  does not appear because its value is found to be independent from the other exponents. Figure 1.3 summarizes the predictions on the values or bounds of scaling exponents based on the linking relationships (1.46a-e), for different possible values of the independent exponents.

Notice that there are only 5 independent exponents (e.g.  $\gamma$ ,  $\delta$ ,  $\Phi$ ,  $\alpha$  and  $\lambda$ ), whereas the observable laws amount to 10: Kleiber's law and Eqs. (1.17), (1.18), (1.28)–(1.30), (1.33) and (1.35). Note that Eq. (1.18) contains three laws because it describes the scaling of the average abundance of a species with its typical body mass and with the area of the ecosystem and the scaling of its cut-off with the area of the ecosystem. Eq. (1.46b) implies that, in any ecosystem where  $z > 0$ , as typically observed,  $\Phi < 1$  and therefore species' densities decrease with increasing area. In Chapter 2, this prediction will be shown to be verified by a dataset of lizard species encompassing several islands worldwide [Novosolov et al., 2015]. Eq. (1.46e) is compatible with the linking relationship derived in Southwood et al. [2006], which is shown here to be one component of a broader set of linking relationships (see next paragraph).

While some of the obtained scaling relationships coincide with the results of heuristic calculations performed assuming pure power-law behavior, their derivation within the framework allows us to assess the role of finite-size effects. In particular, while the combination of Eqs (1.46b) and (1.46d) can be obtained by the heuristic reasoning explained in the introduction, the effect of the cut-off on masses with exponent  $\lambda$  on the values of  $z$  and  $\nu$  cannot be obtained by that reasoning. However, such information gives interesting predictions, as Eq. (1.46d) fixes a constraint on the exponents  $\alpha$ ,  $\delta$  and  $\gamma$  such that  $\nu$  assumes its most likely value of 1,  $1 + \alpha < \delta + \gamma$ . Additionally, this very general formulation allows to study the consequences of any change in the initial hypotheses, including non-power law behavior, as will be shown in section 1.6.

### 1.4.7 Compatibility with previous works

Southwood et al. [2006] derived a linking relationship that is equivalent to Eq. (1.46)e. Here, their result is briefly reviewed and corrected due to a few miscalculations in the original paper.

The authors start from the observation [May, 1988] that the total number of species of length class  $L$  scale as  $S_L(L) \propto SL^{-\Delta}$ , where  $\Delta = 3/2$  and  $S$  is the total number of species in the ecosystem. The estimate  $\Delta = 3/2$  made in that study [Southwood et al., 2006] is incorrect, because the correction needed to account for logarithmic binning [Stegen and White, 2008, White et al., 2008] is missing (see also Box 2.2 in the next chapter). The correct estimate that accounts for logarithmic binning [Stegen and White, 2008] is  $\Delta = 5/2$ . First, Southwood et al. [2006] say that the maximum length is obtained by imposing  $S_L(L_{\max}) = SL_{\max}^{-\Delta} = 1$ , i.e.

## Chapter 1. Covariations in ecological scaling laws

General case					SSI
$\eta = \delta + \gamma$ $\left. \begin{aligned} z &= 1 - \Phi - \max\{0, \lambda(1 + \alpha - \eta)\} \\ v &= 1 - \max\{0, \lambda(1 + \alpha - \eta)\} \end{aligned} \right\} z + \Phi = v$ $\mu = 1 + \max\{0, \lambda(2 - \eta)\} - \max\{0, \lambda(1 + \alpha - \eta)\}$ $z = \xi(\delta - 1)$					
If $\eta > 2$	LIZ	If $\eta \leq 2$			
		$1 + \alpha \leq \eta \leq 2$		$1 \leq \eta \leq 1 + \alpha$	
		$\lambda > 0$	$\lambda = 0$	$\lambda > 0$	$\lambda = 0$ Forests
$z = 1 - \Phi$		$z = 1 - \Phi$	$z = 1 - \Phi$	$z = 1 - \Phi - \lambda(1 + \alpha - \eta)$	$z = 1 - \Phi$
$\mu = 1$		$\mu = 1 + \lambda(2 - \eta)$	$\mu = 1$	$\mu = 1 + \lambda(1 - \alpha)$	$\mu = 1$
$v = 1$		$v = 1$	$v = 1$	$v = 1 - \lambda(1 + \alpha - \eta)$	$v = 1$
		$\gamma < 1, \delta < 2$	$\gamma < 1, \delta < 2$	$\gamma < \alpha, \delta < 1 + \alpha$	$\gamma < \alpha, \delta < 1 + \alpha$
		$z < \xi$	$z < \xi$	$z < \xi\alpha, \Phi < 1 - z$	$z < \xi\alpha$
		If $\xi = 1/2, \alpha = 3/4$		If $\xi = 1/2, \alpha = 3/4$	
		$z < \frac{1}{2}$		$z < \frac{3}{8}$	

Figure 1.3 – Scheme of predictions on the values or bounds of scaling exponents based on the linking relationships (1.46a-e). Names of datasets which are analyzed in Chapter 2 are located in different columns according to the available information on their exponents' values. Forests include Barro Colorado Island [Condit et al., 2012], Luquillo [Zimmerman et al., 2010], SSI stands for Sunda Shelf Islands [Okie and Brown, 2009], LIZ for the dataset of lizard densities on islands worldwide [Novosolov et al., 2015]. See Chapter 2 for details on the datasets. Note that the relationship  $z < \xi\alpha$  is valid for forests only before the physiological constraint has been attained (see section 1.6.3).

$S \propto L_{\max}^{\Delta}$ . This calculation is also incorrect, because it makes an improper use of probability distributions. In fact,  $S_L(L)$  is the fraction of species with length in  $[L, L+dL]$  and the maximum species' length is found by imposing that the probability of finding a species with length larger than  $L_{\max}$  be equal to  $1/S$  (as explained in section 1.4.3 and Box 1.4.4) which results in  $S \propto L_{\max}^{\Delta-1}$ . Incidentally, these two miscalculations compensate each other and together lead to the estimate  $S \propto L_{\max}^{3/2}$ . However, and this is crucial to derive the correct linking relationship, the correct equation is  $S \propto L_{\max}^{\Delta-1}$  and not the one suggested in [Southwood et al., 2006], i.e.  $S \propto L_{\max}^{\Delta}$ . The equation relating  $S$  and  $L$  converts into a species-mass relationship via the scaling  $L \propto m^{1/3}$ , where we have assumed that body density does not scale with body mass. Specifically, one finds  $S \propto m_{\max}^{\delta-1}$  with  $\delta = (\Delta + 2)/3$ . Then, one can use the finding by [Burness et al., 2001],  $m_{\max} \propto A^{\xi}$ , to derive the linking relationship between  $z$ ,  $\xi$  and  $\delta$ . In fact, by comparing  $S \propto m_{\max}^{\delta-1} \propto A^{\xi(\delta-1)}$  with the SAR  $S \propto A^z$  one obtains the linking relationship  $z = \xi(\delta - 1)$ , which coincides with Eq. (1.46)e. However, it should be noted that, in [Southwood et al., 2006],  $z$  was meant as the exponent of the nested species-area relationship, differently from here. Nonetheless, the linking relationship is the same. The original results presented in section 1.4 thus agree with the earlier result by Southwood et al. [2006], where one of the linking relationships was discovered.

As mentioned in the introduction, an approach based on a scaling hypothesis for  $p(n, m|A)$  was also presented in Banavar et al. [2007]. The original work presented in this chapter has nonetheless important differences with respect to the approach in Banavar et al. [2007], namely: 1) the enforcement and the implications of resource limitation, 2) the validation based on a broad class of community dynamics models, and 3) the richer set of macroecological laws that the scaling framework accounts for, most importantly Damuth's law, which allows us to reconcile the predicted linkages with empirical data (see Chapter 2) and community dynamics models (see, e.g. Fig. 1.4). The framework in [Banavar et al., 2007], expressed by Eqs (1.5) and (1.6), leads to different predictions for the pattern covariations (e.g. the linking relationship  $\eta = \delta$ ) that are falsified by empirical data and by the model simulations (Fig. 1.4, see section 1.5). Note that the introduction of a constraint on the total community consumption rate in the framework of Banavar *et al.* would not affect the relationship  $\eta = \delta$ , which is instead a byproduct of the assumptions on  $p(n, m|A)$ .

## 1.5 Stochastic models of community dynamics

The scaling hypothesis for the fundamental distribution  $p(n, m|A)$  presented in section 1.3 was justified by the fact that it is the simplest hypothesis compatible with the empirically observed ecological scaling laws. The hypothesis in Eq. (1.24) might not be the only one satisfying such empirical constraint, therefore an additional corroboration of its appropriateness is useful to strengthen the tenet. This section is devoted to showing that a broad class of stochastic models for the dynamics of a community limited by resource supply supports Eq. (1.24). In fact, despite major changes in the speciation dynamics and regardless of parametrization, all models which reproduce empirical scaling laws are compatible with the finite-size scaling

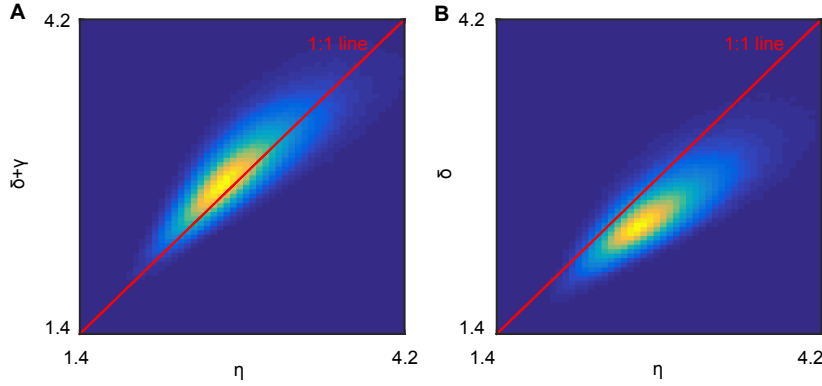


Figure 1.4 – Density-scatter plot of  $\delta + \gamma$  (panel a) and  $\delta$  (panel b) versus  $\eta$  in simulations of the stochastic community dynamics model (model *a*, see section 1.5), with the exponents estimated at each sampling time-point. The parameters of the stochastic community dynamics model are reported in Fig. 1.14; shown are simulation data for the largest simulated area  $A = 10^3$ . Density histograms are normalized to one, with blue representing the value zero and yellow the value one. Note that the small deviations from the 1:1 line are also due to statistical errors in the estimation of the exponents at each time step. Panel b) shows that the prediction [Banavar et al., 2007] that  $\eta = \delta$  is not supported by simulation data. See section 1.5.1 for details on how the exponents were estimated.

structure of Eq. (1.24).

Several variants of a basic community dynamics model are considered, all accounting for the constraint on resource supply rate and incorporating empirically observed allometric relationships for the dependence of vital rates on individuals' body sizes [Brown et al., 2004]. In all the models, the birth and death rates at which an individual of a species of mass  $m_i$  and abundance  $n_i$  is born or dies are, respectively:

$$\begin{aligned}
 u_i &= m_i^{-\theta} n_i, \\
 v_i &= \left[ v_0 + (1 - v_0) c \frac{\sum_j n_j m_j^\alpha}{\mathcal{R}} \right] m_i^{-\theta} n_i,
 \end{aligned}
 \tag{1.48}$$

and thus the per-capita growth rate of species  $i$  is  $\frac{u_i - v_i}{n_i} = (1 - v_0) \left[ 1 - \frac{c}{\mathcal{R}} \sum_j n_j m_j^\alpha \right] m_i^{-\theta}$ , which is equal to zero when  $c \sum_j n_j m_j^\alpha = \mathcal{R} \propto A$ , where  $\mathcal{R}$  is the resource supply rate. At the stationary state, therefore, the total rate of resource consumption of the community fluctuates around  $\mathcal{R}$  but the ecological dynamics continues and determines species' abundances through the constraints imposed by resources and by physiological rates. Sections 1.5.1 to 1.5.4 present several model variants. Starting from a basic model, where the total number of species is fixed to  $S \propto A^z$ , variants are produced by changing the speciation mechanism or by letting  $S$  free to fluctuate, in which case the relationship  $\langle S \rangle \propto A^z$  is an emergent property of the community dynamics. Testing several model variants allows assessing the robustness of the results

## 1.5. Stochastic models of community dynamics

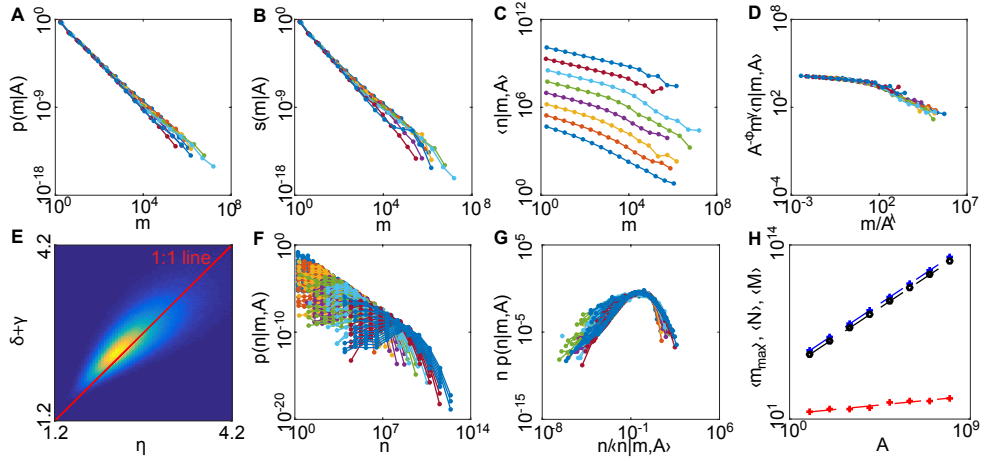


Figure 1.5 – Scaling patterns from the basic community dynamics model. Different colors refer to different values of  $A = 10^i$ , from  $i = 1$  (lower blue curve in panel C) to  $i = 8$  (upper blue curve in panel C). Panels A, B, C and F show respectively  $p(m|A)$ ,  $s(m|A)$ ,  $\langle n|m, A \rangle$  and  $p(n|m, A)$  at stationarity. Panels D and G show collapses of  $\langle n|m, A \rangle$  and  $p(n|m, A)$ . Eqs. (1.17), (1.18) and (1.20) are verified because the curves  $np(n|m, A)$  versus  $n/\langle n|m, A \rangle$  collapse on the same curve for different  $A$  (panel G), and so do the curves  $m^\gamma A^{-\Phi} \langle n|m, A \rangle$  versus  $m/A^\lambda$  (panel D). Panel E: density histogram plot of  $\eta$  vs  $\delta + \gamma$  at different times. Panel H: scaling of  $\langle m_{\max} \rangle$  (red crosses and dashed lines),  $\langle N \rangle$  (black dots and dashed lines) and  $\langle M \rangle$  vs  $A$  (blue crosses and dashed lines).

to changes in the models' assumptions. Data-collapses of  $p(m|A)$ ,  $p(n|m, A)$  and  $\langle n|m, A \rangle$  calculated using model data (Fig. 1.5) show that they all comply with Eqs. (1.17), (1.18) and (1.20). Note that the values of the scaling exponents depend on model specifications, but the scaling properties of the fundamental distribution  $p(n, m|A)$  always hold.

### 1.5.1 Basic community dynamics model

This section concerns the simplest model of community dynamics that reproduces the set of empirically-observed macroecological laws reported in section 1.4. This model will be referred to as “basic model” in the following.

In the basic model, each species speciates with probability  $w$  per unit time (i.e., species-specific speciation events are Poisson-distributed with rate  $w$ ). At each speciation event, a species is selected at random and a random fraction of individuals from such species is assigned to a new species  $j$ . This speciation mechanism, called ‘fission’ speciation, is intended to mimic the mechanism of allopatric speciation (also called vicariant speciation), when a fraction of a species becomes isolated from the rest of the individuals and therefore experiences different selective pressures, eventually bringing to speciation. The mass of the new species is obtained from the mass of the parent species as  $m_j = \max\{m_0; qm_i\}$  where  $q$  is extracted from a lognormal distribution with mean and variance equal to unity so that the descendant has, on

average, the same mass of the parent species. The maximum in the expression for  $m_j$  ensures that the bound on the minimum mass  $m_0$  that a species can attain is satisfied. The mass of the parent species is left unchanged. Species' masses thus undergo a process that is a combination of this evolutionary dynamics of masses and of the birth/death dynamics. Note that the evolutionary dynamics of body masses resembles a multiplicative bounded process, known to produce power-laws [Solomon and Levy, 1996, Sornette and Cont, 1997] (see Box 1.5.2 for a proof). However, the correspondence is not exact. In fact, when a new mass is introduced in the system, the parent mass also remains and can still speciate giving birth to new masses. As shown by the results of the model simulations reported in the following, the outcome of this process, in combination with the ecological dynamics, is also a power-law, although the exponent differs from that expected in the case of a simple bounded multiplicative process. In fact, the latter predicts an exponent  $\delta = 3/2$  when  $q$  has mean and variance equal to unity, while the numerical simulations always produce  $\delta > 2$ .

The number of species  $S$  is set to a constant value proportional to the area:  $S = 10A^z$ . Although the number of species in natural ecosystems may fluctuate in time, fixing it in the basic model allows us to vary the scaling exponent  $z$  to effectively account for relevant ecological and evolutionary processes not included in the model which may affect the value of  $z$  in natural ecosystems (see also section 1.7 for further comments on this). Note that fixing the number of entities in the model (here,  $S$ ) is a common approximation in many related fields, such as population genetics (e.g., the Wright-Fisher model [Wright, 1931] with fixed population size  $N$ ) and neutral and metacommunity theory [Azaele et al., 2016]. To maintain  $S$  constant, each extinction event is followed by a speciation event. Viceversa, at each speciation event, extinction is enforced on a species selected at random with probability inversely proportional to its abundance (i.e., more abundant species are less likely to go extinct) and proportional to the power  $-\theta$  of its mass, which accounts for the fact that ecological rates are faster for smaller species. A variation on this extinction rule is discussed in section 1.5.3.

The total number of individuals  $N = \sum_{i=1}^S n_i$  and the total biomass  $M = \sum_{i=1}^S n_i m_i$  are not fixed in the basic model (nor in the other variants discussed), but fluctuate in time around mean values that depend on the models' parameters and, most importantly, on the ecosystem area  $A$ . In other words, the mean biomass and the mean total abundance are given by a balance between birth, death and speciation events, with the constraint of resource supply limitation set by the ecosystem area  $A$ . The model thus allows to study the scaling of the total number of individuals and the total biomass as functions of  $A$ .

The distribution  $p(m|A)$  exhibits power-law behavior in  $m$ , as prescribed by Eq. (1.17) (Fig. 1.5A). The size spectrum is also a power-law across several orders of magnitude (Fig. 1.5B). The exponents  $\delta$  and  $\eta$  were estimated, respectively, as the average of the exponents  $\delta$  and  $\eta$  computed at every sampling time in the simulation data for the largest area with a maximum-likelihood algorithm [Clauset et al., 2009] (see Box 2.2 in Chapter 2). The curves  $\langle n|m, A \rangle$  exhibit power-law behavior in  $m$  and  $A$  with a cutoff at large  $m$  (Fig. 1.5C). Data collapse (Fig. 1.5D) shows that its functional form is the one given by Eq. 1.18. In fact, the curves

## 1.5. Stochastic models of community dynamics

$m^\gamma A^{-\Phi} \langle n|m, A \rangle$  plotted versus  $m/A^\lambda$  collapse onto the same curve for different values of  $A$ . Estimates of  $\Phi$ ,  $\gamma$  and  $\lambda$  were computed by minimizing a measure of data collapse [Bhattacharjee and Seno, 2001]. Moreover, Fig 1.5G shows that the curves  $n p(n|m, A)$  versus  $n/\langle n|m, A \rangle$  collapse onto the same curve for different values of  $m$  and  $A$ , implying that Eq. 1.20 holds. Fig 1.5F shows the uncollapsed curves. The mean total biomass  $\langle M \rangle$  and the mean total abundance  $\langle N \rangle$  were measured for each value of  $A$  as the mean of the total biomass and the mean of the total abundance across sampling times. They show power-law behavior (Fig. 1.5H) and the exponents  $\mu$  and  $\nu$  were measured via least-squares fitting of log-transformed data. To measure the scaling of the mean maximum mass  $\langle m_{\max} \rangle$  with the area  $A$ , we verified through data-collapse that the distribution of  $m_{\max}$  in the simulations (estimated by measuring the maximum mass at each sampling time point) is of the form  $P(m_{\max}|A) = m_{\max}^{-\chi} L\left(\frac{m_{\max}}{A^\omega}\right)$ , and the scaling function  $L$  is such that  $\int_0^\infty x^{j-\chi} L(x)$  converges for  $j = 0, 1$ . This scaling form of  $P(m_{\max}|A)$  implies that the mean value of  $m_{\max}$  satisfies  $\langle m_{\max} \rangle \propto A^{\omega(2-\chi)} = A^\xi$  (Fig. 1.5H), that is, a power-law as predicted by the theoretical scaling framework. All the estimated values for the exponents and their 68% confidence intervals are reported in table 1.1.

Parameter values used to generate the simulation data reported in Fig. 1.5 are:  $w = 10^{-3}$ ,  $z = 1/4$ ,  $\alpha = 3/4$ ,  $\theta = 1/4$ ,  $v_0 = 1/2$  and  $c = 10^{-5}$ . The stochastic model was simulated via a Gillespie tau-leap algorithm with estimated midpoint technique [Gillespie, 2001], with time step  $\tau = 1$ .

Because the *ansatz* for the fundamental distribution  $p(n, m|A)$  given by Eqs. (1.17),(1.18) and (1.20) holds, the linking relationships among exponents (Eqs. 1.46) are satisfied at steady state by the basic model. The linking relationship  $\eta = \delta + \gamma$  is satisfied by the mean values of the exponents, and the density scatter-plot computed counting the occurrences of the pairs  $(\eta, \delta + \gamma)$  during the temporal evolution of the community dynamics model (Fig. 1.5E, shown are simulation data for the largest area value) is peaked along the 1:1 line. Thus, Eq. 1.46a is satisfied, on average, during the temporal evolution of the community dynamics model.

A thorough exploration of the parameters' space is computationally unfeasible. Nonetheless, a partial exploration, performed varying the values of the parameters that are most meaningful for the dynamics (i.e. Kleiber's law exponent  $\alpha$ , the speciation rate  $w$ , the SAR exponent  $z$  and the exponent  $\theta$  that describes the scaling of vital rates with body mass), suggests that the scaling characterization of the stationary state always holds and the linking relationships in Eqs. 1.46 are always satisfied. Such exploration of the parameter space was performed starting from the set of parameters  $w = 10^{-3}$ ,  $z = 1/4$ ,  $\alpha = 3/4$  and  $\theta = 1/4$  (parameters used to generate Fig. 1.5) and varying one or two parameters at a time, keeping the other ones fixed. The parameters  $v_0$  and  $c$  which appear in Eq. 1.48 were fixed to  $v_0 = 1/2$  and  $c = 10^{-5}$ . The figures showing the ecological patterns computed at stationarity for each set of parameters and the corresponding tables reporting the estimates of the scaling exponents are in Appendix 1.8.3. For each set of parameters, the relationships in Eqs. 1.46 are satisfied within errors, the data collapses predicted by our scaling framework hold and the density scatter-plot of  $\eta$  versus  $\delta + \gamma$  estimated at each time-step (Figs. 1.6–1.12, panel e) is peaked along the 1:1 line,

implying that the linking relationship (1.46a) is satisfied, on average, at all times.

### 1.5.2 Bounded multiplicative process

The content of this box is adapted from [Sornette and Cont, 1997].

A multiplicative process  $w$  is described by

$$w_{t+1} = q_t w_t \quad (1.49)$$

where  $q_t > 0$  is a stochastic variable with probability distribution  $\Pi(q_t)$ . When the process has no bound (unless 0),  $w$  is distributed according to a log-normal distribution at stationarity. In fact,

$$\log w_{t+1} = \log q_t + \log w_t = \sum_{i=0}^t \log q_i + \log w_0. \quad (1.50)$$

The variable  $x := \log w$  therefore evolves according to a random walk with steps  $l := \log q$  distributed according to the probability distribution  $\pi(l) = e^l \Pi(e^l)$ . Therefore, thanks to the central limit theorem, and assuming  $\pi(l)$  has finite variance,  $x = \log w$  is a gaussian variable, making  $w$  lognormal.

Consider now the case when the process  $w$  has a lower bound  $w_0$ . This corresponds to putting back  $w_t$  to  $w_0$  as soon as it would become smaller. In terms of the process  $x = \log w$ , we then find the following behavior:

- If  $\langle l \rangle > 0$ , the random walk of  $x$  is biased towards  $+\infty$ , therefore the lower boundary has no effect in the long term (apart from small boundary effects at  $x_0$ ), therefore the result is still a lognormal distribution.
- If  $\langle l \rangle < 0$ , the random walk drifts toward the barrier. At the stationary state, the drift is balanced by reflection. The stationary solution can be found by writing the master equation for the process  $x$

$$P(x, t+1) = \int_{-\infty}^{\infty} \pi(l) P(x-l, t) dl. \quad (1.51)$$

When  $\pi(l)$  is a narrow enough distribution (for example, a gaussian) one can approximate the master equation by expanding  $P(x-l)$  as

$$P(x-l) = P(x, t) - l \frac{\partial P}{\partial x} \Big|_{x,t} + \frac{1}{2} l^2 \frac{\partial^2 P}{\partial x^2} \Big|_{x,t} \quad (1.52)$$



which, substituted into the master equation, leads to the Fokker-Plank equation:

$$\frac{\partial P(x, t)}{\partial t} = -\langle l \rangle \frac{\partial P(x, t)}{\partial x} + D \frac{\partial^2 P(x, t)}{\partial x^2} \quad (1.53)$$

where  $D = \langle l^2 \rangle - \langle l \rangle^2$ . The stationary solution is of the form  $P_{stat}(x) = A - \frac{B}{s} e^{-sx}$ , where  $s = \frac{\langle l \rangle}{D}$  and  $A, B$  are integration constants. The lower boundary is imposed as

$$\int_{x_0}^{\infty} P_{stat}(x) dx = 1, \quad (1.54)$$

which implies  $A = 0, B = s^2 e^{sx_0}$ , therefore  $P_{stat}(x) = s e^{s(x-x_0)}$ . Translating back to the variable  $w$ , this gives

$$P_{stat}(w) = \frac{sw_0^{-s}}{w^{1+s}}, \quad (1.55)$$

that is, a power-law with exponent  $-(1 + s)$ .

To compare this prediction with the results of the simulations of the stochastic model in section 1.5.1, let us remind that in the model  $q$  is distributed as a lognormal with mean and variance 1. Therefore,  $l$  is a gaussian variable with mean  $\langle l \rangle = \langle \log q \rangle = -\frac{\log 2}{2}$  and variance  $var(l) = var(\log q) = \log 2$ . Therefore,  $s = \frac{\log 2 / 2}{\log 2} = \frac{1}{2}$ , and the exponent of the power law is  $-3/2$ . However, in the results of the stochastic models presented in the text, the exponent  $-\delta$  of the species' body mass distribution is always found to be  $< -2$ . This difference is due to the not exact correspondence of the two process of mass evolution following speciation and to the interplay with ecological dynamics.

### 1.5.3 Variation on the speciation dynamics

A variation of the basic model is investigated in which the species that goes extinct at each speciation event (to maintain  $S$  constant) is chosen randomly with a weight inversely proportional to its abundance, but independent of its mass. The model was run with the same parameter values reported in section 1.5.1 for the basic model. Such modified model is compatible with the scaling framework, and thus with the predicted pattern covariations, which are verified within the errors. Table 1.1 in Appendix 1.8.3 reports the corresponding exponents values and Fig. 1.13 displays the macroecological patterns in this model.

### 1.5.4 Fluctuating number of species

In order to further investigate the sensitivity of the results to changes in the dynamic model assumptions, in this section the constraint of a fixed number of species  $S$  is relaxed. This is achieved this by maintaining the ecological dynamics of births and deaths as in the basic

model and by modifying the speciation dynamics in two different ways:

- a. At each time step, the number of species that undergo speciation is drawn from a Poisson distribution with rate  $w^4$ . The species that undergo speciation are selected randomly. At speciation, a random number  $n'_i < n_i$  of individuals of species  $i$  maintains the original mass  $m_i$ , whereas a number  $n_j$  of individuals are assigned to a new species  $j$  with mass  $m_j = \max\{m_0, qm_i\}$ , where  $q$  is drawn from a lognormal distribution with constant mean and variance. To avoid instability (i.e. extreme fluctuations, see Chapter 3), the sum of the consumption rates of species  $i$  and  $j$  after speciation is imposed to be equal to the consumption rate of species  $i$  before speciation. This is done by setting the abundance of species  $j$  such that  $n'_i m_i^\alpha + n_j m_j^\alpha = n_i m_i^\alpha$ , i.e.  $n_j = (n_i - n'_i) m_i^\alpha / m_j^\alpha$ .
- b. Same as in model a, but the species that undergo speciation are selected randomly with a weight proportional to their abundance, so that more abundant species are more likely to speciate.

Both these models give rise to the empirically observed set of macroecological laws, with exponents values depending both on the model specifications and on the model parameters. Most importantly, despite differences in the speciation dynamics, these models are also compatible with the scaling framework. Thereby, macroecological patterns in these models comply with the predicted pattern covariations (Eqs. 1.46). Tables 1.3 and 1.4 in Appendix 1.8.3 report the exponents values measured in these models and Figs. 1.14–1.17 display the corresponding macroecological patterns. Parameter values used to run the models are reported in the figures captions.

### 1.5.5 Value of $\eta$ in the community dynamics models

The size spectrum exponent  $\eta$  in natural ecosystems typically assumes values  $\eta \in (1, 2]$  (see Tables 2.2, 2.3 and 2.4 in Chapter 2), although values of  $\eta > 2$  can also be found in marine environments (Table 2.2). All the community dynamics models investigated in the previous sections yield values of  $\eta$  that are on average larger than 2 (the average is performed over time, see Tables 1.1, 1.2 and 1.3), although panels *e* in Figs. 1.6–1.17 show that  $\eta$  can assume values smaller than 2 at any fixed time point (i.e. in snapshots of the ecosystem). Unfortunately, a suitably broad exploration of the parameters space in the models is computationally unfeasible, as the estimation of scaling exponents requires several hours of computation in a high-performance computer in order to properly estimate the tails of the distribution  $p(n, m|A)$ . However, based on the exploration of parameters' space,  $\langle \eta \rangle = 2$  does seem to be a lower limit in the community dynamics models. Increasing the mean  $\bar{q}$  of the multiplicative factor  $q$  that specifies the mass of the descendant species at a speciation event (this may be

---

<sup>4</sup>Note that having species' specific speciation rate to be a constant, as in the basic model, or total speciation rate to be a constant does not change qualitatively the results. It only changes the value of the overall speciation rate, which in the first case is  $\tilde{w} = wS$ . Since in this model  $S$  fluctuates in time, the second option was chosen to have a constant overall speciation rate.

seen as an implementation of Cope's rule [Rensch, 1948], which postulates that descendant lineages tend to increase in body size) causes a reduction of the mean size spectrum exponent  $\langle \eta \rangle$  (Figs. 1.16 and 1.17, and Table 1.4). Nonetheless, parameter sets that yield  $\langle \eta \rangle < 2$  lead to communities that are very unstable and that rapidly go towards extinction. It thus appears that the community dynamics models considered here are missing processes that would allow multiple species to coexist at a stable equilibrium with  $\langle \eta \rangle < 2$ . An hypothetical reason for this behavior may be the fact that the models assume a well-mixed system, unlike terrestrial ecosystems such as forests. In this sense, it may not be coincidental that values of  $\eta > 2$  are typically found in aquatic ecosystems (Table 2.2) rather than terrestrial ones (Tables 2.2, 2.3 and 2.4).

## 1.6 Generalizations of the scaling framework

In this section, some generalizations of the scaling framework are presented. Specifically, section 1.6.1 introduces an intra-specific size distribution in the framework, presenting two alternative shapes for such distribution based on empirical evidences. Section 1.6.2 accounts for a curvature in Kleiber's law, which has been occasionally observed empirically. Section 1.6.3 shows that the framework can be generalized to alternative shapes of the distribution of species' typical body masses  $p(m|A)$ , either having a cut-off or displaying an internal mode. Finally, section 1.6.4 presents the generalization to a two-trophic level system.

### 1.6.1 Intra-specific size distribution

In the framework presented in section 1.3, each individual of a species is assumed to have the same mass, thereby neglecting intra-specific size variability. In this section, the effects of accounting for intra-specific size distributions are analyzed. Two types of size distributions with different scaling properties found in empirical datasets are considered: distributions with an internal mode (e.g. lognormal) and power-laws.

#### Internal mode intra-specific size distributions

Let  $p(m|\bar{m})dm$  be the probability that an individual of a species with typical body mass  $\bar{m}$  has an individual body mass  $m$ . Assume that  $m$ , similarly to  $\bar{m}$ , has a lower limit  $m_0$ . Giometto et al. [2013] found that protist species belonging to four different phyla and covering five orders of magnitude in mass have size distributions compatible with the scaling form:

$$p(m|\bar{m}) = \frac{1}{m} \mathcal{F}_1\left(\frac{m}{\bar{m}}\right), \quad (1.56)$$

where  $\mathcal{F}_1(x) \rightarrow 0$  suitably fast for  $x \rightarrow 0$  and  $x \rightarrow \infty$ . In Giometto et al. [2013],  $p(m|\bar{m})$  was experimentally found to be compatible (within accuracy) with a log-normal distribution, i.e.  $\mathcal{F}_1(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(\log x - \mu)^2}{2\sigma^2}}$ . By assuming that intra-specific size distributions are described by Eq.

## Chapter 1. Covariations in ecological scaling laws

(1.56), the results of the scaling frameworks (section 1.4) hold exactly, as shown in this section. The moment  $(j,k)$  is given by:

$$\begin{aligned} I_{j,k} &= \int_0^\infty \int_1^\infty \int_1^\infty n^j m^k p(n, \bar{m}|A) p(m|\bar{m}) dm d\bar{m} dn \\ &= \int_0^\infty \int_1^\infty n^j p(n, \bar{m}|A) \langle m^k | \bar{m} \rangle d\bar{m} dn \end{aligned} \quad (1.57)$$

where :

$$\langle m^k | \bar{m} \rangle := \int_1^\infty m^k p(m|\bar{m}) dm. \quad (1.58)$$

Using Eq.(1.56), one obtains:

$$\begin{aligned} \langle m^k | \bar{m} \rangle &= \int_1^\infty m^{k-1} \mathcal{F}_1\left(\frac{m}{\bar{m}}\right) dm = \bar{m}^k \int_{1/\bar{m}}^\infty x^{k-1} \mathcal{F}_1(x) dx \\ &\simeq \bar{m}^k \int_0^\infty x^{k-1} \mathcal{F}_1(x) dx \propto \bar{m}^k \end{aligned} \quad (1.59)$$

where  $x = m/\bar{m}$  and  $\int_0^\infty x^{k-1} \mathcal{F}_1(x) dx$  is a constant. The prolongation of the integral to 0 is an acceptable approximation when  $\mathcal{F}_1$  decays sufficiently fast. Substituting this result in Eq. (1.57), we have

$$I_{j,k} = \int_0^\infty \int_1^\infty n^j \bar{m}^k p(n, \bar{m}|A) d\bar{m} dn, \quad (1.60)$$

which corresponds to Eq. (1.26). Therefore, introducing an intra-specific size distribution of the type in equation (1.56) does not change the scaling of the moments  $I_{j,k}$  with the area. The linking relationships (1.46b–d), whose derivation relies on the scaling of  $I_{j,k}$  with  $A$ , are thus unchanged.

Furthermore, the linking relationship in Eq. (1.46a) is also unchanged. In fact the size spectrum is obtained as:

$$\begin{aligned} s(m|A) &= \frac{S}{N} \int_0^\infty n \int_1^\infty p(n, \bar{m}|A) m^{-1} \mathcal{F}_1\left(\frac{m}{\bar{m}}\right) d\bar{m} dn \\ &= \frac{S}{N} \int_0^\infty \int_1^\infty \bar{m}^{-\delta} G\left[\frac{n\bar{m}^\gamma}{A^\Phi h(\bar{m}/A^\lambda)}\right] m^{-1} \mathcal{F}_1\left(\frac{m}{\bar{m}}\right) d\bar{m} dn \\ &= m^{-\delta} \frac{S}{N} \int_0^\infty \int_{1/m}^\infty x^{-\delta} G\left[\frac{n(xm)^\gamma}{A^\Phi h(xm/A^\lambda)}\right] \mathcal{F}_1\left(\frac{1}{x}\right) dx dn \\ &= m^{-\delta-\gamma} A^\Phi \frac{S}{N} \int_{1/m}^\infty x^{-\delta-\gamma} h\left(\frac{xm}{A^\lambda}\right) \mathcal{F}_1\left(\frac{1}{x}\right) dx \int_0^\infty G(y) dy, \end{aligned} \quad (1.61)$$

where  $x = \bar{m}/m$ ,  $y = [n(xm)^\gamma] / [A^\Phi h(xm/A^\lambda)]$  and  $\int_0^\infty G(y) dy$  is a constant. Note that the scaling of  $s(m|A)$  with  $m$  cannot be computed analytically from (1.61), because  $m$  appears both at the lower limit of the integral in  $x$  and in the argument of  $h$ . However, for large  $m$ , the lower limit of the integral in  $x$  tends to 0 and thus  $s(m|A) \propto m^{-\delta-\gamma} \tilde{h}(m/A^\lambda)$  in the limit of large  $m$ , where  $\tilde{h}(y) = \int_0^\infty x^{-\delta-\gamma} \mathcal{F}_1(1/x) h(xy) dx$  has the same limiting behavior of  $h(y)$  at

$y \rightarrow 0$  and  $y \rightarrow \infty^5$ , and thus the linking relationship  $\eta = \delta + \gamma$  (Eq. 1.46a) still holds.

Eq. (1.46e) is obtained as in section 1.4, with  $\xi$  being the exponent of the scaling of the largest species' typical mass (not of its largest organism') with area  $A$ . Nevertheless, for an intraspecific distribution of the form in Eq. (1.56), the largest organism's mass scales with the area with the same exponent as the largest species' mass. In fact, for each species, the body mass of its largest individual, being the largest sampled value in  $n$  extractions from the pdf in Eq. (1.56), scales linearly with  $\bar{m}$ . This is seen by noting that the expected maximum value extracted in  $n$  extractions is the value  $m_{\max}$  such that, on average, only one extracted value is  $\geq m_{\max}$ :

$$\frac{1}{n} = \int_{m_{\max}}^{\infty} p(m|\bar{m}) dm = \int_{m_{\max}}^{\infty} \frac{1}{m} \mathcal{F}_1\left(\frac{m}{\bar{m}}\right) dm = \int_{m_{\max}/\bar{m}}^{\infty} \frac{1}{x} \mathcal{F}_1(x) dx. \quad (1.62)$$

Differentiating with respect to  $\bar{m}$  gives

$$0 = \left[ m_{\max} \bar{m}^{-2} - \bar{m}^{-1} \frac{dm_{\max}}{d\bar{m}} \right] \frac{\bar{m}}{m_{\max}} \mathcal{F}_1\left(\frac{m_{\max}}{\bar{m}}\right), \quad (1.63)$$

that is,

$$0 = \frac{m_{\max}}{\bar{m}} - \frac{dm_{\max}}{d\bar{m}} \quad (1.64)$$

giving

$$m_{\max} \propto \bar{m}. \quad (1.65)$$

Therefore the size of the largest organism of a species with average size  $\bar{m}$  will have the same scaling as the average. Consequently, Eq. (1.36) still holds with the exponent  $\xi$  therein being both the exponent of the scaling of the largest organism's mass and of the largest species' mass with the area.

In conclusion, introducing intra-specific variability in mass according to Eq. (1.56) does not alter the linking relationships (1.46).

### Power-law intra-specific size distributions

In tree communities, intra-specific size distributions are much wider than the distribution described by (1.56). An analysis of intra-specific tree size-distribution of Barro Colorado Island (BCI, Panama) and the Luquillo forest (Puerto Rico), presented in section 2.3 of next chapter,

<sup>5</sup> Because  $\lim_{x \rightarrow \infty} h(x) = 0$ ,  $\lim_{x \rightarrow 0} h(x) = h_0$ ,  $h(x) \leq h_0$  and  $\int_0^{\infty} x^{-\delta-\gamma} \mathcal{F}_1(1/x) dx < \infty$  if  $\delta + \gamma > 1$ , then  $\lim_{y \rightarrow \infty} \tilde{h}(y) = 0$  and  $\lim_{y \rightarrow 0} \tilde{h}(y) = \text{const}$  follows from the Lebesgue's dominated convergence theorem.

showed that such distributions are characterized by the following finite-size scaling form:

$$p(m|\bar{m}) = m^{-\Delta} \mathcal{F}_2\left(\frac{m}{\bar{m}^\Omega}\right), \quad (1.66)$$

where  $\Omega = 1/(2 - \Delta)$  ensures that  $\int m p(m|\bar{m}) dm = \bar{m}$  and  $\mathcal{F}_2(x)$  is a scaling function with limiting behaviors  $\mathcal{F}_2(x) \rightarrow \text{const}$  for  $x \rightarrow 0$  and  $\mathcal{F}_2(x) \rightarrow 0$  more rapidly than any power of  $x$  for  $x \rightarrow \infty$ . The exponent  $\Delta$ , fitted as explained in section 2.3, has value  $\Delta = 1.12 \pm 0.06$  for the BCI forest and  $\Delta = 1.14 \pm 0.03$  for the Luquillo forest. Note that, although the shape of Eq. (1.66) seems similar to Eq. (1.56), the different properties of the finite-size functions render the two distribution very different. Eq. (1.56), in fact, describes a distribution with an internal mode around which most of the individuals' sizes are found, since the tails of the distribution are thin. The sizes of a tree species described by Eq. (1.66) are still characterized by a finite mean, thanks to the cut-off function, but have no internal mode as they show a power-law behavior before the cut-off.

Equation (1.66) implies some small corrections to the relationships linking the scaling exponents, some of which can be computed exactly. Using Eq.(1.66) one has:

$$\begin{aligned} \langle m^k | \bar{m} \rangle &= \int_1^\infty m^{k-\Delta} \mathcal{F}_2\left(\frac{m}{\bar{m}^\Omega}\right) dm = \bar{m}^{\Omega(1+k-\Delta)} \int_{1/\bar{m}^\Omega}^\infty x^{k-\Delta} \mathcal{F}_2(x) dx \\ &= \bar{m}^{\Omega(1+k-\Delta)} \left[ c \int_{1/\bar{m}^\Omega}^1 x^{k-\Delta} dx + \int_1^\infty x^{k-\Delta} \mathcal{F}_2(x) dx \right] \\ &= \bar{m}^{\Omega(1+k-\Delta)} \left[ \tilde{c} \bar{m}^{-\Omega(1+k-\Delta)} + \int_1^\infty x^{k-\Delta} \mathcal{F}_2(x) dx \right] \\ &\propto \bar{m}^{\max\{0, \Omega(1+k-\Delta)\}}, \end{aligned} \quad (1.67)$$

where  $c, \tilde{c}$  are constants and we used the properties of  $\mathcal{F}_2(x)$  to ensure that the integral  $\int_1^\infty x^{k-\Delta} \mathcal{F}_2(x) dx$  converges and to substitute  $\mathcal{F}_2(x) \sim c$  when  $x < 1$ . The  $(j, k)$ th moment  $I_{j, k}$  is obtained by substituting (1.67) in (1.57):

$$\begin{aligned} I_{j, k} &= \int_0^\infty \int_1^\infty n^j \bar{m}^{\max\{0, \Omega(1+k-\Delta)\}} p(n, \bar{m}|A) d\bar{m} dn \\ &\propto A^{j\Phi + \max\{0, \lambda(1-\delta - j\gamma + \max\{0, \Omega(1+k-\Delta)\})\}}, \end{aligned} \quad (1.68)$$

where the calculations have been performed similarly to the ones in section 1.4.1. Equation (1.68) differs only slightly from (1.26), the corresponding equation in the case without an intraspecific size distribution. As a result, a small correction applies to the linking relationship (1.46b): the term  $\alpha$  in equation (1.46b) is substituted by  $\frac{1+\alpha-\Delta}{2-\Delta}$ , which for the empirical value  $\Delta = 1.12 \pm 0.06$  found for BCI results in a very small correction  $C = \alpha - \frac{1+\alpha-\Delta}{2-\Delta} = (\alpha - 1) \left(\frac{1-\Delta}{2-\Delta}\right) \simeq 0$  which is compatible with zero, given that  $\Delta$  is compatible with one (within two standard deviations). Equations (1.46c, 1.46d and 1.46e) are unchanged.

The size-spectrum cannot be computed exactly. However, it is possible to derive an approximation that holds for large  $m$  (see Appendix 1.8.2), which is a power-law with exponent:

$$\tilde{\eta} = \Delta - (2 - \Delta)(1 - \delta - \gamma), \quad (1.69)$$

The correction to the linking relationship (1.46a) is small because the difference between  $\eta = \delta + \gamma$  computed via (1.46a) and  $\tilde{\eta}$  computed via (1.69) is  $\eta - \tilde{\eta} = (\Delta - 1)(\delta + \gamma - 2)$  which, again, is compatible with zero within two standard deviations. Using the values of  $\delta$  and  $\gamma$  estimated in for the Barro Colorado Island forest in Chapter 2 (see Table 2.3), we find  $\eta - \tilde{\eta} \simeq -0.07 \pm 0.03$ .

