

## TABLE OF CONTENTS

	Page
<b>1 Introduction</b>	<b>1</b>
1.1 Motivation . . . . .	1
1.2 Literature review for immigration . . . . .	2
1.3 Community ecology . . . . .	4
1.3.1 Species interactions . . . . .	5
1.3.2 The role of structure . . . . .	6
1.3.3 Network generation . . . . .	8
1.4 Ecology <i>in silico</i> . . . . .	8
1.5 Habitat loss . . . . .	9
1.5.1 Modelling HL . . . . .	10
1.5.2 Beyond species richness . . . . .	10
1.6 Outline of thesis . . . . .	10
1.6.1 Habitat loss . . . . .	13
1.6.2 Communities of single and multiple interaction types . . . . .	13
1.6.3 Spatially explicit model and metrics . . . . .	15
1.6.4 Modelling Habitat Loss . . . . .	16
1.6.5 The effects of habitat loss . . . . .	16
<b>2 Methodology</b>	<b>17</b>
2.1 Introduction . . . . .	17
2.2 The interaction network . . . . .	17
2.2.1 The niche model . . . . .	18
2.2.2 Trophic constraints . . . . .	21
2.2.3 Link replacement . . . . .	21
2.3 Individual-based model . . . . .	23
2.3.1 Local rules . . . . .	23
2.3.2 Model Parameters . . . . .	26
2.4 Modelling habitat loss . . . . .	27
2.5 Implementation . . . . .	27

---

**TABLE OF CONTENTS**

---

2.6	Dynamics of the model . . . . .	29
2.7	Ecological metrics and analysis methods . . . . .	30
2.7.1	Biodiversity metrics . . . . .	31
2.7.1.1	Rank abundance distributions . . . . .	33
2.7.2	Stability metrics . . . . .	34
2.7.2.1	Invariability . . . . .	36
2.7.3	Spatial metrics . . . . .	37
2.7.4	Network metrics . . . . .	38
2.7.4.1	Qualitative network descriptors . . . . .	38
2.7.4.2	Quantitative network descriptors . . . . .	39
2.7.4.3	Interaction strength metrics . . . . .	41
<b>3</b>	<b>Habitat loss with high immigration</b>	<b>43</b>
3.1	Introduction . . . . .	43
3.2	Methods . . . . .	44
3.2.1	Simulation procedure . . . . .	44
3.2.2	Sampling and analysis . . . . .	45
3.3	Results . . . . .	45
3.3.1	Diversity . . . . .	46
3.3.2	Network properties . . . . .	54
3.3.3	Stability and spatial metrics . . . . .	59
3.3.4	Invariability . . . . .	64
3.4	Discussion . . . . .	66
3.5	Conclusion . . . . .	73
<b>4</b>	<b>Persistence and Stationarity</b>	<b>75</b>
4.1	Introduction . . . . .	75
4.2	Persistence in closed communities . . . . .	76
4.2.1	Mutualistic-to-antagonistic interaction ratio . . . . .	77
4.2.2	Reproduction rate . . . . .	79
4.2.3	Landscape size . . . . .	83
4.2.4	Number of initial species . . . . .	84
4.2.5	Network structure . . . . .	86
4.2.6	Summary . . . . .	86
4.3	Temporal variability and immigration . . . . .	89
4.4	Second-order stationarity . . . . .	90
4.4.1	Tests for stationarity . . . . .	90
4.4.2	Characterising the tests . . . . .	91
4.4.3	General stationarity results . . . . .	99