### 1.6.2 Kleiber's law with curvature

The constraint on the total resource consumption of the community has been imposed assuming that an organism of body mass  $m$  has a metabolic rate determined by a power of its body mass,  $b \propto m^\alpha$ . Nevertheless, empirical evidence exists supporting that Kleiber's law may differ from a simple power-law, as it purportedly displays curvatures in a metabolic rate-mass log-log plot [Dodds et al., 2001, Kolokotronis et al., 2010, Mori et al., 2010, Marañón et al., 2013, Banavar et al., 2014, Ballesteros et al., 2018]. Mori et al. [2010], for example, found that trees' respiration rates scale linearly with  $m$  until  $m_1 \simeq 40$  g and as  $m^{3/4}$  above, which might be explained by the change in the ratio of photosynthetically active biomass to total biomass from saplings to adult trees. For mammals, respiration rates have been claimed [Dodds et al., 2001] to scale as  $m^{2/3}$  (reflecting surface-to-volume scaling for constant organismic body density) until  $m \simeq 400$  g and as  $m^{3/4}$  above and in general a variety of  $\alpha$  values have been observed, ranging from  $2/3$  to  $1$ . A proposed explanation [Ballesteros et al., 2018] is that the energy consumption of an organism is the sum of the energy dispersed as heat, proportional to  $m^{2/3}$ , and the energy necessary to sustain vital processes, proportional to  $m$ . The resulting energy consumption would be the sum of two power-laws with different exponents, resulting in an apparent exponent in the interval  $(2/3, 1)$ , with variable value depending on the observed range of  $m$ , as one or the other term might in turn dominate. The implications of a departure of Kleiber's law from a pure power-law on the other macroecological patterns can be inferred by plugging in a modified Kleiber's law in the resource constraint. As suggested in [Ballesteros et al., 2018], let us consider the sum of two powers of  $m$ :

$$b(m) = c_0 m^{\alpha_0} + c_1 m^{\alpha_1} \quad (1.70)$$

where  $c_0, c_1$  are positive constants and  $0 < \alpha_0 < \alpha_1$ . Eq. (1.70) describes a convex curvature in Kleiber's law. Using Eq. (1.70) one can compute the total community consumption rate per

species:

$$\begin{aligned}
 B/S &= \int_1^\infty \int_0^\infty n b(m) p(n, m|A) dn dm \\
 &= (\delta - 1) \int_1^\infty \int_0^\infty b(m) m^{-\delta} G \left[ \frac{nm^\gamma}{A^\Phi} \frac{1}{h(m/A^\lambda)} \right] dn dm \\
 &\propto A^\Phi \int_1^\infty b(m) m^{-\gamma-\delta} h \left( \frac{m}{A^\lambda} \right) dm \int_0^\infty G(x) dx \\
 &\propto A^\Phi \int_1^\infty c_0 m^{\alpha_0-\gamma-\delta} h \left( \frac{m}{A^\lambda} \right) dm + A^\Phi \int_1^\infty c_1 m^{\alpha_1-\gamma-\delta} h \left( \frac{m}{A^\lambda} \right) dm \\
 &= \tilde{c}_0 A^{\Phi+\max\{0, \lambda(1+\alpha_0-\gamma-\delta)\}} + \tilde{c}_1 A^{\Phi+\max\{0, \lambda(1+\alpha_1-\gamma-\delta)\}},
 \end{aligned} \tag{1.71}$$

where  $\tilde{c}_0$  and  $\tilde{c}_1$  are constants and the properties of  $h$  have been used to carry out the integrations. Therefore, the specific community consumption rate  $B/S$  scales as the sum of two powers of  $A$ . Imposing the resource constraint  $B \propto A$ , we find that the SAR has the following behavior:

$$S \propto \frac{A^{1-\Phi}}{\tilde{c}_0 A^{\max\{0, \lambda(1+\alpha_0-\delta-\gamma)\}} + \tilde{c}_1 A^{\max\{0, \lambda(1+\alpha_1-\delta-\gamma)\}}} \tag{1.72}$$

Depending on the value of  $\alpha_0$  and  $\alpha_1$ , therefore, the situation differs. If  $1 + \alpha_0 - \delta - \gamma, 1 + \alpha_1 - \delta - \gamma > 0$

$$S \propto \frac{A^{1-\Phi-\lambda(1+\alpha_1-\delta-\gamma)}}{\tilde{c}_0 A^{\lambda(\alpha_0-\alpha_1)} + \tilde{c}_1}. \tag{1.73}$$

Recalling that  $\alpha_1 > \alpha_0$ , in the large  $A$  limit  $\tilde{c}_1 \gg \tilde{c}_0 A^{\lambda(\alpha_0-\alpha_1)}$  therefore asymptotically  $z = 1 - \Phi - \lambda(1 + \alpha_1 - \delta - \gamma)$ . For smaller areas, where  $\tilde{c}_1 < \tilde{c}_0 A^{\lambda(\alpha_0-\alpha_1)}$ , the scaling exponent would be  $\tilde{z} = 1 - \Phi - \lambda(1 + \alpha_0 - \delta - \gamma) > z$ , therefore the SAR would appear concave. The initial slope differs from the asymptotic slope by  $\Delta z = \lambda(\alpha_1 - \alpha_0)$ . If  $1 + \alpha_1 - \delta - \gamma > 0$  but  $1 + \alpha_0 - \delta - \gamma < 0$ , the situation is similar but the exponents differ by  $\Delta z = \lambda(1 + \alpha_1 - \delta - \gamma)$ . In both these cases the SAR is not a pure power-law but for large  $A$  the smaller exponent dominates, therefore one has, asymptotically,

$$z = 1 - \Phi - \max\{0, \lambda(1 + \alpha_0 - \gamma - \delta)\}. \tag{1.74}$$

For intermediate values of  $A$  the slope will appear steeper, with an exponent intermediate between the two. The value of  $A$  at which Eq. (1.74) begins to be valid depends on the values of  $c_0, c_1, \alpha_0, \alpha_1$ . Therefore, under these conditions a convex curvature in the relationship between individual metabolic rates and body mass causes a concave curvature in the scaling of  $S$  with  $A$ .

If, instead,  $1 + \alpha_0 - \delta - \gamma, 1 + \alpha_1 - \delta - \gamma < 0$ , one simply has  $S \propto A^{1-\Phi}$ , therefore:

$$z = 1 - \Phi. \tag{1.75}$$



In this case, the curvature in Kleiber's law has no effect on the SAR. In conclusion, a convex curvature in Kleiber's law might or might not imply a concave curvature in the SAR, depending on the values of the scaling exponents.

### 1.6.3 Alternative hypotheses on $p(m|A)$

In section 1.3 we considered  $p(m|A)$  to be a pure power function, with only the cut-off on abundances (Damuth's law) to hinder the presence of organisms larger than sustainable sizes. In this section, we will analyze possible relaxations of this hypothesis which do not compromise analytical tractability.

#### **A-dependent cut-off**

Here we consider the presence of an  $A$ -dependent cut-off at large sizes, similar to the one imposed in Damuth's law:

$$p(m|A) = m^{-\delta} H_1 \left( \frac{m}{A^\lambda} \right) \quad (1.76)$$

with  $\delta > 1$  (for normalization purposes), where the exponent  $\lambda$  is the same as the one of the cut-off in Damuth's law (1.18) and where  $H_1(x)$  is such that  $\lim_{x \rightarrow 0} H_1(x) = \text{const}$  and  $\lim_{x \rightarrow \infty} H_1(x) = 0$ . The presence of this additional cut-off does not prevent the computation of moments and of the community size spectrum. Both calculations are performed as in section 1.4, the only difference consisting in the product  $H_1 \left( \frac{m}{A^\lambda} \right) h^j \left( \frac{m}{A^\lambda} \right)$ ,  $j = 1, 2$ , playing the role of  $h^j \left( \frac{m}{A^\lambda} \right)$ . The linking relationships (1.2), (1.32a–1.32c) remain therefore valid. The cut-off in the community size-spectrum is now given by the product of the two functions  $H_1$  and  $h$ . However, it is not possible to compute analytically the size of the largest species, therefore we cannot provide the equivalent of Eq. (1.36) in this scenario.

#### **A-independent cut-off**

Eq. (1.35) predicts that, if  $z > 0$ , the maximum species' mass increases with  $A$  as  $m_{\max} = A^\xi$ . When  $A$  is very large, area-independent constraints could settle in to limit the maximum body size, either due to physiological limits or due to competitive ecological dynamics making larger body sizes unfavorable. Because  $\xi = z/(\delta - 1)$ , the critical value of  $A$  above which the maximum body size is independent of  $A$  is equal to  $A_c = M_0^{(\delta-1)/z}$ , with  $M_0$  the maximum size allowed independent of  $A$ . This observation can be reconciled with our framework by generalizing Eq. (1.17) as:

$$p(m|A) = m^{-\delta} H_2 \left( \frac{m}{M_0} \right), \quad (1.77)$$

## Chapter 1. Covariations in ecological scaling laws

---

where the cutoff function  $H_2$  is such that  $H_2(x) \simeq \text{const}$  when  $x \ll 1$  and  $\lim_{x \rightarrow \infty} H_2(x)/x^{\max\{2-\delta-\gamma, 1-\delta\}} = 0$ . This generalization of Eq. (1.17) does not affect our results for  $A < A_c$ . In fact, the joint probability distribution in such a generalized setting reads:

$$p(n, m|A) = n^{-1} m^{-\delta} H_2\left(\frac{m}{M_0}\right) G\left[\frac{nm^\gamma}{A^\Phi h(m/A^\lambda)}\right] \quad (1.78)$$

and integrals of this distribution (e.g. marginals and moments) depend on which of the two finite-size cutoffs ( $h$  and  $H_2$ ) sets in at the lowest value of  $m$ . We show this by calculating the moment  $I_{j,k}$  with  $j, k > 0$ :

$$\begin{aligned} I_{j,k} &= \int_1^\infty \int_0^\infty n^{j-1} m^{k-\delta} G\left[\frac{nm^\gamma}{A^\Phi h(m/A^\lambda)}\right] H_2\left(\frac{m}{M_0}\right) dn dm \\ &= A^{j\Phi} \int_1^\infty m^{k-\delta-j\gamma} h^j\left(\frac{m}{A^\lambda}\right) H_2\left(\frac{m}{M_0}\right) dm \int_0^\infty x^{j-1} G(x) dx, \end{aligned} \quad (1.79)$$

where the properties of  $G$  ensure the convergence of the integral in  $x$ . The scaling of  $I_{j,k}$  with  $A$  is thus determined by which of the two functions  $h^j(m/A^\lambda)$  and  $H_2(m/M_0)$  decays earlier in  $m$ . For  $A < M_0^{1/\lambda}$ , the function  $h$  decays before  $H_2$ , and therefore nothing changes with respect to the case in section 1.4. For  $A > M_0^{1/\lambda}$ , the function  $H_2$  decays before  $h$ , therefore  $h$  can be neglected, and we obtain:

$$I_{j,k} \propto A^{j\Phi} \int_1^\infty m^{k-\delta-j\gamma} H_2\left(\frac{m}{M_0}\right) dm \propto A^{j\Phi}, \quad (1.80)$$

where the integral in  $m$  converges owing to the properties of  $H_2$ . Therefore, for large areas the linking relationships (1.46b–1.46e) are substituted by

$$z = 1 - \Phi \quad (1.81a)$$

$$\nu = 1 \quad (1.81b)$$

$$\mu = 1 \quad (1.81c)$$

$$\xi = 0. \quad (1.81d)$$

This could be the situation for the two forest datasets analyzed in section 2.3, where one should have  $\lambda = 0$  to verify the reasonable assumption  $\mu = \nu = 1$  (see section 2.3 for details). Eq. (1.46a), instead, remains unchanged although the cutoff in the size spectrum is now given by the function  $H_2\left(\frac{m}{M_0}\right)$ . Furthermore, the cutoff in Damuth's law would not be observable due to the extremely low probability to observe a species with  $m > M_0$ . In conclusion, when maximum body size is limited by a constraint independent from the ecosystem area, the scaling exponents related to area ( $z$ ,  $\nu$  and  $\mu$ ) have a crossover at the area where the constraint is attained,  $A_c$ . Note, however, that if  $\eta \geq 2$  Eqs (1.46b–1.46d) reduce to Eqs (1.81a–1.81c), therefore there is no crossover.

**Internal mode**

Some empirical studies [Siemann et al., 1996, Labra et al., 2015] found a  $p(m|A)$  showing an internal mode, well-fitted by a log-normal. This case can be described as

$$p(m|A) = m^{-\delta} H_3\left(\frac{m}{A^\lambda}\right) \quad (1.82)$$

where  $H_3(x)$  is such that  $\lim_{x \rightarrow 0, \infty} x^j H_3(x) = 0 \forall j$ . Normalization requires  $\delta = 1$ .  $p(m|A)$  described by Eq. (1.82) has therefore a peak around a size  $m_p \propto A^\lambda$ , where  $\lambda$  is taken to be the same exponent appearing in Damuth's law cut-off for ease of calculations. Note that this does not mean that  $p(m|A)$  peaks at the same size at which Damuth's law cut-off starts to be effective, but only that the two points scale with the same power of  $A$ . The scaling of the  $(j, k)$ th-moment ( $j = 1, 2, k \in [0, 1]$ ) with  $A$ , for large  $A$ , can be computed as:

$$\begin{aligned} I_{j,k} &= \int_1^\infty \int_0^\infty n^j m^k p(n, m|A) dn dm \\ &= \int_1^\infty m^{k-1} H_3\left(\frac{m}{A^\lambda}\right) \int_0^\infty n^{j-1} G\left[\frac{nm^\gamma}{A^\Phi} \frac{1}{h(m/A^\lambda)}\right] dn dm \\ &= A^{j\Phi} \int_1^\infty m^{k-1-j\gamma} H_3\left(\frac{m}{A^\lambda}\right) h^j\left(\frac{m}{A^\lambda}\right) dm \int_0^\infty x^{j-1} G(x) dx \\ &\propto A^{j\Phi} \int_1^\infty m^{k-1-j\gamma} H_3\left(\frac{m}{A^\lambda}\right) h^j\left(\frac{m}{A^\lambda}\right) dm \\ &\propto A^{j\Phi+\lambda(k-j\gamma)} \int_{1/A^\lambda}^\infty y^{k-1-j\gamma} H_3(y) h^j(y) dy \propto A^{j\Phi+\lambda(k-j\gamma)}, \end{aligned} \quad (1.83)$$

where  $x = \frac{nm^\gamma}{A^\Phi} \frac{1}{h(m/A^\lambda)}$  and  $y = \frac{m}{A^\lambda}$ . We used the property (1.21) to ensure that the integral  $\int_0^\infty x^{j-1} G(x) dx$  converges for  $j = 1, 2$  and, in the last passage, the properties of  $H_3$  to note that the integral depends only weakly on the lower extreme  $1/A^\lambda$ . Therefore, using  $I_{j,k} \propto A^{j\Phi+\lambda(k-j\gamma)}$  to obtain the scalings of  $S$ ,  $N$  and  $M$  as was done in section 1.4.1, the linking relationships (1.46b–1.46d) are replaced by:

$$z = 1 - \Phi - \lambda(\alpha - \gamma) \quad (1.84a)$$

$$v = 1 - \lambda\alpha \quad (1.84b)$$

$$\mu = 1 + \lambda(1 - \alpha) \quad (1.84c)$$

The community size-spectrum is computed as:

$$\begin{aligned} s(m|A) &= \frac{S}{N} \int_0^\infty np(n, m|A) dn = \frac{S}{N} m^{-1} H_3\left(\frac{m}{A^\lambda}\right) \int_0^\infty dn G\left(\frac{nm^\gamma}{A^\Phi h(m/A^\lambda)}\right) \\ &\propto m^{-1} \left(\frac{A^\lambda}{m}\right)^\gamma H_3\left(\frac{m}{A^\lambda}\right) h\left(\frac{m}{A^\lambda}\right) = m^{-1} \mathcal{H}\left(\frac{m}{A^\lambda}\right) \end{aligned} \quad (1.85)$$

where  $\mathcal{H}(x) := x^{-\gamma} H_3(x) h(x)$  is such that  $\lim_{x \rightarrow 0, \infty} x^j \mathcal{H}(x) = 0 \forall j$ . Therefore,  $s(m|A)$  is not a power-law, but has an internal mode, similarly to  $p(m|A)$ . The generalization of Eq. (1.46e) in

this scenario cannot be computed analytically.

#### 1.6.4 Multiple trophic levels

The metabolic constraint imposed,  $B \propto A$ , refers to a community competing for the same resources, i.e. a single trophic level. The results obtained apply, therefore, to primary producers, e.g. forests or phytoplanktonic communities, or to consumers competing for the same resource like herbivorous mammals or predators sharing the same preys.

Consider now a simple foodweb made of two trophic levels, with one level feeding on abiotic factors like light (i.e. the producers) and the other feeding on the lower trophic level (i.e. the consumers). In such an ecosystem, producers would be described by the framework in section 1.3 where resources are limited by ecosystem area, that is,  $\mathcal{R} \propto A$ , while consumers would be described by the same framework where the limiting resource is now the total producers' biomass, which scales as  $A^{\mu_p}$  (the subscript  $p$  identifies the producers). For the producers' trophic level, Eqs (1.46) are therefore valid. At the consumers' level, instead, the different resource scaling affects the SAR exponent directly, as:

$$z_c = \mu_p - \Phi_c - \max\{0, \lambda_c(1 + \alpha_c - \eta_c)\}, \quad (1.86)$$

which differs from (1.46b) only for the exchange of 1 with  $\mu_p$  (the subscript  $c$  identifies the consumers). Such modification propagates to equations (1.46c) and (1.46d) (recall that  $N = SI_{1,0}$  and  $M = SI_{1,1}$ ) which become:

$$\begin{aligned} \mu_c &= \mu_p + \max\{0, \lambda_c(2 - \eta_c)\} - \max\{0, \lambda_c(1 + \alpha_c - \eta_c)\}, \\ \nu_c &= \mu_p - \max\{0, \lambda_c(1 + \alpha_c - \eta_c)\}. \end{aligned} \quad (1.87)$$

Note, however, that  $\mu_p$  is possibly equal to one in many ecosystems. In such a case, the consumers level is described exactly by the framework in section 1.3. Analogous generalizations to multiple trophic levels are possible.

In summary, in a multi-trophic level system, the linking relationships Eqs (1.46) are valid within single trophic levels. Power law distributions and relationships, however, are usually observed empirically even across trophic levels. For example, power-law size spectra are routinely observed in marine microbial ecosystems over several orders of magnitude in organismic size [Sheldon et al., 1972] comprising both primary producers and consumers (and measurable regardless of species below a threshold mass, say, via flow cytometry e.g.[Cavender-Bares et al., 2001]). If the scaling exponents exhibit the same value (within accuracy) in all trophic levels, global scaling laws will have the same exponents as the single-level ones, therefore the relationships Eqs (1.46) are valid globally. On the other hand, if the exponents differ across trophic levels the situation becomes more difficult to sort out. Suppose, for example, that in the simple two-level system described above the numbers of species in the two levels scale respectively as  $S_p = c_p A^{z_p}$  and  $S_c = c_c A^{z_c}$ , where the subscripts  $p$  and  $c$  refer to producers and

consumers. The total number of species  $S = S_p + S_c$  is not a pure power-law, although when  $z_p - z_c$  is small it will look very similar to one. If fitted to a power-law, the apparent exponent will be comprised between  $z_p$  and  $z_c$ , with the exact value determined by the weights  $c_p$  and  $c_c$ . The other scaling laws behave similarly, but unfortunately it is not possible to find simple relationships linking them.

## 1.7 Discussion

The theoretical framework proposed here produces model-free testable predictions on the relationships among ecological laws, rationalizing the observed variability of ecological exponents across ecosystems. Jointly with empirical evidence (which will be presented in Chapter 2), the framework supports the tenet that scaling exponents may vary across ecosystems but must satisfy consistency relationships that result in exact covariations of ecological patterns. One of the main conclusions of this theoretical work is that, when applying scaling laws, for example in conservation, care should be exerted not to combine exponents measured in different settings, which may not satisfy the relationships (1.46) leading to incorrect predictions for unmeasured patterns.

The investigation of dynamic birth, death and speciation models presented in section 1.5 corroborates the generality of the scaling framework and the predicted pattern covariations. In fact, all the models investigated that are compatible with the empirically observed macroecological patterns are all characterized by the same scaling properties of  $p(n, m|A)$ , which are encapsulated in the scaling framework and univocally specify the pattern covariations in Eqs. 1.46. Therefore, Eqs. (1.17), (1.18) and (1.20) do not rely on specific assumptions about the population and speciation dynamics of a community, but rather specify the universal scaling properties that possibly any dynamic model compatible with the empirically observed macroecological laws must satisfy. Furthermore, the pattern covariations predicted by the scaling framework agree with broad empirical evidence (see Chapter 2) and with heuristic arguments, like for example the back-of-the-envelope calculations reported in section 1.1. Thus, one can speculate that such scaling framework describes not only the community dynamics models studied in section 1.5, but more generally any ecosystem subject to the constraint of finite resource supply rate. The basic model is thus arguably the simplest of a class of models that share the same scaling properties of the fundamental distribution, which in turn imply the same covariations of ecological patterns. This is akin to the concept of universality class [Stanley, 1999, Stanley and Amaral, 2000], applied to the scaling form rather than to the exponents of the joint probability distribution and of ecological scaling laws.

It must be understood, however, that the scaling framework does not predict the values of scaling exponents, but rather their covariations. The various community dynamic models studied here, instead, do predict scaling exponents values, which emerge from the rates and assumptions concerning birth, death and speciation events. However, there may exist several dynamic models capable of reproducing quantitatively one specific set of exponents'

values, as it is often the case that several processes lead to the same pattern [Newman, 2007]. Furthermore, the dynamic models presented here cannot be claimed to describe any real ecosystem in all its complexity, as such models are of course overly simplified to encompass the broad range of processes that may set the values of macroecological scaling exponents (e.g. species' interactions, landscape structure and ecological disturbances, to name a few). When modelling any natural process, the first step is that of abstraction: unnecessary details are removed, until one reaches the simplest model that is still compatible with the observed patterns, which for the purpose of this investigation are the functional forms of the various scaling relationships and distributions. In this sense, the studied models of birth, death and speciation arguably capture the essential ingredients that produce the empirically-observed functional forms of macroecological laws and that set the scaling properties of the joint distribution of mass and abundance, and thus the pattern covariations. The exact values of the macroecological scaling exponents, instead, are most likely determined by several processes that are not included in the community dynamics models considered here, but can be properly described by the scaling framework.

The proposed framework adopts the minimum set of hypotheses allowing to reproduce widespread macroecological patterns found in empirical data, without compromising analytical tractability. Such analytical tractability is important in this context because it highlights the relationships among macroecological patterns in simple terms, i.e. via algebraic relationships among their scaling exponents. In section 1.6, generalization to the framework were presented which prove its flexibility in assimilating novel empirical ecological evidence. Specifically, the framework may account for two different intra-specific size distributions, one which applies to marine microorganisms and one which applies to tropical tree species, and for curvatures affecting Kleiber's law [Kolokotronis et al., 2010, Mori et al., 2010, Marañón et al., 2013, Banavar et al., 2014]. These generalizations imply small (or no) corrections to the linking relationships obtained in section 1.4, suggesting that macroecological laws are quite robust with respect to change in ecosystem details. In particular, there are no corrections to the linking relationships highlighted in 1.4 if the intra-specific body size distributions has an internal mode (e.g., as found for unicellular protists [Giometto et al., 2013]). Conversely, small deviations to such linking relationships arise if the intra-specific body size distribution is a power law, as shown for tree species. Additionally, a concave curvature in the SAR might stem from a convex curvature in Kleiber's law, depending on the value of the scaling exponents. The framework proved also capable of including a cut-off in the species' size distribution  $p(m|A)$  or its possible non-scaling form. The effects of these assumptions on the other scaling laws were derived, showing in particular how an internal mode in  $p(m|A)$  translates into a corresponding internal mode in the community size-spectrum. It is argued that such flexibility makes our framework widely adaptable to describe diverse empirical settings and to find the implications of particular ecological assumptions. In the most general scenario in which the dependence of  $p(m|A)$  on  $m$  and  $\langle n|m, A \rangle$  on  $m$  and  $A$  cannot be expressed as in (1.17) and (1.18) (which, however, are compatible with several empirical case studies) nor described by the generalizations treated in section 1.6, one would have to rely on numerical methods to

derive the covariations between macroecological patterns, following the same route adopted in our theoretical investigation. Generalizations of Eqs. (1.46) would hold in this scenario, although they would be expressed as integral equations in terms of the probability distributions introduced above.

## 1.8 Appendix

### 1.8.1 Relative species abundance (RSA)

The RSA is the distribution of species abundances  $p(n|A)$ . In our framework it can be obtained by marginalizing  $p(n, m|A)$  over  $m$ :

$$\begin{aligned} p(n|A) &= \int_1^\infty p(n, m|A) dm \\ &= (\delta - 1)n^{-1} \int_1^\infty m^{-\delta} G\left(\frac{n}{\langle n|m, A \rangle}\right) dm. \end{aligned} \quad (1.88)$$

This integral cannot be computed in the general case, that is, without specifying  $G$  and  $h$ . We compute it here for the particular choice  $G(x) = \frac{1}{\sqrt{\pi\sigma}} e^{-\frac{1}{\sigma}(\log x + \sigma/4)^2}$ , with  $\sigma > 0$  constant. Note that  $\int_0^\infty G(x) dx = \int_0^\infty G(x)/x dx = 1$  as prescribed in sections 1.3 and 1.4. For  $h(x)$ , it is sufficient to know that  $h(x)$  is monotonically decreasing to carry out the calculations, but to simplify the expressions we take  $h(x) = h_0 e^{-x}$  with  $h_0 > 0$  constant. With these assumptions,  $p(n|A)$  reads:

$$p(n|A) = \frac{1}{n} \frac{\delta - 1}{\sqrt{\pi\sigma}} \int_1^\infty m^{-\delta} e^{f(m)} dm, \quad (1.89)$$

where we defined

$$f(m) = -\frac{1}{\sigma} \left[ \log\left(\frac{nm^\gamma}{h_0 A^\Phi}\right) + \frac{m}{A^\lambda} + \frac{\sigma}{4} \right]^2. \quad (1.90)$$

The integral in (1.89) cannot be computed analytically. However, noticing that the contribution to the integral is maximum when  $m = m^*$  where  $m^*$  maximizes  $f(m)$ , we can approximate the integral for certain values of  $n$ . The approximation is akin to the Laplace method, but it is not possible to give an upper bound on the error made by the approximation. Nonetheless, the approximation can be compared to the numerical computation of  $p(n|A)$  (see Fig. 1.2). The derivative of  $f(m)$  reads:

$$f'(m) = -\frac{2}{\sigma} \left[ \log\left(\frac{nm^\gamma}{h_0 A^\Phi}\right) + \frac{m}{A^\lambda} + \frac{\sigma}{4} \right] \left( \frac{\gamma}{m} + \frac{1}{A^\lambda} \right). \quad (1.91)$$

Note that the derivative  $f'(m)$  is negative for any  $m \in [1, \infty]$  if  $n > A^\Phi h_0 e^{-1/A^\lambda} e^{-\sigma/4} \simeq A^\Phi h_0 e^{-\sigma/4}$ . Thus, for  $n \gg A^\Phi h_0 e^{-\sigma/4}$  (i.e. in the tail of the distribution)  $f(m)$  is maximum at



$m_{\text{tail}}^* = 1$  and the approximation gives:

$$\begin{aligned}
 p(n|A)_{\text{tail}} &= \frac{1}{n} \frac{\delta - 1}{\sqrt{\pi\sigma}} \frac{e^{-\frac{1}{\sigma} \left[ \log\left(\frac{n}{h_0 A^\Phi}\right) + \frac{1}{A^\lambda} + \frac{\sigma}{4} \right]^2}}{\frac{2}{\sigma} \left[ \log\left(\frac{n}{h_0 A^\Phi}\right) + \frac{1}{A^\lambda} + \frac{\sigma}{4} \right] \left( \gamma + \frac{1}{A^\lambda} \right)} \\
 &\simeq \frac{1}{n} \frac{\delta - 1}{\sqrt{\pi\sigma}} \frac{e^{-\frac{1}{\sigma} \left[ \log\left(\frac{n}{h_0 A^\Phi}\right) + \frac{\sigma}{4} \right]^2}}{\frac{2\gamma}{\sigma} \left[ \log\left(\frac{n}{h_0 A^\Phi}\right) + \frac{\sigma}{4} \right]}.
 \end{aligned} \tag{1.92}$$

Note that the tail of the RSA resembles that of a lognormal distribution, which is typically found empirically [Azaele et al., 2016], plus a correction of the form  $C_1 + C_2 \log n$  at the denominator, where  $C_1$  depends on  $A$ .

If  $n < A^\Phi h_0 e^{-1/A^\lambda} e^{-\sigma/4} \simeq A^\Phi h_0 e^{-\sigma/4}$ , the maximum of  $f(m)$  occurs at a value  $\hat{m} > 1$ . However, one cannot solve  $f'(m) = 0$  analytically to determine  $\hat{m}$ . We can approximate the RSA at small and intermediate values of  $n$  as follows. The behavior of  $p(n|A)$  can be characterized for  $n \simeq A^\Phi h_0 e^{-\sigma/4}$  by recognizing that, at such values of  $n$ , the value  $m_{\text{body}}^*$  maximizing  $f(m)$  is close to 1. Therefore, we approximate  $\log\left(\frac{nm^\gamma}{h_0 A^\Phi} e^{m/A^\lambda}\right) \simeq \log\left(\frac{nm^\gamma}{h_0 A^\Phi} e^{1/A^\lambda}\right)$  in  $f'(m) = 0$  (see Eq. 1.91) and solve for  $m$ , yielding  $m_{\text{body}}^* \simeq \left[ \frac{A^\Phi h_0}{n} e^{-1/A^\lambda} e^{-\sigma/4} \right]^{\frac{1}{\gamma}}$ . By applying the approximation method, one finds that the approximation for the RSA for  $n \simeq A^\Phi h_0 e^{-\sigma/4}$  is:

$$p(n|A)_{\text{body}} = \frac{\delta - 1}{\gamma} \left[ A^\Phi h_0 e^{-\frac{\sigma}{4}} \right]^{\frac{1-\delta}{\gamma}} n^{-1 - \frac{1-\delta}{\gamma}}, \tag{1.93}$$

which is a power law with exponent  $-1 - \frac{1-\delta}{\gamma}$ . The extents of the tail and body of the RSA distribution depend on the values of  $h_0$  and  $\sigma$ .

### 1.8.2 Community size-spectrum with intra-specific size-distribution

In the case of section 1.6.1, the size spectrum is given by:

$$\begin{aligned}
 s(m|A) &= \frac{S}{N} \int_0^\infty n \int_1^\infty p(n, \bar{m}|A) m^{-\Delta} \mathcal{F}\left(\frac{m}{\bar{m}^\Omega}\right) d\bar{m} dn \\
 &= \frac{S}{N} \frac{1}{m^\Delta} \int_1^\infty \frac{1}{\bar{m}^\delta} \mathcal{F}\left(\frac{m}{\bar{m}^\Omega}\right) \int_0^\infty G\left[\frac{n\bar{m}^\gamma}{A^\Phi h(\bar{m}/A^\lambda)}\right] dn d\bar{m} \\
 &= \frac{S}{N} m^{-\Delta + (1-\delta)/\Omega} \int_0^m x^{(\delta-1)/\Omega - 1} \mathcal{F}(x) \\
 &\quad \times \int_0^\infty G\left[\frac{n(m/x)^{\gamma/\Omega}}{A^\Phi h[(m/x)^{1/\Omega}/A^\lambda]}\right] dn dx \\
 &= \frac{S}{N} A^\Phi m^{-\Delta + (1-\delta-\gamma)(2-\Delta)} \int_0^\infty G(y) dy \\
 &\quad \times \int_0^m x^{(2-\Delta)(\delta+\gamma-1)-1} \mathcal{F}(x) h\left[\frac{(m/x)^{1/\Omega}}{A^\lambda}\right] dx
 \end{aligned} \tag{1.94}$$

## Chapter 1. Covariations in ecological scaling laws

---

where  $x = m/\bar{m}^\Omega$ ,  $\Omega = 1/(2 - \Delta)$ ,  
 $y = [n(m/x)^{1/\Omega}] / \{A^\Phi h[(m/x)^{1/\Omega}/A^\lambda]\}$  and  $\int_0^\infty G(y)dy$  is a constant. Note that we cannot compute the scaling exponent of  $s(m|A)$  with  $m$  exactly, because  $x$  and  $m$  are found in the arguments of both  $\mathcal{F}$  and  $h$ . We can, however, derive an approximation that holds for large  $m$ . This is most easily seen if we assume that no finite-size effect is found in Damuth's law, i.e.  $h = \text{const}$ , for which we have:

$$s(m|A) \propto \frac{S}{N} A^\Phi m^{(2-\Delta)(1-\delta-\gamma)-\Delta} \int_0^m x^{-(2-\Delta)(1-\delta-\gamma)-1} \mathcal{F}(x) dx. \quad (1.95)$$

When  $m \gg 1$ , the fast decay of  $\mathcal{F}$  makes the integral depend only weakly on  $m$ . Therefore  $s(m|A)$  is, to a good approximation, a power-law with exponent:

$$\eta = \Delta - (2 - \Delta)(1 - \delta - \gamma), \quad (1.96)$$

which generalizes Eq. (1.2) to the case of intra-specific size distributions as in (2.14). When  $h$  is not constant, going back to Eq. (1.94), for  $m \gg 1$  we make a small mistake if we extend the upper limit of the integral in  $x$  to  $\infty$ . In analogy to the calculations performed in section 1.6.1, we find  $s(m) \propto m^{(2-\Delta)(1-\delta-\gamma)-\Delta} \tilde{h}(m/A^{\lambda\Omega})$ , where the function  $\tilde{h}(y) = \int_0^\infty x^{-(2-\Delta)(1-\delta-\gamma)} \mathcal{F}(x) h[(y/x)^{1/\Omega}] dx$  has the same limiting behavior of  $h(y)$  for  $y \rightarrow 0$  and  $y \rightarrow \infty$ <sup>6</sup>. Therefore, the linking relationship in (1.96) holds also when  $h$  is not identically constant.

---

<sup>6</sup> Because  $\lim_{x \rightarrow \infty} h(x) = 0$ ,  $\lim_{x \rightarrow 0} h(x) = h_0$ ,  $h(x) \leq h_0$  and  $\int_0^\infty x^{-\delta-\gamma} F(1/x) dx < \infty$  if  $\delta + \gamma > 1$ , then  $\lim_{y \rightarrow \infty} \tilde{h}(y) = 0$  and  $\lim_{y \rightarrow 0} \tilde{h}(y) = \text{const}$  follows from the Lebesgue's dominated convergence theorem.

## 1.8.3 Supplementary figures and tables

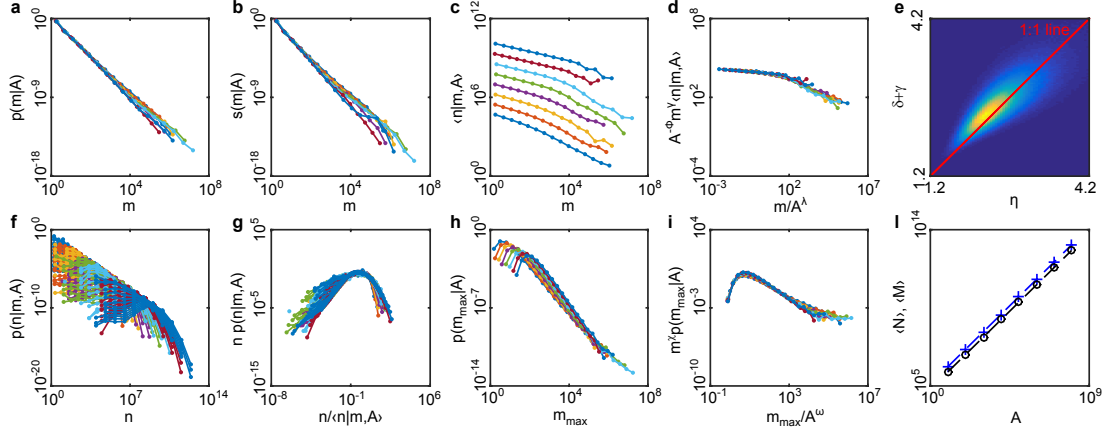


Figure 1.6 – Basic model statistics with parameters  $z = 1/4$ ,  $w = 10^{-3}$ ,  $\alpha = 3/4$ ,  $\theta = 1/4$  (simulation results shown in Fig. 1.5). Different colors refer to different values of  $A = 10^i$ , from  $i = 1$  (lower blue curve in panel c) to  $i = 8$  (upper blue curve in panel c). Panels a-c, f and h show respectively  $p(m|A)$ ,  $s(m|A)$ ,  $\langle n|m, A \rangle$ ,  $p(n|m, A)$  and  $p(m_{\max}|A)$  estimated at stationarity. Panels d, g and i show collapses of simulation data for  $\langle n|m, A \rangle$ ,  $p(n|m, A)$  and  $p(m_{\max}|A)$ , respectively. Panel e shows the density scatter-plot of  $\delta + \gamma$  versus  $\eta$ . The density histogram is normalized to one, with blue representing the value zero and yellow the value one. Shown are simulation data for the largest area  $A = 10^8$ . Panel i shows the scaling of the average total biomass  $\langle M \rangle$  (blue crosses and dashed lines) and average total abundance  $\langle N \rangle$  (black dots and dashed lines) with  $A$ . See Table 1.1 for estimates of exponents' values.

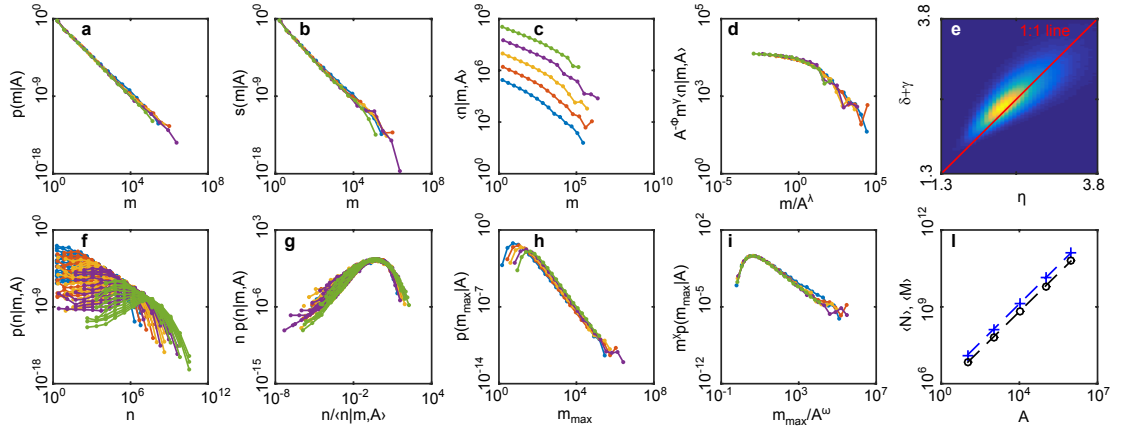


Figure 1.7 – Basic model statistics with parameters  $z = 1/4$ ,  $w = 10^{-4}$ ,  $\alpha = 3/4$ ,  $\theta = 1/4$ . Different colors refer to different values of ecosystem area  $A = 10^i$ , with  $i = 2$  (cyan); 3 (orange); 4 (yellow); 5 (purple); 6 (green). Panels a-c, f and h show respectively  $p(m|A)$ ,  $s(m|A)$ ,  $\langle n|m, A \rangle$ ,  $p(n|m, A)$  and  $p(m_{\max}|A)$  estimated at stationarity. Panels d, g and i show collapses of simulation data for  $\langle n|m, A \rangle$ ,  $p(n|m, A)$  and  $p(m_{\max}|A)$ , respectively. Panel e shows the density scatter-plot of  $\delta + \gamma$  versus  $\eta$ . The density histogram is normalized to one, with blue representing the value zero and yellow the value one. Shown are simulation data for the largest area  $A = 10^6$ . Panel i shows the scaling of the average total biomass  $\langle M \rangle$  (blue crosses and dashed lines) and average total abundance  $\langle N \rangle$  (black dots and dashed lines) with  $A$ . See Table 1.1 for estimates of exponents' values.

## Chapter 1. Covariations in ecological scaling laws

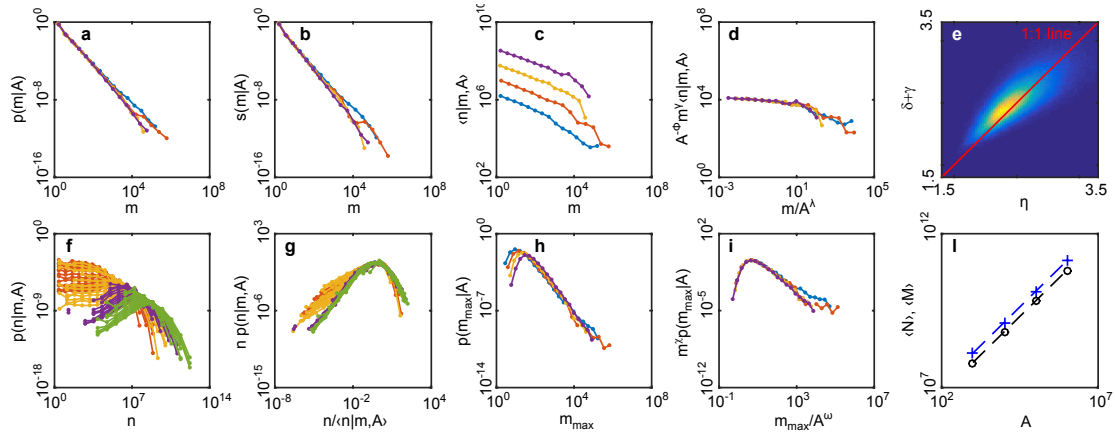


Figure 1.8 – Basic model statistics with parameters  $z = 1/4$ ,  $w = 10^{-5}$ ,  $\alpha = 3/4$ ,  $\theta = 1/4$ . Different colors refer to different values of ecosystem area  $A = 10^i$ , with  $i = 3$  (cyan); 4 (orange); 5 (yellow); 6 (purple). Panels a-c, f and h show respectively  $p(m|A)$ ,  $s(m|A)$ ,  $\langle n|m, A \rangle$ ,  $p(n|m, A)$  and  $p(m_{\max}|A)$  estimated at stationarity. Panels d, g and i show collapses of simulation data for  $\langle n|m, A \rangle$ ,  $p(n|m, A)$  and  $p(m_{\max}|A)$ , respectively. Panel e shows the density scatter-plot of  $\delta + \gamma$  versus  $\eta$ . The density histogram is normalized to one, with blue representing the value zero and yellow the value one. Shown are simulation data for the largest area  $A = 10^6$ . Panel l shows the scaling of the average total biomass  $\langle M \rangle$  (blue crosses and dashed lines) and average total abundance  $\langle N \rangle$  (black dots and dashed lines) with  $A$ . See Table 1.1 for estimates of exponents' values.

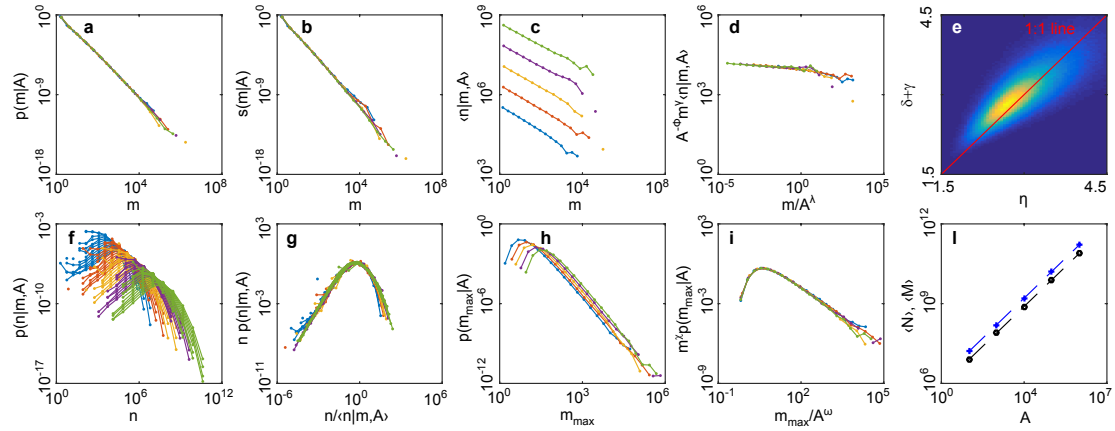


Figure 1.9 – Basic model statistics with parameters  $z = 1/4$ ,  $w = 10^{-3}$ ,  $\alpha = 1/2$ ,  $\theta = 1/4$ . Different colors refer to different values of ecosystem area  $A = 10^i$ , with  $i = 2$  (cyan); 3 (orange); 4 (yellow); 5 (purple); 6 (green). Panels a-c, f and h show respectively  $p(m|A)$ ,  $s(m|A)$ ,  $\langle n|m, A \rangle$ ,  $p(n|m, A)$  and  $p(m_{\max}|A)$  estimated at stationarity. Panels d, g and i show collapses of simulation data for  $\langle n|m, A \rangle$ ,  $p(n|m, A)$  and  $p(m_{\max}|A)$ , respectively. Panel e shows the density scatter-plot of  $\delta + \gamma$  versus  $\eta$ . The density histogram is normalized to one, with blue representing the value zero and yellow the value one. Shown are simulation data for the largest area  $A = 10^6$ . Panel l shows the scaling of the average total biomass  $\langle M \rangle$  (blue crosses and dashed lines) and average total abundance  $\langle N \rangle$  (black dots and dashed lines) with  $A$ . See Table 1.1 for estimates of exponents' values.

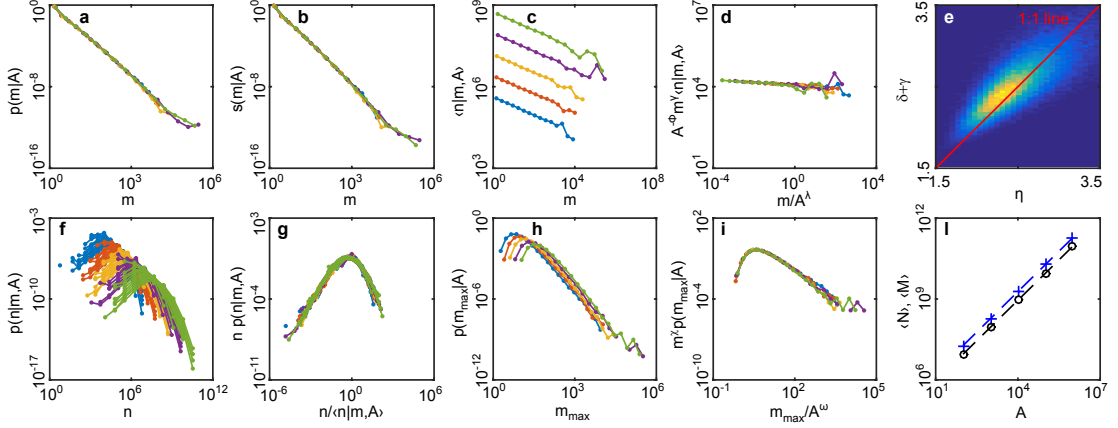


Figure 1.10 – Basic model statistics with parameters  $z = 1/4$ ,  $w = 10^{-3}$ ,  $\alpha = 1/4$ ,  $\theta = 1/4$ . Different colors refer to different values of ecosystem area  $A = 10^i$ , with  $i = 2$  (cyan); 3 (orange); 4 (yellow); 5 (purple); 6 (green). Panels a-c, f and h show respectively  $p(m|A)$ ,  $s(m|A)$ ,  $\langle n|m, A \rangle$ ,  $p(n|m, A)$  and  $p(m_{\max}|A)$  estimated at stationarity. Panels d, g and i show collapses of simulation data for  $\langle n|m, A \rangle$ ,  $p(n|m, A)$  and  $p(m_{\max}|A)$ , respectively. Panel e shows the density scatter-plot of  $\delta + \gamma$  versus  $\eta$ . The density histogram is normalized to one, with blue representing the value zero and yellow the value one. Shown are simulation data for the largest area  $A = 10^6$ . Panel i shows the scaling of the average total biomass  $\langle M \rangle$  (blue crosses and dashed lines) and average total abundance  $\langle N \rangle$  (black dots and dashed lines) with  $A$ . See Table 1.1 for estimates of exponents' values.

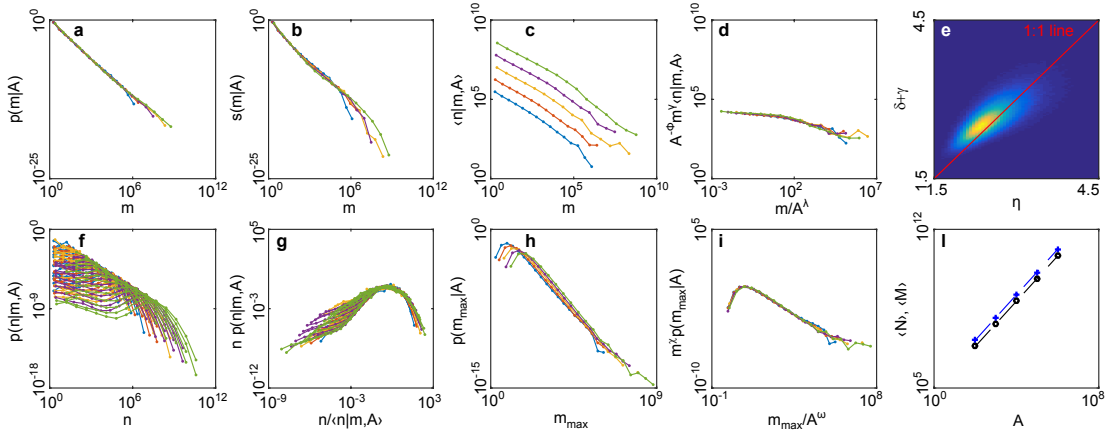


Figure 1.11 – Basic model statistics with parameters  $z = 1/4$ ,  $w = 10^{-3}$ ,  $\alpha = 3/4$ ,  $\theta = 1/2$ . Different colors refer to different values of ecosystem area  $A = 10^i$ , with  $i = 2$  (cyan); 3 (orange); 4 (yellow); 5 (purple); 6 (green). Panels a-c, f and h show respectively  $p(m|A)$ ,  $s(m|A)$ ,  $\langle n|m, A \rangle$ ,  $p(n|m, A)$  and  $p(m_{\max}|A)$  estimated at stationarity. Panels d, g and i show collapses of simulation data for  $\langle n|m, A \rangle$ ,  $p(n|m, A)$  and  $p(m_{\max}|A)$ , respectively. Panel e shows the density scatter-plot of  $\delta + \gamma$  versus  $\eta$ . The density histogram is normalized to one, with blue representing the value zero and yellow the value one. Shown are simulation data for the largest area  $A = 10^6$ . Panel i shows the scaling of the average total biomass  $\langle M \rangle$  (blue crosses and dashed lines) and average total abundance  $\langle N \rangle$  (black dots and dashed lines) with  $A$ . See Table 1.1 for estimates of exponents' values.

## Chapter 1. Covariations in ecological scaling laws

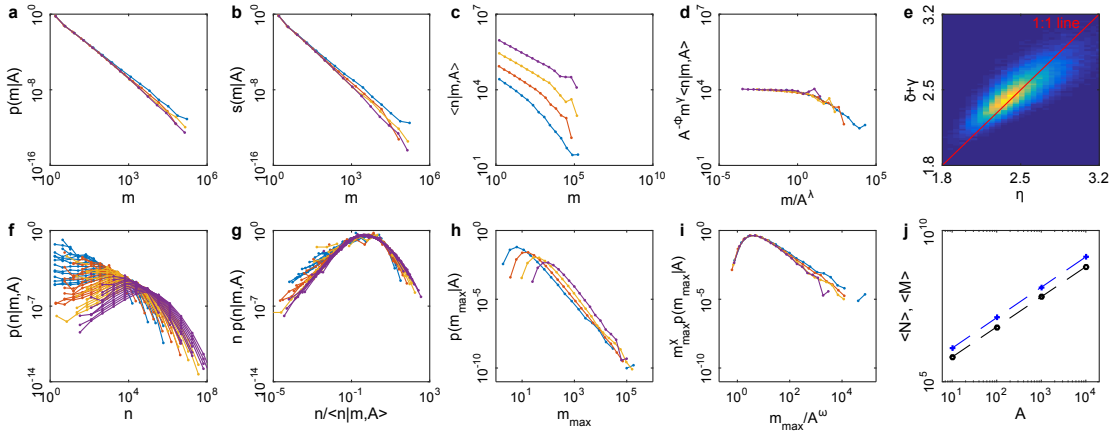


Figure 1.12 – Basic model statistics with parameters  $z = 1/2$ ,  $w = 10^{-3}$ ,  $\alpha = 3/4$ ,  $\theta = 1/4$ . Different colors refer to different values of ecosystem area  $A = 10^i$ , where  $i = 1$  (cyan); 2 (orange); 3 (yellow); 4 (purple). Panels a-c, f and h show respectively  $p(m|A)$ ,  $s(m|A)$ ,  $\langle n|m, A \rangle$ ,  $p(n|m, A)$  and  $p(m_{\max}|A)$  estimated at stationarity. Panels d, g and i show collapses of simulation data for  $\langle n|m, A \rangle$ ,  $p(n|m, A)$  and  $p(m_{\max}|A)$ , respectively. Panel e shows the density scatter-plot of  $\delta + \gamma$  versus  $\eta$ . The density histogram is normalized to one, with blue representing the value zero and yellow the value one. Shown are simulation data for the largest area  $A = 10^4$ . Panel j shows the scaling of the average total biomass  $\langle M \rangle$  (blue crosses and dashed lines) and average total abundance  $\langle N \rangle$  (black crosses and dashed lines) with  $A$ . See Table 1.1 for estimates of exponents' values.

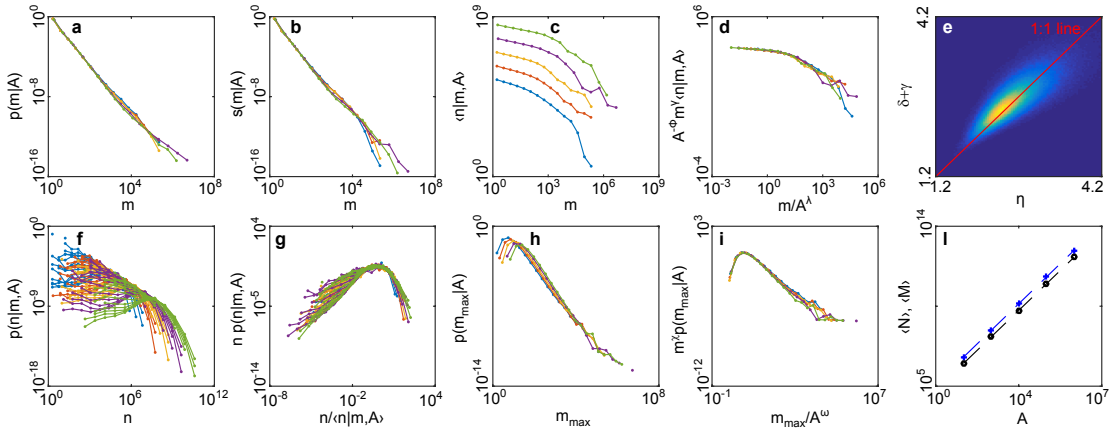


Figure 1.13 – Variation of the basic model, described in section 1.5.3. Statistics computed with parameters  $z = 1/4$ ,  $w = 10^{-6}$ ,  $\alpha = 3/4$ ,  $\theta = 1/2$ . Different colors refer to different values of ecosystem area  $A = 10^i$ , where  $i = 1$  (cyan); 3/2 (orange); 2 (yellow); 5/2 (purple); 3 (green). Panels a-c, f and h show respectively  $p(m|A)$ ,  $s(m|A)$ ,  $\langle n|m, A \rangle$ ,  $p(n|m, A)$  and  $p(m_{\max}|A)$  estimated at stationarity. Panels d, g and i show collapses of simulation data for  $\langle n|m, A \rangle$ ,  $p(n|m, A)$  and  $p(m_{\max}|A)$ , respectively. Panel e shows the density scatter-plot of  $\delta + \gamma$  versus  $\eta$ . The density histogram is normalized to one, with blue representing the value zero and yellow the value one. Shown are simulation data for the largest area  $A = 10^3$ . Panel j shows the scaling of the average total biomass  $\langle M \rangle$  (blue crosses and dashed lines) and average total abundance  $\langle N \rangle$  (black crosses and dashed lines) with  $A$ . See Table 1.2 for estimates of exponents' values.

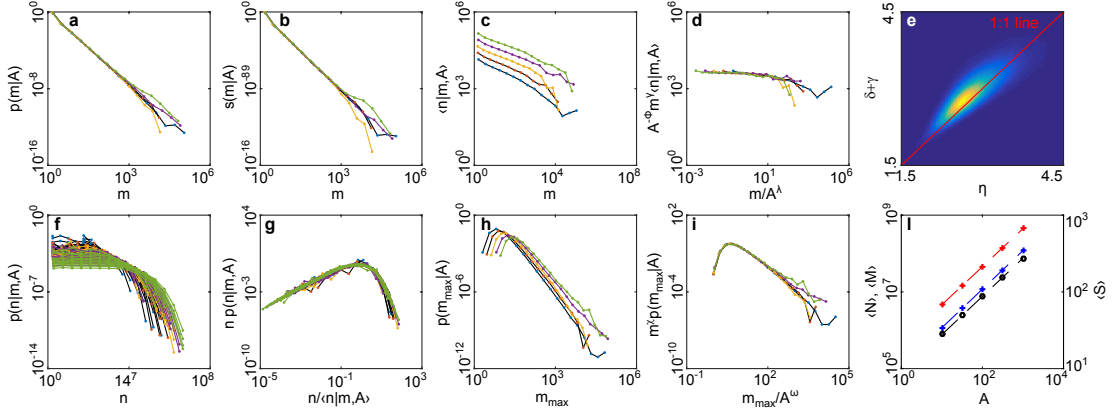


Figure 1.14 – Model *a* (section 1.5.4) statistics computed with the parameter set:  $\alpha = 3/4$ ,  $\theta = 1/4$ ,  $w = 10^{-2}$ ,  $\nu_0 = 1/2$  and  $c = 10^{-5}$ . Different colors refer to different values of ecosystem area  $A = 10^i$ , where  $i = 1$  (cyan);  $3/2$  (orange);  $2$  (yellow);  $5/2$  (purple);  $3$  (green). Panels a-c, f and h show respectively  $p(m|A)$ ,  $s(m|A)$ ,  $\langle n|m, A \rangle$ ,  $p(n|m, A)$  and  $p(m_{\max}|A)$  estimated at stationarity. Panels d, g and i show collapses of simulation data for  $\langle n|m, A \rangle$ ,  $p(n|m, A)$  and  $p(m_{\max}|A)$ , respectively. Panel e shows the density scatter-plot of  $\delta + \gamma$  versus  $\eta$ . The density histogram is normalized to one, with blue representing the value zero and yellow the value one. Shown are simulation data for the largest area  $A = 10^3$ . Panel l shows the scaling of the average total biomass  $\langle M \rangle$  (blue crosses and dashed lines), average total abundance  $\langle N \rangle$  (black crosses and dashed lines) and average number of species  $\langle S \rangle$  (red crosses and dashed lines) with  $A$ . See Table 1.3 for estimates of exponents' values.

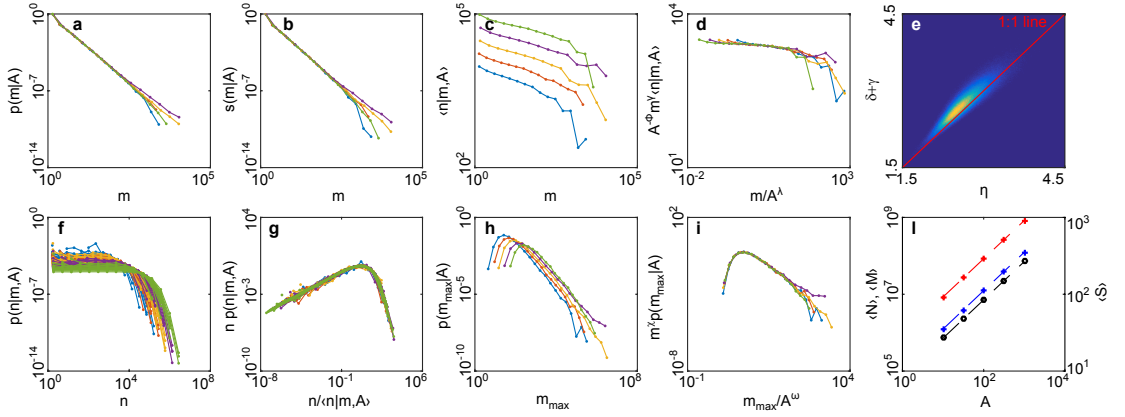


Figure 1.15 – Model *b* (section 1.5.4) statistics computed with the parameter set:  $\alpha = 3/4$ ,  $\theta = 1/4$ ,  $w = 10^{-2}$ ,  $\nu_0 = 1/2$  and  $c = 10^{-5}$ . Different colors refer to different values of ecosystem area  $A = 10^i$ , where  $i = 1$  (cyan);  $3/2$  (orange);  $2$  (yellow);  $5/2$  (purple);  $3$  (green). Panels a-c, f and h show respectively  $p(m|A)$ ,  $s(m|A)$ ,  $\langle n|m, A \rangle$ ,  $p(n|m, A)$  and  $p(m_{\max}|A)$  estimated at stationarity. Panels d, g and i show collapses of simulation data for  $\langle n|m, A \rangle$ ,  $p(n|m, A)$  and  $p(m_{\max}|A)$ , respectively. Panel e shows the density scatter-plot of  $\delta + \gamma$  versus  $\eta$ . The density histogram is normalized to one, with blue representing the value zero and yellow the value one. Shown are simulation data for the largest area  $A = 10^3$ . Panel l shows the scaling of the average total biomass  $\langle M \rangle$  (blue crosses and dashed lines), average total abundance  $\langle N \rangle$  (black crosses and dashed lines) and average number of species  $\langle S \rangle$  (red crosses and dashed lines) with  $A$ . See Table 1.3 for estimates of exponents' values.

## Chapter 1. Covariations in ecological scaling laws

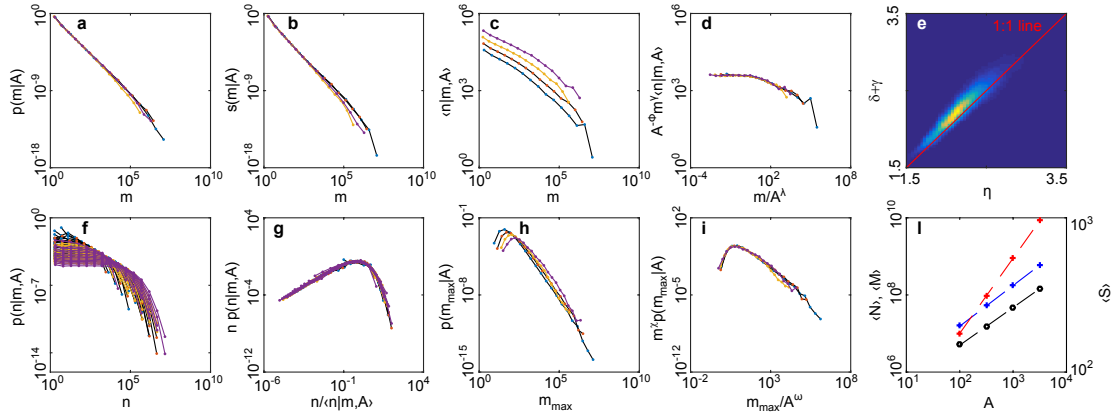


Figure 1.16 – Model *a* (section 1.5.4) statistics computed with the parameter set:  $\alpha = 3/4$ ,  $\theta = 1/4$ ,  $w = 10^{-2}$ ,  $\nu_0 = 1/2$ ,  $c = 10^{-5}$  and  $\bar{q} = 1.2$  ( $\bar{q}$  is the mean of the multiplicative factor  $q$  that defines the descendant species' mass at each speciation event, cfr. section 1.5.4). Different colors refer to different values of ecosystem area  $A = 10^i$ , where  $i = 1$  (cyan);  $3/2$  (orange);  $2$  (yellow);  $5/2$  (purple);  $3$  (green). Panels a-c, f and h show respectively  $p(m|A)$ ,  $s(m|A)$ ,  $\langle n|m, A \rangle$ ,  $p(n|m, A)$  and  $p(m_{\max}|A)$  estimated at stationarity. Panels d, g and i show collapses of simulation data for  $\langle n|m, A \rangle$ ,  $p(n|m, A)$  and  $p(m_{\max}|A)$ , respectively. Panel e shows the density scatter-plot of  $\delta + \gamma$  versus  $\eta$ . The density histogram is normalized to one, with blue representing the value zero and yellow the value one. Shown are simulation data for the largest area  $A = 10^3$ . Panel j shows the scaling of the average total biomass  $\langle M \rangle$  (blue crosses and dashed lines), average total abundance  $\langle N \rangle$  (black crosses and dashed lines) and average number of species  $\langle S \rangle$  (red crosses and dashed lines) with  $A$ . See Table 1.4 for estimates of exponents' values.

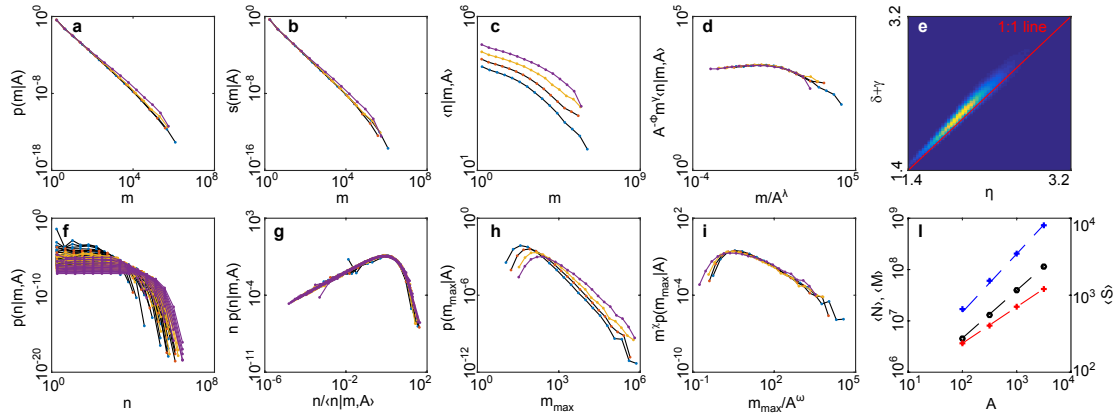


Figure 1.17 – Model *b* (section 1.5.4) statistics computed with the parameter set:  $\alpha = 3/4$ ,  $\theta = 1/4$ ,  $w = 10^{-2}$ ,  $\nu_0 = 1/2$ ,  $c = 10^{-5}$  and  $\bar{q} = 1.2$  ( $\bar{q}$  is the mean of the multiplicative factor  $q$  that defines the descendant species' mass at each speciation event, cfr. section 1.5.4). Different colors refer to different values of ecosystem area  $A = 10^i$ , where  $i = 1$  (cyan);  $3/2$  (orange);  $2$  (yellow);  $5/2$  (purple);  $3$  (green). Panels a-c, f and h show respectively  $p(m|A)$ ,  $s(m|A)$ ,  $\langle n|m, A \rangle$ ,  $p(n|m, A)$  and  $p(m_{\max}|A)$  estimated at stationarity. Panels d, g and i show collapses of simulation data for  $\langle n|m, A \rangle$ ,  $p(n|m, A)$  and  $p(m_{\max}|A)$ , respectively. Panel e shows the density scatter-plot of  $\delta + \gamma$  versus  $\eta$ . The density histogram is normalized to one, with blue representing the value zero and yellow the value one. Shown are simulation data for the largest area  $A = 10^3$ . Panel j shows the scaling of the average total biomass  $\langle M \rangle$  (blue crosses and dashed lines), average total abundance  $\langle N \rangle$  (black crosses and dashed lines) and average number of species  $\langle S \rangle$  (red crosses and dashed lines) with  $A$ . See Table 1.4 for estimates of exponents' values.



Table 1.1 – Scaling exponents measured in the basic model for different sets of parameters' values. Each row (delimited by horizontal lines) refers to a set, indicated by the parameter which value has been changed with respect to the parameters set described in text (indicated as "Base set", specified in section 1.5.1. Under each value, the lower and upper ends of the confidence intervals are reported. The estimates and the confidence intervals were obtained as described in the text.

	$\delta$	$\eta$	$\gamma$	$\Phi$	$\lambda$	$\chi$	$\omega$	$\mu$	$\nu$
Base set Fig. 1.6	2.23	2.54	0.26	0.74	0.38	0.99	0.18	1.0016	0.9982
	2.05	2.28	0.20	0.72	0.32	0.87	0.17	1.0016	0.9980
	2.41	2.79	0.34	0.76	0.42	1.03	0.21	1.0017	0.9983
$w = 10^{-4}$ Fig. 1.7	2.27	2.64	0.33	0.76	0.52	0.98	0.19	1.0030	0.9968
	2.01	2.22	0.28	0.74	0.50	0.92	0.17	1.0027	0.9967
	2.54	2.94	0.38	0.78	0.57	1.04	0.20	1.0034	0.9969
$w = 10^{-5}$ Fig. 1.8	2.28	2.63	0.30	0.75	0.46	0.94	0.18	1.000	0.998
	2.00	2.21	0.25	0.74	0.41	0.73	0.16	0.996	0.995
	2.54	2.94	0.36	0.79	0.52	1.05	0.22	1.005	1.001
Fig. 1.9	2.00	2.21	0.26	0.74	0.73	0.93	0.18	1.0029	0.9989
	2.54	2.93	0.31	0.78	0.87	1.03	0.21	1.0031	0.9989
$\alpha = 1/4$ Fig. 1.10	2.38	2.64	0.29	0.76	0.63	0.98	0.19	1.000	1.000
	2.01	2.22	0.25	0.73	0.60	0.94	0.19	0.998	0.998
	2.54	2.94	0.35	0.80	0.69	1.03	0.20	1.002	1.000
$\theta = 1/2$ Fig. 1.11	2.13	2.62	0.49	0.78	0.47	0.95	0.22	1.002	0.997
	1.88	2.26	0.26	0.75	0.44	0.87	0.21	1.002	0.996
	2.36	2.94	0.56	0.86	0.53	1.03	0.26	1.003	0.998
$z = 1/2$ Fig. 1.12	2.23	2.41	0.32	0.50	0.93	0.98	0.38	1.0	0.99
	2.05	2.28	0.31	0.49	0.92	0.96	0.36	1.00	0.99
	2.54	2.79	0.34	0.51	0.96	1.05	0.39	1.01	1.00

## Chapter 1. Covariations in ecological scaling laws

Table 1.2 – Scaling exponents measured in the variation of the basic model described in section 1.5.3, Fig. 1.13. Under each value, the lower and upper ends of the confidence intervals are reported. The estimates and the confidence intervals were obtained as described in the text.

$\delta$	$\eta$	$\gamma$	$\Phi$	$\lambda$	$\chi$	$\omega$	$\mu$	$\nu$
2.54	2.78	0.13	0.76	0.36	0.94	0.14	1.0021	0.998
2.22	2.29	0.04	0.74	0.35	0.7	0.14	1.0021	0.998
2.83	3.11	0.28	0.78	0.40	1.1	0.17	1.0022	0.998

Table 1.3 – Scaling exponents measured in the community dynamics models with fluctuating numbers of species. Each row (delimited by horizontal lines) refers to a different model (see section 1.5.4). Model parameters are reported in Figs. 1.14 and 1.15. Under each value, the lower and upper ends of the confidence intervals are reported. The estimates and the confidence intervals were obtained as described in the text.

Model	$\delta$	$\eta$	$\gamma$	$\Phi$	$\lambda$	$\chi$	$\omega$	$\mu$	$\nu$	$z$
a Fig. 1.14	2.52	2.83	0.28	0.51	0.92	1.00	0.32	1.01	1.00	0.50
	2.26	2.50	0.24	0.48	0.82	0.91	0.29	1.00	0.99	0.49
	2.77	3.15	0.34	0.56	0.99	1.08	0.33	1.01	1.00	0.50
b Fig. 1.15	2.52	2.69	0.13	0.50	0.60	0.98	0.32	1.00	1.00	0.50
	2.27	2.41	0.10	0.48	0.56	0.92	0.30	0.98	0.99	0.49
	2.76	2.96	0.18	0.52	0.66	1.11	0.34	1.02	1.01	0.50

Table 1.4 – Scaling exponents measured in the models with fluctuating numbers of species and  $\bar{q} > 1$  (see section 1.5.5). Each row (delimited by horizontal lines) refers to a model and parameter set. Model parameters are reported in Figs. 1.16 and 1.17. Under each value, the lower and upper ends of the confidence intervals are reported. The estimates and the confidence intervals were obtained as described in the text. The parameter  $\bar{q}$  is the mean of the multiplicative factor  $q$  that defines the descendant species' mass at each speciation event (cfr. section 1.5.4).

Model	$\delta$	$\eta$	$\gamma$	$\Phi$	$\lambda$	$\chi$	$\omega$	$\mu$	$\nu$	$z$
a with $\bar{q} = 1.2$ Fig. 1.16	1.96	2.22	0.31	0.52	0.80	0.94	0.50	1.06	0.98	0.49
	1.75	1.98	0.26	0.50	0.75	0.75	0.45	0.91	0.95	0.48
	2.16	2.44	0.34	0.55	0.90	1.18	0.58	1.20	1.00	0.50
b with $\bar{q} = 1.2$ Fig. 1.17	1.93	2.06	0.25	0.53	0.93	0.91	0.64	1.13	0.93	0.47
	1.71	1.83	0.19	0.50	0.89	1.65	0.57	0.97	0.87	0.43
	2.14	2.27	0.28	0.55	1.10	0.46	0.70	1.30	0.99	0.52

## 2 Ecological scaling laws in macroecological data

*Empirical laws portraying patterns in ecology are routinely measured in marine and terrestrial environments, and observations suggest that their power-law shape is ubiquitous. However, their scaling exponents do not seem to have a universal value. For example, the abundance of individuals decreases as a power-law of their body mass, but the rate of decrease may vary across different ecosystem types or environmental conditions. The theoretical framework presented in Chapter 1 predicts linking relationships among the scaling exponents of several ecological laws, which describe precisely how their variations are related. In this Chapter, these predictions are tested on available empirical datasets. Tropical forests censuses permit to estimate the exponents of the community size spectrum, of Damuth's law and of the species' typical body mass distribution. The latter pattern is also measured in a dataset of mammal communities on the Sunda Shelf archipelago, together with the SAR and the scaling of the mass of the largest organism with ecosystem area. Finally, island communities of lizards allow estimating the SAR, Damuth's law and the species' typical body mass distribution. The data permit the verification of three of the five linking relationships derived in Chapter 1.<sup>1</sup>*

### 2.1 Introduction

Starting from the second half of the last century, the mathematical quantification of empirical observations has provided abundant examples of power-law patterns in ecology. Figure 2.1 summarizes some of the most important milestones in the identifications of ecological patterns, starting from the influential work of MacArthur and Wilson [1963] on the equilibrium theory of island biogeography. The ubiquity of the power-law functional form, observed for several different patterns and in diverse types of ecosystems, has fascinated and challenged scientist thence.

The scaling patterns considered in this Chapter are either functional relationships between two relevant ecological quantities or probability distributions characterizing their occurrence.

---

<sup>1</sup>This Chapter features contents from [Zaoli et al., 2017].

## Chapter 2. Ecological scaling laws in macroecological data

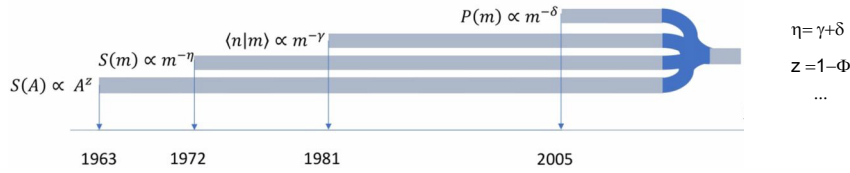


Figure 2.1 – Time line associated with the description of different macroecological patterns and their integration through statistical mechanic approaches: The species area relationship  $S \propto A^z$ , the size spectra  $s(m) \propto m^{-\eta}$ , the scaling of density and size  $\langle n|m \rangle \propto m^{-\gamma}$ , and the power law distribution of body sizes  $p(m) \propto m^{-\delta}$ . At the far right are some of the relationships among exponents derived in Chapter 1. Figure modified from [Marquet, 2017].

Power-law functional relationships have the form

$$y = ax^b \quad (2.1)$$

where  $y$  is the dependent variable,  $x$  the independent one,  $b$  the scaling exponent and  $a$  a proportionality constant. Power-law probability distributions, instead, have the form

$$p(x) = ax^{-b}. \quad (2.2)$$

Eq. (2.2) can be normalized only if  $x \in [x_0, \infty)$ , with  $x_0 > 0$ , and  $b > 1$ . The constant  $a$  is then fixed by normalization. Power-laws have been observed in very diverse fields of science, from biology to physics. To name a few examples, though many more exist, power-law distributions describe the magnitude of earthquakes, the population of cities, the intensity of solar flares, the frequency of words [Newman, 2005], the size of marine organisms [Sheldon et al., 1972, Cavender-Bares et al., 2001, Marañón, 2015, Quinones et al., 2003, Rodriguez and Mullin, 1986] and of trees in a forest [Muller-Landau et al., 2006, Condit et al., 2012]. The commonness of this type of distributions has been explained by the many different processes which produce them [Newman, 2005]. Power-law functional relationships are also observed quite frequently, for example many properties of living organisms (metabolic rate, life span, heart rate, ...) are proportional to a power of their body mass.

Power-laws are often regarded as elegant and beautiful patterns due to their simplicity and their lack of a preferred scale ( $p(bx) = g(b)p(x)$ , see [Newman, 2005]), which might partly explain their large success as a candidate hypothesis against which to test empirical data. On top of their use for prediction and modeling, power-laws have received much interest in ecology and biology because they are considered as a sign that, underneath the complexity of organisms and ecosystems, there exist some general properties and processes which dominate and determine the emergent patterns. In fact, power-law trends, like the probability distribution of organismic size and size-related allometries, hold across very different organisms, from the smallest microbes to much more complicated organisms,

from autotrophs to heterotrophs. This cross-category validity is thought to result from some fundamental properties or processes which associate all living organisms.

For some patterns, a candidate mechanism explaining the power-law shape can be identified. For example, network structure and optimization principles have been invoked to justify the scaling behaviour of metabolic rate with size [West et al., 1997, West, 1999, Banavar et al., 2002, 2010]. Nevertheless, their precious role in the description of empirical patterns, in prediction and in modeling is independent of their mechanistic explication. In some cases, power-laws might be just an approximation of the real behaviour of data but, still, a useful one. First, because they typically capture well the order of magnitude of the dependent variable, which is often sufficient. Second, because they describe the overall behavior of the pattern with just two parameters (one in the case of probability distributions, where the proportionality constant is fixed by normalization), making comparisons straightforward. Third, because their simplicity is convenient when the relationship they describe needs to be used in a wider modeling framework. Clearly, limits of validity of the description must be taken into consideration. For example, power-laws might only describe the pattern in a certain range of values of the independent variable and deviations at small or large values can be explained either by different processes driving the system at small scales or by finite-size arguments. Additionally, power-law functional relationships might only describe the average trend of the pattern, around which potentially important fluctuations occur.

The literature reporting empirical measurements of power-law patterns for the ecological and biological laws considered in Chapter 1 is large. Table 2.1 contains a non-exhaustive list of relevant references divided by type of ecosystems: forests, terrestrial (typically mammals) and aquatic. Note that, for the sake of completeness, the table contains also references for the RSA, although this pattern is usually not a power-law (but see Ser-Giacomi et al. [2018]). A comparison of empirical results highlights the non-universality of the values of scaling exponents. Fig. 2.2 shows evidences of Kleiber's law (panel A), Damuth's law (panel B), the Species-Area Relationship (SAR) (panel C) and the community size-spectrum  $s(m)$  (panel D) for the three types of ecosystems listed above. Regression lines in Fig. 2.2 are fits provided in the original papers (see legends), except for the patterns for forests and terrestrial ecosystems in panel B, which were fitted by linear least-squares fits on log-transformed data, and for the community size-spectra from BCI and Niwot Ridge datasets in panel D, which were fitted with maximum likelihood [Clauset et al., 2009]. Table 2.2 reports the estimates for the scaling exponents. It is clear from these results that exponents values have considerable variations across different types of ecosystems. As shown later in this chapter for consecutive censuses of a tropical forest, they can also vary in time. According to the results obtained in Chapter 1, these variations must be such that at all times the scaling exponents satisfy a set of five linking relationships descending from consistency arguments and from the finite resources constraint present in a finite ecosystem. Specifically, the linking relationships that the exponents have to

satisfy are the following:

$$\eta = \gamma + \delta, \quad (2.3a)$$

$$z = 1 - \Phi - \max\{0, \lambda(1 + \alpha - \eta)\}, \quad (2.3b)$$

$$\mu = 1 + \max\{0, \lambda(2 - \eta)\} - \max\{0, \lambda(1 + \alpha - \eta)\}, \quad (2.3c)$$

$$v = 1 - \max\{0, \lambda(1 + \alpha - \eta)\}, \quad (2.3d)$$

$$\xi = \frac{z}{\delta - 1}. \quad (2.3e)$$

This chapter is devoted to the testing of the relationships in Eqs. (2.3a-e) in empirical datasets. As each of the exponents appearing in Eqs. (2.3a-e) could assume different values in different ecosystems and environmental conditions, in order to verify the validity of Eqs. (2.3a-e) each relationship must be verified on a single dataset, where all the exponents appearing in that particular relationship have been measured simultaneously. A dataset where all the scaling exponents implied in Eqs. (2.3a-e) does not seem to exist to date, but single relationships could be verified on separate datasets, as showed in the following sections.

Box 2.2 explains how to recognize and fit power-laws in empirical data, with particular attention to the cases encountered in the data analysis presented in the following. In section 2.3, Eq. (2.3a) is successfully tested in two tropical forests datasets, in section 2.4 Eq. (2.3b) is tested in a dataset of lizard species living on several islands worldwide and finally in section 2.5 Eq. (2.3e) is tested on a dataset of mammal species living on islands of the Sunda Shelf archipelago.

## 2.2 Power-laws in empirical data

This box presents methods to fit power-laws in empirical data, focusing on the tools used in the following sections. The content of the box is adapted from [Newman, 2005] and [Clauset et al., 2009]. All computations concern the case of continuous variables, relevant for the applications in this chapter.

### 2.2.1 Plotting a power-law

Power-laws enjoy the useful property of being linear when plotted in a log-log scale. In fact, if  $y = ax^b$ , then  $\log y = \log a + b \log x$ . The slope of the line in the log-log plot is the power-law exponent. Double logarithmic scale plots are therefore a first tool to reveal power-law behavior. Nevertheless, in the case of power-law distributions matters are complicated because of binning. If a simple linear binning is used, the tail of the distribution will appear very noisy, as few data points will fall in each bin. Often, the tail of the distribution contains useful information which one does not want to discard. One possible solution is then to use bins whose size grows with  $x$ . In particular, a common

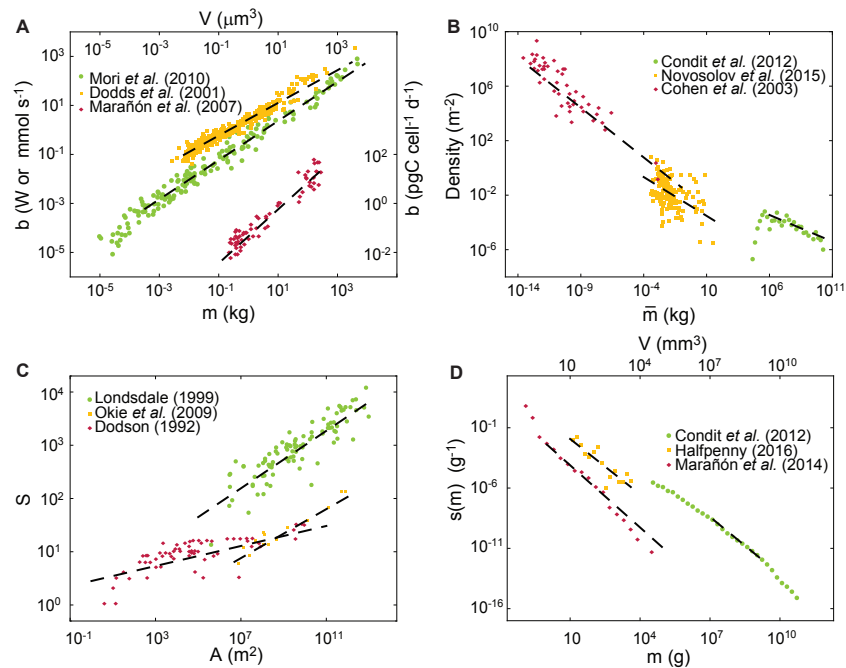


Figure 2.2 – Empirical evidence of scaling ecological patterns in different ecosystems: forests (green), terrestrial (yellow) and aquatic ecosystems (magenta). Regression lines are linear least square fits of log-transformed data. A) Kleiber's law: metabolic rates in  $\mu\text{mol s}^{-1}$  (forests), W (terrestrial ecosystems),  $\text{pgC cell}^{-1} \text{d}^{-1}$  (aquatic ecosystems). Size in kg (forests and terrestrial ecosystems) and in  $\mu\text{m}^3$  (aquatic ecosystems); B) Damuth's Law ( $\bar{m}$  is a species' mean mass). See section 2.4 for a comment on the scattering of points of the dataset in Novosolov et al. [2015]; C) SAR; D) Community size-spectrum: size in g (forests and terrestrial ecosystems) and  $\mu\text{m}^3$  (aquatic ecosystems). See section table 2.2 for scaling exponents estimates/errors.

choice is logarithmic binning, where bins are chosen such that each is a fixed multiple wider than the one before it. This means that the bins in the tail of the distribution get more samples than they would if bin sizes were fixed, and this reduces the statistical errors in the tail. It also has the nice side-effect that the bins appear to be of constant width when the histogram is plotted on log scales. Probability distributions throughout this thesis are plotted with this method. Note that a correction to the counts is necessary to take into account the increasing bin size. In order to obtain counts per unit interval of  $x$ , the counts in a bin of size  $\Delta x$  must be divided by  $\Delta x$ . In the case of power-laws, failure to correct for the logarithmic binning causes a biased estimation of the exponent, which results increased by 1. In fact, if bins have all the same size in a logarithmic scale, a bin centered around  $x$  has width  $\Delta x \propto x$  in the linear scale. Therefore, its counts will be multiplied by a factor  $x$ , modifying the slope of the plotted line. This issue has often caused some confusion on the empirical values of scaling exponents reported in the literature.

Another method for reducing the noise is plotting cumulative distributions, that is, the probability that the variable exceeds a value  $x$ ,

$$C(x) = \int_x^{\infty} p(x') dx'. \quad (2.4)$$

If  $p(x) = ax^{-b}$ , with  $b > 1$  (otherwise the distribution cannot be normalized),  $C(x)$  will also be a power-law with a shallower exponent  $-b + 1$ :

$$C(x) = \int_x^{\infty} ax'^{-b} dx' = \frac{a}{b-1} x^{-(b-1)}. \quad (2.5)$$

This method does not require any binning, solving all problems related to bins size.

### 2.2.2 Fitting a power-law

In the case of power-law functional relationships, a linear least square fit of log transformed data is a reliable method to estimate the parameters  $a$  and  $b$ . In the case of a power-law distribution  $p(x) = ax^{-b}$ , instead, such method is known to introduce systematic bias in the estimated value of the exponent [Goldstein et al., 2004]. A more reliable method to estimate the parameters is maximum likelihood. To explain how that works, let us begin by noting that a pure power-law distribution can only be normalized on  $[x_{\min}, \infty]$  when  $x_{\min} > 0$ . Therefore, either  $p(x) = 0$  for  $x < x_{\min}$ , or  $p(x)$  has a different behavior before  $x_{\min}$  (for example, flat or increasing). Consider first the simpler case in which the value  $x_{\min}$  is known. The normalized form of  $p(x)$  is then

$$p(x) = \frac{b-1}{x_{\min}} \left( \frac{x}{x_{\min}} \right)^{-b}. \quad (2.6)$$



Given a set of  $n$  values  $x_i$ , assuming they are independent, the probability that they were generated from this distribution is

$$P(x|b) = \prod_{i=1}^n p(x_i) = \prod_{i=1}^n \frac{b-1}{x_{\min}} \left( \frac{x}{x_{\min}} \right)^{-b}. \quad (2.7)$$

This quantity is called the *likelihood* of the dataset. The probability that the power-law exponent has value  $b$ , given the dataset  $\{x_i\}$ , is related to  $P(x|b)$  by Bayes' theorem:

$$P(b|x) = P(x|b) \frac{P(b)}{P(x)} \quad (2.8)$$

where  $P(x)$  is the prior probability of the data, and  $P(b)$  is the prior probability of the exponent, assumed uniform. Therefore, maximizing  $P(b|x)$  with respect to  $b$  is equivalent to maximizing the likelihood  $P(x|b)$  with respect to  $b$ . One can equivalently maximize the log likelihood

$$\begin{aligned} \mathcal{L} = \log P(x|b) &= \sum_{i=1}^n [\log(b-1) - \log x_{\min} - b \log \frac{x_i}{x_{\min}}] \\ &= n \log(b-1) - n \log x_{\min} - b \sum_{i=1}^n \log \frac{x_i}{x_{\min}}. \end{aligned} \quad (2.9)$$

Setting  $\partial \mathcal{L} / \partial b = 0$ , one obtains the estimate

$$b = 1 + n \left[ \sum_i \log \frac{x_i}{x_{\min}} \right]^{-1}. \quad (2.10)$$

One can also give an error to this estimate by computing the standard deviation of  $P(b|x)$ ,  $\sigma = \sqrt{\langle b^2 \rangle - \langle b \rangle^2}$  (see Newman [2007]), which when  $n \gg 1$  is approximated by

$$\sigma = \frac{b-1}{\sqrt{n}} \quad (2.11)$$

where  $b$  is the maximum likelihood estimate.

In practice,  $x_{\min}$  is usually not known exactly and needs to be estimated before one can estimate  $b$ . Wrongly choosing  $x_{\min}$  can create bias in the estimated  $b$ , either because data from the non power-law regime are included in the estimate, or because useful data are discarded, increasing the statistical error. An objective method to estimate  $x_{\min}$  is proposed in [Clauset et al., 2009]. The fundamental idea behind this method is that the estimate of  $x_{\min}$  is that such that the best-fit power-law model with that value of  $x_{\min}$  is as similar as possible to the measured data above  $x_{\min}$ . An algorithm to perform the estimation is provided in [Clauset et al., 2009]. The algorithm tests increasing values of  $x_{\min}$ , for each computes the best-fit power law and compares it to the data by a Kolmogorov-Smirnov test. The estimated value of  $x_{\min}$  is the one minimizing the KS

distance. The algorithm is used to fit power-law distributions in this chapter when no cut-off at large size is present.

### 2.2.3 Power-laws with cut-off at large value

In some cases, probability distributions studied in this chapter show a power-law behavior followed by a faster decay at large  $x$ . The exponent of the power-law part of the distribution can still be estimated by maximum likelihood if the shape of the cut-off is known. For example, let us consider the case where the cut-off is exponential, i.e.,

$$p(x) = ax^{-b}e^{-cx} \quad (2.12)$$

for  $x > x_{min}$ . The constant  $a$  is fixed by normalization to  $a = \frac{c^{1-b}}{\Gamma(1-b, x_{min}b)}$  where  $\Gamma$  is the incomplete Gamma function. The log likelihood is then

$$\log \mathcal{L}(b, c) = n \log \left( \frac{c^{1-b}}{\Gamma(1-b, x_{min}b)} \right) - b \sum_{i=1}^n \log x_i - c \sum_{i=1}^n x_i, \quad (2.13)$$

which can be maximized numerically with respect to  $b$  and  $c$ .

## 2.3 Test of the relationship $\eta = \gamma + \delta$ in tropical forests

### 2.3.1 The datasets

The three exponents involved in Eq. (2.3a) are simultaneously measurable in censuses of tropical forests, where individual sizes of trees are measured jointly with the species to which they belong. Here, two datasets are considered: the Barro Colorado Island (BCI) forest [Condit et al., 2012] (Fig. 2.3) and the Luquillo forest [Zimmerman et al., 2010] (Fig. 2.4). The BCI forest is situated in an island of 1560 ha in the Panama canal which was formed when engineers dammed the Chagres River in 1914 to create Gatun Lake. Censuses of a 50 ha plot were performed regularly starting from 1985 every 5 years. The Luquillo forest is located within the U.S. El Yunque National Forest, in the northeastern Puerto Rico. The Luquillo Forest Dynamics Plot, 16 ha, is censused regularly. The analysis reported here concerns the fifth, sixth and seventh censuses of BCI and the five censuses of the Luquillo forest available online in the Center for Tropical Forest Science (CTFS) dataset collection. These datasets report the trunk diameter and the species' identity of every tree having a diameter at breast height (dbh) > 10 mm contained in the censused plot. Diameters were converted into mass using an established allometric relationship between mass and diameter [Enquist and Niklas, 2001, Simini et al., 2010],  $m = 0.124d^{8/3}$  kg, with  $d$  expressed in cm.

### 2.3.2 Estimating the scaling exponents $\delta$ and $\eta$

The community size-spectrum and the species typical body mass distribution are shown in Figs. 2.3A and B and 2.4A and B. To compute  $p(\bar{m}|A)$ , where  $\bar{m}$  is the typical species' mass, the mean mass of the species' individuals was used as an estimate of  $\bar{m}$ . To obtain estimates of  $\delta$  and  $\eta$ , the probability distributions  $p(\bar{m}|A)$  and  $s(m|A)$  were fitted via maximum likelihood to the functional forms  $p(\bar{m}|A) = a_1 \bar{m}^{-\delta} e^{-b_1 \bar{m}}$  and  $s(m|A) = a_2 m^{-\eta} e^{-b_2 m}$ , where  $a_1$  and  $a_2$  are the normalization constants (which can be expressed in terms of  $b_1$  and  $b_2$ ),  $b_1$  and  $b_2$  are constants that accounts for possible finite-size effects, and  $\delta$  and  $\eta$  are the power-law exponents (see Box 2.2.2). Section 1.6.3 justifies the possible presence of an upper cutoff in  $p(\bar{m}|A)$ . To account for deviations from the power-law behavior at low values of  $m$  or  $\bar{m}$  (these may arise for various reasons, like e.g. sampling protocols affecting the estimates of mean masses and mean abundances at small masses, as described later in the section)  $\delta$  and  $\eta$  were estimated via maximum-likelihood by considering only the data with  $m > m_k$  at various values of  $m_k = a^k 0.124$  kg in the range  $(0.124 - 10^2)$  kg (with  $1 < a < 2$  and  $k$  integer). Note that 0.124 kg is the mass of a tree with dbh=10 mm, i.e. the lower limit of the sampling protocol. If the data were distributed according to a pure power-law with no finite-size effects, such procedure would return approximately the same value of the exponent for any  $m_k$ . If the data were distributed according to a power-law with finite-size effects at small and large values of  $m$ , instead, one would observe an approximately constant estimate of the exponent at intermediate  $m_k$  and deviations from such estimate at small and large values of  $m_k$  (see e.g. Fig. 2.5). For each fit, the extent of the power-law regime was identified and the estimate of the exponent and the associated error are, respectively, the mean and standard deviation of the maximum-likelihood exponent at different values of  $\bar{m}_k$  in the power-law regime.

### 2.3.3 Estimating the scaling exponents $\gamma$

Damuth's law in the two datasets is plotted respectively in Figs. 2.3C and 2.4C. The estimate of Damuth's law exponent  $\gamma$ , describing the decay of mean abundances with species' typical masses  $\bar{m}$ , requires a correction for a bias introduced by the sampling protocol on the estimates of mean abundances and mean masses at small values of  $\bar{m}$ . In fact, the sampling protocol in tropical forests censuses instructs to sample only the trees with dbh larger than 10 mm. The measured abundances of small species (i.e. those with typical diameter close to 10 mm) are therefore lower than the true ones because individuals with diameter  $d \leq 10$  mm were not censused. As a result, the average abundance as a function of a typical species' mass initially increases with  $\bar{m}$  and is followed by the decreasing power-law regime where the effect of the sampling protocol becomes unimportant (Fig. 2.6A). The initial increase is a sampling artifact. One can verify that this is the case by creating an artificial forest dataset where species' mean abundances follow Damuth's law exactly. Within such artificial forest, species mean masses  $\bar{m}$  are distributed according to the pure power-law  $p(\bar{m}) = (\delta - 1) \bar{m}^{-\delta}$ . The abundance of each species is drawn from a Poisson distribution with mean  $\bar{m}^{-\gamma}$ . Finally, each individual of each species needs to be assigned a mass. To do so, a characterization of the intra-specific mass

## Chapter 2. Ecological scaling laws in macroecological data

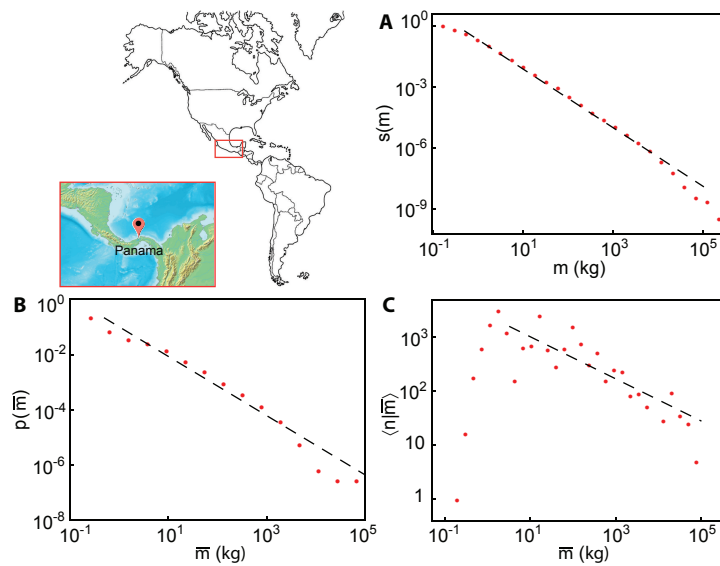


Figure 2.3 – Empirical evidence of scaling patterns in Barro Colorado Island [Condit et al., 2012], seventh census: A) Community size-spectrum  $s(m)$ , B) Distribution of species mean masses  $p(\bar{m})$ , C) Damuth's law  $\langle n|\bar{m} \rangle$ , where each point is the average abundance over bins of logarithmic size. Red dots represent empirical data, dashed black lines show power-functions with exponents reported in Table 2.3. Details on exponents' estimates are reported in the text. Note that finite-size effects may be present both at small and large values of  $m$  and  $\bar{m}$ , for example due to the sampling protocol (see text).

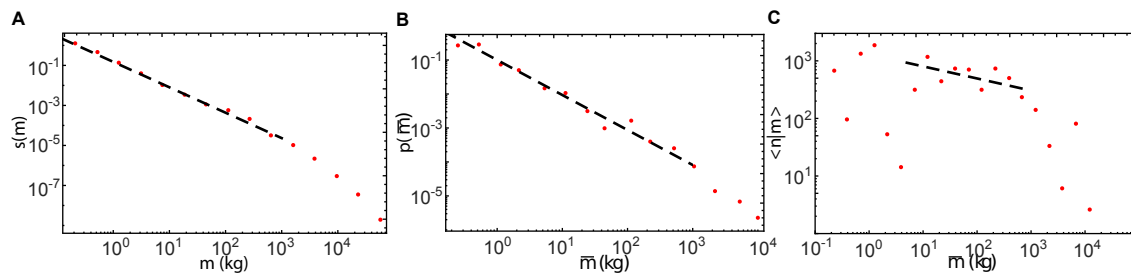


Figure 2.4 – Empirical evidence of scaling patterns in the Luquillo forest, first census. A) Community size-spectrum  $s(m)$ , B) Distribution of species mean masses  $p(\bar{m})$ , C) Damuth's law  $\langle n|\bar{m} \rangle$ , where each point is the average abundance over bins of logarithmic size. Red dots represent empirical data, dashed black lines show power-functions with exponents reported in Table 2.4. Details on exponents' estimates are reported in the text. Note that finite-size effects may be present both at small and large values of  $m$  and  $\bar{m}$ , for example due to the sampling protocol (see text).

### 2.3. Test of the relationship $\eta = \gamma + \delta$ in tropical forests

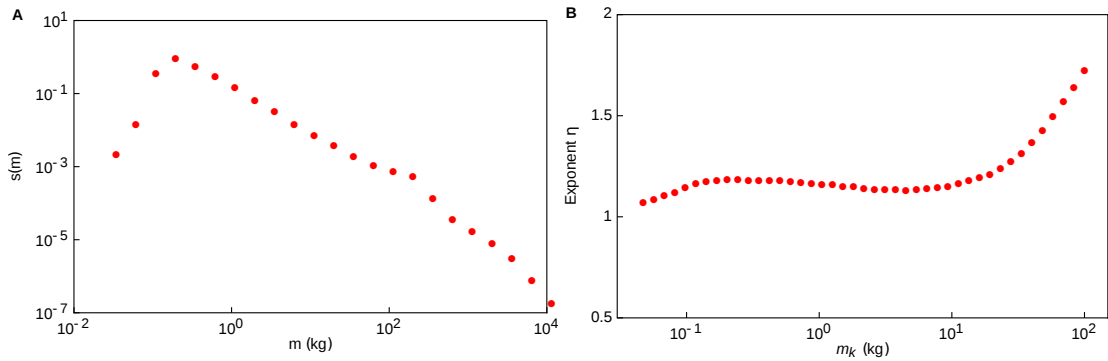


Figure 2.5 – A) Size spectrum in the Luquillo forest, second census. Finite-size effects are present both at small (i.e.  $m < 0.2$  kg) and large (i.e.  $m > 10^2$  kg) values of  $m$ . B) Size-spectrum exponent  $\eta$  estimated using only data with mass  $m > m_k$ . The estimated exponent initially increases (until  $m_k \approx 0.2$  kg) due to a finite-size effect, then is rather stable until the statistics is not sufficient to properly estimate it ( $m_k > 20$  kg).

distributions of tropical trees  $p(m|\bar{m})$  was used, i.e. the probability that an individual has mass in  $(m, m + dm)$  given that it belongs to a species with mean mass  $\bar{m}$ . Such characterization is presented in section 2.3.4. The masses of individuals belonging to each species in the artificial forest can then be randomly sampled from such distributions. Finally, the sampling protocol is mimicked by eliminating all individuals with mass lower than 0.124 kg (corresponding to a dbh of 10 mm) and Damuth's law is computed in such a filtered dataset. Fig. 2.6C shows that, despite the fact that mean species abundances in the artificial forest follow Damuth's law exactly, the sampling protocol causes the emergence of a new regime at small  $\bar{m}$  where the relationship between  $\langle n|\bar{m} \rangle$  and  $\bar{m}$  is monotonically increasing. This demonstrates that the sampling protocol introduces an artificial deviation from the power-law regime which has to be considered with care while interpreting empirical data.

To derive the estimate for  $\gamma$  and the associated error, the typical species masses were binned logarithmically and the mean abundance of all species within each bin was computed. Then, the exponent of Damuth's law was computed for different choices of the number of bins  $n_{bin}$  via least-squares fitting of log-transformed data, weighted by the standard deviation of abundances within each bin, yielding estimates  $\gamma_{n_{bin}}$ . The estimate for  $\gamma$  is the mean  $\gamma = \langle \gamma_{n_{bin}} \rangle$  across several values of  $n_{bin}$ . To correct for the bias caused by the sampling protocol, such computation was repeated by considering only the species with mean mass  $\bar{m} > \bar{m}_k$ , with  $\bar{m}_k = a^k \cdot 0.124$  kg in the range  $(0.124 - 10^2)$  kg (with  $1 < a < 2$  and  $k$  integer). If mean abundances followed Damuth's law exactly (in the absence of sampling bias and with sufficient statistics), such procedure would return the same value of  $\gamma$  for any  $\bar{m}_k$ . If finite-size effects were present at small and large values of  $\bar{m}$ , instead, one would observe an approximately constant estimate of the exponent at intermediate  $\bar{m}_k$  and deviations from such estimate at small and large values of  $\bar{m}_k$  (see e.g. Fig. 2.6B). For each fit, the extent of the power-law regime was identified and the estimate of  $\gamma$  and the associated error are, respectively, the mean and standard deviation of the exponents estimated at the different

## Chapter 2. Ecological scaling laws in macroecological data

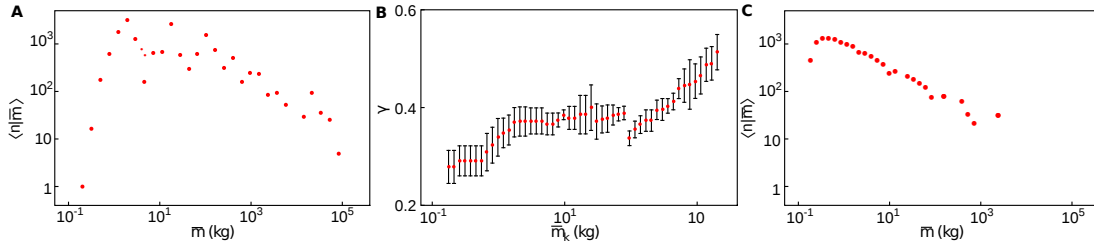


Figure 2.6 – A) Damuth's law ( $\langle n|\bar{m} \rangle$  vs  $\bar{m}$ ) in BCI, seventh census. Finite-size effects are present both at small (i.e.  $\bar{m} < 10$  kg) and large (i.e.  $\bar{m} > 10^3$  kg) values of  $\bar{m}$ . B) Damuth's law exponent  $\gamma$  estimated using only the species with typical mass  $\bar{m} > \bar{m}_k$ . The estimated exponent initially increases, then is stable until the statistics is not sufficient to properly estimate it (or finite-size effects become relevant). C) Damuth's law in an artificial forest after mimicking the sampling protocol (see text for details). The effect of the sampling bias is visible at small masses.

values of  $\bar{m}_k$  in the power-law regime.

### 2.3.4 Intra-specific body mass distribution for trees

The intra-specific mass distributions were computed in the BCI and Luquillo forests for species with more than 1500 (BCI) and 400 (Luquillo) individuals and are shown in Fig. 2.7. Most species have intra-specific size distributions characterized by the finite-size scaling form:

$$p(m|\bar{m}) = m^{-\Delta} \mathcal{F}\left(\frac{m}{\bar{m}^\Omega}\right), \quad (2.14)$$

where  $\Omega = 1/(2 - \Delta)$  ensures that  $\int m p(m|\bar{m}) dm = \bar{m}$ ,  $\Delta = 1.12 \pm 0.06$  for the BCI forest and  $\Delta = 1.14 \pm 0.03$  for the Luquillo forest and  $\mathcal{F}(x)$  is a scaling function with limiting behaviors  $\mathcal{F}(x) \simeq \text{const}$  when  $x \ll 1$  and  $\lim_{x \rightarrow \infty} \mathcal{F}(x) = 0$ . The estimate of the exponent  $\Delta$  is obtained by optimizing the data collapse of  $p(m|\bar{m})$  of several tree species: plotting  $m^\Delta p(m|\bar{m})$  vs  $m/\bar{m}^\Omega$ , all the curves collapse on the same curve  $\mathcal{F}$  for the appropriate value of  $\Delta$ , see insets in Fig. 2.7. The optimization is performed according to the method described in [Bhattacharjee and Seno, 2001]. The error is computed as the value of the exponent at which the error functional  $P_b$  defined in [Bhattacharjee and Seno, 2001] is 1% larger than its value at the minimum. The cut-off function obtained by data collapse is well fitted by  $\mathcal{F}(x) = q_0 e^{-q_1 x}$ , where  $q_0$  and  $q_1$  are constants. The fitted values are  $q_0 = 0.17$ ,  $q_1 = 0.21$  for the BCI forest, and  $q_0 = 0.14$ ,  $q_1 = 0.15$  for the Luquillo forest.

### 2.3.5 Test of the linking relationship

The estimated exponents in each of the considered censuses of the two forests, reported in tables 2.3 and 2.4, satisfy the linking relationship (2.3a) within the errors. Whereas BCI censuses appear very similar to each other (and therefore also the exponent values estimated

### 2.3. Test of the relationship $\eta = \gamma + \delta$ in tropical forests

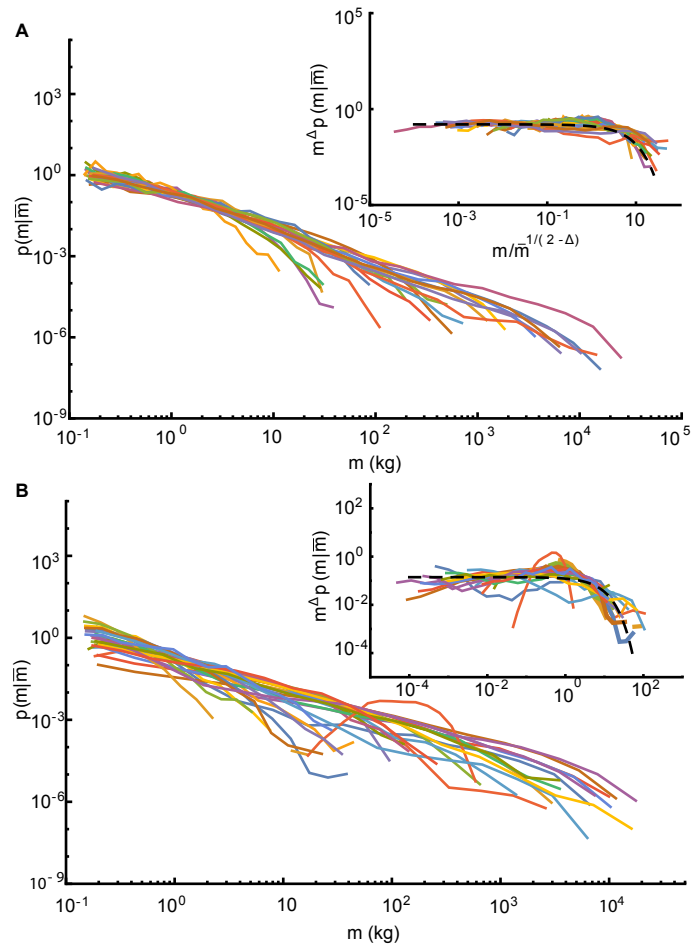


Figure 2.7 – Intra-specific tree size-distribution in BCI forest (A) and Luquillo forest (B). Each color corresponds to a different species. Insets: Intra-specific tree size distributions collapse according to (2.14) onto the same universal curve. The dashed black line is the best fit of  $\mathcal{F}(x) = q_0 e^{-q_1 x}$  (see text).

in different censuses), the Luquillo forest appears to be more dynamic (note that the forest was hit by a major hurricane between censuses 2 and 3), with the values of  $\gamma$  decreasing in time after 1998 (census 2, see Table 2.4). Because the estimate of  $\delta$  remains constant suggesting that climatic, ecological or anthropogenic dynamics affected only species' abundances in this forest, the framework would predict via Eq. (2.3a) that  $\eta$  would also decrease in time, and in fact this is also found in the data, with Eq. (2.3a) being verified in all censuses. Note that both the BCI and the Luquillo datasets reject the linking relationship  $\eta = \delta$  predicted by the previous theoretical work in [Banavar et al., 2007].

Although there is no estimate of the exponents  $\mu$  and  $\nu$  for tropical forests, a reasonable assumption is  $\mu = 1 = \nu$ . Given that  $\eta < 2$  for the BCI and the Luquillo datasets, such assumption is only verified if  $\lambda = 0$ , i.e. if the maximum body size does not scale with the ecosystem area. An analysis of this situation is provided in section 1.6.3.

## **2.4 Test of equation $z = 1 - \Phi - \max\{0, \lambda(1 + \alpha - \eta)\}$ on a dataset of lizard species**

The validity of (2.3b) was tested on a dataset gathering population densities of several species of lizards on 64 islands worldwide (LIZ) [Novosolov et al., 2015], with areas ranging from  $10^{-1}$  to  $10^5$  km<sup>2</sup>. In this dataset, the SAR was fitted with linear least-squares regression on log-transformed data and  $p(\bar{m}|A)$  via maximum-likelihood [Clauset et al., 2009] (Fig. 2.8A and B), where  $\bar{m}$  are species' mean masses.  $p(\bar{m})$  is computed gathering together species from all the islands in the dataset. The exponent  $\Phi = 0.78$  (describing the dependence of  $\langle n|\bar{m}, A \rangle$  on  $A$ ) was obtained by maximizing the coefficient of determination  $R^2$  of the linear least-squares regression of the pairs  $(\log \bar{m}, \log \frac{n}{A^\Phi})$  obtained by varying  $\Phi$  in the interval  $[0, 2]$  (Fig. 2.8C and D). The estimate of  $\gamma$  is obtained from the pairs  $(\log \bar{m}, \log \frac{n}{A^\Phi})$  computed with the optimal value of  $\Phi$  with the same methods used for forests, with  $\bar{m}_k = 2c^k 10^{-1}$  kg in the range  $(10^{-1} - 10^2)$  kg (with  $1 < c < 2$  and  $k$  integer). Note that this estimate of  $\gamma$  is different from the one given in Table 2.2 for the same dataset, which was obtained by plotting densities versus typical masses (equivalent to taking  $\Phi = 1$ ) in order to allow comparison with other data from the literature. Accounting for the different area of the islands through the factor  $A^\Phi$  reduces the scatter of the points (compare the yellow dots in Fig. 2.2B and Fig. 2.8D). The remaining scatter in Fig. 2.8D, as well as that in Fig. 2.8A, can be explained, first, by the different environmental conditions of the different islands, possibly implying a different proportionality factor in front of the scaling with  $A$  of the species densities and of the number of species. This cause of scattering is minimized when data come from islands of the same archipelago, sharing the same environmental conditions (see, for example, the data used in section 2.5). Secondly, scatter might be caused by measurement errors. Finally, a certain amount of scatter is expected even in optimal measuring conditions, being the number of species and their densities the outcome of the stochastic processes of community dynamics,



## 2.5. Test of the relationship $z = \xi(\delta - 1)$ on a dataset of mammal species

---

and as such subject to random fluctuations (see Chapter 3 for a more thorough discussion of this point). However, the scatter does not prevent the estimation of the average trends of the SAR and of Damuth's law in Figs 2.8A and D. The estimates for the scaling exponents in this dataset are reported in Table 2.5. Because in this dataset  $\eta = \delta + \gamma$  is compatible with 2, Eqs. (2.3c) and (2.3d) imply  $\mu = \nu$ . Furthermore, because in general  $\alpha \leq 1$ , one has  $\max\{0, \lambda(1 + \alpha - \eta)\} = 0$  and therefore our framework predicts

$$z = 1 - \Phi, \tag{2.15a}$$

$$\nu = \mu = 1. \tag{2.15b}$$

Eq. (2.15a) is satisfied within the errors by the estimates of  $z$  and  $\Phi$  (see Table 2.5). Eq. (2.15a) implies that in order to have  $z > 0$ , as always found empirically,  $\Phi$  needs to be smaller than one. Since  $\Phi$  describes the scaling of  $\langle n | \bar{m}, A \rangle$  with  $A$ , the framework predicts that species' densities should decrease with increasing ecosystem area, as found in the LIZ dataset. Note that Eq. (2.3a) could not be tested on this dataset because the body size of single individual was not measured, therefore  $\eta$  cannot be estimated. Also Eqs (2.3c) and (2.3d) could not be tested, as total abundance and biomass information is not available. As individual body masses are not known, Eq. (2.3e) could not be tested either.

## 2.5 Test of the relationship $z = \xi(\delta - 1)$ on a dataset of mammal species

The validity of (2.3e) was tested on a dataset of terrestrial mammals species presence/absence data on several islands in Sunda Shelf (SSI) [Okie and Brown, 2009], covering a wide range of island areas ( $10^1$  to  $10^6$  km<sup>2</sup>). The dataset includes extant species and species that went extinct during historic times because of human impacts and excludes human-introduced species. For each species, mean body mass is known (see Okie and Brown [2009] for details). The SAR and the scaling of the maximum body mass with the area were fitted with linear least-squares regression on log-transformed data, while  $p(\bar{m} | A)$  was fitted with maximum-likelihood [Clauset et al., 2009].  $p(\bar{m})$  is computed gathering together species from all the islands in the dataset. Scaling exponents in this dataset, reported in Table 2.6, are consistent with Eq. (2.3e) within one standard error, as  $z = 0.23 \pm 0.02$  and  $\xi(\delta - 1) = 0.29 \pm 0.1$ . Note that Eqs (2.3a) and (2.3b) could not be tested in this dataset because abundances are missing. Eqs (2.3c) and (2.3d) could not be tested, as total abundance and biomass information is not available.

## 2.6 Discussion

In this chapter, linking relationship predicted by the theoretical framework in Chapter 1 were successfully tested on empirical data on tropical forests, and island communities of lizard and mammals. The empirical tests performed are summarized in Table 2.7. Since the framework

## Chapter 2. Ecological scaling laws in macroecological data

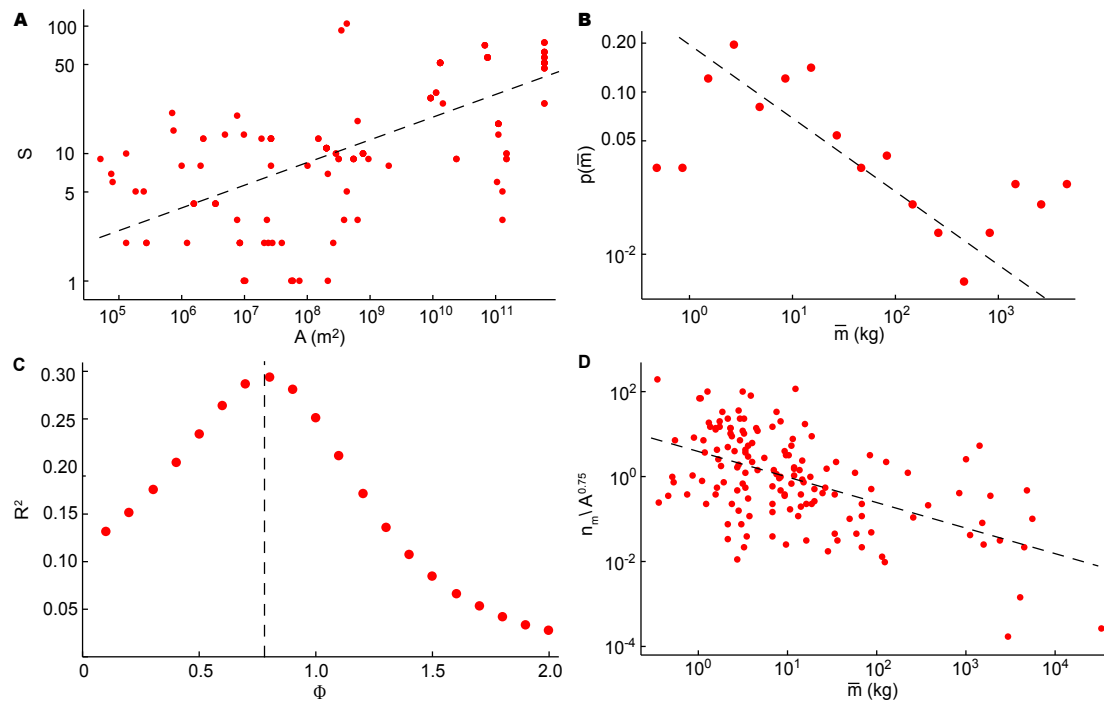


Figure 2.8 – Empirical evidence of scaling patterns in LIZ dataset [Novosolov et al., 2015]: A) SAR; B)  $p(\bar{m})$ ; C) Coefficient of determination  $R^2$  of the linear least-squares fit of  $\left(\log \bar{m}, \log \frac{n(\bar{m}, A)}{A^\Phi}\right)$  for  $\Phi \in [0, 2]$ . Dashed line in correspondence of the value  $\Phi = 0.75$  giving the best fit; D) Non-binned Damuth's law plotted using the estimated value of  $\Phi$  ( $\bar{m}$  are species' mean masses). Best fit parameters are reported in Table 2.5, details on the fit in the text.

assumptions on the scaling form of  $p(n, m|A)$  (Eq. (1.24) in Chapter 1) cannot be verified directly on empirical data, due to the practical difficulty in gathering the necessary type and amount of data in the field, verifying its consequences on the linkage of scaling exponent is the best available test of the theory. While single ecological laws have often been measured, datasets allowing the simultaneous measurement of several scaling exponents are rare. This limits the possibilities to test the framework. Additionally, the measurement of the exponents  $\nu$  and  $\mu$  in Eqs (2.3c) and (2.3d) would be technically very difficult, as it would require counting the number of individuals present in the whole ecosystem area, and their total biomass, for several ecosystems of different areas, preventing the verification of these two laws. As an outlook of the present work, the hope is to guide future empirical efforts in the direction of more comprehensive observations of ecosystems, allowing the simultaneous measurement of several exponents. Note that exponents describing scalings with ecosystem area ( $z, \nu, \mu, \xi, \Phi$ ) must be estimated comparing entire ecosystems of different areas, and not sub-patches of the same ecosystem, as the scaling might differ in the two cases. Consider for example the SAR. In the limit of a well-mixed ecosystem, where there is no species clustering, a large-enough sub-patch of the ecosystem would already contain as many species as the entire ecosystem. On the contrary, a smaller ecosystem of the same size of this sub-patch would contain less species, as the evolutionary process would have produced less diversification in this smaller ecosystem, with a smaller number of individuals.

The knowledge of linking relationships among the different patterns could also guide the interpretation of empirical observations. As an example, the equation  $\eta = \gamma + \delta$  tells us that a change in the slope of the size-spectrum (observed, e.g., in marine microbial communities with different resource supply rates [Cavender-Bares et al., 2001, Marañón, 2015]) can be due either to a change in species composition, affecting the slope  $\delta$  of the species size distribution, or to a modified balance in the abundance of large/small species, affecting the slope  $\gamma$  of Damuth's law, or both. Therefore, a mechanistic explanation of the observed size-spectrum and of the changes in its slope should also reproduce the observed behavior of the other two patterns.

In conclusion, the set of linking relationships resulting from the theoretical framework proposed in this thesis and supported by the available empirical data are intended as a guide for the coherent interpretation and rationalization of empirical observations from different settings and scales.

## Chapter 2. Ecological scaling laws in macroecological data

Table 2.1 – References to empirical measurements of the ecological patterns discussed in the main text. RSA stands for Relative Species Abundance and Max. body mass stands for the scaling of the maximum body mass with the area of the ecosystem.

Law	Forests	Terrestrial	Aquatic
Kleiber's law	Mori et al. [2010]	Kleiber [1932] Dodds et al. [2001]	Nielsen and Sand-Jensen [1990] Finkel et al. [2004] Marañón et al. [2007]
SAR	Lomolino [1982]	MacArthur and Wilson [1963] Newmark [1986] Okie and Brown [2009] Preston [1962]	Dodson [1992] Lonsdale [1999] Smith et al. [2005]
Damuth's law	Cohen et al. [2012]	Damuth [1981] Nee et al. [1991] Novosolov et al. [2015]	Cyr et al. [1997] Cohen et al. [2003]
$s(m)$	Muller-Landau et al. [2006] Stegen and White [2008] Condit et al. [2012]	White et al. [2007] Halfpenny [2018]	Sheldon et al. [1972] Rodriguez and Mullin [1986] Cavender-Bares et al. [2001] Rinaldo et al. [2002] Quinones et al. [2003] Marañón [2015]
$p(m)$		Marquet and Taper [1998] Smith et al. [2003] Marquet et al. [2005] Southwood et al. [2006]	
Max. body mass		Burness et al. [2001] Okie and Brown [2009]	
RSA	Zillio et al. [2008]	Preston [1948]	Ser-Giacomi et al. [2018]
Taylor's law	Giometto et al. [2015]	Taylor [1961] Taylor et al. [1980] Anderson et al. [1982]	

Table 2.2 – Estimates for the scaling exponents of the ecological patterns depicted in Fig. 2.2. Errors are SEM, CI stands for confidence interval. If no error is reported, none was given in the original paper.

Law	Forests	Terrestrial	Aquatic
Kleiber's law (panel a)	0.80 ± 0.01	0.67 ± 0.2	1.10, CI 95%: [0.94, 1.21]
Damuth's law (panel b)	0.26 ± 0.05	0.57 ± 0.08	0.73, CI 95%: [0.73, 0.92]
SAR (panel c)	0.27 ± 0.01	0.23 ± 0.02	0.094
$s(m)$	1.59, CI 95%: [1.57, 1.63]	1.5 ± 0.2	2.11

Table 2.3 – Estimates of scaling exponents  $\eta$ ,  $\delta$  and  $\gamma$  in the BCI forest. Errors are computed as reported in the text.

BCI forest	Census 5	Census 6	Census 7
$s(m)$	$\eta = 1.43 \pm 0.04$	$\eta = 1.43 \pm 0.03$	$\eta = 1.44 \pm 0.04$
$p(m)$	$\delta = 1.03 \pm 0.03$	$\delta = 1.05 \pm 0.03$	$\delta = 1.07 \pm 0.05$
Damuth's law	$\gamma = 0.41 \pm 0.07$	$\gamma = 0.40 \pm 0.06$	$\gamma = 0.38 \pm 0.06$

Table 2.4 – Estimates of scaling exponents  $\eta$ ,  $\delta$  and  $\gamma$  in the Luquillo forest [Zimmerman et al., 2010]. Errors are computed as reported in the text.

Luquillo forest	Census 1	Census 2	Census 3	Census 4	Census 5
$s(m)$	$\eta = 1.27 \pm 0.04$	$\eta = 1.18 \pm 0.03$	$\eta = 1.09 \pm 0.03$	$\eta = 1.09 \pm 0.04$	$\eta = 0.95 \pm 0.07$
$p(m)$	$\delta = 1.02 \pm 0.02$	$\delta = 1.01 \pm 0.03$	$\delta = 1.02 \pm 0.05$	$\delta = 1.02 \pm 0.03$	$\delta = 1.02 \pm 0.01$
Damuth's law	$\gamma = 0.21 \pm 0.08$	$\gamma = 0.23 \pm 0.04$	$\gamma = 0.16 \pm 0.04$	$\gamma = 0.09 \pm 0.06$	$\gamma = 0.03 \pm 0.04$

Table 2.5 – Estimates of scaling exponents  $z$ ,  $\delta$ ,  $\Phi$  and  $\gamma$  for the LIZ dataset [Novosolov et al., 2015]. Errors on  $z$  and  $\delta$  are SEM, the error on  $\gamma$  is the SD, the error on  $\Phi$  was obtained by bootstrapping.

SAR	$z = 0.17 \pm 0.01$	$R^2 = 0.46$
Damuth's law	$\Phi = 0.78 \pm 0.08$	
Damuth's law	$\gamma = 0.53 \pm 0.03$	$R^2 = 0.89$
$p(m)$	$\delta = 1.45 \pm 0.06$	

Table 2.6 – Estimates of scaling exponents  $z$ ,  $\xi$  and  $\delta$  for the SSI dataset [Okie and Brown, 2009]. Errors on  $z$ ,  $\delta$  and  $\xi$  are SEM.  $R^2$  is the coefficient of determination.

SAR	$z = 0.23 \pm 0.02$	$R^2 = 0.93$
$M_{\max}$	$\xi = 0.49 \pm 0.09$	$R^2 = 0.76$
$p(m)$	$\delta = 1.6 \pm 0.2$	

Table 2.7 – Summary of the empirical tests performed. References are to equations in the text.

Dataset	Measured exponents	Relationship that was verified
BCI	$\eta, \gamma, \delta$ (Table 2.3)	Eq. (2.3a): $\eta = \delta + \gamma$
Luquillo	$\eta, \gamma, \delta$ (Table 2.4)	Eq.(2.3a): $\eta = \delta + \gamma$
LIZ	$z, \Phi, \gamma, \delta$ (Table 2.5)	Eq.(2.3b): $z = 1 - \Phi - \max\{0, \lambda(1 + \alpha - \gamma - \delta)\}$
SSI	$z, \xi, \delta$ (Table 2.6)	Eq.(2.3e): $\xi = \frac{z}{\delta - 1}$



### 3 Fluctuations scaling of the species-area relation

*The species–area relationship (SAR) is under intense scrutiny in contemporary ecology to probe its reliability in predicting the numbers of species going extinct as a direct result of habitat loss. Here, SARs are considered as portraying average trends in ecosystems, around which fluctuations are expected due to the intrinsic stochasticity of ecosystem dynamics and to external perturbations. This probabilistic interpretation of SARs, though implicit in the original theory of island biogeography, has not been equally well studied and poses significant theoretical and practical problems. In this Chapter, a suite of theoretical models of ecosystem dynamics is explored where the number of species  $S$  emerges from diverse ecological and evolutionary processes, and stationary predictions for the scaling with the mean of the relative fluctuations of  $S$ , i.e its coefficient of variation, are compared. Different models are found to diverge radically in their predictions. In island biogeography and in neutral frameworks, where fluctuations are only driven by the stochasticity of diversification and extinction events, relative fluctuations decay when the mean increases, therefore the average value of  $S$  is representative of point measurements for ecosystems with large enough  $S$ . For small systems, relative fluctuations remain non-negligible, and point measurements may deviate significantly from the SAR describing the mean trend. Computational evidence suggests that such result is robust in the presence of competition for space or resources. When species compete for finite resources, and mass is introduced as a trait determining species' resource consumption, relative fluctuations emerge whose size does not decay with increasing system size due to the occasional arrival of new species with large resource demand. The new arrivals, by immigration or speciation, prompt extinctions in the community. Habitat stochasticity, in the context of metapopulation models, causes small variability in the number of coexisting species, which proves negligible with respect to that caused by stochasticity in community dynamics. The effect of generic external perturbations in neutral dynamics is also investigated. If the number of species going extinct during perturbations increases linearly with  $S$ , the predicted stationary average of  $S$  becomes progressively less informative about the system state. This study shows that modeling does not provide a unique description of the scaling with the average of the relative fluctuations of  $S$ . Therefore, empirical investigations should sort out the fluctuation scaling scenario that applies to their specific context, thus pointing at the modeling approach that includes the processes controlling community assembly.*<sup>1</sup>

<sup>1</sup>This chapter features contents from [Zaoli et al., 2018] (in review).

### 3.1 Introduction

Among the many regular patterns discussed in Chapters 1 and 2, emerging across scales of space, time and organizational complexity, one of the most well-known and studied is the Species–Area Relationship (SAR). The SAR quantifies the observation that species richness  $S$  tends to increase with increasing sampling area  $A$ , a relationship firmly placed at the origins of quantitative ecology [Arrhenius, 1921, Gleason, 1922, MacArthur and Wilson, 1967]. The SAR formulation most supported by data states that the number of species  $S$  inhabiting an ecosystem increases as a power of its area,  $A$ , such that  $S = cA^z$ , where  $c$  is a constant and  $z \leq 1$  is the SAR’s scaling exponent.

A distinction should be made between types of SARs under study, which radically differ in the way area and species sampling are chosen [Drakare et al., 2006] thus reflecting different measures of biodiversity. Notably, ‘island’ SARs are obtained by counting species inhabiting disjoint, isolated patches in the same biogeographical region (e.g. islands, lakes, mountain tops or any set of areas separated by environmental barriers) [MacArthur and Wilson, 1967]. They stem from the eco-evolutionary dynamics shaping communities on ecological and evolutionary timescales. For ‘island’ SARs, field evidence points at power-law functional relationships as the most explanatory [Triantis et al., 2012]. Nested SARs are obtained by considering sub-patches of different size within a larger, encompassing domain [Harte et al., 2009]. Compared to island SARs, they are affected by the spatial distribution of individuals (say, the degree of species clustering) within the domain. This Chapter is focused on island SARs.

The interest in SARs and in the value of their exponent has been broad from all of ecology, in particular for their implied predictive use to forecast the effects of large-scale environmental or climatic change on biodiversity [Thomas et al., 2004], in particular as a consequence of habitat loss or fragmentation [Durrett and Levin, 1996, Hanski, 2013, Borile et al., 2013]. In fact, within a deterministic power-law framework, the fraction of species surviving a habitat reduction from area  $A$  to area  $A_{new}$  would simply be equal to  $(A_{new}/A)^z$ . Note that only the scaling exponent  $z$  matters for such a prediction (and not the proportionality constant), explaining the interest in the exponent’s value, see e.g. [Rybicki and Hanski, 2013]. Such estimate certainly neglects some important factors which have an impact on the number of surviving species, e.g. the spatial configuration of the remaining habitat [Rybicki and Hanski, 2013] and the temporal dynamics of extinctions [Pimm and Raven, 1995]. However, SARs still provide an order-of-magnitude estimate of species loss and can be used as a base on which to build more specific estimates. Against this background, assessing theoretically the nature of SARs and the reliability of their predictions is an issue of longstanding interest to community ecology, biogeography, and macroecology.

A factor seemingly overlooked both in the modeling and in the interpretation of SARs concerns the fluctuations of the number of species  $S$  that the ‘law’ is meant to address. Deterministic approaches, in fact, only address the average biodiversity, neglecting any fluctuations due to



the intrinsic stochasticity of the processes taking place in ecosystems and to external sources of variability. In models of ecosystem dynamics, however, the number of species  $S$  in stationary conditions results from the balance of competing stochastic processes, which according to neutral theory [Hubbell, 2001] are speciation, immigration and drift. Therefore, the number of species  $S$ , as well as other variables describing the state of the ecosystem, should be seen as stochastic variables and thus subject to fluctuations. If it were possible to capture its evolution in time within one single ecosystem, or compare it in exact replicas of the same ecosystem with the same area (assuming ergodicity of the process),  $S$  would assume different values according to its stationary distribution  $p(S, t \rightarrow \infty) = p(S)$ . Past theoretical work [MacArthur and Wilson, 1963, 1967, Rosenzweig, 1995] focused mainly on the deterministic SAR emerging for its average value  $\langle S \rangle$  by determining its functional dependence on ecosystem area  $A$  and related coefficients [Durrett and Levin, 1996, Lomolino, 2000, Azaele et al., 2016], and studying its behavior following changes in speciation rate, typical dispersal length and system size [Pigolotti and Cencini, 2009, Cencini et al., 2012, Shem-Tov et al., 2017]. One exception has been provided by Diamond and Gilpin [1980], who tackled the issue of the scaling of the relative fluctuations of  $S$  with ecosystem size  $A$  in the case of the equilibrium theory of island biogeography [MacArthur and Wilson, 1967].

While the interest in studying average behaviours is not questionable, fluctuations of  $S$  may not be negligible with respect to their mean (Figure 3.1) and thus they must be considered as an important component of the ecosystem's dynamics. Scatter in measured SARs is usually disregarded because either related to measuring errors or to factors impossible to quantify. However, the residual noise attributable to ecological dynamics (defined 'turnover noise' by Diamond and Gilpin [1980]) reflects a theoretically interesting problem. In fact, there would exist cases where, depending on prevailing ecosystem's dynamics, the instantaneous value of  $S$ , the target of any field measurements, may differ significantly from its expected value. Technically, the amplitude of the relative fluctuations of the (random) number of species  $S$  is described by its coefficient of variation:

$$\frac{\Delta S}{\langle S \rangle} = \frac{\sqrt{\langle (S - \langle S \rangle)^2 \rangle}}{\langle S \rangle} = \frac{\sqrt{\text{var}(S)}}{\langle S \rangle}, \quad (3.1)$$

When relative fluctuations decay with the mean,  $S$  can be considered deterministic for ecosystems with large enough  $\langle S \rangle$  because observable values lie within a negligible distance from the mean. Whether (and why) relative fluctuations do not decay is an interesting question. In fact, in the latter scenario, when using a deterministic SAR one should be aware of its reduced predictive power due to the potential large distance of observations from the average. Note that theoretically the scaling of the variance with the mean of a random variable has been often found to conform to a power-law with exponent between 1 and 2 [Taylor, 1961, Cohen, 2014, Giometto et al., 2015],  $\text{var}(S) \propto \langle S \rangle^\beta$ , an observation that is often referred to as 'Taylor's law' or 'fluctuation scaling'. The value of the exponent  $\beta$  discriminates over the behavior of

the relative fluctuations (Figure 3.1), because:

$$\frac{\Delta S}{\langle S \rangle} \propto \langle S \rangle^{\frac{\beta}{2}-1}. \quad (3.2)$$

If  $\beta \geq 2$ , relative fluctuations do not decay for increasing  $\langle S \rangle$ . As an example, for a Poisson random variable  $X$ ,  $\text{var}(X) = \langle X^2 \rangle - \langle X \rangle^2 = \langle X \rangle$ , therefore  $\beta = 1$  and the relative fluctuations vanish in the limit  $\langle X \rangle \rightarrow \infty$ . On the other hand, for a random variable  $X$  distributed as a power-law with exponent  $\alpha \in (1, 2)$ , the sample mean and variance yield  $\beta = (3 - \alpha)/(2 - \alpha) \geq 2$ , i.e. relative fluctuations diverge in the  $\langle X \rangle \rightarrow \infty$  limit. The value of  $\beta$  is therefore a useful quantity to characterize the fluctuations of  $S$ , and classify different models of ecosystem dynamics as done below. Note that increasing ecosystem size  $A$  corresponds to increasing the average biodiversity  $\langle S \rangle$  when the SAR is monotonically growing, which is almost always found empirically, in particular for island SARs. Therefore, when relative fluctuations decay with increasing  $\langle S \rangle$ , one can equivalently state that they decay with increasing ecosystem size.

Here, the statistics of  $S$  are examined in a number of community dynamics models which implement distinct ecological processes possibly influencing the  $S$  dynamics in real ecosystems. The scaling of the relative fluctuations of  $S$  with the mean is characterized to determine how their behavior changes in response to the introduction of different ecological processes. Exact results, reviewed in section 3.2, exist for the simplest models, i.e. the model of MacArthur and Wilson and the non-interacting neutral model, offering a natural starting point for the study of the behavior of fluctuations. Species competition will be investigated and shown, under certain circumstances, to be a cause of important perturbations to the system, leading to relative fluctuations which do not decay with the mean. Furthermore, the effects of different types of external perturbations on the system will be examined, addressing the effects of their size and frequency in relation to system size and recovery rate (due to speciation/immigration). The scope of such model comparison is to further our understanding of the possible statistical properties of  $S$  which should be expected in field observations.

### **3.2 Species richness from minimalist models of community dynamics**

The equilibrium theory of island biogeography [MacArthur and Wilson, 1963] has been the first attempt to explain mechanistically SARs. Even in its original formulation, it has theoretical significance to the problem at hand because the stationary state can be understood in probabilistic terms [Diamond and Gilpin, 1980]. In this theory, the stationary number of species emerges from the balance between immigration from a species pool, containing  $P$  species, and extinction, with rates  $I(S)$  and  $E(S)$ , respectively. In this approach, the variance of  $S$  is exactly [MacArthur and Wilson, 1967, Diamond and Gilpin, 1980]:

$$\text{var}(S) = \frac{I(\langle S \rangle)}{dE/dS - dI/dS}, \quad (3.3)$$

### 3.2. Species richness from minimalist models of community dynamics

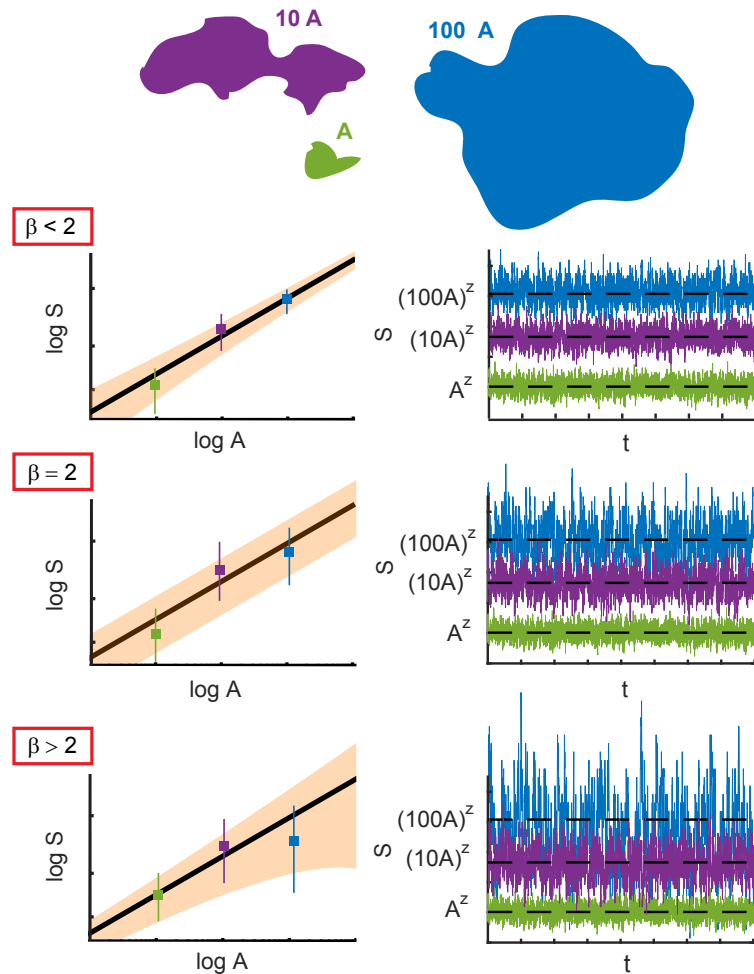


Figure 3.1 – A conceptual scheme highlighting the probabilistic structure of SARs: (top) Three equivalent (in terms of environmental conditions), independent ecosystems of area  $A$ ,  $10A$ ,  $10^2 A$  are sketched. The ecosystems are meant to comply with the stipulations of the theory of island biogeography; (bottom) Three SAR scenarios are shown, differing for the value of  $\beta$ , respectively by plotting (left) the mean number of species with its confidence interval (the expected one and an example of the outcome of an hypothetical measurement for each of the three areas). The plot, in double logarithmic scale, is based on a power-law SAR with slope equal to  $z$ . (right) A sketch of what a time-series of measured total number of species would look like in the three test ecosystems, should a perfect replicated field study be staged therein. The parameter  $\beta$  subsumes the properties of the scaling of the variance of the random variable  $S$ , number of observable species within an ecosystem of area  $A$ , with its mean  $\langle S \rangle$  (Taylor's law). From top to bottom, the case of the Taylor's law exponent  $\beta < 2$  would support a deterministic validity of the SAR in some large area  $A$  limit. A notable exception is the case  $\beta = 2$  for which the variance scales with the mean, and thus fluctuations cannot be neglected whatever the size of the ecosystem. Fluctuations grow instead without bounds in the last case  $\beta > 2$  and in such cases deterministic SARs are not particularly informative.

### Chapter 3. Fluctuations scaling of the species-area relation

---

where the average number of species  $\langle S \rangle$  is obtained from the condition

$$I(\langle S \rangle) = E(\langle S \rangle). \quad (3.4)$$

Diamond and Gilpin [1980] showed that, for a specific choice of  $I(S)$  and  $E(S)$ , at stationarity  $\text{var}(S)$  increases linearly with  $\langle S \rangle$  ( $\beta = 1$ ) when  $\langle S \rangle \ll P$ , reaches a maximum around  $\langle S \rangle = P/2$  and then decays to zero as  $\langle S \rangle$  approaches  $P$ . Such result implies, therefore, that relative fluctuations decay as  $1/\sqrt{\langle S \rangle}$  when  $\langle S \rangle \ll P$  and even faster afterwards. This behavior is related to the lack of speciation in the model, limiting the maximum number of species to the number of species in the pool.

Such predictions can be compared with the ones of a neutral model where new species are introduced by speciation and immigration without a fixed upper limit on  $S$ . Consider a community dynamics where new species appear as a point Poisson process with rate  $\lambda$  and are characterized by a birth rate  $b$  and a death rate  $d$  that are identical for all species, i.e. the model is neutral. For this model with non-interacting species, the number of species  $S$  at stationarity is a Poisson random variable [Suweis et al., 2012b] with mean

$$\langle S \rangle = \lambda \langle \tau \rangle, \quad (3.5)$$

where  $\langle \tau \rangle$  is the average lifetime of a species, also called persistence time. An alternative formulation of this model is the following: the arrival of new species in the system is a Poisson process of rate  $\lambda$  and their lifetimes  $\tau$  are i.i.d. variables with mean  $\langle \tau \rangle$ . With this formulation, the fact that  $S$  is Poisson-distributed with mean given by Eq. (3.5) corresponds to a known result of queuing theory [Benes, V, 1965]. As a consequence of  $p(S)$  being a Poisson distribution, one has  $\beta = 1$ . Therefore, relative fluctuations decay with the mean and  $S$  is effectively deterministic in the limit of large  $S$ . The same result can be proved for the mean-field multi-species Voter model [Suweis et al., 2012b], where species interact through the competition for space, in the large  $A$  limit (when interactions become negligible). Note that, with the introduction of space,  $\langle \tau \rangle$  becomes a function of the ecosystem area  $A$ . Therefore, Eq. (3.5) provides a link between persistence times and the SAR [Bertuzzo et al., 2011].

The introduction of more realistic features such as a finite dispersal length, non-negligible interactions or the relaxation of the neutral assumption prevent exact results to be obtained for  $p(S)$ . In fact, interactions make persistence times  $\tau$  not independent, while introducing a trait differentiating species, e.g. mass, makes the persistence times  $\tau$  not identically distributed across species and the above results do not necessarily apply. Therefore, the statistical properties of  $S$  in a variety of stochastic models of ecosystem dynamics were explored numerically, varying the assumption of neutrality and the presence or absence of competition. The stationary distributions  $p(S)$  was studied in different generalizations of the simple neutral model of non-interacting species presented above, comparing different mechanisms of mean-field competition for resources (section 3.3), neutral and non-neutral models and different speciation mechanisms (section 3.4). For each model and each of the simulated system sizes  $A$ , the

following quantities were computed:  $p(S|A)$  (i.e., the probability that in an ecosystem of size  $A$ , at stationarity,  $S$  species coexist), its average  $\langle S \rangle$ , its variance  $\text{var}(S) = \langle S^2 \rangle - \langle S \rangle^2$  and the relative fluctuations  $\Delta S / \langle S \rangle$ ,  $p(\tau|A)$  (i.e., the probability that a species in an ecosystem of area  $A$  has a persistence time  $\tau$ ) and its average  $\langle \tau \rangle$ . In addition, the effect of habitat stochasticity was also explored by considering a metapopulation model [Rybicki and Hanski, 2013], where the biodiversity level is determined by the habitat diversity present in the landscape, allowing species with different niches to survive (section 3.6). Also for this model, for each simulated system size  $A$ ,  $p(S|A)$  was computed under randomization of the landscape, along with its average, variance and relative fluctuations.

### 3.3 Neutral models with competition

Two different ways of introducing competition for a shared resource, i.e. interaction among species, are considered: a spatially explicit one in section 3.3.1 and a spatially implicit one in section 3.3.2.

#### 3.3.1 Multi-species Voter model (MSV)

In the multi-species Voter model (MSV) with speciation and nearest-neighbors dispersal, each individual occupies a node of an  $L \times L$  lattice and  $N = L \times L$  is the total number of individuals, therefore space is the common resource. At each time step, one individual chosen at random dies and is replaced with probability  $\lambda$  by an individual of a new species or with probability  $1 - \lambda$  by an offspring of one of its nearest neighbors. The probability  $\lambda$  should be interpreted as the sum of the probability of speciation and of the probability of immigration of a new species from outside the lattice, and will be called ‘diversification rate’ hereafter. While for the well-mixed case (or equivalently for infinite dispersal length), in the limit  $A \rightarrow \infty$ , a parallel can be established with the non-interacting case [Suweis et al., 2012b] allowing to extend the analytical results, no analytical results are known when the dispersal length is finite. A semi-analytical approach gives an approximation for  $\langle S \rangle$  [Shem-Tov et al., 2017], but not for  $p(S)$ . The MSV model was simulated on periodic lattices of size  $A = L \times L$  with  $L = 50, 100, 200$  and diversification rate  $\lambda = 10^{-3}$ . Results are expected to be robust with respect to the diversification rate  $\lambda$ . To compute  $p(S)$  for the different areas, 1000 independent realizations of  $S$  at the stationary state were obtained for each value of  $A$  by simulating the model with the coalescent method [Etienne and Olf, 2004], taking advantage of the dual representation of the Voter model [Holley and Liggett, 1975]. This method assures that the realizations of  $S$  are independent. Since persistence times cannot be measured with the coalescent method, one forward simulation of the model was also performed for each value of  $A$ . Species’ persistence times were measured as the interval between the appearance of a species and its extinction. The simulations were continued until the estimate of  $\langle \tau \rangle$  did not increase further with simulation length.

### 3.3.2 Neutral-Spatially Implicit model (N-SI)

Competition for resources can also be introduced in a spatially-implicit model, in which a space-dependent resource constraint is implemented in the birth and death rates, as in the community dynamic model used in Chapter 1. In this section, a neutral version of that model is considered. Let  $A$  be the area of the ecosystem, and  $\mathcal{R} \propto A$  the resources supply rate. Individual births and deaths are Poisson-distributed events. An individual of a species with abundance  $n_i$  is born or dies with rates, respectively:

$$\begin{aligned} u_i &= n_i, \\ v_i &= \left[ v_0 + (1 - v_0) c \frac{N(t)}{\mathcal{R}} \right] n_i, \end{aligned} \tag{3.6}$$

where  $c$  is the resource consumption rate of an individual and the per-capita birth rate for an individual of unit mass is taken equal to one without loss of generality. A proportionality constant  $u_0$  between  $u_i$  and  $n_i$  would simply correspond to re-scaling time by  $1/u_0$ . When the resource consumption of the community  $cN(t)$  is equal to the resource supply rate  $\mathcal{R}$ , the birth rate is equal to the death rate,  $u = v$ , and the community has no net growth (this is considered to be stationarity). However,  $N(t)$  will continue to fluctuate with time. Diversification is implemented as a Poisson process with rate  $\lambda$ . At each diversification event, a species is chosen at random and a random fraction of individuals from such species is assigned to a new species (*fission speciation*). In this model, space is introduced implicitly via the assumption  $\mathcal{R} \propto A$ , which determines the number of individuals that the ecosystem can sustain. The model was simulated for  $A = 10^i$  with  $i = 2, 3, 4, 5$ ,  $\lambda = 10^{-4}$ ,  $v_0 = 1/2$  and  $c = 10^{-4}$ . Also in this case, the results are expected to be robust with respect to a broad range of parameters' variations. The stationary state was considered attained when the number of species  $S$  did not show a net change in time, but only fluctuations around a stationary mean value. To compute  $p(S)$  at the stationary state, the number of species  $S$  was sampled with frequency  $f = 10^{-5}$ . Such a low sampling frequency guarantees the independence of the sampled  $S$  values. Persistence times were measured as explained in section 3.3.1. In the following, this model is referred to as N-SI (neutral-spatially implicit). Differently from the MSV model, where each individual always occupies one node, i.e. consumes a fixed amount of resources, here the neutral assumption for the resource consumption rate can easily be relaxed (see section 3.4).

### 3.3.3 Comparison with the neutral model without competition

Numerical results suggest that the exact results valid in the non-interacting case (i.e.,  $S$  being a Poisson variable with mean given by Eq. (3.5)) are still valid for the two models with competition in the large area limit. Figs. 3.2 and 3.3 show  $p(S)$  obtained from simulations of, respectively, the the MSV and the N-SI model for different values of  $A$  and the corresponding theoretical predictions for the non-interacting case, i.e. Poisson distributions with mean  $\lambda \langle \tau \rangle$ . The results match closely the predictions and, as expected for a Poisson random variable,  $\text{var}(S)$  scales as a power of  $\langle S \rangle$  with an exponent  $\beta$  compatible with 1, i.e  $S$  is asymptotically de-

terministic. The estimates of  $\beta$  for the two models are, respectively  $\beta = 0.985 \pm 0.005$  ( $R^2 = 1.0$ ) and  $\beta = 1.08 \pm 0.06$  ( $R^2 = 0.99$ ). Fig. 3.4A displays the time-evolution of  $S$  for two different areas in the N-SI model (a similar dynamics is found for the MSV model). The average number of species is found to scale as a power-law of  $A$ . The values of  $z$  estimated by linear least square fitting of  $(\log A, \log S)$  for the MSV and N-SI models are, respectively,  $z = 0.998 \pm 0.002$  ( $R^2 = 1.0$ ) and  $z = 0.496 \pm 0.001$  ( $R^2 = 0.99$ ). Note that the value  $z = 1$  for the MSV model is expected from the results in [Shem-Tov et al., 2017] to hold independently of  $\lambda$ . Some considerations regarding the distribution of persistence times in the two models are presented in Appendix 3.8.1.

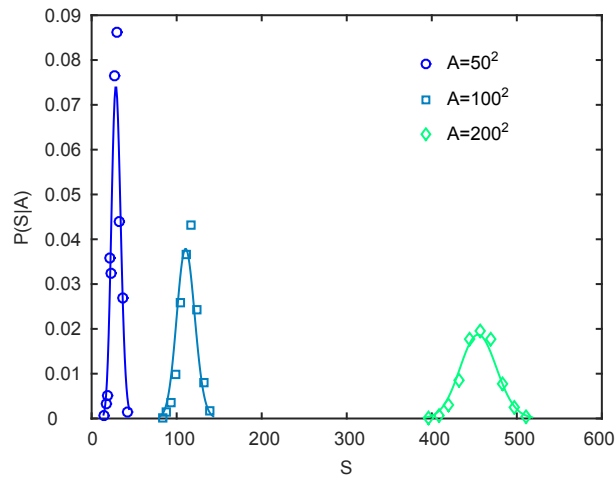


Figure 3.2 –  $p(S|A)$  for the MSV model with  $\lambda = 10^{-3}$  and  $A = L^2$  with  $L = 50, 100, 200$  obtained from 1000 simulations of the model with the coalescent method. Continuous lines are Poisson distribution with mean  $\lambda\langle\tau\rangle$ , where  $\langle\tau\rangle$  was obtained from a forward model simulation.

### 3.4 Non-neutral models

The neutral assumption can be relaxed by distinguishing different species via specific traits. Here, mass is considered as the master species' trait, given its importance in determining a species' physiology and ecology [Kleiber, 1932, Brown et al., 2004, Giometto et al., 2013, Marañón, 2015]. Two variants of the N-SI model are studied, implementing empirically observed allometric relationships for the dependence of vital rates on individuals' body sizes [Brown et al., 2004] and resource consumption rates [Kleiber, 1932]. Models of this class are of interest as they are capable of reproducing the empirically observed scaling laws related to abundance and size (see Chapter 1, section 1.5). In these models, an individual of a species

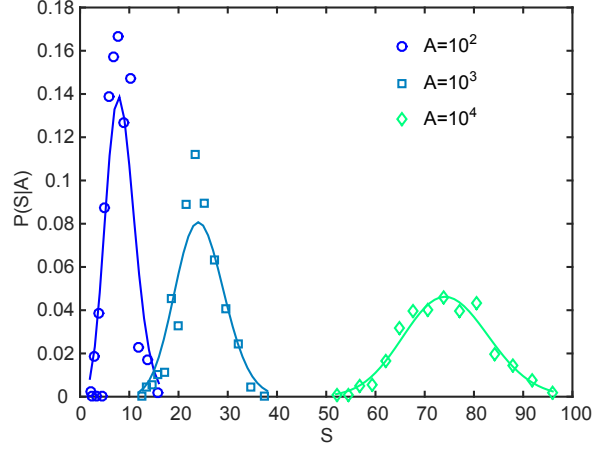


Figure 3.3 –  $p(S|A)$  for the N-SI model. Colors refer to different values of  $A$ . Continuous lines are Poisson distributions with mean  $\lambda\langle\tau\rangle(A)$ .

with abundance  $n_i$  and body mass  $m_i$  is born or dies with rates, respectively:

$$\begin{aligned}
 u_i &= n_i m_i^{-1/4}, \\
 v_i &= \left[ v_0 + (1 - v_0) c \frac{\sum_j n_j m_j^{3/4}}{\mathcal{R}} \right] n_i m_i^{-1/4},
 \end{aligned}
 \tag{3.7}$$

where  $c$  is the resource consumption rate of an individual with unit mass and the per-capita birth rate for an individual of unit mass is taken equal to one without loss of generality. The exponent values  $-1/4$  and  $-3/4$  are the typical values for the corresponding allometries and it was shown in section 1.5 that they affect the scaling exponents of macroecological patterns, but not their functional form and covariations. Similarly to the corresponding neutral model (N-SI), the community net grow is zero when the community consumes all the available resources, i.e. when  $c \sum_j n_j m_j^{3/4} = \mathcal{R} \propto A$ . The two variants investigated here differ in the diversification mechanism: in model M1, diversification is implemented as a Poisson process with rate  $\lambda$ . This model corresponds to the basic model in section 1.5.1, but with no constraint on  $S^2$ . In model M2, the overall rate of diversification  $\lambda$  depends on the area of the ecosystem as  $\lambda = \lambda_0 A^{-\xi}$ . This choice accounts for the finding of Bertuzzo et al. [2011], which showed that the species diversification rate scales with the area as  $\lambda \propto A^{-\xi}$  with  $\xi = 0.84$ , a result that was obtained by contrasting the prediction of a neutral model for species persistence times,  $p(t) \propto t^{-\alpha} e^{-k\lambda t}$  with constant  $k$ , with empirical data on breeding birds and herbaceous plants [Suweis et al., 2012a]. This result agrees with the interpretation of  $\lambda$  as the sum of the rate of immigration and that of speciation, as the arrival of new species by immigration is expected to diminish as the ecosystem area increases [Chisholm and Lichstein, 2009].

<sup>2</sup>The results presented here concern the case where overall speciation is a Poisson process with rate  $\lambda$ , but the case where species' specific speciation has rate  $\lambda$  produces similar results.



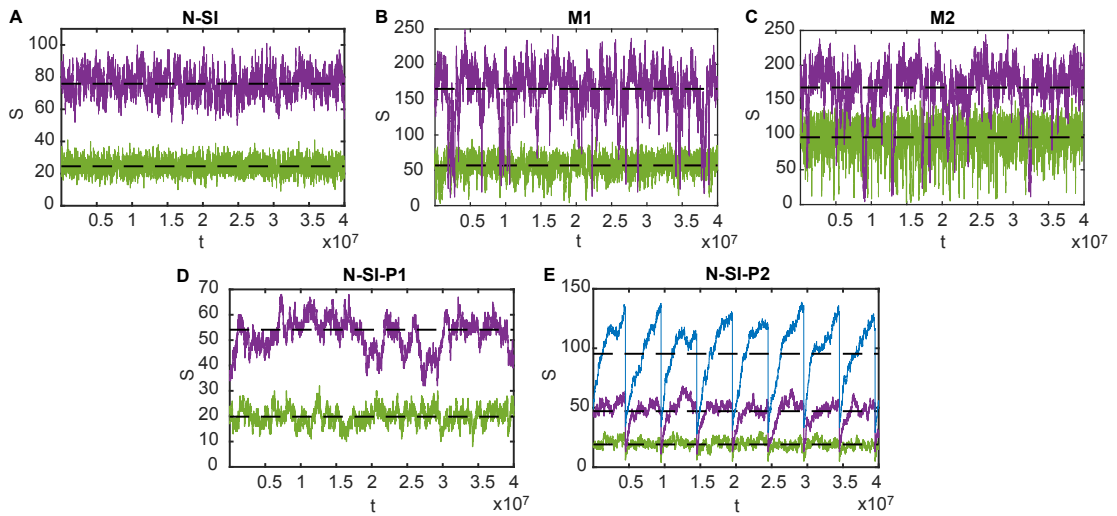


Figure 3.4 – Dynamics of the number of species  $S$  in the models presented in the text. A) Model N-SI with  $\lambda = 10^{-4}$ , areas  $10^3$  (green) and  $10^4$  (purple); B) model M1 with  $\lambda = 10^{-2}$ , areas  $10^3$  (green) and  $10^4$  (purple); C) model M2 with  $\lambda = A^{-0.5}$ , areas  $10^3$  (green) and  $10^4$  (purple); D) model N-SI-P1 with  $\lambda = 10^{-4}$ , areas  $10^3$  (green) and  $10^4$  (purple); E) model N-SI-P2 with  $\lambda = 10^{-4}$ , areas  $10^3$  (green),  $10^4$  (purple) and  $10^5$  (blue). Note that the decrease in the number of species at fixed frequencies is due to the specific perturbations that we added by design to the neutral dynamics (see text). Black dotted lines mark the average  $S$ . The metapopulation dynamics model is not shown because its temporal dynamics is not explicitly simulated to compute the statistics of  $S$ .

### Chapter 3. Fluctuations scaling of the species-area relation

---

In models M1 and M2, diversification is implemented via a fission mechanism as in model N-SI, and the evolution of mass due to a speciation event is the same described in section 1.5. The descendant species  $j$  conserves, on average, the mass  $m_i$  of the parent species  $i$ , being obtained as the product of  $m_i$  and a random factor with mean 1. More precisely,  $m_j = \max\{m_0; q \cdot m_i\}$  where  $q$  is extracted from a lognormal distribution with mean and variance equal to unity. The maximum in the expression for  $m_j$  introduces a bound on the minimum mass  $m_0$  that a species can attain. The mass of the parent species is left unchanged. The distribution of species' masses in the system is determined by the combination of this multiplicative bounded process, known to produce power-laws (as was shown in Box 1.5.2), and the birth/death dynamics. Models M1 and M2 were simulated for  $A = 10^i$  with  $i = 1.5, 2, 2.5, 3, 3.5, 4$  and parameters values  $c = 10^{-3}$ ,  $m_0 = 1$ . The diversification rate was set to  $\lambda = 10^{-2}$  for M1, while for M2  $\lambda_0 = 1$  and  $\xi = 0.5$  were used. The number of species and the persistence times were measured as for the N-SI model.

The dynamics of the random variable  $S$  in the two models, displayed in Fig. 3.4B and C, shows strikingly different features in comparison with the neutral model. In particular, the ratio  $\Delta S / \langle S \rangle$  does not decrease with  $A$ . The species' number distributions, shown in Figs. 3.5A and 3.6A, deviate strongly from a Poisson distribution, with fluctuations quantified, respectively, by  $\beta = 1.88 \pm 0.05$  and  $\beta = 1.93 \pm 0.14$ . Rather than being Poisson distributed, species number distributions in models M1 and M2 are described by the following finite-size scaling form:

$$p(S|A) = S^{-\sigma} \mathcal{F}\left(\frac{S}{A^z}\right) \quad (3.8)$$

where  $\mathcal{F}$  is a function such that  $\mathcal{F}(x) \rightarrow 0$  for  $x \rightarrow 0, \infty$ . Normalization requires that  $\sigma = 1$  (see Box 1.2). The validity of Eq. (3.8) is shown by plotting, in Fig. 3.5B and 3.6B the curves  $Sp(S)$  versus  $\frac{S}{A^z}$  for different values of  $A$  and noting that they all collapse. Note that an attempt to collapse the distributions obtained with the N-SI model according to Eq. (3.8) fails (Fig. 3.7), due to the different behavior of the fluctuations of  $S$ . The visual comparison, in Fig. 3.8, of Poisson distributions and lognormal distributions of the same mean, which satisfy Eq. (3.8), clearly shows this difference. The values of  $z$  yielding the best collapses for M1 and M2, computed with the algorithm in [Bhattacharjee and Seno, 2001], are, respectively,  $z = 0.46$  with 95% confidence interval [0.44, 0.52], and  $z = 0.222$  with 95% confidence interval [0.216, 0.244]. A consequence of the validity of Eq. (3.8) (see Box 1.2) is that  $\langle S \rangle \propto A^z$  (Fig. 3.9). The values of the exponent  $z$  estimated by performing a least-squares linear fit of the pairs  $(\log A, \log \langle S \rangle)$  are compatible with the values obtained by collapse of the distributions. Estimates are, respectively:  $z = 0.464 \pm 0.001$ , and  $z = 0.233 \pm 0.002$ . Equation (3.5) proves still valid: figures 3.14C and 3.15C show the points  $(\lambda \langle \tau \rangle, \langle S \rangle)$  for different values of  $A$  falling on the 1:1 line. As expected from equation (3.8) (see Box 1.2), consecutive moments of  $p(S)$  satisfy  $\langle S^{j+1} \rangle / \langle S^j \rangle \propto \langle S \rangle$  (see Figs 3.5C and 3.6C). Therefore,  $\text{var}(S) = \langle S^2 \rangle - \langle S \rangle^2 \propto \langle S \rangle^2$ , i.e.  $\beta = 2$ . Using Eq. (3.2), this implies that the relative fluctuations of  $S$  remain constant as the system grows. As a consequence, the two non-neutral models M1 and M2 do not support a deterministic approach to biodiversity prediction, as it is not possible to define a threshold

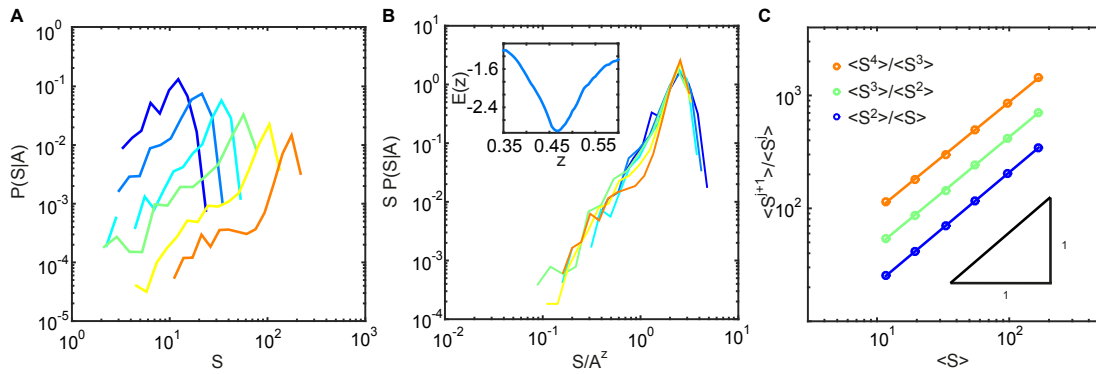


Figure 3.5 – Statistics of  $S$  for model M1. A)  $p(S|A)$ , different colors refer to different values of  $A = 10^i$ , from  $i = 1.5$  to  $i = 3.5$ . B) Collapse of  $p(S|A)$ . Equation (3.8) is verified because  $Sp(S|A)$  versus  $S/A^z$  collapse on the same curve for different values of  $A$ . Inset: the minimum of the functional  $E(z)$  provides the best estimate for the exponent and the associated error [Bhattacharjee and Seno, 2001]. C) Scaling of the ratio of consecutive moments with  $\langle m \rangle$ .

area above which the deviation of  $S$  from the deterministic prediction is smaller than any prescribed value. The effect of the different behavior of fluctuations of  $S$  on the uncertainty of field measurements is exemplified in the inset of Fig. 3.8, which compares the scattering of single values of  $S$  around the  $S = A^z$  line in a  $(A, S)$  plot for values distributed according to Eq. (3.8) and to a Poisson distribution.

One may speculate that the reason for the difference between the neutral and the non-neutral cases lies in the different characteristics of the perturbations caused to the system by a diversification event. In the neutral case, the appearance or disappearance of a species causes a fixed change in the available resources (and therefore in the death rates) because each species has the same consumption rate (or occupies the same space, in the case of the MSV model). When  $A$  increases (and so does  $\mathcal{R}$ ), this change becomes relatively smaller and so do relative fluctuations around stationary mean values. In the non-neutral models, instead, the appearance or disappearance of a species brings a change in the total resources consumption which depends on its mass. As  $A$  increases, the ecosystem can sustain larger and larger species, therefore the potential amplitude of fluctuations increases with  $A$ . To further investigate how the presence and characteristic of perturbations influence the behavior of  $S$ , section 3.5 is devoted to examining the dynamics of neutral models where perturbations are imposed, to disentangle them from the diversification mechanism and to better understand the role of perturbations in shaping the stationary probability distribution of  $S$ .

### 3.5 The role of perturbations

In section 3.4, perturbations were hypothesized to be the most important factor in determining the fluctuations of  $S$ , and in particular the dependence of their amplitude on ecosystem size to

### Chapter 3. Fluctuations scaling of the species-area relation

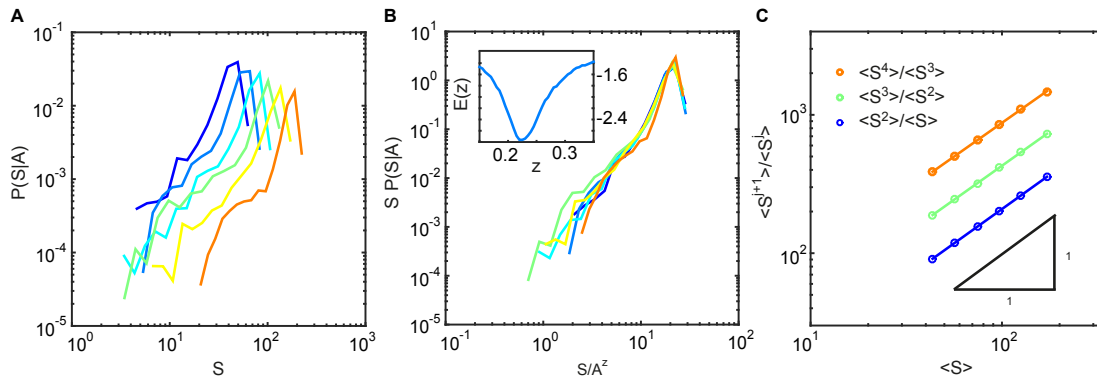


Figure 3.6 – Statistics of  $S$  for model M2. A)  $p(S|A)$ , different colors refer to different values of  $A = 10^i$ , from  $i = 1.5$  to  $i = 3.5$ . B) Collapse of  $p(S|A)$ . Equation (3.8) is verified because  $Sp(S|A)$  versus  $S/A^z$  collapse on the same curve for different values of  $A$ . Inset: the minimum of the functional  $E(z)$  provides the best estimate for the exponent and the associated error [Bhattacharjee and Seno, 2001]. C) Scaling of the ratio of consecutive moments with  $\langle m \rangle$ .

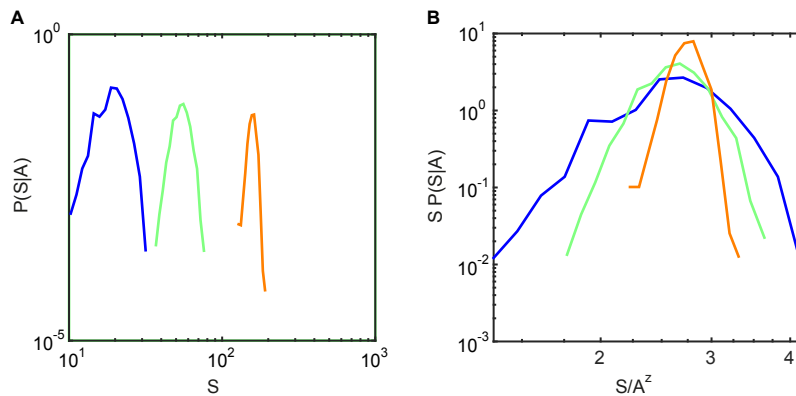


Figure 3.7 – Failure of the collapse in Eq. (3.8) for the N-SI model. A)  $p(S|A)$  (log-log scale). Colors refer to different values of  $A$ . B) Attempt to collapse the distributions as suggested by Eq. (3.8). The failure of the collapse shows that, in the neutral models, the fluctuations behave in a remarkably different way than in the non-neutral models.

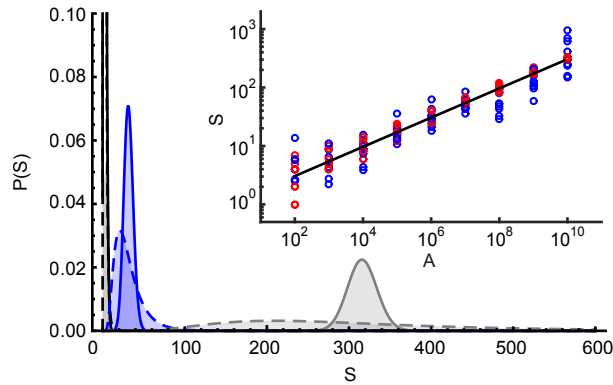


Figure 3.8 – Comparison of Poisson distributions (continuous lines) and lognormal distributions (dashed lines) with means  $A^z$ , namely with  $z = 1/4$  and  $A = 10^2$  (black),  $10^6$  (blue) and  $10^{10}$  (grey). The lognormal distribution satisfies Eq. (3.8) with  $F(x) = 1/\sqrt{2\pi\sigma^2} e^{-(\log(x)-\mu)^2/(2\sigma^2)}$ . To have  $\langle S \rangle = A^z$ , one needs  $\mu = -\sigma^2/2$ . The parameter  $\sigma$  is set such that, for  $A = 10^2$ , the variance of the lognormal is equivalent to the variance of the Poisson distribution. Inset: values of  $S$  extracted from a Poisson distribution (red) and from a lognormal distribution (blue) with mean  $A^z$ , with  $z = 1/4$  and  $A = 10^i$  for  $i = 2, \dots, 10$ . The parameter  $\sigma$  of the lognormal distribution was set as explained above. For each value of the area, 10 extracted values are shown. The black line is  $S = A^z$ .

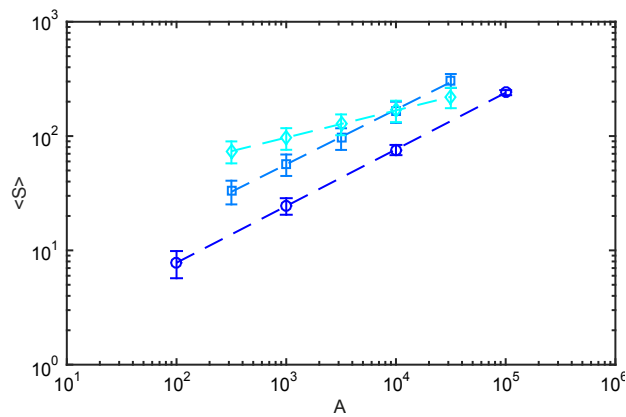


Figure 3.9 – Scaling of average number of species  $\langle S \rangle$  vs the ecosystem size  $A$  for models N-SI (circles), M1 (squares) and M2 (diamonds). Lines are least square fit of logarithmically transformed data, see text for exponent estimates. Error bars are SEM.

### Chapter 3. Fluctuations scaling of the species-area relation

---

be key in discriminating between the asymptotically deterministic scenario and the strongly stochastic one. In the two non-neutral models M1 and M2, perturbations arise naturally from the diversification dynamics, where a new species randomly arriving by speciation or immigration would increase resource consumption and therefore determine a resource debt causing extinctions, i.e. downward fluctuations of  $S$ . Recovery from perturbations, on the other hand, is controlled by the diversification rate. In models M1 and M2, the rate at which perturbations happen is related to the recovery rate, rendering the exploration of their interplay impossible. Also, it is not possible to control the size of such perturbations *a priori*. In general, however, the rates of perturbations and diversification may be unrelated, for example when perturbations are caused by the environment. This section is devoted to examining the effects of imposing external perturbations of a given size and frequency, which can be controlled independently from the diversification rate. The frequency of the perturbations is chosen to be constant with respect to ecosystem area  $A$ , to assess in a paradigmatic case the effects of the same type of perturbation on ecosystems of different areas. The N-SI model dynamics is an appropriate base to study the effect of external perturbation as it is, in itself, stable.

Perturbations may have different origins, e.g. climatic or environmental events or temporary shifts in resource availability. As such, they vary in frequency and in their way in which they affect ecosystems. Here, the effect of transient disturbances is modeled directly in terms of species loss. For simplicity, disturbances are assumed to occur with a fixed frequency, but equivalent results are expected if they arrive as a fixed-rate Poisson process. Two types of perturbations are compared. In model N-SI-P1, perturbations of fixed magnitude are added to the dynamics of model N-SI, regardless of  $S$ : with a fixed frequency  $\nu_p$ , a number  $S_p$  of species is removed from the system. In model N-SI-P2, instead, perturbations grow with system size: with a fixed frequency  $\nu_p$ , a fraction  $f_p$  of species is removed from the system. This second type of perturbations could describe, for example, the case of response to an environmental change. Imagine in fact that each species has a certain probability  $q$  to survive, adapting to the new environmental conditions. Then, on average  $(1 - q)S$  species will go extinct. The two models were simulated for  $A = 10^i$  with  $i = 2.5, 3, 3.5, 4, 4.5, 5, 5.5$ ,  $S_p = 5$ ,  $f_p = 3/4$  and  $1/\nu_p = 5 \cdot 10^6$ . Figure 3.4D and E display the temporal dynamics of  $S$  in the two cases. As expected, the statistics of  $S$  in model N-SI-P1 do not differ strongly from the unperturbed case, yielding  $\beta = 0.96 \pm 0.07$  ( $R^2=0.975$ ). Relative fluctuations therefore vanish rather fast with increasing ecosystem size, allowing the identification of a threshold area over which a deterministic SAR is an acceptable approximation, as in the neutral unperturbed case. The SAR shows a power-law pattern with a slope comparable with the value of  $z$  measured in the unperturbed case (Fig. 3.10),  $z = 0.436 \pm 0.005$  ( $R^2=0.99$ ). The dynamics of model N-SI-P2, shown in figure 3.4E, is instead rather different. For small areas, between two perturbations the system is able to return to the stationary state of the unperturbed model, with Poisson fluctuations around the mean value of  $S$ . Although  $\beta \approx 2$ , indicating non-negligible stochasticity, if the system is observed at timescales that are smaller than the perturbation timescale  $1/\nu_p$ , one would see no differences with the unperturbed case. However, increasing the area causes the recovery time

### 3.6. Effect of habitat stochasticity in a metapopulation dynamics model

to increase, and the system will spend less and less time in the stationary state corresponding to the unperturbed system, until a threshold value of  $A$  is reached, above which the system is unable to ever reach such a stationary state. When this situation occurs, the observed value of  $S$  might be far from the average predicted by assuming ecosystem stationarity even at small time scales, and strongly dependent on the ecosystem history. The area at which this behavior starts depends on the interplay of diversification rate and perturbation rate. This non-stationary regime shows deviations from a pure power-law both in the SAR (Fig. 3.10), whose slope decreases gradually towards zero, and in Taylor's law. The decreasing slope of the SAR is qualitatively explained by a simplified deterministic description of the model dynamics, where  $S$  grows linearly in time with rate  $r$  with a hard limit at  $\bar{S}_{stat} = cA^z$ , the stationary average of  $S$  in the unperturbed model N-SI, and with rate  $v_p$  a perturbation makes  $fS$  species go extinct, with  $S$  the number of species when the perturbation occurs. As shown in appendix 3.8.2, one can distinguish between two cases: when  $\eta = \frac{v_p f c A^z}{r} \leq 1$  the stationary state is reached, while when  $\eta > 1$  it is never reached. One can show (see appendix 3.8.2) that the slope of the scaling of  $\langle S \rangle$  with increasing  $A$  is  $z$  when  $\eta \ll 1$  but decreases as  $\eta$  increases, reaching zero when  $\eta \geq 1$ . For Taylor's law, a value of  $\beta = 3$  is predicted at small  $\eta$  values, decreasing with increasing  $\eta$ . Taylor's law in model N-SI-P2 indeed shows a slope  $\beta \sim 3$  for middle-large areas, but the decrease is not seen at the simulated areas.

This simple model identifies the factors playing a role in determining the dynamics of biodiversity in the presence of perturbations. When perturbations size increases with ecosystem size (as in model N-SI-P2), the ratio  $\eta = \frac{v_p f c A^z}{r}$  determines whether the system spends most of the time in its unperturbed stationary state (small  $\eta$ ) or whether it spends most or all the time recovering from perturbations ( $\eta \geq 1$ ). In particular, all the other factors being fixed, a larger system will spend more time recovering, as is seen clearly in Fig. 3.4E. This differs from the case where perturbation size and diversification rate are independent of ecosystem size, as in model N-SI-P1, in which ecosystems of all sizes have the same behavior: either they all return to the stationary state, or none does.

### 3.6 Effect of habitat stochasticity in a metapopulation dynamics model

In a given ecosystem, the availability of different niches might change in time, affecting the number of species surviving in it and therefore creating additional variability. It might also differ between ecosystems of the same size. Here, a spatially explicit metapopulation model is considered, a well-established tool to study biodiversity patterns in a spatially heterogeneous environment [Hanski and Ovaskainen, 2000, Rybicki and Hanski, 2013, Grilli et al., 2015] which allows assessing the role of habitat stochasticity in the variability of  $S$ . The metapopulation model is based on the interplay between extinction and colonization dynamics. Each species has a probability  $p_i$  to be present (with one or more individuals) in a site  $i$  of the landscape, which is a  $L \times L$  lattice. Species do not interact, thus they have independent dynamics. Let  $E_i$  be the rate at which a species becomes extinct in patch  $i$  and  $C_i$  the rate at which the patch is

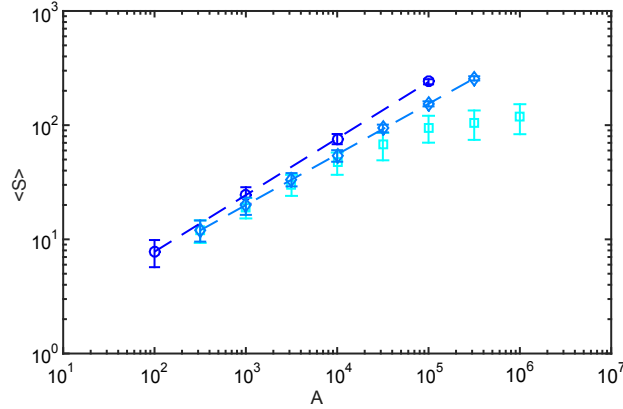


Figure 3.10 – Scaling of average number of species  $\langle S \rangle$  versus ecosystem size  $A$  in models N-SI (circles), N-SI-P1 (diamonds) and N-SI-P2 (squares). Lines are least square fit of logarithmically transformes data, see text for exponent estimates. Error bars are SEM.

colonized. The probability of occupancy  $p_i$  of patch  $i$  is then governed by

$$\frac{dp_i}{dt} = C_i(1 - p_i) - E_i p_i. \quad (3.9)$$

Each patch is characterized by a habitat type, determined by the value of a parameter  $h$ ,  $0 \leq h \leq 1$ . A species is characterized by 5 parameters: the colonization rate  $c$ , the extinction rate  $e$ , the phenotype  $\phi$ , the niche width  $\gamma$  and the average dispersal distance  $1/\alpha$ . For a focal species, the fitness in a patch  $i$  with habitat parameter  $h_i$  is  $q = \exp[-(h_i - \phi)^2 / (2\gamma^2)]$ . The fitness, which is larger when  $\phi$  is similar to  $h_i$ , determines how well the species will perform in that patch, i.e. the probability that it will become extinct, the extinction rate being defined as  $E_i = e/q_i$ . In turn, the colonization rate, given by  $C_i = c / (2\pi) \sum_{j \neq i} \alpha^2 e^{-d_{ij}\alpha} q_j p_j$ , governs the interactions between the different patches. Species parameters are extracted at random from a uniform distribution in a suitable interval. The values of  $h$  characterizing the landscape are also extracted at random, but with a fixed spatial correlation (see below for details) and are constant in time. It was shown [Hanski and Ovaskainen, 2000] that the condition under which Eq. ((3.9)) has an equilibrium solution different from  $p_i = 0 \forall i$  is

$$\lambda_M > e/c, \quad (3.10)$$

where  $\lambda_M$  is the metapopulation capacity, i.e. the leading eigenvalue of an appropriate matrix, which depends on the landscape and on the species phenotype  $\Phi$ , niche width  $\gamma$  and dispersal range  $1/\alpha$ . It is therefore said that a species is persistent in the landscape if it satisfies the condition in Eq. ((3.10)).

To study the variability in the number of species persisting in a landscape of a given size in the presence of habitat stochasticity, a pool of  $S_0 = 500$  species was considered, then the number of species satisfying the condition in Eq. (3.10) were evaluated and the operation was



### 3.6. Effect of habitat stochasticity in a metapopulation dynamics model

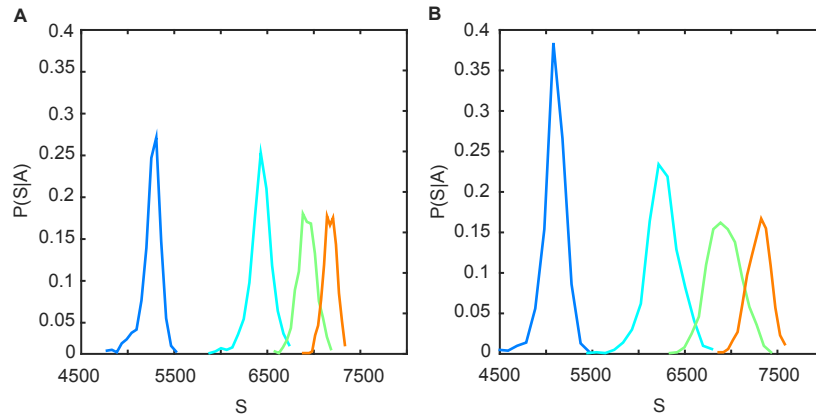


Figure 3.11 –  $p(S|A)$  for the metapopulation model with landscapes generated according to Rybicki and Hanski [2013] A) and with the HYDRO\_GEN algorithm (see text) B). Colors, from blue to orange, refer to landscape sizes  $A = 25^2, 50^2, 100^2, 200^2$ .

repeated in a set of 1000 different random landscapes at fixed spatial correlation with the same pool of species. Spatially correlated random landscapes were generated with two alternative methods. The first is the one used in Rybicki and Hanski [2013], while the second uses the HYDRO\_GEN algorithm [Bellin and Rubin, 1996] for the generation of correlated random fields, which has the advantage of allowing us to set a prescribed correlation. Landscape areas  $A = 25^2, 50^2, 100^2, 200^2$  were considered. Initially, 1000 landscapes of size  $200^2$  were generated, then the 1000 smaller landscapes were obtained as subsets. Species parameter ranges were taken as follow:  $c \in [0.25, 2]$ ,  $e \in [0.025, 0.4]$ ,  $\phi \in [0.1, 0.9]$ ,  $\gamma \in [0.1, 0.5]$ ,  $\alpha \in [0.07, 1]$ . To increase the species pool without increasing the computational time (mainly dependent on the  $\lambda_M$  computation), for each of the 500 sets of  $\phi$ ,  $\gamma$  and  $\lambda$  15 additional values of  $c$  and  $e$  were extracted, effectively enlarging the species pool to  $S_0 = 8000$ .

This model differs from the ones presented above in that the number of persisting species is not determined by the equilibrium of the competing stochastic processes of speciation and extinction, but to the diversity of habitat types present in the landscape. A larger landscape tends to have more habitat diversity, therefore more species (on average) are able to persist in it. The  $p(S|A)$  resulting from the metapopulation model with the two different methods of landscape generation are shown in figures 3.11A and B. The variance decreases with the mean with a pattern that is not described by a power-law, and therefore an estimation of  $\beta$  is not possible. However, relative fluctuations decrease with the average much faster than in all other considered models, at least as  $1/\langle S \rangle$ . Note that by using the condition in Eq. (3.10) to count the number of surviving species, the stochasticity of the dynamics in each individual landscape results effectively averaged. Therefore, the observed variability is only that due to the habitat stochasticity. In conclusion, the variability generated by habitat stochasticity at fixed spatial correlation is negligible with respect to the one possibly caused by community dynamics, at least within the assumptions of the present model.

### 3.7 Discussion

When SARs are used to predict biodiversity, and biodiversity responses to habitat loss or fragmentation, it is implicitly assumed that the average value,  $\langle S \rangle$ , described by the SAR is representative of what could be actually measured empirically, and that all ecosystem features of interest are stationary. The work presented in this Chapter draws attention on the fact that such conditions are not always met, and that the scaling of the relative fluctuations with the mean is a useful tool to tell apart different scenarios. Specifically, when relative fluctuations decay with increasing mean (or, equivalently, increasing ecosystem size), they eventually become negligible in a large enough ecosystem, and the deterministic SAR becomes an appropriate prediction tool (Figure 3.1). Nevertheless, fluctuations might still be important for small systems. When relative fluctuations do not decay (or decay very slowly), ecosystems of any size maintain the same level of relative stochasticity. If this level is non-negligible, the predictions of the SAR will not be representative of the behavior of the ecosystem, where  $S$  may fall far apart from the predicted average. These results are based on an extensive survey of models of ecosystem dynamics reflecting different types of internal dynamics and external perturbations driving the system. Even though the analysis does not encompass all the models of community dynamics ever explored, it shows that different models can lead to very different predictions for the fluctuations of the number of species  $S$ . This study shows that the fluctuations of the number of species  $S$  underpinning the applicability of a deterministic SAR do not have a universal behavior but rather depend on the main ecological processes reflected in various model assumptions. Specifically:

- In a neutral framework where fluctuations are only driven by the stochasticity of diversification and of extinction events, biodiversity is asymptotically deterministic, i.e. its relative fluctuations decay with increasing ecosystem size (the case where Taylor's  $\beta = 1$ ). Fluctuations are non-negligible for small areas, where single realizations of  $S$  may deviate significantly from the deterministic prediction  $S = cA^z$ . As a consequence, the empirical verification of theoretically predicted small-scale patterns would require a time or ensemble average of biodiversity measurements. It is possible to define a threshold area above which the expected deviation of  $S$  from  $\langle S \rangle$  is smaller than a fixed threshold  $\epsilon_t$ . More precisely, if  $\langle S \rangle = cA^z$  is the deterministic prediction,

$$\frac{\Delta S}{\langle S \rangle} = \frac{1}{\sqrt{\langle S \rangle}} = \frac{1}{\sqrt{cA^z}} \leq \epsilon_t \iff A \geq \left( \frac{1}{\epsilon_t \sqrt{c}} \right)^{2/z}$$

Computational evidence was produced showing that such result is also valid in the presence of competition for space or resources, where the analytical results valid for the non-interacting case still proves applicable;

- The introduction of mass as a trait determining resource consumption of individuals, following Kleiber's law, causes large fluctuations in the number of species due to the occasional introduction by speciation/immigration of new species with large resource

consumption capabilities. The over-exploitation of the ecosystem's resources following such events triggers extinctions. The related stationary distribution  $p(S)$  displays relative fluctuations which remain constant as the area  $A$  increases ( $\beta = 2$ ). Therefore, it is not possible to define a threshold area above which deviations of  $S$  from the deterministic predictions are smaller than a prescribed value. In such a scenario, deterministic SARs should be considered only as an average trend of ecosystems, with the understanding that actual field observations might differ significantly from it. Note that non-neutral resource consumption does not necessarily cause this behavior of the fluctuations. In fact, if the introduction of new species in the system (by immigration or speciation) is strictly constrained by supply limitation, i.e. the total community consumption rate can never exceed the supply rate, fluctuations would behave similarly to the neutral case. Nevertheless, it is deemed possible that no similar constraint apply and the scenario described by our model holds;

- Habitat stochasticity, e.g. as introduced by Rybicki and Hanski [2013], causes small variability in the number of coexisting species, which decays relatively fast with ecosystem area. It is overall negligible with respect to the variability caused by the stochasticity of community dynamics. Remark that habitat stochasticity is implemented at fixed spatial correlation scale of habitat type. If habitat stochasticity were allowed to uniform/diversify habitat types present in the ecosystem, fluctuations would increase accordingly;
- Peaks in resource consumption caused by the arrival of new species are just one possible cause of driving noise for the ecosystem. The effects of generic perturbations with fixed frequency on the dynamics of biodiversity were explored and found to depend on the specific properties of the perturbations. In particular, perturbations whose magnitude (the number of species going extinct in one burst of activity) does not depend on the number of species present in the ecosystem become progressively less important for increasing ecosystem size. The dynamics thus becomes asymptotically similar to the unperturbed one. Instead, when the magnitude of perturbations is larger for species-richer communities, larger systems need progressively longer times to recover, and therefore the theoretically predictable stationary average is progressively less informative about the ecosystem's state. In such conditions, empirically measured values of  $S$  will often be far from the expected stationary average. Therefore, a SAR estimated from empirical measurements will have little predictive power. Depending on the ratio between the frequency of perturbations and the recovery rate, on the perturbations intensity and on the ecosystem size, the system can still be considered stationary at short timescale and away from perturbation events, and the conclusions found for the non-perturbed neutral dynamics apply. Note that perturbations are modeled phenomenologically. In fact, no explicit description of the response of individual species to environmental events was given, but only the overall response of the set of species was considered. A further step could be to investigate what type of individual response would cause a certain overall effect of perturbations.

### Chapter 3. Fluctuations scaling of the species-area relation

---

The different conclusions obtained with different models call for further research to clarify which scenario applies to a specific real ecosystems. That would imply estimating the variance of  $S$  measurements from islands or isolated ecosystems of different areas but similar environmental conditions (e.g., similar climate), like e.g. the Solomon archipelago [Diamond and Gilpin, 1980], and testing its scaling with the mean. Not only the scaling of the mean, but also the amplitude of fluctuations should be considered. In fact, relative fluctuations could not decay but be small enough so as not to prevent deterministic predictions. The variance could be measured across different replicate ecosystems (that is, ecosystems of very similar size) or from long-term time series of a single ecosystem [Vellend et al., 2013, Dornelas et al., 2014].

Long-term time series would also prove useful to spot trends highlighting non-stationarity. For example, Vellend et al. [2013] and Dornelas et al. [2014] provide meta-analyses of several time-series with various durations up to tens of years. While overall there is no dominating trend, a number of single time series show an increasing or decreasing trend, signaling non-stationary conditions [Dornelas et al., 2014]. Moreover, through repeated observations in time, recovery dynamics after a perturbation may be assessed, say, after a fire or the arrival of an invasive species. Identifying the time required to return to stationarity after perturbation and comparing it to the expected time between consecutive perturbation events would allow assessing whether the system's biodiversity should be expected to spend most of its time around its stationary average or else in transient states possibly far from it. The study of the dynamics following a perturbation would also indicate whether its effects scale with system size, supporting the scenario of model N-SI-P2. One could for example observe if the arrival of an invasive species causes more extinctions in a more diverse community than in a less diverse one.

The analysis presented in this Chapter proves that modeling does not provide a unique answer on how the fluctuations in the number of species scale with the mean. Rather, models reflecting different ecological assumptions and processes support different scalings. Therefore, comparison of model predictions with empirical observations of such scaling, jointly with empirical investigation of the effect of perturbations on systems of different size, shed light on the processes controlling community assembly. This work provided an overview of the effect of specific modeling choices and ecological assumptions on fluctuation scaling, and identified which mechanisms could cause relative fluctuations which do not decay with system size. In conclusion, conventional SARs describe a mean trend in biodiversity which may or may not be appropriate to predict measurable patterns of biodiversity, depending on the processes acting within the ecosystem and the nature of the external sources of variability.

## 3.8 Appendix

### 3.8.1 Persistence time distribution

The persistence time distributions of the analyzed models, as shown in Figs. 3.12, 3.13, 3.14A and B and 3.15A and B, satisfy

$$p(\tau|A) = \frac{1}{A^\theta} G\left(\frac{\tau}{A^\theta}\right) \quad (3.11)$$

with  $\theta = z$  for models MSV, N-SI and M1 and  $\theta = z + \xi$  for model M2, where  $z$  is the exponent of the SAR. The function  $G(x)$  is such that  $G(x) \sim c$  when  $x < 1$ , with  $c > 0$  a constant and  $\lim_{x \rightarrow \infty} G(x) = 0$ . This result is in agreement with Eq. (3.5), in fact, when  $A$  is large:

$$\begin{aligned} \langle \tau \rangle &= \int_1^\infty \tau p(\tau|A) d\tau = \frac{1}{A^\theta} \int_1^\infty \tau G\left(\frac{\tau}{A^\theta}\right) d\tau = A^\theta \int_{1/A^\theta}^\infty x G(x) dx \\ &\approx A^\theta \left[ c \int_{1/A^\theta}^1 x dx + \int_1^\infty x G(x) dx \right] = A^\theta \left[ \frac{c}{2} A^{-2\theta} + c' \right] \propto A^\theta \end{aligned} \quad (3.12)$$

where  $x = \tau/A^z$ ,  $c' > 0$  is a constant and the properties of  $G(x)$  were used to evaluate the integrals. Therefore, substituting in Eq. (3.5) and recalling that for M2 one has  $\lambda = \lambda_0 A^{-\xi}$ , one finds  $\langle S \rangle \propto A^z$ .

In the MSV model, a more precise characterization of  $p(\tau|A)$  is possible. In fact, as found in Bertuzzo et al. [2011], the distribution of persistence time  $p(\tau)$  is a power-law with exponential cutoff when  $\tau > A$ , where  $A$  the number of time-steps in a “generation”, i.e. the time interval when on average every individual dies and gets replaces. For  $\tau < A$ ,  $p(\tau)$  is roughly constant (Fig. 3.13). Therefore, one can write:

$$p(\tau) = \mathcal{C} \begin{cases} 1 & \text{for } \tau \leq A \\ \left(\frac{\tau}{A}\right)^{-\alpha} e^{-\lambda\left(\frac{\tau}{A}-1\right)} & \text{for } \tau > A \end{cases} \quad (3.13)$$

where  $\mathcal{C} = \left[ A + A^\alpha e^\lambda \left(\frac{\lambda}{A}\right)^{\alpha-1} \Gamma(1-\alpha, \lambda) \right]^{-1}$  is the normalization constant. The value of  $\alpha$  depends on the topology of the lattice (dimension and connectivity, see Bertuzzo et al. [2011]). For the 2D lattice considered here, a maximum likelihood fit of the tail of the distribution to a power-law with exponential cutoff gives  $\alpha \sim 3/2$ . The fitted curves are the dashed lines in Fig. 3.13.

### 3.8.2 Deterministic model of a system subject to perturbations

This appendix describes a deterministic model describing the dynamics of the number of species  $S$  in a system subject to periodical perturbations causing the extinction of species. The model is intended to mimic the stochastic models N-SI-P1 and N-SI-P2 in a simplified way which allows obtaining analytic results explaining qualitatively the behavior of the two

### Chapter 3. Fluctuations scaling of the species-area relation

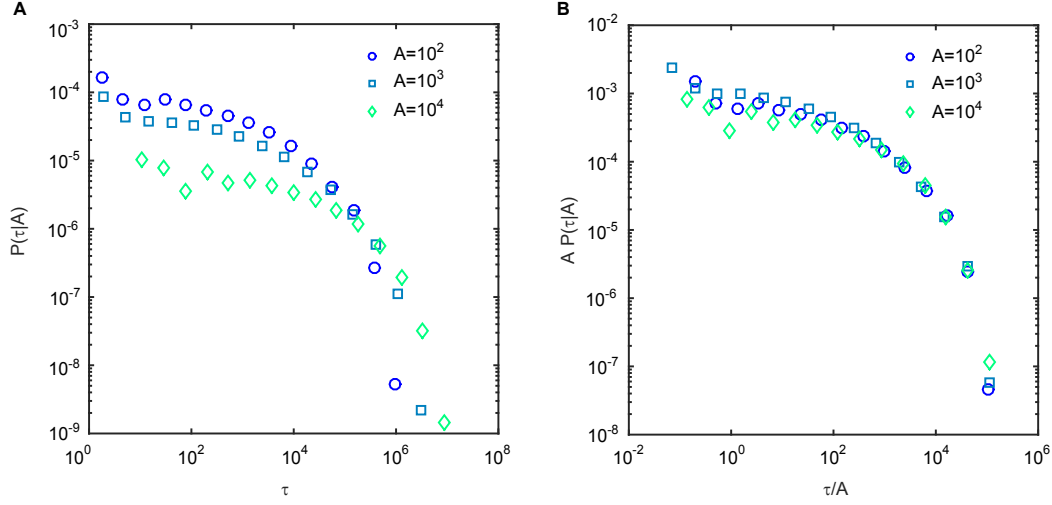


Figure 3.12 – Persistence time distributions for the N-SI model. Colors refer to different values of  $A$ . A)  $p(\tau|A)$  for the N-SI model with  $\lambda = 10^{-3}$  and  $A = 10^i$  with  $i = 2, 3, 4$ . B)  $A p(\tau|A)$  is a function of  $\tau/A^z$ , where  $z$  is the SAR exponent (see text).

stochastic models. As shown later, this is possible only in the cases where the perturbations dominate over intrinsic stochasticity in determining the statistics of  $S$ .

Let  $\nu_p$  be the frequency of perturbations and  $fS^\alpha$  the number of species going extinct due to a perturbation, where  $S$  is the number of species present at the moment when the perturbation occurs and  $\alpha \in [0, 1]$ . The values  $\alpha = 0, 1$  correspond, respectively, to the situations in models N-SI-P1 and N-SI-P2. Let us assume, for simplicity, that between two perturbations  $S$  grows linearly with rate  $r$ , but analogous computations can be performed also for other types of growth. The stationary value of  $S$  in the unperturbed N-SI model,  $\bar{S}_{st} = cA^z$ , is taken as a hard boundary to the growth of  $S$ . The constant  $c$  is set to 1 in the following. Supposing that the upper boundary  $\bar{S}_{st}$  was reached before a perturbation, the time  $T$  needed to return to it is such that

$$A^z - fA^{\alpha z} + rT = A^z, \quad (3.14)$$

therefore  $T = \frac{fA^{\alpha z}}{r}$ . Two cases must therefore be distinguished: i) when  $T \leq \frac{1}{\nu}$ , the upper boundary is reached between two perturbations, and the system stays at  $S = \bar{S}_{st}$  for an interval  $\frac{1}{\nu} - T$  (see Fig. 3.16A); ii) when  $T > \frac{1}{\nu}$  the upper boundary is never reached, and the system eventually stabilizes on a cycle where the growth in time  $1/\nu$  is equal to the loss of species due to the perturbation (see Fig. 3.16B). Defining  $\eta = \frac{\nu f A^{\alpha z}}{r}$ , the two cases are identified by  $\eta \leq 1$  and  $\eta > 1$ . Note that, when  $\alpha = 0$ ,  $\eta$  does not depend on  $A$ . Therefore, all other parameter being fixed, systems of different sizes all belong to the same case: either they all reach the upper boundary or none of them does.

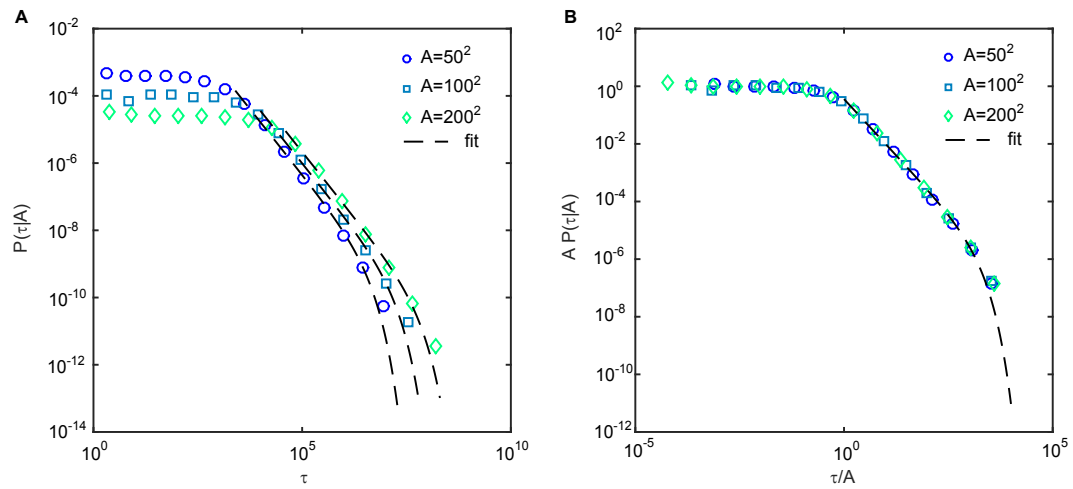


Figure 3.13 – Persistence-time distribution for the MSV model. A)  $p(\tau|A)$  for the MSV model with  $\lambda = 10^{-3}$  and  $A = L^2$  with  $L = 50, 100, 200$ . The dotted lines are maximum likelihood fits of the tail of the distribution ( $\tau > A$ ) to a power-function with exponential cutoff (Eq. (3.13)). B)  $A p(\tau|A)$  is a function of  $\tau/A$  (see text).

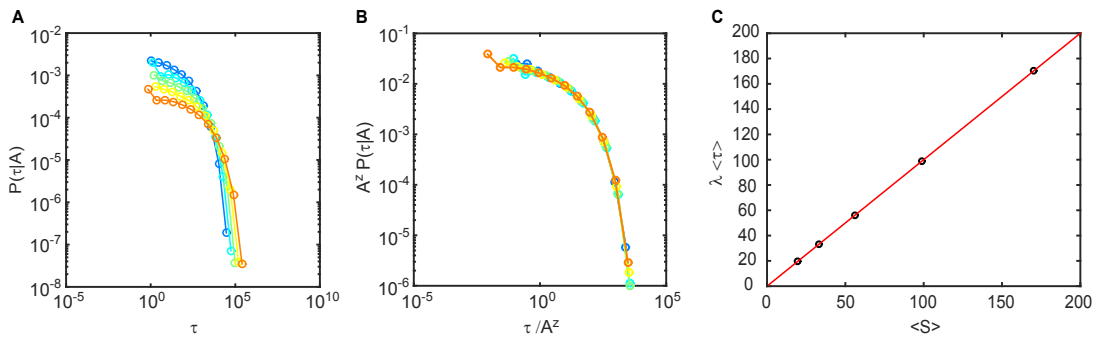


Figure 3.14 – Persistence time distributions for model M1. A)  $p(\tau|A)$ , colors from blue to orange correspond to areas  $A = 10^i$  with  $i = 2, 2.5, 3, 3.5, 4$ . B)  $A^z p(\tau|A)$  is a function of  $\tau/A^z$ , where  $z$  is the SAR exponent (see text). C) Plot of  $\langle S \rangle$ , vs  $\lambda \langle \tau \rangle$ , showing that Eq. (3.5) is verified. The continuous line has slope one.

### Case $\eta \leq 1$

After an initial phase where there is net growth, the upper boundary is reached and  $S(t)$  becomes periodic with period  $1/\nu$ . Within a period, the evolution of  $S$  is described by

$$S(t) = \begin{cases} A^z - f A^{\alpha z} + r t & t \leq T \\ A^z & t > T. \end{cases} \quad (3.15)$$

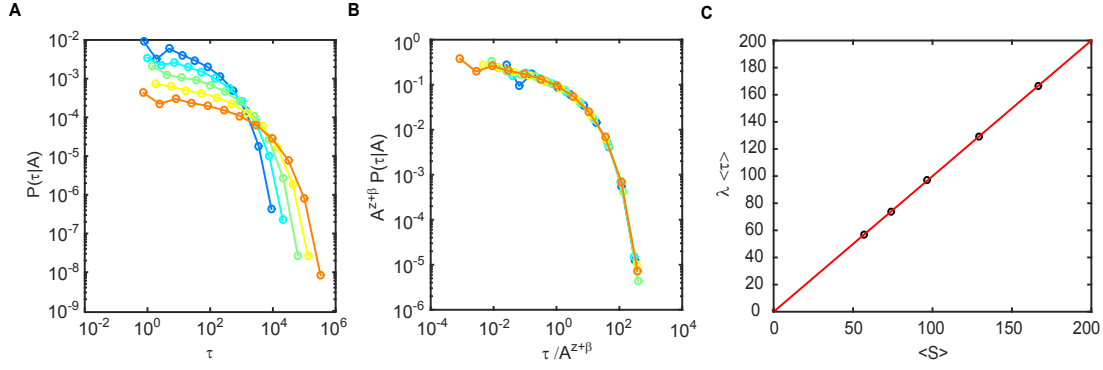


Figure 3.15 – Persistence time distribution for model M2. A)  $p(\tau|A)$ , colors from blue to orange correspond to areas  $A = 10^i$  with  $i = 2, 2.5, 3, 3.5, 4$ . B)  $A p(\tau|A)$  is a function of  $\tau/A^{z+\beta}$ , where  $z$  is the SAR exponent and  $\beta$  the exponent of the dependence of the diversification rate on  $A$  (see text). C) Plot of  $\langle S \rangle$ , vs  $\lambda\langle\tau\rangle$ , showing that Eq. (3.5) is verified. The continuous line has slope one.

The average of  $S$  over a period can be computed as follows:

$$\langle S \rangle = v \int_0^{1/v} S(t) dt = v \left[ \int_0^T (A^z - fA^{\alpha z} + rt) dt + \int_T^{1/v} A^z dt \right] = A^z \left( 1 - \frac{\eta f}{2} A^{-(1-\alpha)z} \right). \quad (3.16)$$

One immediately sees that, for  $\eta \ll 1$ , the SAR is a power-law with slope  $z$ . For non-negligible  $\eta$  values, the slope of the SAR is obtained as

$$\frac{d \log \langle S \rangle}{d \log A} = z \left( 1 - \frac{2\alpha - 1}{\frac{2r}{f^2 v A^{2\alpha-1}} - 1} \right). \quad (3.17)$$

Therefore, when  $\alpha > 1/2$  the slope changes from  $z$  to 0 as  $A$  increases (all other parameters being equal, valid until  $\eta \leq 1$ ). When  $\alpha < 1/2$ , instead, the slope is  $z$  in the large  $A$  limit. This result explains why model N-SI-P1 has a power-law SAR similar to the one in the absence of perturbations, while the N-SI-P2 model has a SAR with gradually decreasing slope. The plateau reached at larger areas for model N-SI-P2 is explained by the case  $\eta > 1$ .

To compute the exponent  $\beta$ , we first compute the second moment of  $S$ :

$$\begin{aligned} \langle S^2 \rangle &= v \int_0^{1/nu} S(t)^2 dt = v \left[ \int_0^T (A^z - fA^{\alpha z} + rt)^2 dt + \int_T^{1/v} A^{2z} dt \right] \\ &= A^{2z} - f\eta A^{(\alpha+1)z} + \frac{f^2 \eta}{3} A^{2\alpha z}. \end{aligned} \quad (3.18)$$

Therefore, we have

$$\text{var}(s) = \langle S^2 \rangle - \langle S \rangle^2 = A^{2\alpha z} \frac{f^2 \eta}{3} \left( 1 - \frac{3}{4} \eta \right) = A^{2\alpha z} \frac{f^3 v}{3r} \left( 1 - \frac{3}{4} \eta \right), \quad (3.19)$$



which, when  $\eta \ll 1$ , gives  $\beta = 3\alpha$ . For  $\alpha = 1$ , this explains correctly the results of model N-SI-P2 for middle-large values of the area. For  $\alpha = 1/3$  the fluctuations given by the perturbations are characterized by  $\beta = 1$ , therefore they have the same scaling as the fluctuations due to the intrinsic stochasticity of the dynamics. For  $\alpha < 1/3$ , therefore, the predictions of the deterministic model are not valid, because the intrinsic stochasticity will dominate on the perturbations, giving  $\beta = 1$ , as observed for the N-SI-P1 model ( $\alpha = 0$ ).

**Case  $\eta > 1$**

When  $\eta > 1$ , the process becomes periodical before reaching the upper boundary  $S = A^z$ . To show this, let us consider the case  $\alpha = 1$  for simplicity and call  $S_i$  the number of species after the  $i$ -th perturbation. The process is therefore expressed by

$$S_{i+1} = \min\{(1-f)(S_i + r/\nu), (1-f)A^z\}. \quad (3.20)$$

This process has two fixed points,  $\tilde{S}_1 = \frac{(1-f)r}{\nu f}$  and  $\tilde{S}_2 = (1-f)A^z$ . The fixed point  $\tilde{S}_1$ , corresponding to the situation in which the growth in an interval of length  $1/\nu$  is equaled by the loss of species due to a perturbation, is attractive, i.e.  $|\tilde{S}_1 - S_{i+1}| < |\tilde{S}_1 - S_i|$ . Therefore, when  $\tilde{S}_1 < \tilde{S}_2$  (which is, when  $\eta > 1$ ), the process will tend to this fixed point  $\forall S_0 < A^z$ . This situation corresponds to a periodic dynamics described by

$$S(t) = \frac{(1-f)r}{\nu f} + r t \quad t \leq 1/\nu \quad (3.21)$$

which does not depend on  $A$ . The average over a period therefore is also independent of  $A$ :

$$\langle S \rangle = \nu \int_0^{1/\nu} S(t) dt = \frac{r(2-f)}{2\nu f}, \quad (3.22)$$

which explains the plateau observed in the SAR of model N-SI-P2 in Fig. 3.10. Being  $S(t)$  independent of  $A$ , all moments are independent of  $A$  and therefore also the relative fluctuation  $\frac{\Delta S}{S}$  are constant.

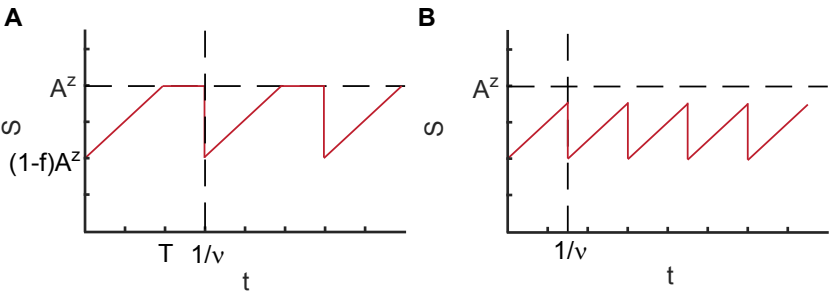


Figure 3.16 – Dynamics of the deterministic model of a system subject to perturbations described in section 3.5. A) Case  $\eta \leq 1$ , the system reaches the upper boundary; B) case  $\eta > 1$ , the system never reaches the upper boundary.

## 4 Single-cell metabolic rates size-scaling in freshwater phytoplankton

*The scaling of metabolic rate with organismic body mass holds a central role in structuring communities, given its link with size scaling of growth rates and other organismic properties. Typically, this relationship is measured for mean metabolic rates and characteristic body masses of species. However, when considering a population in detail, individual metabolic rates and body sizes show variabilities, possibly affecting inter-species average allometric scaling and explaining its documented anomalies. The present chapter fills a gap by providing evidence for the statistical features of intra-specific metabolic rate distributions using experiments on freshwater phytoplanktonic cultures measuring jointly nutrient uptake rates and size at the single-cell level. Measurements were conducted for monocultures of strains of *Synechococcus* sp., *Scenedesmus obliquus* and *Cryptomonas ovata* using NanoSIMS, a technique based on secondary ions mass spectrometry in combination with stable isotope labeling. Joint measurements of cell volume and uptake rates of carbon and nitrogen were carried out. Although limited by the sample sizes that current technology allows probing, results highlight relevant statistical features of marginal probability distributions of uptake rate and size, showing universality across different taxa covering three orders of magnitude in volume. Such results prove the feasibility of a general description of the joint distribution of metabolic rate and size independent of most biological details and suggests the scaling form required of it to be consistent with measured marginals.*

### 4.1 Introduction

Decades of research in ecology and biology have unraveled stunning regularities in the distribution of species, their abundances and metabolic requirements [Brown et al., 2004, Marquet et al., 2005]. These ubiquitous patterns hold for most ecosystems on Earth and help us predict how ecological communities assemble in the environment and how they respond to environmental change. Kleiber's law (KL) is regarded as one of the most important and widespread of such patterns [Kleiber, 1932, 1947], although the extent to which it holds has been questioned [Kolokotronis et al., 2010, Marañón et al., 2013, Banavar et al., 2014, DeLong et al., 2010]. KL

#### Chapter 4. Single-cell metabolic rates size-scaling in freshwater phytoplankton

---

states that the metabolic rate  $b$  of an organism scales with its body mass,  $m$ , according to a universal power-law  $b = cm^\alpha$  (where  $\alpha$  is a scaling exponent,  $c$  a constant). Supported by an array of physiological studies [Kleiber, 1932, Calder, 1984, McMahon and Bonner, 1983, Peters, 1986], it purportedly holds across more than 27 orders of magnitude in body mass – from metabolic machinery (e.g. mitochondria) to whales and sequoias [West and Brown, 2004]. The metabolic power required to support life across that range spans over 21 orders of magnitude [West and Brown, 2004].

The majority of experimental studies supports the claim that KL applies across species with  $\alpha \sim 3/4$  in many groups of organisms [Brown, 1995, Bartels, 1982, Feldman and MacMahon, 1983, Savage et al., 2004], a predominance [Savage et al., 2004] that has been referred to as the central paradigm of comparative physiology [White and Seymour, 2005]. Metabolic demand per unit mass is thus agreed to decrease as body mass increases. Theoretical explanations have been proposed to support the  $3/4$  exponent based on some general features of metabolic networks that hold across different species and levels of organismic complexity [West et al., 1997, Banavar et al., 1999b]. The 'universality' of the exponent  $3/4$  has however been challenged based on robust empirical evidence of a wealth of exponents  $\alpha$  deviating from  $3/4$  [Dodds et al., 2001, Glazier, 2005, Hubert, 2010, DeLong et al., 2010, Marañón et al., 2013, Marañón, 2015, Ahluwalia, 2017] and by arguments suggesting curvatures altering the power law relation [Clarke et al., 2010, Kolokotronis et al., 2010, Mori et al., 2010].

Metabolic requirements of organisms are claimed to influence many fundamental organismic properties [Brown et al., 2004, West et al., 2002] such as growth rate, mortality rate or carrying capacity, attributing KL a central role in several ecological processes and in the understanding of patterns related to body size and abundance [Zaoli et al., 2017, Cavender-Bares et al., 2001, Marañón, 2015] (its use in Chapter 1 being an example). Notably, the change of slope of the community size-spectra of marine phytoplankton between oligotrophic and eutrophic conditions [Cavender-Bares et al., 2001, Marañón, 2015] could be explained by a change of slope in KL across different resource supply conditions, modifying the relative advantages of organism of different sizes in the uptake and use of resources, although the precise mechanisms at work have not been clarified theoretically yet. Therefore, understanding the limits of KL and its validity is of paramount importance.

Most studies investigating KL focused on average metabolic rates and body masses of species, therefore addressing a mean trend. Nevertheless, individuals within a species show heterogeneity in both body size and metabolic rate, with the latter ranging from basal to maximum and varying in relation to individual life cycle stage and individual age. In its classic form, KL explains only a part of the total metabolic rate variability, relating it deterministically to body size. A generalized KL, acknowledging the variability, should describe the joint probability  $p(m, b | \langle m \rangle) dmdb$  that an individual has a metabolic rate in  $[b, b + db]$  and a body mass in  $[m, m + dm]$ , given that it comes from a species with typical body mass  $\langle m \rangle$ . The feasibility of such a general description of correlated fluctuation of  $m$  and  $b$ , independent of most biological details of species, except for their average body size, would be meaningful at a

fundamental level. In fact, first it would prove that processes underlying the heterogeneity of body size and metabolic requirements of organisms are general and valid across biological scales. Second, it would provide a theoretical framework synthesizing correlated fluctuations of size and metabolic rates of different origins at the intra- and inter-specific levels. Such a synthesis is necessary to test the potential role of intra-population variability in individual traits in shaping size-related patterns and their response to different environmental conditions, e.g. resource supply. For example, characterizing the joint fluctuations of body size and metabolic rates would help generalizing the theoretical framework in Chapter 1 and predict possible effects of such fluctuation on patterns covariations predicted therein. As a matter of fact, intrapopulation variability in individual traits (i.e. phenotypic heterogeneity) is being increasingly recognized as an important factor for the functionality of microbial populations and communities [Ackermann, 2015]. Additionally, the knowledge of  $p(m, b|m)$  would allow assessing if the maximum or minimum metabolic rate expected for a species scale differently from the average, possibly explaining the different exponents measured empirically, and if intra-specific scaling differs from inter-specific scaling. The latter information would clarify the effect of a sheer change in organismic size, other organismic properties being equal, on the metabolic rate.

The intra-specific probability distribution of body sizes has been studied previously [Giometto et al., 2013]. It obeys general statistical properties across size scales and taxa, allowing its full description on the basis of a minimum of species-specific detail (see Methods, section 4.2.1). Concerning metabolic rate variability, Dodds et al. [2001] analyzed the fluctuations of the average metabolic rate of mammal and bird species around the mean trends but did not investigate intra-specific heterogeneity. Single-cell variability was examined in single species (e.g. [Schreiber et al., 2016, Kennard et al., 2016, Welkenhuysen et al., 2017, Zimmermann et al., 2015]), often looking at species-specific molecular pathways. A previous study [Labra et al., 2007], investigating the temporal fluctuations of metabolic rates of single individuals, found that such fluctuations follow a universal distribution. However, no systematic measures of intra-specific metabolic rate heterogeneity and of its covariation with body size have been performed across taxa and scales of body size and metabolic rate with a uniform measuring method, as required to answer the research question addressed in this chapter.

This Chapter presents the results of an experiment designed to measure simultaneously body size and metabolic rate at the single-cell level for three species of freshwater phytoplankton, covering three orders of magnitude in volume ( $1 \mu\text{m}^3$  to  $1000 \mu\text{m}^3$ ). The choice of phytoplankton as a study organism is twofold: on the one hand, it allows covering a significant range of scales while maintaining taxonomic similarity of the studied species, which allow uniform measuring methods and direct comparison of results; on the other, such choice is motivated by the importance of understanding the mechanisms driving size-related patterns in marine and freshwater primary producers [Marañón, 2015], an aim towards which the formulation of a generalized KL heads. Recently, methods for single-cell investigation of metabolism and other cell properties have become available (see Vasdekis and Stephanopoulos [2015] for a review), allowing the investigation of the intra-population heterogeneity which classical

population-level measurement average out. Here, Nano Secondary Ion Mass Spectrometry (NanoSIMS), in combination with stable isotope labeling, is used to permit the simultaneous measurement of cell size and of rates of nutrient uptake (see section 4.2), a proxy of metabolic activity of particular relevance for phytoplanktonic communities. Specifically, nitrogen and carbon uptakes were targeted (see section 4.2.4). Cell size is measured by the volume  $v$ . Note that, when cell density is constant, volume and mass only differ by a proportionality constant, therefore KL and its generalization can equivalently be formulated for mass  $m$  or volume  $v$ . In the following, therefore, the variable  $v$  is used (see section 4.4 for further comments on the suitability of this assumption for phytoplankton).

The experiment tests the viability of a description of metabolic rate fluctuations holding across taxa and in which biological detail only enters in the form of the dimensionless variable  $b/\langle v \rangle^\alpha$  (see section 4.2.1 for an exact statement of the hypothesis). It also gathers preliminary yet significant information on the covariation of metabolic rate and size at the single cell level. The long-term scope of this line of research is thus to experimentally validate a generalized KL formulation embedding fluctuations in both body size and metabolic rates.

## 4.2 Methods

### 4.2.1 Generalized Kleiber's law

The scope of this chapter is to characterize the joint probability distribution of body size  $v$  and metabolic rate  $b$  for the individuals of a species  $S$ . Such distribution is indicated with  $p(v, b|S)$ , where the symbol  $|$  indicates the conditioning on species-specific traits, and quantifies the probability that an individual of a species of typical body size  $\langle v \rangle$  has metabolic rate in  $[b, b + db]$  and body size in  $[v, v + dv]$ . Because mass is a strong determinant of a species' physiology and ecology [McMahon and Bonner, 1983, Peters, 1986, Marañón, 2015], and in line with previous investigations on intra-specific size distributions [Giometto et al., 2013], species' typical body size  $\langle v \rangle$  is assumed to be the most relevant species trait for this problem. The form of the joint probability distribution  $p(v, b|\langle v \rangle)$  is subject to some empirically driven constraints: i) it must imply the observed scaling of the average  $b$  with the average  $v$ , that is, the classical KL, ii) it must yield marginal distributions of body size and metabolic rate (i.e.,  $p(v|\langle v \rangle) = \int db p(v, b|\langle v \rangle)$  and  $p(b|\langle v \rangle) = \int dv p(v, b|\langle v \rangle)$ ) akin to those measured experimentally. Its two marginal distributions  $p(v|\langle v \rangle)$  and  $p(b|\langle v \rangle)$  describe respectively the intra-specific variability in body size and metabolic rate. The marginal distributions of body sizes of several microbial species spanning four orders of magnitude in volume have been shown to satisfy [Giometto et al., 2013]

$$p(v|\langle v \rangle) = \frac{1}{v} F\left(\frac{v}{\langle v \rangle}\right), \quad (4.1)$$

where  $F$  a species-independent function such that  $\lim_{x \rightarrow 0, \infty} F(x) = 0$ . Note that the prefactor  $v^{-1}$  is required for normalization (see Appendix 4.5.1 for details). The viability of such a general description for the variability of intra-specific body mass might be explained by the

processes of cell growth and division shaping it [Giometto et al., 2013], common to different species.

The hypothesis made here is that the intra-specific metabolic rate variability might be described by an analogous scaling form i.e.

$$p(b|\langle v \rangle) = \frac{1}{b} G\left(\frac{b}{\langle v \rangle^\alpha}\right), \quad (4.2)$$

with  $\alpha > 0$  and  $G$  a species-independent function such that  $\lim_{x \rightarrow 0, \infty} G(x) = 0$ . The meaning of this hypothesis is that the species average body size constrains the typical scale of the metabolic rate of its individuals, proportional to  $\langle v \rangle^\alpha$ , and the fluctuations around this typical value are described by a function  $G$ , which is the same for all the species. Therefore, the metabolic rate distributions of different species, with different typical body mass  $\langle v \rangle$ , hold the same shape when rescaled by a factor  $\langle v \rangle^\alpha$  (see Box 1.2 in Chapter 1). The hypothesis in Eq. (4.2) implies classical KL, i.e.  $\langle b \rangle \propto \langle v \rangle^\alpha$  (see Appendix 4.5.1). The prefactor  $b^{-1}$  is required, again, by normalization. The scaling behavior of Eqs (4.1) and (4.2) also implies that all the moments of the distributions scale with the average, i.e.  $\langle v^k \rangle \propto \langle v \rangle^k$  and  $\langle b^k \rangle \propto (\langle v \rangle^\alpha)^k \propto \langle b \rangle^k$  (see Appendix 4.5.1). A remarkable implication of this is that the maximum and minimum metabolic rates observable within a given species scale with size with the same exponent as the average (see Appendix 4.5.1). For comparison with the experiments, two exponents  $\alpha_C$  and  $\alpha_N$  and two functions  $G_C$  and  $G_N$  describing the variability of nutrient uptake rates of C and N are considered.

The heterogeneity of metabolic rates within a species will derive partly from the body size heterogeneity and partly from the intrinsic causes of variation mentioned in the introduction. The former component of the heterogeneity is expected to have the general statistic features hypothesized in Eq. (4.2), as it descends from (4.1). The hypothesis that also the latter component enjoys such general features is based on the conjecture that the processes leading to that intrinsic variability are quite general as well, and that their biological specificity can be summarized by the average species' body size. Some sort of 'universal' behavior of the fluctuations is a necessary condition for the formulation of a generalized KL based on little species-specific details, and is what the experiment presented in the following intends to test.

#### 4.2.2 Maintenance of phytoplanktonic cultures

Monocultures of the cyanobacterium *Synechococcus sp.* (SYN), the chlorophyte *Scenedesmus obliquus* (SCE) and the cryptophyte *Cryptomonas ovata* (CRY) obtained from the Culture Collection of Algae and Protozoa (CCAP) were grown semicontinuously in WC medium at 20°C under cool white light with a photon flux of 25-50  $\mu\text{Em}^{-2}\text{s}^{-1}$  with a 14L:10D photoperiod. Samples for the measurement of single-cell size and metabolic rate were obtained during the exponential phase.

### 4.2.3 Labeling experiment

In order to assess the C and N uptake rate of single cells, samples were incubated in enriched WC medium. The enriched medium was prepared by adding 1 ml of  $K^{15}NO_3$  1M stock solution and 1ml of  $NaH^{13}CO_3$  1M stock solution to 1 L of WC medium prepared without  $Na^{14}NO_3$ , which was substituted by  $K^{15}NO_3$ . Cultures were diluted in a 1:1 proportion with the enriched medium and incubated for 3h. The time interval for incubation was chosen such that the uptake would be above detection levels while still allowing to neglect cell division. At the end of the incubation period, the cultures were fixed in 1% paraformaldehyde. A comparison of cell volumes between fixed and unfixed cultures showed that fixation did not cause cell shrinkage. Note that the enriched medium contained the same amount of N as the standard WC medium but a higher quantity of C. Nevertheless, the pH difference was verified to be negligible, therefore growth conditions during the incubation can be considered analogous to the maintenance conditions.

### 4.2.4 NanoSIMS measurements

NanoSIMS is an ion microprobe which performs mass spectrometry on secondary ions sputtered from a solid target by the impact of a primary beam of charged particles. Secondary ions are sputtered from the top few atomic monolayers of the sample. The high spatial precision of the ion beam ( $\sim 100nm$ ) allows the creation of an ion image of the sample through a raster of the primary beam on the sample surface. See Box 4.2.8 for more details on the principles of NanoSIMS.

The measurement of isotope ratios ( $^{13}C:^{12}C$  and  $^{15}N:^{14}N$ ) in single cells which have been incubated in enriched medium allows to compute their carbon and nitrogen uptake rates, while size (volume) can be inferred from the image, as detailed in section 4.2.6. The measurement of cellular uptake rates *a posteriori* of the labeling experiment allows to probe also the variability due to heterogeneous microenvironments cells might experience in populations, for example due to spatial and temporal gradients of nutrients or irradiance. This is an advantage of NanoSIMS with respect to other single-cell measurement techniques which require cell isolation (e.g. microfluidic techniques). The NanoSIMS technique has been previously used with success on biological samples [Pernice et al., 2012, Schreiber et al., 2016] and specifically on phytoplanktonic cells [Bonnet et al., 2016].

For each strain, a control sample before incubation and an enriched sample after incubation were prepared for NanoSIMS analysis by filtering an appropriate amount of fixed culture on polycarbonate filters of diameter 5mm (Isopore<sup>TM</sup> Membrane Filters from Merck Millipore, cat. no. RTTP01300) with pore size  $1.2 \mu m$ , which had been previously coated in gold. The quantity of culture to be filtered was estimated such that a single layer of cell would deposit on the filter, so that all analyzed cells would lie on the same plane. The prepared filters were once more coated in gold previous to NanoSIMS analyses. The samples were imaged with the NanoSIMS ion microprobe of the Laboratory for Biological Geochemistry at EPFL, in Lausanne (Switzerland), in order to measure their  $^{13}C:^{12}C$  and  $^{15}N:^{14}N$  ratios. Cells were



located with the camera of the NanoSIMS instrument and bombarded with a primary beam of  $\text{Cs}^+$  ions of intensity 40 nA focused to a spot of about 180 nm on the sample surface. An initial pre-sputtering to remove the surface of the sample and permit the imaging of internal layers of the cells was performed using diaphragm D1-1 or D1-2, resulting in an intensity of 3.5 pA on the sample surface. For the imaging, the beam was focused to a spot of about 180 nm on the sample surface (diaphragm D1-3), resulting in a beam of intensity 1.5 pA on the sample. Ion images of the sample were obtained by rastering the primary beam across the sample with a dwell-time of 5 ms per pixel. Secondary ion images for  $^{12}\text{C}^{14}\text{N}^-$ ,  $^{12}\text{C}^{15}\text{N}^-$ ,  $^{12}\text{C}_2$  and  $^{13}\text{C}^{12}\text{C}$  were simultaneously recorded from analysis areas of  $15 \times 15 \mu\text{m}$  to  $50 \times 50 \mu\text{m}$  with a resolution of  $256 \times 256$  pixels. Ratio images were obtained taking the ratio of the  $^{12}\text{C}^{15}\text{N}^-$  and  $^{12}\text{C}^{14}\text{N}^-$  image and of the  $^{13}\text{C}^{12}\text{C}^-$  and  $^{12}\text{C}_2^-$  image (see examples of ratio images in Fig. 4.1). Mass resolution of the detectors ( $\delta M/M$ ) was  $>8000$ , sufficient to resolve the interference between  $^{13}\text{C}^{12}\text{C}$  and  $^{12}\text{C}_2\text{H}$  and between  $^{12}\text{C}^{15}\text{N}$  and  $^{13}\text{C}^{14}\text{N}$ . The measured ratios  $^{13}\text{C}^{12}\text{C}:^{12}\text{C}_2$  were converted to  $^{13}\text{C}:^{12}\text{C}$  ratios by halving them [Nuñez et al., 2018], while the  $^{12}\text{C}^{15}\text{N}:^{12}\text{C}^{14}\text{N}$  ratios was considered equivalent to the  $^{15}\text{N}:^{14}\text{N}$  ratios. These approximations are sufficient for the present scopes.

#### 4.2.5 C-N analysis

To measure the average C and N content of cells, necessary to compute the total C and N uptake during the incubation period (as explained in section 4.2.6), an adequate quantity of each culture was filtered on a pre-ashed glass microfiber filter (VWR, cat. no. 516-0343, particle retention:  $0.7 \mu\text{m}$ ). The number of cells on each filter was estimated by flow cytometric measurement of the cultures. The filters were stored in ashed aluminum foil, dried at  $100^\circ\text{C}$  for 30 minutes and analyzed for TOC and TN by gas chromatography. The average C and N content per cell,  $\langle C \rangle$  and  $\langle N \rangle$ , were then obtained by dividing the TOC and TN values by the number of cells on the filter.

#### 4.2.6 Computation of volumes and uptake rates

Regions of interest (ROI) were defined for each single cell imaged following its contour (Fig. 4.1). Individual cell volumes were inferred from the images by measuring axes of the ROIs (cells' cross sections) and using appropriate geometrical assumptions. A spherical shape was assumed for *Synechococcus* and an ellipsoidal one for the other two strains. The average isotope ratios  $r_C = ^{13}\text{C}/^{12}\text{C}$  and  $r_N = ^{15}\text{N}/^{14}\text{N}$  of each single cell were obtained by averaging the value of all the ROI pixels in the corresponding ratio image. In doing so, it was assumed that the analyzed layer is representative of the entire cell. As superficial layers show lower enrichment than internal ones, this assumption is an approximation, causing an over- or under-estimation of the real average enrichment depending on the position of the analyzed layer within the cell. Nevertheless, given that the expected variability in the analyzed layer position is low, all cells lying on the same plane, this error is expected not to compromise results.

## Chapter 4. Single-cell metabolic rates size-scaling in freshwater phytoplankton

For each single cell imaged, the uptake rate  $b_i$  of element  $i$ , where  $i=C$  or  $N$ , is computed as follows:

$$b_i = \frac{(\bar{r}_i - \bar{r}_{i,0})C_i}{\bar{r}_{i,med}} \frac{1}{\Delta t} \quad (4.3)$$

where  $\bar{r}_i = \frac{r_i}{1+r_i}$  are, respectively, the ratios  $^{13}\text{C}:C$  and  $^{15}\text{N}:N$ ,  $r_{i,0}$  are the natural ratios  $^{13}\text{C}:C$  and  $^{15}\text{N}:N$  measured in the control sample,  $\bar{r}_{i,med}$  are the  $^{13}\text{C}:C$  and  $^{15}\text{N}:N$  ratios in the incubation medium,  $\Delta t$  is the incubation time, and  $C_i$  is the total cell content of, respectively,  $C$  and  $N$ .  $C_i$  is computed from the average  $C$  and  $N$  contents measured for each strain, by noting that the average  $C$  and  $N$  content of a cell scales sublinearly with the average cell volume of that strain,  $\langle v \rangle$ , i.e.  $\langle C_i \rangle \propto \langle v \rangle^{\gamma_i}$  with  $\gamma_i < 1$ . For carbon, a similar result had already been observed in several studies, as reviewed in Gosselain et al. [2000]. The observed value of the scaling exponent  $\gamma_C$  measured in the studies cited therein, for tens of taxa covering several order of magnitude in cell volume, ranges from 0.76 to 1.04. The measurements on the three strains used in this experiment, covering three orders of magnitude in cell volume, yield, respectively,  $\gamma_C = 0.8 \pm 0.1$  and  $\gamma_N = 0.8 \pm 0.1$  (see Fig. 4.3a). Making the most parsimonious assumption, that is, that the same scaling holds also at the individual cell level within a species,  $C_i$  is obtained as

$$C_i = k_i \langle C_i \rangle \left( \frac{v}{\langle v \rangle} \right)^{\gamma_i}, \quad (4.4)$$

where  $k_i \simeq 1$  is a constant not depending on individuals or species identity (see Appendix 4.5.2).

### 4.2.7 Hypothesis testing

The hypotheses that the body size distributions satisfy Eq. (4.1) and that the uptake rate distributions satisfy Eq.(4.2) was tested as follows. First, the exponents  $\alpha_C$  and  $\alpha_N$  are estimated by optimizing the data collapse of  $p(b_C|\langle v \rangle)$  and  $p(b_N|\langle v \rangle)$  for the three species. When  $p(B_i|\langle V \rangle)$  satisfies eq. (4.2), in fact, the curves  $B_i p(B_i|\langle V \rangle)$  vs  $B_i/\langle V \rangle^{\alpha_i}$  collapse on the same curve  $G_i$ <sup>1</sup> for the appropriate value of  $\alpha_i$ . The value of  $\alpha_i$  is therefore estimated optimizing the collapse with the method in Bhattacharjee and Seno [2001] (separately for  $i = C, N$ ). Secondly, note that if the body sizes and the uptake rates satisfy respectively Eq. (4.1) and Eq. (4.2) then the rescaled variables  $X = \log\left(\frac{v}{\langle v \rangle}\right)$  and  $Y_i = \log\left(\frac{b_i}{\langle v \rangle^{\alpha_i}}\right)$  satisfy, respectively,  $p(X) = F(X)$  and  $p(Y_i) = G_i(Y_i)$ . That is, the probability distributions of the rescaled variables  $X$  and  $Y_i$  are the same in the three strains. For each variable ( $v$ ,  $b_C$  and  $b_N$ ), a K-Sample Anderson-Darling test was used to test the null hypothesis that the rescaled samples from the three different strains come from the same distribution. If the null hypothesis cannot be discarded, it follows that the hypotheses in Eqs. (4.1) and (4.2) also cannot be discarded.

<sup>1</sup>The function is allowed to be different for carbon and nitrogen uptake, thence the index  $i$ .

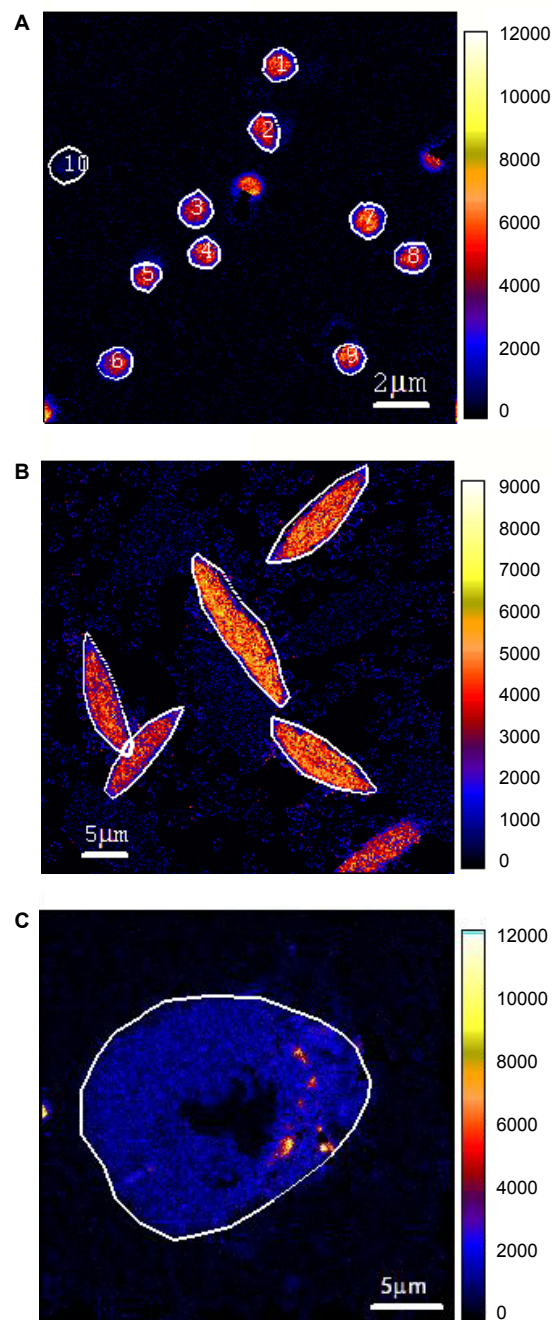


Figure 4.1 – Examples of ratio images obtained with the NanoSIMS. Each pixel's color refers to the  $^{13}\text{C}$  enrichment with respect to the control sample in that pixel, measured as  $\delta = \frac{r_{\text{mes}} - r_{\text{ctrl}}}{r_{\text{ctrl}}} \times 1000$ , with  $r = {}^{13}\text{C}/{}^{12}\text{C}$ , indicated by the color bar adjacent to each picture. White contours are the defined ROI over which the isotopic ratio is averaged, details in section 4.2.6. A) *Synechococcus* sp., black circles are pores in the filter; B) *Scenedesmus obliquus*; C) *Cryptomonas ovata*.

### 4.2.8 Principles of NanoSIMS

This Box contains a description of the NanoSIMS technique which scope is to allow a better understanding of the experimental methods and of the data. The content of the Box has been adapted from Hoppe et al. [2013].

#### **Sputtering, ionization, transmission and ion detection**

Figure 4.2, reproduced from Hoppe et al. [2013], illustrates the physical principle of secondary ions mass spectroscopy (SIMS). The surface of a solid sample, contained in an ultra-high-vacuum environment, is hit by a primary beam of ions with energies typically in the range of several kilo-electron volts (keV). When a primary ion hits the surface of the target, it triggers a collision cascade, causing atoms and small molecules from the upper layers of the sample (typically down to 5–20 nm depth) to be ejected. This process is referred to as ‘sputtering’. Of these sputtered particles, only a small fraction is ionized, typically in the range between  $10^{-5}$  and  $10^{-2}$ . These ions, referred to as ‘secondary ions’, leave the sample surface with relatively low kinetic energies and are extracted by an electrostatic field. They are then transferred to a magnetic sector mass spectrometer, where secondary ions of different mass-to-charge ratio are physically separated by the Lorentz force as they pass through a magnetic field arranged perpendicular to the velocity vector of the secondary ions. A precise isotope measurement requires that this process has high transmission that is, low loss of secondary ions between the sample surface and the detector, and high mass resolution, that is, the capability to separate the secondary ions or molecules of interest from other ions/molecules of very similar mass. The design of NanoSIMS meets both requirements.

The high lateral resolution of NanoSIMS, allowing to probe samples at the cellular and sub-cellular level, requires a primary beam of small intensity, typically 1 pA. This beam is focused on an area of, typically,  $100 \text{ nm} \times 100 \text{ nm}$ . As a consequence, a smaller number of primary ions hit the sample with respect to conventional SIMS, consequently producing a smaller number of secondary ions and making technically more difficult to obtain precise measurements. Primary beam ions are chosen to maximize the secondary ion yield (that is, the fraction of the atoms released by a primary ion hitting the surface which is ionized). For the measurement of isotopes of H, C, N, O, Si, and S ions of Caesium ( $\text{Cs}^+$ ) are used.

Secondary ions separated by the mass spectrometer are then detected by electron multipliers. Technical difficulties encountered in the detection process, described in detail in Hoppe et al. [2013], include isobaric interference, quasi-simultaneous arrival of ions on the detector, detector dead time.

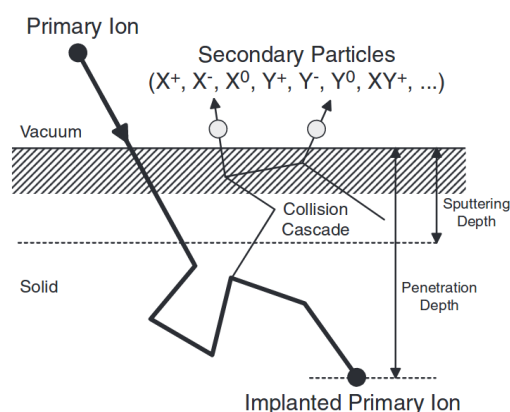


Figure 4.2 – The physical principle of SIMS. Impacting primary ions on a sample surface creates sputtering or ejection of atoms and small molecules, a fraction of which are ionised. These secondary ions are subsequently transferred through a mass spectrometer and counted. Figure reproduced from Hoppe et al. [2013].

### Design and optics

The fundamental property distinguishing the NanoSIMS from other types of magnetic-sector SIMS instruments is its ability to produce a small primary beam diameter (defined as encompassing 68% of the primary beam flux assuming a Gaussian density distribution) on the surface of a flat sample. Under optimal tuning condition, minimum beam diameters for  $\text{Cs}^+$  ion beams are  $\sim 50$  nm. The beam is produced by a  $\text{CsCO}_3$  source. An optical path (details in Hoppe et al. [2013]) focuses the beam, corrects it for aberrations and allows to raster it across the sample with a maximum area of  $200 \times 200 \mu\text{m}^2$ . This raster allows ion images to be created. Ion yields are largely constant across relatively large areas in ion images. An aperture of variable size allows to control the beam diameter by limiting its angular dispersion. The beam hits the sample normally, and the ions extraction and initial focusing occurs co-axially. Owing to this feature of NanoSIMS, the ion optical elements can be placed much closer to the sample surface than in conventional ion probe instruments. This allows obtaining a smaller beam diameter and a higher ion yield. Due to this short distance between the optics and the sample, it is not possible to do real-time viewing of the sample in an optical microscope. Nevertheless, the sample can be inspected by a high-resolution reflected light image if it is moved away from its analysis position.

The secondary ions are separated from the oppositely charged primary ions by a deflection plate, centered and focused before entering the detector. The masses are separated by a magnetic field, in which focal plane seven detectors sit, of which six can be moved to detect different isotopes. The maximum mass separation that can be detected by the NS50L instrument is 21. Mass resolution is defined as  $m/\Delta m$ , where  $m$  is the (mean)

mass and  $\Delta m$  the mass difference between the two species to be separated. Only when the achieved mass resolution is larger than this quantity, the peaks corresponding to the two species can be separated.

## 4.3 Results

### 4.3.1 Scaling of uptake rates with cell volume

Figures 4.3C) and D) display the measured values ( $v$ ,  $b_i$ ), respectively for  $i = C, N$ , for the three strains. Each point represents a cell and the strain averages ( $\langle v \rangle$ ,  $\langle b_i \rangle$ ) are plotted in black. The average uptake rates scale with the average strain cell volume, respectively, with exponents  $\alpha_C = 0.69 \pm 0.01$  for carbon and  $\alpha_N = 0.59 \pm 0.01$  for nitrogen (obtained by linear least squares fitting of log-transformed values). These exponents, close to  $2/3$ , suggest that the cell uptake might be limited by membrane transport, rather than by diffusion (see Box 4.3.7), which is realistic given the nutrient rich medium.

Fig. 4.3C) and D) also shows that fluctuations of  $b$  and  $v$  are correlated at the intraspecific level. However, the observed pattern rests on the assumption that the inter-specific scaling of C and N content also holds at the intra-specific level, an assumption which could not be verified experimentally due to the impossibility of measuring absolute C and N content of single cells with the experimental methods used here. On the other hand, the C- and N-specific uptake rates, obtained as

$$\tilde{b}_i = \frac{b_i}{C_i} = \frac{(\bar{r}_i - \bar{r}_{i,0})}{\bar{r}_{i,med}} \frac{1}{\Delta t} \quad (4.5)$$

and shown respectively in figures 4.3E) and F), are not affected by that assumption. Within a single species, the correlation of  $\tilde{b}_i$  with volume is very small, that is, fluctuations of size and C- and N-specific uptake for cells of the same species are roughly independent. The correlation of total uptake rates  $b_C$  and  $b_N$  with  $v$  is therefore mediated by the cellular content of C and N. Note, finally, that a small number of *Synechococcus* cells show a very small C uptake, much lower than average. Some of these cells are visible in Fig. 4.3E), while some do not appear in the plot in log-scale as their enrichment is  $\approx 0$ . The same cells exhibit an N uptake within expectation, with the exception of one cell which has both uptake rates far below average. These cells were likely affected by errors during manipulation and/or SIMS analysis and were excluded from the following analysis.

### 4.3.2 Correlation of C and N uptake rates

Carbon-specific and nitrogen-specific uptake rates show a strong positive correlation (Pearson coefficient 0.83, see Fig. 4.3B), suggesting that the two uptake processes are strongly coupled in the present experimental conditions. Note that the presence of correlation is tested on the nutrient-specific rates because the multiplication by cell volume necessary to obtain total uptake rate introduces a strong but trivial correlation.

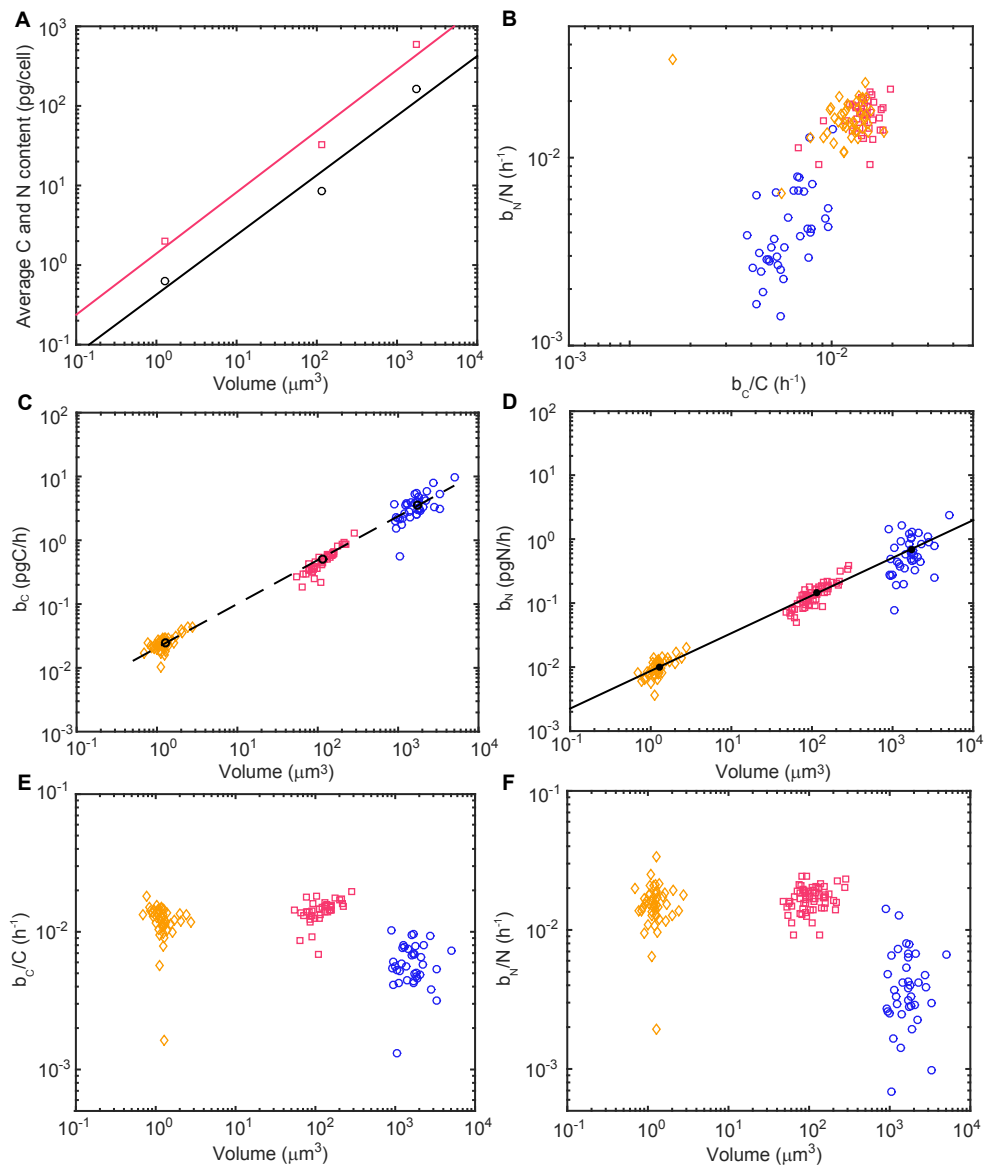


Figure 4.3 – A) Scaling of the average cell content of carbon (red squares) and nitrogen (black circles) with average cell volume for the three strains. Lines are linear least squares fit of log transformed values; B) Scatter plot of single-cell specific uptake rates of carbon and nitrogen,  $b_C/C$  and  $b_N/N$ ; C) Scaling of the single cell carbon uptake rate with single cell volume for the three strains; D) Scaling of the single cell nitrogen uptake rate with single cell volume for the three strains. In C) and D) black dots are species averages ( $\langle v \rangle, \langle b_i \rangle$ ),  $i=C,N$ , and black lines are linear least squares fit of ( $\log \langle v \rangle, \log \langle b_i \rangle$ ); E) Scaling of single cell carbon-specific carbon uptake rate  $b_C/C$  with single cell volume for the three strains; F) Scaling of single cell nitrogen-specific nitrogen uptake rate  $b_N/N$  with single cell volume for the three strains. Diamonds: SYN, Squares: SCE, Circles: CRY.

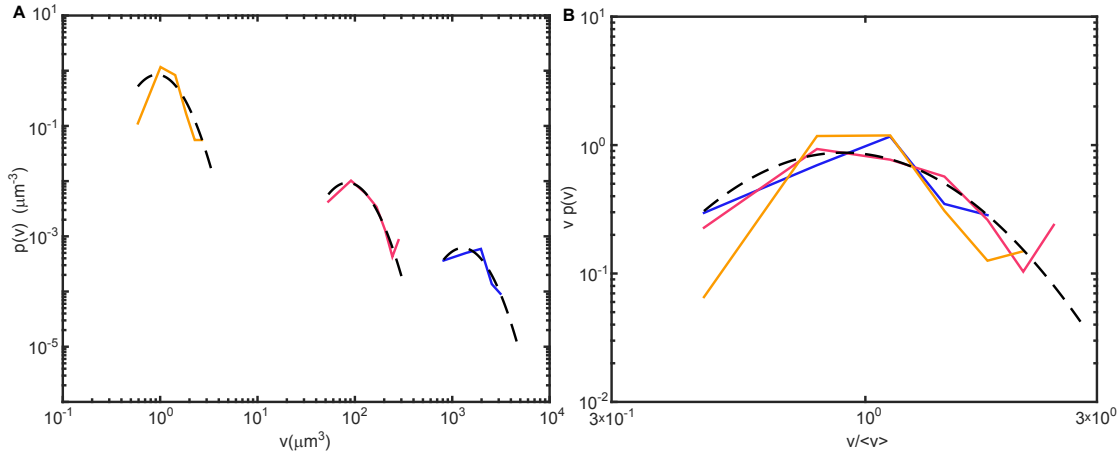


Figure 4.4 – A) Body size distributions of the three strains. Dashed lines are lognormal curves [Giometto et al., 2013],  $p(v|\langle v \rangle) = \frac{1}{\sqrt{2\pi}\sigma v} e^{-\frac{(\log(v/\langle v \rangle) + \sigma^2/2)^2}{2\sigma^2}}$ ; B) Collapse of the three distributions in A) according to Eq. (4.1). The dashed line in B) is a quadratic least square fit of the average of the three collapsed curves, yielding the estimate of the parameter  $\sigma$  of the lognormal fit. Orange: SYN, Pink: SCE, Blue: CRY.

### 4.3.3 Marginal probability distributions of body size

The intra-specific volume distributions  $p(v|\langle v \rangle)$ , shown in Fig. 4.4A). Fig. 4.4B) shows their successful collapse according to eq. (4.1). The function  $F$  is well fitted by a parabola in a log-log plot ( $R^2=0.86$ ), corresponding to a lognormal distribution of body sizes [Giometto et al., 2013],  $p(v|\langle v \rangle) = \frac{1}{\sqrt{2\pi}\sigma v} e^{-\frac{(\log(v/\langle v \rangle) + \sigma^2/2)^2}{2\sigma^2}}$  with  $\sigma = 0.46$  (95% confidence interval [0.38, 0.53]). The parameter  $\sigma$  was estimated by fitting a parabola to the average of the three rescaled volume distributions obtained by plotting  $\log(vp(v))$  versus  $\log(v/\langle v \rangle)$ . Note that this is a one-parameter fit, because a log-normal derived from a normal distribution with mean  $\mu$  and variance  $\sigma^2$  satisfies Eq. (4.1) only if  $\mu = -\sigma^2/2$  [Giometto et al., 2013]. The resulting lognormal fits of the size distribution are plotted in Fig. 4.4A). The K-Sample Anderson-Darling test on the three rescaled samples (see methods, section 4.2.7) does not reject the hypothesis that the samples come from the same distribution at the 5% confidence level (p-value: 0.17). Therefore, the data support the hypothesis in Eq. (4.1), in accordance with the previous results of Giometto et al. [2013].

### 4.3.4 Marginal probability distributions of uptake rates

The probability distributions of  $b_C$  and  $b_N$  for the three species are plotted in Figs. 4.5A) and C). Figs. 4.5B) and D) show the successful collapse of the three curves according to Eq. (4.2) for the exponent values optimizing the collapse, respectively,  $\alpha_C = 0.685 \pm 0.002$  and  $\alpha_N = 0.585 \pm 0.005$ . The error is computed as the value of the exponents at which the error



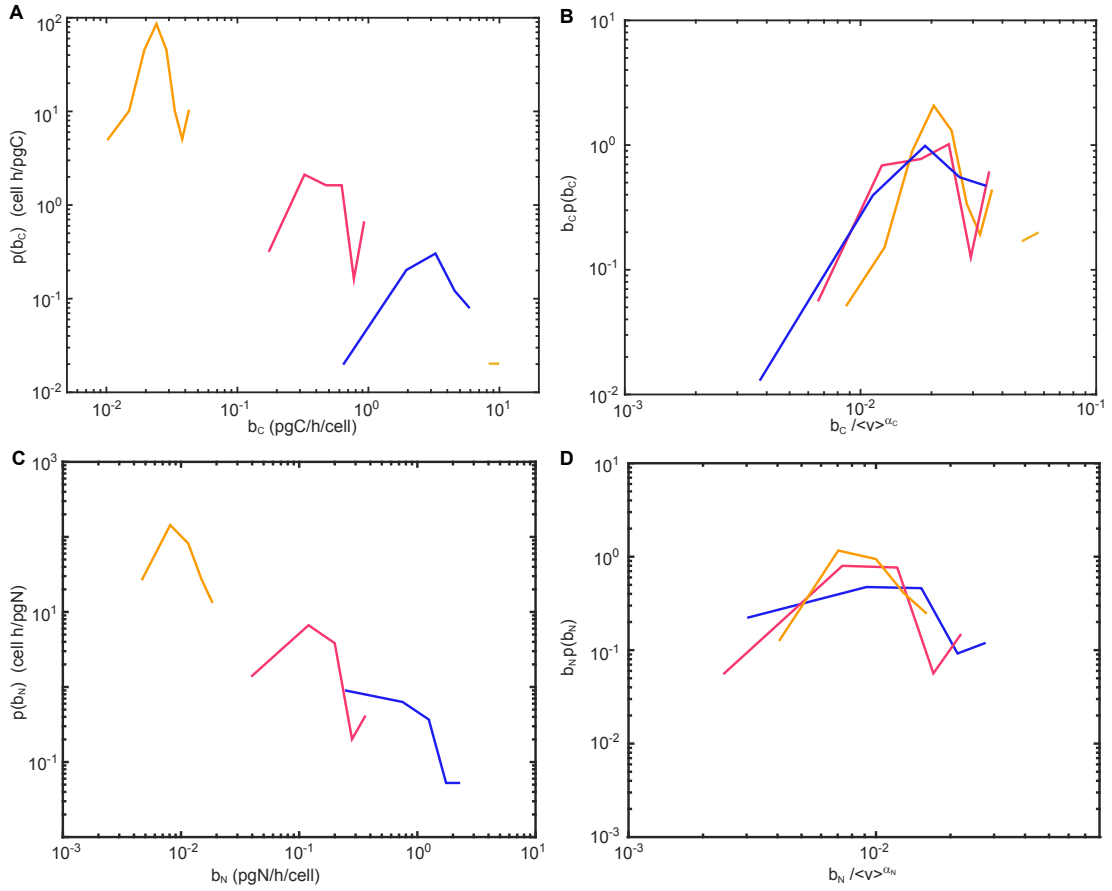


Figure 4.5 – A) Carbon uptake rate distributions of the three strains; B) Collapse of the three distributions in A) according to Eq. (4.2); C) Nitrogen uptake rate distributions of the three strains; D) Collapse of the three distributions in C) according to Eq. (4.2). Orange: SYN, Pink: SCE, Blue: CRY.

functional  $P_b$  (defined in [Bhattacharjee and Seno, 2001] by the sum of the areas enclosed between all pairs of rescaled curves) is 1% larger than its value at the minimum. Both estimates are compatible with the ones obtained from the scaling of the average uptake rate with the average body size, as expected for consistency (see section 4.2.1). For both C and N uptake rates, the K-Sample Anderson-Darling on the rescaled samples cannot reject the hypothesis that the three samples come from the same distribution at the 5% confidence level (p-values: 0.09, 0.07 respectively). Therefore, the data support the hypothesis in Eq. (4.2).

#### 4.3.5 Joint distribution of body size and uptake rates

As  $b$  and  $m$  are two correlated random variables, their joint probability distribution is not simply given by the products of the two marginals. Rather,  $p(m, b | \langle m \rangle) = p(m | \langle m \rangle) p(b | m, \langle m \rangle)$  where  $p(b | m, \langle m \rangle)$  is the metabolic rate distribution conditional both on typical species' mass

and individual mass. To study  $p(b|m, \langle m \rangle)$ , or equivalently the full joint distribution, one needs to bin the experimental data both in body size and in the metabolic rate. This procedure requires a number of analyzed cells for each species much larger than that achievable with the experimental methods used here. Nevertheless, based on the observed properties of marginals, a reasonable assumption for the scaling properties of the joint distribution is

$$p(v, b|\langle v \rangle) = \frac{1}{vb} H\left(\frac{v}{\langle v \rangle}, \frac{b}{\langle v \rangle^\alpha}\right), \quad (4.6)$$

where  $H$  is a species-independent function. See Appendix 4.5.3 for additional detail.

#### 4.3.6 Contribution of intrinsic variability of $b$ to the total observed variability

The observed variability of metabolic rate within one species is partly caused by the intra-specific heterogeneity in cell volume and partly by the intrinsic variability of metabolic rates, related to cell cycle stage, cell age, physiological conditions, etc. The importance of the intrinsic variability can be estimated by computing the fraction of the total observed intra-specific variance of  $b$  explained only by the intra-specific variance of  $m$  as a consequence of classical KL. Specifically, if there was no intrinsic variability and the metabolic rate of an individual of volume  $v$  was simply obtained as

$$b = b_0 v^\alpha, \quad (4.7)$$

the intra-specific distribution of  $b$  would be expected to have a variance

$$\text{var}(b) = b_0^2 \text{var}(v^\alpha) \quad (4.8)$$

where  $\text{var}(v^\alpha)$  is computed within each species. Computing this expected variance for each species, both for  $b_C$  and  $b_N$ , using the values of  $b_0$  and  $\alpha$  estimated by fitting the points  $(\langle v \rangle, \langle b \rangle)$ , one finds that the variance of the intraspecific distribution of volume explains only 34 to 58 % of the observed variance of C uptake rates within one species and only 14 to 31% of the variance of N uptake rates. Intrinsic variability is therefore shown to be a non-negligible contribution.

#### 4.3.7 Diffusion-limited and transport-limited uptake

Consider a spherical cell of radius  $R$  immersed in a medium containing nutrients. The cell uptakes the nutrient through the membrane, creating a nutrient gradient and therefore a flow towards itself, due to diffusion. The flow of nutrient through the entire membrane per unit time is the cell uptake rate.

Let  $\bar{\rho}$  be the nutrient concentration very far from the cell. If diffusion is the only process at work, the concentration  $\rho(r, t)$  at distance  $r$  from the center of the cell and at time  $t$

satisfies

$$\frac{\partial \rho}{\partial t} = D \nabla^2 \rho = D \frac{1}{r^2} \frac{d}{dr} \left[ r^2 \frac{d\rho}{dr} \right], \quad (4.9)$$

where the fact that  $\rho$  is spherically symmetric was used. At stationarity, this gives

$$0 = D \frac{1}{r^2} \frac{d}{dr} \left[ r^2 \frac{d\rho}{dr} \right] \implies r^2 \frac{d\rho}{dr} = c \quad (4.10)$$

with  $c$  a constant. Integrating,

$$\rho(r) = -\frac{c}{r} + c'. \quad (4.11)$$

The integration constants can be fixed by imposing boundary conditions  $\rho(\infty) = \bar{\rho}$  and  $\rho(R) = \rho_R$ , yielding

$$\rho(r) = \frac{R}{r} (\rho_R - \bar{\rho}) + \bar{\rho}. \quad (4.12)$$

Now, using Fick's law, the flux through the membrane (incoming nutrient per unit time and surface) is

$$J = -D \frac{d\rho}{dr} \Big|_R = -D \frac{\bar{\rho} - \rho_R}{R}. \quad (4.13)$$

where the minus sign indicates that it is directed towards the interior of the cell.

Now, two cases must be distinguished:

- if the membrane is able to uptake all the nutrient it comes in contact with,  $\rho(R) = 0$ . Then, the flux is  $J = -D \frac{\bar{\rho}}{R}$ . This case is called “diffusion limited”;
- if the flux potentially crossing the membrane is larger than what the membrane can sustain (due to limited number and efficiency of transport sites),  $J = J_{\max}$ , independent of  $R$ . This case is called “transport limited”.

Summarizing,  $J = \min\{J_{\max}, -D \frac{\bar{\rho}}{R}\}$ . For a fixed nutrient concentration, the crossover between the two situations happens for cells of radius  $R_c = -D \frac{\bar{\rho}}{J_{\max}}$ . For a fixed cell size, it happens for a nutrient concentration  $\bar{\rho}_C = -\frac{R J_{\max}}{D}$ . The uptake rate is obtained multiplying the flux by the cell surface:

$$b = 4\pi R^2 |J| = \begin{cases} 4\pi R^2 J_{\max} \propto \nu^{2/3} & \text{transport – limited} \\ 4\pi \frac{D}{\bar{\rho}} R \propto \nu^{1/3} & \text{diffusion – limited.} \end{cases} \quad (4.14)$$

The exponents of the scaling of nutrient uptake rate with cell volume found experimentally (sec. 4.3.1) suggest that the cells are not limited by diffusion. The higher values of

exponents measured ( $\sim 2/3$ ) could be due to transport limits or to other mechanisms not included in the simple analysis presented in this Box.

### 4.4 Discussion

For the development of a theoretical framework of ecosystem dynamics accounting for individual heterogeneity of relevant biological traits, empirical characterization of the scaling properties of metabolic rates and body size at the individual level is needed. The results of this Chapter support the validity of an approach based on general statistical features valid across biological scales. In fact, they suggest that intra-specific metabolic rate heterogeneity is successfully described by a scaling probability distribution in which the average body size is the only biological trait differentiating the distributions of different species. When confronting the enormous diversity of phytoplankton form and function, two broadly divergent (although not mutually exclusive) approaches exist, one emphasizing the existence of common constraints and master traits, such as cell size, that explain much of that diversity, the other emphasizing phylogenetic variability. The latter implies that taxonomic differences would be crucial to explain functional differences. The results obtained here show consistent patterns in the intra-specific variability of size-dependent metabolic rates of species that differ widely in their phylogenetic affiliation, therefore supporting body size as a fundamental trait to understand the structure and function of microbial ecosystems.

In Giometto et al. [2013], biotic and abiotic stresses were shown to act on the average  $\langle v \rangle$ , but not on the shape of the intra-specific body size distribution. One may expect that the same holds true for the metabolic rate distribution, although the experiments presented here did not investigate the effects of stress.

The properties of the statistical characterization of intra-specific metabolic rate variability supported by the experiments would suggest that maximum or minimum metabolic rates do not scale differently from the average, inter-specifically. This implies that discrepancies in the measured value of the scaling exponent  $\alpha$  could not be explained by discrepancies in the measuring method (say, targeting basal, field or maximum metabolic rates) but would rather reflect real shifts in the exponent value reflecting different degrees of allometry as a consequence of varying degrees of resource supply limitation [Rinaldo et al., 2002, Finkel et al., 2004].

Limitations of the current measurements ought to be pointed out at this point. In particular, further work should expand the size range considered and increase the sample size analyzed for each species. In fact, in terms of theoretical deductions, the limited number of individual cells that can be analyzed with the current preparation technology precludes proper identification of joint probability distributions. Also, the specific study organisms considered (freshwater phytoplankton) and a few experimental circumstances limit the claims of generality that may be deduced by these early trials, that should therefore be seen as a

proof of concept of the need for broader investigations. First, cultures were grown at likely sub-saturating irradiances, which may explain the fact that the observed mean size-scaling exponent for the whole data set is lower (0.69 for C and 0.59 for N) than earlier values obtained under light-saturating conditions ( $\sim 0.8 - 0.9$ ). In fact, low light is known to induce a shallower slope in the cell size-metabolic rate relationship [Finkel et al., 2004], one possible reason being that the larger the cells, the more they suffer from the package effect (a reduction in chlorophyll-specific light absorption), which becomes progressively more important as light levels decrease (low light induces higher pigment content and therefore a lower absorption efficiency). Nutrient concentration can also influence the scaling of uptake rate (Box 4.3.7). Further experiments should therefore vary both light intensity and nutrient concentration. Secondly, determining C and N content at the single-cell level would be a major advance. In fact, the results presented here rest on the assumption that individual elemental content can be predicted by cell volume, but their relationship at the intra-specific level has not been tested to date. Combining electron-probe X-ray microanalysis (XRMA) [Segura-Noguera et al., 2012], which allows the measurement of elemental content in single cells, with NanoSIMS analyses is a possible solution. Related to this, previous observations [Marañón et al., 2013] have shown that elemental content is a much better predictor of metabolic rate than cell volume. Therefore, relating metabolic rate to elemental content rather than volume would be a promising avenue for future investigations.

Finally, it should be noted that a prediction of Kleiber's rule is that mass-specific metabolic rates (units of  $\text{h}^{-1}$ ) scales as the  $-1/4$ -power of body mass. The fact that *Synechococcus* and *Scenedesmus* have similar mass-specific metabolic rates, despite having a  $\sim 15$ -fold difference in cell C and N, supports a significant deviation of phytoplankton metabolism from Kleiber's rule (which predicts that the mass-specific metabolic rate of *Scenedesmus* should be roughly half of that of *Synechococcus*). Such a deviation was found recently in a comprehensive analysis of the size-dependence of phytoplankton growth [Kremer et al., 2017].

In conclusion, the analysis reported in this Chapter suggests that the intra-specific variability in metabolic rate and size can be described in terms of general statistical features, potentially providing a key to a generalization of Kleiber's rule of metabolic rate size-scaling, currently assumed to play a central role in shaping ecosystems. Although these results should be seen as a proof of concept owing to the limited samples that current technology allows for single cell measurements, they are suggestive of scaling features of the marginal distributions, in turn pointing at the characters likely required of a joint probability distribution of mass and metabolic rates operating inter-species across scales.

## 4.5 Appendix

### 4.5.1 Properties of $p(v|\langle v \rangle)$ and $p(b|\langle v \rangle)$

This section is devoted to showing some properties of probability distributions of the form

$$p(x|C) = \frac{1}{x^\sigma} f\left(\frac{x}{C^\delta}\right), \quad (4.15)$$

with  $\sigma, \delta > 0$  and  $C$  a positive constant setting a typical scale for  $x$ . Eq. (4.15) applies both to the intra-specific distribution of body mass in Eq. (4.1) and to the intra-specific distribution of metabolic rate in Eq. (4.2). The experimental results in figures 4.4 B) and 4.5 B), D) as well as those in [Giometto et al., 2013] show that both  $F$  and  $G$  in Eqs. (4.1) and (4.2) are functions with an internal mode, decaying fast when their argument tends to 0 or  $\infty$ . Therefore, here the case in which the function  $f(x)$  decays to 0 both when  $x \rightarrow 0$  and when  $x \rightarrow \infty$  is considered. In 0, a decay  $f(x) \sim x^\epsilon$  with  $\epsilon > 0$  is sufficient to assure the integrability of  $x^k f(x)$  for  $k \geq -1$ , while in  $\infty$  a behavior  $f(x) \sim o(x^{-(k+1)})$  assures the integrability of  $x^k f(x)$ . Here, for simplicity,  $f(x)$  is assumed to decay faster than any power of  $x$  in  $\infty$ , a condition which assures the convergence of all moments and which is realized for the lognormal form of  $p(v|\langle v \rangle)$  found in [Giometto et al., 2013] and supported by the data presented in this Chapter.

First, let us show that normalization requires  $\sigma = 1$ . The condition to impose is

$$1 = \int_0^\infty p(x|C) dx = \int_0^\infty \frac{1}{x^\sigma} f\left(\frac{x}{C^\delta}\right) dx = C^{\delta(1-\sigma)} \int_0^\infty \frac{1}{y^\sigma} f(y) dy, \quad (4.16)$$

where the substitution  $y = x/C^\delta$  was performed within the integral. For the normalization constant not to depend on  $C$ , one must impose  $\sigma = 1$ . Additionally, the function  $f$  must be such that  $\int_0^\infty \frac{1}{y} f(y) dy = 1$ .

The moments of (4.15) are obtained as follows:

$$\langle x^k \rangle = \int_0^\infty x^k p(x|C) dx = \int_0^\infty x^{k-1} f\left(\frac{x}{C^\delta}\right) dx = C^{\delta k} \int_0^\infty y^{k-1} f(y) dy \propto C^{\delta k}, \quad (4.17)$$

where the substitution  $y = x/C^\delta$  was performed. The integral  $\int_0^\infty y^{k-1} f(y) dy$  converges for  $k \geq 0$  thanks to the properties of  $f$ . Considering the case  $k = 1$ , one has

$$\langle x \rangle \propto C^\delta. \quad (4.18)$$

This implies that, in the case  $C = \langle x \rangle$  (case of Eq. (4.1)),  $\delta = 1$  for consistency. Instead, in the case of Eq. (4.2) the above result implies

$$\langle b \rangle \propto \langle v \rangle^\alpha, \quad (4.19)$$

therefore  $\alpha$  can be identified as the exponent of the classical Kleiber's law.

For the moment of order  $k$ , from Eqs. (4.17) and (4.18) follows that all moments of the

distribution are proportional to the mean:

$$\langle x^k \rangle \propto \langle x \rangle^k. \quad (4.20)$$

One can also show that, when Eq. (4.2) holds, the maximum and minimum values of  $b$  in a sample scale with  $\langle v \rangle$  with the same exponent as the mean. In fact, for a species of abundance  $N$ , its individuals' metabolic rates are  $N$  draws from the distribution  $p(b|\langle v \rangle)$ . Therefore, the expected maximum value extracted is the value  $b_{\max}$  such that, on average, only one extracted value is  $\geq b_{\max}$ :

$$\frac{1}{N} = \int_{b_{\max}}^{\infty} p(b|\langle v \rangle) db = \int_{b_{\max}}^{\infty} \frac{1}{b} G\left(\frac{b}{\langle v \rangle^\alpha}\right) db = \int_{b_{\max}/\langle v \rangle^\alpha}^{\infty} \frac{1}{x} G(x) dx. \quad (4.21)$$

Differentiating with respect to  $\langle v \rangle$  gives

$$0 = \left[ \alpha b_{\max} \langle v \rangle^{-\alpha-1} - \langle v \rangle^{-\alpha} \frac{db_{\max}}{d\langle v \rangle} \right] \frac{\langle v \rangle^\alpha}{b_{\max}} G\left(\frac{b_{\max}}{\langle v \rangle^\alpha}\right), \quad (4.22)$$

that is,

$$0 = \alpha \frac{b_{\max}}{\langle v \rangle} - \frac{db_{\max}}{d\langle v \rangle} \quad (4.23)$$

giving

$$b_{\max} \propto \langle v \rangle^\alpha. \quad (4.24)$$

A similar computation gives the corresponding result for  $b_{\min}$ . As a consequence of this, if the variability of metabolic rates satisfy Eq. (4.2), the maximum and minimum metabolic rates measured in a sample of individuals of the same species scales with  $\langle v \rangle$  with the same exponent  $\alpha$  of the mean metabolic rate scaling.

#### 4.5.2 Single-cell content of $C$ ad $N$

Let  $C_i$  be the content of element  $i$  in a single cell of volume  $v$ . Experimental measurements yielded  $\langle C_i \rangle = k_{1,i} \langle v \rangle^{\gamma_i}$ , where  $k_{1,i}$  is a constant and the average is over cells of the same species. The assumption made to estimate single-cell elemental content is  $C_i = k_{2,i} v^{\gamma_i}$ , that is, the same scaling exponent holds at the intra- and inter-specific level. With this assumption, and using Eq. (4.1),

$$\begin{aligned} \langle C_i \rangle &= k_{2,i} \langle v^{\gamma_i} \rangle = k_{2,i} \int_0^{\infty} dv v^{\gamma_i} p(v|\langle v \rangle) = k_{2,i} \int_0^{\infty} dv v^{\gamma_i-1} F\left(\frac{v}{\langle v \rangle}\right) \\ &= k_{2,i} \langle v \rangle^{\gamma_i} \int_0^{\infty} dx x^{\gamma_i-1} F(x), \end{aligned} \quad (4.25)$$

## Chapter 4. Single-cell metabolic rates size-scaling in freshwater phytoplankton

where  $x = v/\langle v \rangle$ . Substituting  $k_{2,i} = C_i/v^{\gamma_i}$ , one has

$$\langle C_i \rangle = C_i \left( \frac{\langle v \rangle}{v} \right)^{\gamma_i} \int_0^\infty dx x^{\gamma_i-1} F(x). \quad (4.26)$$

Calling  $1/k_i := \int_0^\infty dx x^{\gamma_i-1} F(x)$ , finally

$$C_i = k_i \langle C_i \rangle \left( \frac{v}{\langle v \rangle} \right)^{\gamma_i}. \quad (4.27)$$

Using the values of  $\gamma_C = 0.8 = \gamma_N$  estimated experimentally and the function  $F$  fitted to the experimental volume distribution in section 4.3.3,  $k_C = k_N = 0.98$ .

### 4.5.3 Appendix: Possible forms of $p(v, b|\langle v \rangle)$

Experimental results allowed, although with a precision limited by the amount of data, the characterization of the marginal distribution of metabolic rates (Eq. (4.2)). Being  $b$  and  $v$  two correlated variables, their joint distribution is not simply given by the product of the two marginals, but rather by  $p(v, b|\langle v \rangle) = p(v|\langle v \rangle)p(b|v, \langle v \rangle)$ , where  $p(b|v, \langle v \rangle)$  is the distribution of metabolic rates conditional on both the species typical body size  $\langle v \rangle$  and the body size  $v$  of the individual. Now, the scaling properties of  $p(b|v, \langle v \rangle)$  cannot be inferred from the experimental data since they would require binning in both  $v$  and  $b$ . With the amount of data available, this is not possible as very few data would fall in each bin. Nevertheless, knowing that  $p(b|v, \langle v \rangle)$  must be compatible with the form of the marginal  $p(b|\langle v \rangle)$ , reasonable hypotheses can be made on its shape. In particular, two cases can be distinguished, depending on whether the intra-specific size-scaling of  $b$  has the same exponent  $\alpha$  as the inter-specific scaling, or a different one.:

- In the case in which the intra-specific scaling of  $b$  has the same exponent of the inter-specific scaling, a reasonable further assumption is

$$p(b|v, \langle v \rangle) = p(b|v) = \frac{1}{b} \tilde{G}_1 \left( \frac{b}{v^\alpha} \right), \quad (4.28)$$

where  $\tilde{G}_1$  is a function with the same properties of  $G$ . This assumption means that, knowing the individual body size  $v$ , the average species body size  $\langle v \rangle$  does not give me further information on the individual metabolic rate. This hypothesis implies that  $\langle b|v, \langle v \rangle \rangle = \int b p(b|v, \langle v \rangle) db \propto v^\alpha$  (see section 4.5.1), therefore  $\alpha$  is also the exponent of the intra-specific size-scaling of  $b$ . The joint distribution would then have the form

$$p(v, b|\langle v \rangle) = \frac{1}{vb} F \left( \frac{v}{\langle v \rangle} \right) \tilde{G}_1 \left( \frac{b}{v^\alpha} \right) = \frac{1}{vb} H_1 \left( \frac{v}{\langle v \rangle}, \frac{b}{\langle v \rangle^\alpha} \right) \quad (4.29)$$

where  $H_1(x, y) = F(x) \tilde{G}_1(\frac{y}{x^\alpha})$ . The function  $H_1(x, y)$  decays faster than  $F$  when  $x \rightarrow 0, \infty$ .



This form yields a marginal with the form in Eq. (4.2):

$$\begin{aligned} p(b|\langle v \rangle) &= \int_0^\infty p(v, b|\langle v \rangle) dv = \frac{1}{b} \int_0^\infty \frac{1}{v} H_1\left(\frac{v}{\langle v \rangle}, \frac{b}{\langle v \rangle^\alpha}\right) dv \\ &= \frac{1}{b} \int_0^\infty \frac{1}{x} H_1\left(x, \frac{b}{\langle v \rangle^\alpha}\right) dx = \frac{1}{b} G\left(\frac{b}{\langle v \rangle^\alpha}\right) \end{aligned} \quad (4.30)$$

where the substitution  $x = v/\langle v \rangle$  was performed and  $G(y) = \int_0^\infty x^{-1} H_1(x, y) dx$ , converging thanks to the properties of  $H(x, y)$ . The form of the marginal implies (see section 4.5.1) that  $\langle b|\langle v \rangle \rangle \propto \langle v \rangle^\alpha$ , i.e., the exponent of the inter-specific scaling is also  $\alpha$ .

- If the the intra-specific scaling of  $b$  has a different scaling exponent than the inter-specific one, one can hypothesize

$$p(b|v, \langle v \rangle) = \frac{1}{b} \tilde{G}_2\left(\frac{b}{v^\alpha}, \frac{v}{\langle v \rangle}\right), \quad (4.31)$$

where  $\tilde{G}_2(x, y)$  is a function that decays fast enough when  $x \rightarrow 0, \infty$ , to allow for normalization and convergence of moments (conditions similar to those required for the function  $G$  appearing in the marginal distribution). The normalization condition on this distribution implies that  $\int_0^\infty \frac{1}{x} \tilde{G}_2\left(x, \frac{v}{\langle v \rangle}\right) dx = 1$ . This condition is satisfied, for example, by a lognormal distribution of the form

$$p(b|v, \langle v \rangle) = \frac{1}{b} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(\log \frac{b}{v^\alpha} - \eta \log \frac{v}{\langle v \rangle})^2}{2\sigma^2}} \quad (4.32)$$

which is normalized and has mean  $\langle b|v, \langle v \rangle \rangle = e^{\sigma^2/2} v^{\alpha+\eta} \langle v \rangle^{-\eta}$ . In general, Eq. (4.31) implies

$$\langle b|v, \langle v \rangle \rangle = \int_0^\infty b p(b|v, \langle v \rangle) db = v^\alpha Q\left(\frac{v}{\langle v \rangle}\right) \quad (4.33)$$

where  $Q(y) = \int_0^\infty \tilde{G}_2(x, y) dx$  converges thanks to the properties of  $\tilde{G}_2$ . To have a power-law intra-specific size-scaling of  $b$ , one need therefore to ask  $Q(y) \propto y^\eta$ , which gives

$$\langle b|v, \langle v \rangle \rangle \propto v^{\alpha+\eta} \langle v \rangle^{-\eta}. \quad (4.34)$$

Therefore, intra-specifically (for fixed  $\langle v \rangle$ ) the scaling exponent is  $\alpha + \eta$ .

In this case, the joint distribution has the form

$$p(v, b|\langle v \rangle) = \frac{1}{vb} F\left(\frac{v}{\langle v \rangle}\right) \tilde{G}_2\left(\frac{b}{v^\alpha}, \frac{v}{\langle v \rangle}\right) = \frac{1}{vb} H_2\left(\frac{v}{\langle v \rangle}, \frac{b}{v^\alpha}\right) \quad (4.35)$$

where  $H_2(x, y) = F(x) \tilde{G}_2(y, x)$ . Note that this can also be rewritten as  $p(v, b|\langle v \rangle) = \frac{1}{vb} \tilde{H}_2\left(\frac{v}{\langle v \rangle}, \frac{b}{\langle v \rangle^\alpha}\right)$  with  $\tilde{H}_2(x, y) = H_2\left(x, \frac{y}{x^\alpha}\right)$ . Therefore, the scaling form of the joint distribution is formally the same in the two cases, although the behavior of the functions  $H_1$

## Chapter 4. Single-cell metabolic rates size-scaling in freshwater phytoplankton

---

and  $\tilde{H}_2$  differs. This form, again, yields a marginal with the form in Eq. (4.2):

$$\begin{aligned} p(b|\langle v \rangle) &= \int_0^\infty p(v, b|\langle v \rangle) dv = \frac{1}{b} \int_0^\infty \frac{1}{v} H_2\left(\frac{v}{\langle v \rangle}, \frac{b}{v^\alpha}\right) dv \\ &= \frac{1}{b} \int_0^\infty \frac{1}{x} H_2\left(x, \frac{b}{x^\alpha \langle v \rangle^\alpha}\right) dx = \frac{1}{b} G\left(\frac{b}{\langle v \rangle^\alpha}\right) \end{aligned} \quad (4.36)$$

where the substitution  $x = v/\langle v \rangle$  was performed and  $G(y) = \int_0^\infty x^{-1} H_2(x, y/x^\alpha) dx$ . This marginal implies

$$\langle b|\langle v \rangle \rangle \propto \langle v \rangle^\alpha, \quad (4.37)$$

that is, the inter-specific scaling is described by the exponent  $\alpha$ , different from the exponent of the intra-specific.

Both hypotheses are compatible with the  $p(b|\langle v \rangle)$  supported by the data, but the data do not allow distinguishing between the two for two reasons. First, the data points are not sufficient to verify the scaling form of  $p(b|v, \langle v \rangle)$ . Second, the intra-specific scaling of  $b$  observed in the data (figures 4.3 C and D of the main text) is dependent on our assumption that the C and N content of cells scale with volume at the intra-specific level with the same exponent of the inter-specific scaling (see section 4.2.6 of methods). This assumption needs verification before one can draw conclusions regarding the intra-specific scaling of metabolic rates.

# Conclusions

The present Thesis has shown that scaling laws prove a precious tool to describe the properties of complex ecosystems and organisms across scales by proposing a theoretical framework unifying the observed patterns and predicting their covariations and investigated the statistical properties of fluctuations around the average trends described by such patterns.

Chapters 1 and 2 were dedicated to the formulation and empirical validation of the theoretical framework linking several empirically observed patterns. The framework is based on the idea, inspired from statistical physics, that macroscopic observables (i.e., ecological laws) can be derived from the probability distribution of microscopic states of the system, which can be interpreted, in the ecological case, as the joint probability distribution of body mass and abundance of species. In Chapter 1 a scaling hypothesis for such joint probability distribution was proposed and its consequences in terms of linking relationships among the scaling exponents of ecological laws were derived. The scaling hypothesis was shown to be supported by a class of stochastic models of community dynamics including the minimal set of processes allowing to reproduce empirically observed patterns. In Chapter 2, the predictions of the framework were successfully compared to the available empirical data.

While previous attempts to link ecological laws exist, the approach presented in this thesis represents an advancement due to the larger number of patterns linked, the empirical verification of predictions and the flexibility of the proposed framework to the inclusion of deviations from pure power-laws or additional ecological details, e.g. intra-specific size distributions. The necessity of a general theory of ecology encompassing different spatial, temporal and biological scales has been voiced on several occasions [Levin, 1992, Marquet, 2017, Ellison, 2018]. The approach presented in Chapters 1 and 2 is a step in that direction, offering a framework that links organism-level patterns, like the size-scaling of metabolism, to community- and ecosystem-level patterns, but certainly much work is left to do. In fact, the framework presented here takes the shape of the empirical patterns as a fact, but, in the words of Levin [1992], “simple statistical description of patterns is a starting point, but correlations are no substitute for mechanistic understanding”. Therefore, future work will have to investigate the mechanistic basis of empirical patterns and clarify which processes determine the shape of ecological laws and the value of the independent exponents.

In Chapters 3 and 4 fluctuations around the average trends described by the Species-Area Relationship and by Kleiber’s law were studied by numerical modeling and by experiments

## Conclusions

---

at the single-cell level. In Chapter 3, the scaling with the mean of relative fluctuations of the number of species  $S$  coexisting in an ecosystem at stationarity was studied by numerical simulations in several stochastic models of community dynamics. The models include different processes possibly determining the dynamics of  $S$  in real ecosystems. The results obtained showed that including different processes in the dynamics may produce a different behavior of relative fluctuations, in some cases impairing the use of a deterministic SAR for predictions. This non-univocal result of modeling indicates that empirical investigations should sort out which fluctuation scaling scenario applies in real ecosystems, possibly depending on specific contexts. Empirical observations then, through a comparison with modeling results, will point out which processes have a central role in community assembly.

Chapter 4 explored, by theoretical and experimental methods, the possibility to formulate a generalized law describing the size-scaling of metabolic rate which acknowledges the intra-specific variability of body size and metabolic rate. Such variability was measured experimentally at the single-cell level for three species of freshwater phytoplankton. Results support the hypothesis that the intra-specific distributions of body size and metabolic rate enjoy statistical properties that hold across taxa and scales of body size, indicating that they are underlain by fundamental physiological processes. In fact, distributions pertaining to different species, covering three orders of magnitude in volume, are identical upon appropriate rescaling according to the mean species' body size. Body size is thus confirmed in its role of "master trait" having high explanatory power on many organismic properties. Nevertheless, this role is now interpreted in a probabilistic way, rather than in a deterministic one as in the classic Kleiber's law. Experimental results were limited, mainly, in two regards. First, specific nutrient uptake rates were measured (uptake rates per unit carbon or nitrogen present in the cell), therefore the computation of total uptake rate required an estimation of single-cell carbon and nutrient content, introducing uncertainty and, possibly, biases. Secondly, the number of cells which the current experimental method allowed to probe was too small to investigate the joint probability distribution of body size and metabolic rate. While a reasonable scaling hypothesis for such joint distribution could be put forward on the basis of the measured marginals, additional empirical support is needed. Further experimental developments are envisaged, planning to use alternative measurement techniques with higher throughput and which allow either the measurement of absolute uptake rates (or other proxies of metabolism, e.g. respiration rates) or the simultaneous measurement of single-cell nutrient content.

In this thesis, scaling was used in two forms. First, power-laws were used to describe how relevant quantities scale with the size of the system, the organism, or any size characterizing the problem. Secondly, scaling hypotheses were used to describe how the size of the system or a typical size characterizing the problem affect the probability distributions of the variables of interest. This approach provides a statistical characterization of such variables which does not depend on any model but can still contribute useful information, as proved in Chapters 1 and 4. It is therefore a powerful tool to interpret patterns across spatial and biological scales, where different specific mechanisms could be at work preventing an encompassing mechanistic description.

# Bibliography

- Ackermann, M. (2015). A functional perspective on phenotypic heterogeneity in microorganisms. *Nature Rev. Microbiol.*, 13:497–508.
- Ahluwalia, A. (2017). Allometric scaling in-vitro. *Sci. Rep.*, 7(42113).
- Anderson, R. M., Gordon, D. M., Crawley, M. J., and Hassell, M. P. (1982). Variability in the abundance of animal and plant species. *Nature*, 296:245.
- Arrhenius, O. (1921). Species and area. *J. Ecology*, 9:95–99.
- Azaele, S., Suweis, S., Grilli, J., Volkov, I., Banavar, J. R., and Maritan, A. (2016). Statistical mechanics of ecological systems: Neutral theory and beyond. *Rev. Mod. Phys.*, 88:035003.
- Ballesteros, F. J., Martinez, V. J., Luque, B., Lacasa, L., Valor, E., and Moya, A. (2018). On the thermodynamic origin of metabolic scaling. *Sci. Rep.*, 8:1448.
- Banavar, J. R., Cooke, T. J., Rinaldo, A., and Maritan, A. (2014). Form, function, and evolution of living organisms. *Proc. Natl. Acad. Sci. USA*, 111:3332–3337.
- Banavar, J. R., Damuth, J., Maritan, A., and Rinaldo, A. (2002). Supply-demand balance and metabolic scaling. *Proc. Natl. Acad. Sci. U. S. A.*, 99(16):10506–10509.
- Banavar, J. R., Damuth, J., Maritan, A., and Rinaldo, A. (2007). Scaling in ecosystems and the linkage of macroecological laws. *Phys. Rev. Lett.*, 98(6):068104.
- Banavar, J. R., Green, J. L., Harte, J., and Maritan, A. (1999a). Finite size scaling in ecology. *Phys. Rev. Lett.*, 83:4212–4214.
- Banavar, J. R., Maritan, A., and Rinaldo, A. (1999b). Size and form in efficient transportation networks. *Nature*, 399(May 1999):130–132.
- Banavar, J. R., Moses, M. E., Brown, J. H., Damuth, J., Rinaldo, A., Sibly, R. M., and Maritan, A. (2010). A general basis for quarter-power scaling in animals. *Proc. Natl. Acad. Sci. U. S. A.*, 107(36):15816–20.
- Bartels, H. (1982). Metabolic rate of mammals equals the 0.75 power of their body weight. *Exp. Bio. Med.*, 7:1–6.

- Bellin, A. and Rubin, Y. (1996). HYDRO\_GEN: A spatially distributed random field generator for correlated properties. *Stoch. Hydrol. Hydraul.*, 10:253–278.
- Benes, V. E. (1965). *Mathematical Theory of Connecting Network and Telephon Traffic*. Academic Press, New York and London.
- Bertuzzo, E., Suweis, S., Mari, L., Maritan, A., Rodriguez-Iturbe, I., and Rinaldo, A. (2011). Spatial effects for species persistence and implications for biodiversity. *Proc. Natl. Acad. Sci. USA*, 108(11):4346–4351.
- Bhattacharjee, S. M. and Seno, F. (2001). A measure of data collapse for scaling. *J. Phys. A-Mathematical Gen.*, 34(33):6375–6380.
- Blanchard, J. L., Heneghan, R. E., Everett, J. D., Trebilco, R., and Richardson, A. J. (2017). From Bacteria to Whales: Using Functional Size Spectra to Model Marine Ecosystems. *Trends Ecol. Evol.*, 32(3):174–186.
- Bonnet, S., Berthelot, H., Turk-Kubo, K., Cornet-Barthaux, V., Fawcett, S., Berman-Frank, I., Barani, A., Grégori, G., Dekaezemacker, J., Benavides, M., and Capone, D. G. (2016). Diazotroph derived nitrogen supports diatom growth in the South West Pacific: A quantitative study using nanoSIMS. *Limnol. Oceanogr.*, 61(5):1549–1562.
- Borile, C., Maritan, A., and Munoz, M. (2013). The effect of quenched disorder in neutral theories. *J. Stat. Mech*, 4:p04032.
- Brown, J. H. (1995). *Macroecology*. The University of Chicago Press, Chicago.
- Brown, J. H., Gillooly, J. F., Allen, A. P., Savage, V. M., and West, G. B. (2004). Toward a metabolic theory of ecology. *Ecology*, 85(7):1771–1789.
- Burness, G. P., Diamond, J., and Flannery, T. (2001). Dinosaurs, dragons, and dwarfs: the evolution of maximal body size. *Proc. Natl. Acad. Sci. U. S. A.*, 98(25):14518–14523.
- Calder, W. (1984). *Size, Function, and Life History*. Harvard University Press.
- Cavender-Bares, K. K., Rinaldo, A., and Chisholm, S. W. (2001). Microbial size spectra from natural and nutrient enriched ecosystems. *Limnol. Oceanogr.*, 46(4):778–789.
- Cencini, M., Pigolotti, S., and Munoz, M. (2012). What ecological factors shape species-area curves in neutral models? *PLOS ONE*, 7(6):e38232.
- Chisholm, R. A. and Lichstein, J. W. (2009). Linking dispersal, immigration and scale in the neutral theory of biodiversity. *Ecology Letters*, 12(12):1385–1393.
- Clarke, A., Rothery, P., and Isaac, N. (2010). Scaling of basal metabolic rate with body mass and temperature in mammals. *J. Anim. Ecol.*, 79:610–619.
- Clauset, A., Shalizi, C. R., and Newman, M. E. J. (2009). Power-Law Distributions in Empirical Data. *SIAM Rev.*, 51(4):661.

- Cohen, J. (2014). Stochastic population dynamics in a Markovian environment implies Taylor's power law of fluctuation scaling. *Theor. Ecol.*, 7:77–86.
- Cohen, J. E., Cohen, J. E., Jonsson, T., Jonsson, T., Carpenter, S. R., and Carpenter, S. R. (2003). Ecological community description using the food web, species abundance, and body size. *Proc. Natl. Acad. Sci. U. S. A.*, 100(4):1781–1786.
- Cohen, J. E., Xu, M., and Schuster, W. S. F. (2012). Allometric scaling of population variance with mean body size is predicted from Taylor's law and density-mass allometry. *Proc. Natl. Acad. Sci.*, 109(39):15829–15834.
- Condit, R., Lao, S., Pérez, R., Dolins, S. B., Foster, R., and Hubbell, S. (2012). Barro Colorado Forest Census Plot Data (Version 2012).
- Cuesta, J. A., Delius, G. W., and Law, R. (2017). Sheldon spectrum and the plankton paradox: two sides of the same coin—a trait-based plankton size-spectrum model. *Journal of Mathematical Biology*.
- Cyr, H., Peters, R. H., and Downing, J. A. (1997). Population density and community size structure: comparison of aquatic and terrestrial systems. *Oikos*, 80:139–149.
- Damuth, J. (1981). Population density and body size in mammals. *Nature*, 290:699–700.
- Damuth, J. (1987). Interspecific allometry of population-density in mammals and other animals: the independence of body-mass and population energy-use. *Biol. J. Linn. Soc.*, 31:193–246.
- Damuth, J. (1993). Cope's rule, the island rule and the scaling of mammalian population density. *Nature*, 365:748–750.
- DeLong, J., Okie, J., Moses, M., Sibly, R., and Brown, J. (2010). Shifts in metabolic scaling, production, and efficiency across major evolutionary transitions of life. *Proc. Natl. Acad. Sci. USA*, 107:12941–12945.
- Diamond, J. M. and Gilpin, M. E. (1980). Turnover Noise: Contribution to Variance in Species Number and Prediction from Immigration and Extinction Curves. *Am. Nat.*, 115(6):884–889.
- Dodds, P. S., Rothman, D. H., and Weitz, J. S. (2001). Re-examination of the "3/4-law" of metabolism. *J. Theor. Biol.*, 209(1):9–27.
- Dodson, S. (1992). Predicting crustacean zooplankton species richness. *Limnol. Oceanogr.*, 37(4):848–856.
- Dornelas, M., Gotelli, N. J., McGill, B., Shimadzu, H., Moyes, F., Sievers, C., and Magurran, A. E. (2014). Assemblage time series reveal biodiversity change but not systematic loss. *Science*, 344(6181):296–299.

- Drakare, S., Lennon, J., and Hillebrand, H. (2006). The imprint of the geographical, evolutionary and ecological context on species–area relationships. *Ecology Letters*, 9:215–227.
- Durrett, R. and Levin, S. (1996). Spatial models for species-area curves. *J. Theor. Biol.*, 179(2):119–127.
- Ellison, A. M. (2018). A Sense of Scale. *Bull. Ecol. Soc. Am.*, 99(2):173–179.
- Enquist, B. J. and Niklas, K. J. (2001). Invariant scaling relations across tree-dominated communities. *Nature*, 410(6829):655–660.
- Etienne, R. S. and Olff, M. (2004). A novel genealogical approach to neutral biodiversity theory. *Ecol. Lett.*, 7(3):170–175.
- Falkowski, P. G. (1998). Biogeochemical Controls and Feedbacks on Ocean Primary Production. *Science (80-. )*, 281(5374):200–206.
- Feldman, H. and MacMahon, T. (1983). The 3/4 mass exponent for energy metabolism is not a statistical artifact. *Respir. Physiol.*, 52:149–163.
- Finkel, Z. V., Irwin, A. J., and Schofield, O. (2004). Resource limitation alters the 3/4 size scaling of metabolic rates in phytoplankton. *Mar. Ecol. Prog. Ser.*, 273(September 2015):269–279.
- Fisher, M. E. and Barber, M. N. (1972). Scaling theory for finite-size effects in the critical region. *Phys. Rev. Lett.*, 28:1516–1519.
- Gherardi, M., Mandrà, S., Bassetti, B., and Cosentino Lagomarsino, M. (2013). Evidence for soft bounds in Ubuntu package sizes and mammalian body masses. *Proc. Natl. Acad. Sci. U. S. A.*, 110(52):21054–8.
- Gillespie, D. T. (2001). Approximate accelerated stochastic simulation of chemically reacting systems. *J. Chem. Phys.*, 115(4):1716–1733.
- Giometto, A., Altermatt, F., Carrara, F., Maritan, A., and Rinaldo, A. (2013). Scaling body size fluctuations. *Proc. Natl. Acad. Sci. U. S. A.*, 110(12):4646–4650.
- Giometto, A., Formentin, M., Rinaldo, A., Cohen, J. E., and Maritan, A. (2015). Sample and population exponents of generalized Taylor's law. *Proceedings of the National Academy of Sciences*, 112(25):7755–7760.
- Glazier, D. S. (2005). Beyond the 3/4: variation in the intra- and interspecific scaling of metabolic rate in animals. *Biol. Rev.*, 80:611–632.
- Gleason, H. (1922). On the relation between species and area. *Ecology*, 3:156–162.
- Goldstein, M. L., Morris, S. A., and Yen, G. G. (2004). Problems with fitting to the power-law distribution. *Eur. Phys. J. B - Condens. Matter Complex Syst.*, 41(2):255–258.



- Gosselain, V., Hamilton, P. B., and Descy, J.-P. (2000). Estimating phytoplankton carbon from microscopic counts: an application for riverine systems. *Hydrobiologia*, 438(1):75–90.
- Grilli, J., Barabas, G., and Allesina, S. (2015). Metapopulation persistence in random fragmented landscapes. *PLoS Computational Biol*, 11:e100251.
- Halfpenny, J. (2018). Small mammal disturbance data for Niwot Ridge from 1981/6/30 - 1990/8/23, yearly provided by the NSF supported Niwot Ridge Long-Term Ecological Research project and the University of Colorado Mountain Research Station.
- Hanski, I. (2013). Extinction debt at different spatial scales. *Anim. Conserv.*, 16(1):12–13.
- Hanski, I. (2016). *Messages from Islands. A Global Biodiversity Tour*. Univ. Chicago Press, London.
- Hanski, I. and Ovaskainen, O. (2000). The metapopulation capacity of a fragmented landscape. *Nature*, 404(6779):755–758.
- Harte, J., Smith, A. B., and Storch, D. (2009). Biodiversity scales from plots to biomes with a universal species-area curve. *Ecol. Lett.*, 12(8):789–797.
- Holley, R. and Liggett, T. (1975). Ergodic theorems for weakly interacting infinite systems and the voter model. *The Annals of Probability*, 3(4):643–663.
- Hoppe, P., Cohen, S., and Meibom, A. (2013). NanoSIMS: Technical Aspects and Applications in Cosmochemistry and Biological Geochemistry. *Geostand. Geoanalytical Res.*, 37(2):111–154.
- Hubbell, S. (2001). *The Unified Theory of Biodiversity and Biogeography*. Princeton Univ. Press, Princeton.
- Hubert, A. (2010). A sceptic's view: 'Kleiber law' or the '3/4 Rule' is neither a law nor a rule but rather an empirical approximation. *Systems*, 186:186–202.
- Kadanoff, L. P., Götze, W., Hamblen, D., Hecht, R., Lewis, E. A. S., Palciauskas, V. V., Rayl, M., Swift, J., Aspnes, D., and Kane, J. (1967). Static phenomena near critical points: Theory and experiment. *Rev. Mod. Phys.*, 39:395–431.
- Kennard, A. S., Osella, M., Javer, A., Grilli, J., Nghe, P., Tans, S. J., Cicuta, P., and Cosentino Lagomarsino, M. (2016). Individuality and universality in the growth-division laws of single e. coli cells. *Phys. Rev. E*, 93:012408.
- Kleiber, M. (1932). Body size and metabolism. *Hilgardia*, 6(11):315–353.
- Kleiber, M. (1947). Body size and metabolic rate. *Physiol. Rev.*, 6(27):511–541.
- Kolokotronis, T., Van Savage, Deeds, E. J., and Fontana, W. (2010). Curvature in metabolic scaling. *Nature*, 464(7289):753–756.

- Kremer, C., Thomas, M., and Litchman, E. (2017). Temperature- and size-scaling of phytoplankton population growth rates: Reconciling the eppley curve and the metabolic theory of ecology. *Limnol. Oceanogr.*, 62:1658–1670.
- Labra, F. A., Hernández-Miranda, E., and Quiñones, R. A. (2015). Dynamic relationships between body size, species richness, abundance, and energy use in a shallow marine epibenthic faunal community. *Ecology and Evolution*, 5:391–408.
- Labra, F. A., Marquet, P. A., and Bozinovic, F. (2007). Scaling metabolic rate fluctuations. *Proc. Natl. Acad. Sci. U. S. A.*, 104(26):10900–10903.
- Levin, S. (1992). The problem of pattern and scale in ecology. *Ecology*, 76:1943–1963.
- Lomolino, M. V. (1982). Species-area and species-distance relationships of terrestrial mammals in the Thousand Island Region. *Oecologia*, 54(1):72–75.
- Lomolino, M. V. (2000). Ecology's most general, yet protean pattern: the species–area relationship. *J. Biogeogr.*, 27(12):17–26.
- Lonsdale, W. M. (1999). Global Patterns of Plants Invasions and the Concept of Invasibility. *Ecology*, 80(5):1522–1536.
- MacArthur, R. H. and Wilson, E. O. (1963). An Equilibrium Theory of Insular Zoogeography. *Evolution (N. Y.)*, 17(4):373–387.
- MacArthur, R. H. and Wilson, E. O. (1967). *The theory of island biogeography*. Princeton University Press, Princeton, New Jersey, monographs edition.
- Marañón, E. (2015). Cell Size as a Key Determinant of Phytoplankton Metabolism and Community Structure. *Ann. Rev. Mar. Sci.*, 7(1):241–264.
- Marañón, E., Cermeño, P., López-Sandoval, D. C., Rodríguez-Ramos, T., Sobrino, C., Huete-Ortega, M., Blanco, J. M., and Rodríguez, J. (2013). Unimodal size scaling of phytoplankton growth and the size dependence of nutrient uptake and use. *Ecol. Lett.*, 16(3):371–9.
- Marañón, E., Cermeno, P., Rodriguez, J., Zubkov, M., and Harris, R. (2007). Scaling of phytoplankton photosynthesis and cell size in the ocean. *Limnol. Oceanogr.*, 52(5):2190–2198.
- Marquet, P. A. (2000). Invariants, scaling laws, and ecological complexity. *Science*, 289:1487–1488.
- Marquet, P. A. (2017). Integrating macroecology through a statistical mechanics of adaptive matter. *Proc. Natl. Acad. Sci. U. S. A.*, 114:10523–10525.
- Marquet, P. A., Navarrete, S. A., and Castilla, J. C. (1990). Scaling population density to body size in rocky intertidal communities. *Science*, 250(4984):1125–1127.
- Marquet, P. A., Quiñones, R. A., Abades, S., Labra, F., Tognelli, M., Arim, M., and Rivadeneira, M. (2005). Scaling and power-laws in ecological systems. *J. Exp. Biol.*, 208:1749–1769.

- Marquet, P. A. and Taper, M. L. (1998). On size and area: Patterns of mammalian body size extremes across landmasses. *Evol. Ecol.*, 12(2):127–139.
- May, R. M. (1988). How many species are there on earth? *Science*, 241:1441–1449.
- McMahon, P. and Bonner, J. (1983). *On Size and Life*. Scientific American.
- Mori, S., Yamaji, K., Ishida, A., Prokushkin, S. G., Masyagina, O. V., Hagihara, A., Hoque, a. T. M. R., Suwa, R., Osawa, A., Nishizono, T., Ueda, T., Kinjo, M., Miyagi, T., Kajimoto, T., Koike, T., Matsuura, Y., Toma, T., Zyryanova, O. a., Abaimov, A. P., Awaya, Y., Araki, M. G., Kawasaki, T., Chiba, Y., and Umari, M. (2010). Mixed-power scaling of whole-plant respiration from seedlings to giant trees. *Proc. Natl. Acad. Sci. U. S. A.*, 107(4):1447–1451.
- Muller-Landau, H. C., Condit, R. S., Harms, K. E., Marks, C. O., Thomas, S. C., Bunyavejchewin, S., Chuyong, G., Co, L., Davies, S., Foster, R., Gunatilleke, S., Gunatilleke, N., Hart, T., Hubbell, S. P., Itoh, A., Kassim, A. R., Kenfack, D., LaFrankie, J. V., Lagunzad, D., Lee, H. S., Losos, E., Makana, J. R., Ohkubo, T., Samper, C., Sukumar, R., Sun, I. F., Nur Supardi, M. N., Tan, S., Thomas, D., Thompson, J., Valencia, R., Vallejo, M. I., Muñoz, G. V., Yamakura, T., Zimmerman, J. K., Dattaraja, H. S., Esufali, S., Hall, P., He, E., Hernandez, C., Kiratiprayoon, S., Suresh, H. S., Wills, C., and Ashton, P. (2006). Comparing tropical forest tree size distributions with the predictions of metabolic ecology and equilibrium models. *Ecol. Lett.*, 9(5):589–602.
- Nee, S., Read, A. F., Greenwood, J. D., and Harvey, P. H. (1991). The relationship between abundance and body size in British birds. *Nature*, 351:312–313.
- Newman, M. E. J. (2005). Power laws, pareto distributions and zipf's law. *Contemp. Phys.*, 46(5):323–351.
- Newman, M. E. J. (2007). Power laws, pareto distributions and zipf's law. *Contemp. Phys.*, 46:323–351.
- Newmark, W. D. (1986). Species-area relationship and its determinants for mammals in western North American national parks. *Biol. J. Linn. Soc.*, 28(1-2):83–98.
- Nielsen, S. L. and Sand-Jensen, K. (1990). Allometric scaling of maximal photosynthetic growth rate to surface / volume ratio. *Limnol. Oceanogr.*, 35(1):177–181.
- Novosolov, M., Rodda, G. H., Feldman, A., Kadison, A. E., Dor, R., and Meiri, S. (2015). Power in numbers. The evolutionary drivers of high population density in insular lizards. *Glob. Ecol. Biogeogr.*, 25(1):87–95.
- Nuñez, J., Renslow, R., Cliff, J. B., and Anderton, C. R. (2018). NanoSIMS for biological applications: Current practices and analyses. *Biointerphases*, 13(3):03B301.
- Okie, J. G. and Brown, J. H. (2009). Niches, body sizes, and the disassembly of mammal communities on the Sunda Shelf islands. *Proc. Natl. Acad. Sci. U. S. A.*, 106(suppl. 2):19679–19684.

- Pernice, M., Meibom, A., Van Den Heuvel, A., Kopp, C., Domart-Coulon, I., Hoegh-Guldberg, O., and Dove, S. (2012). A single-cell view of ammonium assimilation in coral–dinoflagellate symbiosis. *ISME J.*, 6(7):1314–1324.
- Peters, R. (1986). *The Ecological Implications of Body Size*. Cambridge Univ. Press.
- Pigolotti, S. and Cencini, M. (2009). Speciation-rate dependence in species–area relationships. *Journal of Theoretical Biology*, 260(1):83 – 89.
- Pimm, S. and Raven, P. (1995). Biodiversity – extinction by numbers. *Nature*, 403:843–845.
- Preston, F. W. (1948). The Commonness, And Rarity, of Species. *Ecology*, 29(3):254–283.
- Preston, F. W. (1962). The Canonical Distribution of Commonness and Rarity: Part I. *Ecology*, 43(2):185–215.
- Quinones, R. A., Platt, T., and Rodríguez, J. (2003). Patterns of biomass-size spectra from oligotrophic waters of the Northwest Atlantic. *Prog. Oceanogr.*, 57(3):405–427.
- Rensch, B. (1948). Histological changes correlated with evolutionary changes of body size. *Evolution (N. Y.)*, 2:218–230.
- Rinaldo, A., Maritan, A., Cavender-Bares, K. K., and Chisholm, S. W. (2002). Cross-scale ecological dynamics and microbial size spectra in marine ecosystems. *Proc. R. Soc. B Biol. Sci.*, 269(1504):2051–2059.
- Rodríguez, J. and Mullin, M. (1986). Relation between biomass and body weight of plankton in a steady-state oceanic ecosystem. *Limnol. Oceanogr.*, 316–370:316–370.
- Rosenzweig, M. (1995). *Species Diversity in Space and Time*. Cambridge Univ. Press, New York.
- Rybicki, J. and Hanski, I. (2013). Species-area relationships and extinctions caused by habitat loss and fragmentation. *Ecol. Lett.*, 16:27–38.
- Savage, V. M., Allen, A. P., Brown, J. H., Gillooly, J. F., Herman, A. B., Woodruff, W. H., and West, G. B. (2007). Scaling of number, size, and metabolic rate of cells with body size in mammals. *Proceedings of the National Academy of Sciences*, 104(11):4718–4723.
- Savage, V. M., Gillooly, J. F., Woodruff, W. H., West, G. B., Allen, A. P., Enquist, B. J., and Brown, J. H. (2004). The predominance of quarter-power scaling in biology. *Funct. Ecol.*, 18(2):257–282.
- Schreiber, F., Littmann, S., Lavik, G., Escrig, S., Meibom, A., Kuypers, M., and Ackermann, M. (2016). Phenotypic heterogeneity driven by nutrient limitation promotes growth in fluctuating environments. *Nat. Microbiol.*, 1(6):1–7.
- Segura-Noguera, M., Blasco, D., and Fortuño, J. (2012). An improved energy-dispersive X-ray microanalysis method for analyzing simultaneously carbon, nitrogen, oxygen, phosphorus, sulfur, and other cation and anion concentrations in single natural marine microplankton cells. *Limnol. Oceanogr. Methods*, 10:666–680.

- Ser-Giacomi, E., Zinger, L., Malviya, S., De Vargas, C., Karsenti, E., Bowler, C., and Monte, S. D. (2018). Ubiquitous abundance distribution of non-dominant plankton across the world's ocean. *bioRxiv*.
- Sheldon, R. W., Prakash, A., and Sutcliffe, H. (1972). The size distribution of particles in the ocean. *Limnol. Oceanogr.*, XVII(3):327–340.
- Shem-Tov, Y., Danino, M., and Shnerb, N. M. (2017). Solution of the spatial neutral model yields new bounds on the Amazonian species richness. *Scientific Reports*, 7:42415.
- Siemann, E., Tilman, D., and Haarstad, J. (1996). Insect species diversity, abundance and body size relationships. *Nature*, 380(6576):704–706.
- Simberloff, D. (2004). Community Ecology: Is It Time to Move On?: (An American Society of Naturalists Presidential Address). *Am. Nat.*, 163(6):787–799.
- Simini, F., Anfodillo, T., Carrer, M., Banavar, J. R., and Maritan, A. (2010). Self-similarity and scaling in forest communities. *Proc. Natl. Acad. Sci. U. S. A.*, 107(17):7658–62.
- Smith, F. A., Lyons, K. S., Ernest, M. S. K., Jones, K. E., Kaufman, D. M., Dayan, T., Marquet, P. A., Brown, J. H., and Haskell, J. P. (2003). Body mass of late quaternary mammals. *Ecology*, 84(12):3403.
- Smith, V. H., Foster, B. L., Grover, J. P., Holt, R. D., Leibold, M. a., and Denoyelles, F. (2005). Phytoplankton species richness scales consistently from laboratory microcosms to the world's oceans. *Proc. Natl. Acad. Sci. U. S. A.*, 102(12):4393–4396.
- Solomon, S. and Levy, M. (1996). Spontaneous scaling emergence in generic stochastic systems. *Int. J. Mod. Phys. C*, 07:745–751.
- Sornette, D. and Cont, R. (1997). Convergent multiplicative processes repelled from zero: Power laws and truncated power laws. *J. Phys. I France*, 7:431–444.
- Southwood, T. R. E., May, R. M., and Sugihara, G. (2006). Observations on related ecological exponents. *Proc. Natl. Acad. Sci. U. S. A.*, 103(18):6931–6933.
- Sprules, W. G. and Barth, L. E. (2016). Surfing the biomass size spectrum: some remarks on history, theory, and application. *Can. J. Fish. Aquat. Sci.*, 73(4):477–495.
- Stanley, H. and Amaral, L. (2000). Scale invariance and universality: organizing principles in complex systems. *Phys. A Stat. . . .*, 281(1):60–68.
- Stanley, H. E. (1999). Scaling, universality, and renormalization: Three pillars of modern critical phenomena. *Rev Mod Phys*, 71(2):S358—S366.
- Stegen, J. C. and White, E. P. (2008). On the relationship between mass and diameter distributions in tree communities. *Ecol. Lett.*, (11):1287–1293.

- Suweis, S., Bertuzzo, E., Mari, L., Rodriguez-Iturbe, I., Maritan, A., and Rinaldo, A. (2012a). On species persistence-time distributions. *Journal of Theoretical Biology*, 303:15 – 24.
- Suweis, S., Rinaldo, A., and Maritan, A. (2012b). An exactly solvable coarse-grained model for species diversity. *J. Stat. Mech. Theory Exp.*, 2012(07):P07017.
- Taylor, L., Woiwod, I., and Perry, J. (1980). Variance and the large scale spatial stability of aphids, moths and birds. *J. Anim. Ecol.*, 49(3):831–854.
- Taylor, L. R. (1961). Aggregation, variance and the mean. *Nature*, 189(4766):732–735.
- Thomas, C. D., Cameron, A., Green, R. E., Bakkenes, M., Beaumont, L. J., Collingham, Y. C., Erasmus, B. F. N., de Siqueira, M. F., Grainger, A., Hannah, L., Hughes, L., Huntley, B., van Jaarsveld, A. S., Midgley, G. F., Miles, L., Ortega-Huerta, M. A., Townsend Peterson, A., Phillips, O. L., and Williams, S. E. (2004). Extinction risk from climate change. *Nature*, 427:145.
- Thompson, D'Arcy W. (2001). *The Unified Theory of Biodiversity and Biogeography*. Princeton Univ. Press, Princeton.
- Triantis, K., Guilhaumon, F., and Whittaker, R. (2012). The island species–area relationship: biology and statistics. *Journal of Biogeography*, 39(2):215–231.
- Vasdekis, A. E. and Stephanopoulos, G. (2015). Review of methods to probe single cell metabolism and bioenergetics. *Metab. Eng.*, 27:115–135.
- Vellend, M., Baeten, L., Myers-Smith, I. H., Elmendorf, S. C., Beauséjour, R., Brown, C. D., De Frenne, P., Verheyen, K., and Wipf, S. (2013). Global meta-analysis reveals no net change in local-scale plant biodiversity over time. *Proceedings of the National Academy of Sciences*, 110(48):19456–19459.
- Welkenhuysen, N., Borgqvist, J., Backman, M., Bendrioua, L., Goksör, M., Adiels, C. B., Cvijovic, M., and Hohmann, S. (2017). Single-cell study links metabolism with nutrient signaling and reveals sources of variability. *BMC Syst. Biol.*, 11:59.
- West, G. B. (1999). The origin of universal scaling laws in biology. *Phys. A Stat. Mech. its Appl.*, 263(1-4):104–113.
- West, G. B. and Brown, J. H. (2004). Life's universal scaling laws. *Phys. Today*, 57:36–42.
- West, G. B., Brown, J. H., and Enquist, B. J. (1997). A General Model for the Origin of Allometric Scaling Laws in Biology. *Science*, 276(5309):122–126.
- West, G. B., Woodruff, W. H., and Brown, J. H. (2002). Allometric scaling of metabolic rate from molecules and mitochondria to cells and mammals. *Proc. Natl. Acad. Sci. U. S. A.*, 99 Suppl 1:2473–2478.
- West, G.B. et al. (2003). Why does metabolic rate scale with body size? *Nature*, 421(4):713–714.

- White, C. and Seymour, R. (2005). Allometric scaling of mammalian metabolism. *J. Exp. Biol.*, 208:1611–1619.
- White, E. P., Enquist, B. J., and Green, J. L. (2008). On estimating the exponents of power-law frequency distributions (Appendix A). *Ecology*, 89:905–912.
- White, E. P., Ernest, S. K. M., Kerkhoff, A. J., and Enquist, B. J. (2007). Relationships between body size and abundance in ecology. *Trends Ecol. Evol.*, 22(6):323–330.
- Wright, S. (1931). Evolution in mendelian populations. *Genetics*, 16:97–159.
- Zaoli, S., Giometto, A., Giezendanner, J., Maritan, A., and Rinaldo, A. (2018). On the probabilistic nature of the species-area relation.
- Zaoli, S., Giometto, A., Maritan, A., and Rinaldo, A. (2017). Covariations in ecological scaling laws fostered by community dynamics. *Proceedings of the National Academy of Sciences*, 114(40):10672–10677.
- Zillio, T., Banavar, J. R., Green, J. L., Harte, J., and Maritan, A. (2008). Incipient criticality in ecological communities. *Proceedings of the National Academy of Sciences*, 105(48):18714–18717.
- Zimmerman, J., Comita, L., Thompson, J., Uriarte, M., and Brokaw, N. (2010). Patch dynamics and community metastability of a subtropical forest: Compound effects of natural disturbance and human land use. *Landsc. Ecol.*, 25:1099–1111.
- Zimmermann, M., Escrig, S., Hübschmann, T., Kirf, M. K., Brand, A., Inglis, R. F., Musat, N., Müller, S., Meibom, A., Ackermann, M., and Schreiber, F. (2015). Phenotypic heterogeneity in metabolic traits among single cells of a rare bacterial species in its natural environment quantified with a combination of flow cell sorting and nanosims. *Front. Microbiol.*, 6:243–254.





# Curriculum Vitae

## Personal information

Name: Silvia

Surname: Zaoli

Date of Birth: July 3th, 1990

Nationality: Italian

E-mail: [silvia.zaoli@gmail.com](mailto:silvia.zaoli@gmail.com), [silvia.zaoli@epfl.ch](mailto:silvia.zaoli@epfl.ch)

## Research interests

Theoretical ecology, theoretical biology, mathematical modeling of ecosystems, community dynamics, complex systems, scaling, statistical physics.

## Education

- **Dec 2014-July 2018:** PhD, École Polytechnique de Lausanne (EPFL), Lausanne, Switzerland
- **Oct 2013-July 2014:** Master of Physics, Università degli Studi di Padova, *110 cum laude* (GPA: 29.7/30)
- **Oct 2009- July 2012:** Bachelor of Physics, Università degli Studi di Padova, *110 cum laude* (GPA: 29.0/30)

## Certifications

- **2014:** TOEFL (Test of English as a Foreign Language) (109/120)
- **2008:** FCE (First Certificate in English)

### Talks and posters at international conferences

- **April 2018:** Laws in Ecology: a unifying framework, Poster contribution at the workshop “Stochastic models in ecology and evolutionary biology”, Venice
- **February 2018:** A Finite-size Scaling Framework Uncovers the Covariations of Ecological Scaling Laws, Quantitative Life Science Guest Seminar at ICTP, Trieste
- **June 2016:** Covariations in ecological scaling laws fostered by eco-evolutionary dynamics, Oral contribution presented at the summer school "Quantitative Laws II", Fondazione Alessandro Volta, Como, Italy
- **May 2016:** Covariations in ecological scaling laws fostered by eco-evolutionary dynamics, Special SIAM seminar, EAWAG, Dübendorf
- **September 2015:** Scaling exponents of macroecological laws: are they linked?, Oral contribution presented at the Conference of complex systems (CCS), Tempe, USA
- **September 2015:** Consequences of limited availability of resources on scaling exponents of interrelated macroecological laws, Oral contribution presented at the conference "Living systems, from interaction patterns to critical behaviour", Venice, Italy
- **August 2015:** Consequences of limited availability of resources on scaling exponents of interrelated macroecological laws, Oral contribution presented at the 100<sup>th</sup> Ecological Society of America (ESA) annual meeting, Baltimore, USA
- **October 2014:** A stochastic model for the motility of self-propelled microorganisms, Poster contribution at the Venice meeting on fluctuations in small complex systems II, Istituto veneto di scienze, lettere e arti, Venezia, Italy
- **June 2014:** A stochastic model for the motility of self-propelled microorganisms, Poster contribution at the XIX National Conference on Statistical Physics and Complex Systems, Università degli Studi di Parma, Parma, Italy

### Participation to workshops and summer schools

- **February 2017:** Workshop: Phytoplankton biodiversity, dynamic eco– physiology, and ecosystem function, Leibniz Center for Tropical Marine Ecology, Bremen, Germany
- **June 2016:** Summer school "Quantitative Laws II", Fondazione Alessandro Volta, Como, Italy
- **22 June-3 July 2015:** Doctoral training in statistical physics, École de physique des Houches, Les Houches, France

## List of Publications

Zaoli S., Giometto A., Maritan A., Rinaldo A., Covariations in ecological scaling laws fostered by eco-evolutionary dynamics, PNAS 114(40):10672-10677, 2017

## Forthcoming Publications

*In review:* Zaoli S., Giometto A., Giezendanner J., Maritan A., Rinaldo A., On the probabilistic nature of the species-area relation.

## Teaching activities

- **2017:** Assistant for the Mise à Niveau Course (Algebra and Geometry modules), bachelor of Engineering at EPFL
- **2016:** Assistant for Linear Algebra, bachelor of Engineering at EPFL
- **February 2016:** Educational talk “*A che cosa serve la ricerca scientifica?*” (“*Why do we need scientific research?*”), Scientific High school A. Serpieri, Rimini

## Language skills

- Italian, Mother tongue
- English, C1
- French, C1

## Computer skills

- Programming languages: C++
- Software: Matlab, Wolfram Mathematica, R, Adobe Illustrator, Office (Word, Excel, PowerPoint)

## Laboratory skills

Cell culturing, phytoplankton, flow cytometry, NanoSIMS