---

TABLE OF CONTENTS

4.5	Determinism tests . . . . .	103
4.5.1	Recurrence quantification analysis . . . . .	103
4.5.1.1	Results . . . . .	104
4.6	Convergence and repeatability . . . . .	105
4.6.1	Convergence . . . . .	107
4.6.2	Repeatability . . . . .	108
4.7	Conclusion . . . . .	113
<b>5</b>	<b>Varying Immigration rate</b>	<b>115</b>
5.1	Motivation . . . . .	115
5.2	Experimental approach . . . . .	116
5.3	Initial analysis . . . . .	118
5.3.1	Estimator comparison . . . . .	118
5.3.2	Diversity . . . . .	119
5.3.3	Variability and interaction strengths . . . . .	122
5.3.4	Dependence on immigration . . . . .	125
5.4	Bivariate Analysis . . . . .	125
<b>6</b>	<b>Inferring species interactions</b>	<b>127</b>
6.1	Motivation . . . . .	127
6.2	Methodology . . . . .	128
6.2.1	Interaction strength . . . . .	129
6.2.2	Population models . . . . .	130
6.2.3	Simulation procedure . . . . .	133
6.2.4	Numerical estimation method . . . . .	134
6.2.5	Examples . . . . .	137
6.3	Results . . . . .	140
6.3.1	Linear model . . . . .	140
6.3.2	Holling model . . . . .	143
6.3.3	Range sampling . . . . .	143
6.4	TEMP : other results . . . . .	143
6.5	Application to IBM (optional) . . . . .	144
6.5.1	Testing functional response (and intrinsic growth functions) . . . . .	146
6.5.2	2 Species IBM model . . . . .	147
6.5.3	Extend methodology (3, 4 and 5 species) . . . . .	149
6.5.4	Results . . . . .	149
6.5.5	2 species . . . . .	151
6.5.6	3 species . . . . .	158
6.5.7	5 species . . . . .	166

TABLE OF CONTENTS

---

6.6 Discussion . . . . .	175
<b>Bibliography</b>	<b>177</b>

## INTRODUCTION

### 1.1 Motivation

An ecosystem is a subset of the *biosphere* - the entirety of the living systems on this planet. These biological systems are closely coupled, in a two way relationship, to the abiotic systems of the planet to such an extent that Lovelock, and other proponents of the *Gaia theory*, suggest that the biotic and abiotic components together form a single homeostatic system which maintains conditions that are harmonious to life. In the strongest version of the theory the planet itself is a living system. This is not the place to argue for or against the theory, but regardless of its validity it remains undeniable that living systems generate significant effects at a planetary scale[REFS]. Humans, as one component of the biosphere, are fairly unique not only in the extent of their impact on planetary systems, but also their *potential ability* to make reasoned decisions about collective actions based on knowledge of their impact. Given that the rate of species extinctions in the last century has been conservatively estimated at between 8 and 100 times the background extinction rate [?], and that the major drivers for these extinctions are anthropogenic [REF], it is abundantly clear that humans are failing to realise the aforementioned potential. Arguably the main reason for this failure is the systems of organisation..hard to make reasoned decisions when don't have sound theory!..but also lack of understanding about how ecosystems function..Second paragraph on stewardship, conservation and restoration.. This is my reason for studying the field of ecology....

Habitat loss/alteration and climate change as the main anthropogenic drivers..Focus here on habitat loss. Community ecology - tools for the study of localised groups of species. Computational approach - hypothesis generation - defend bottom-up modelling of complex systems - make clear what it can and cannot do..

The topic of the current thesis falls under the generously sized umbrella that is *complexity*

science. Ladyman, Lambert and Weisner review the various definitions of a complex system [66]. They attempt to synthesise a concrete definition of this concept, which has previously been rather loosely and carelessly applied. The authors arrive at the following tentative definition of physical complexity:

*"A complex system is an ensemble of many elements which are interacting in a disordered way, resulting in robust organisation and memory."*

They assert that this definition contains conditions necessary, but perhaps not sufficient for complexity, and propose an alternative *data driven* definition under which a complex system is one that generates data with high Statistical Complexity [21]. Indeed to define a complex system is no easy task. Depending on who you ask there may be further necessary conditions such as *emergence*, *hierarchical organisation*, *de-centralised control* and even anti-reductionist properties such as *top-down causation*. If a definition may be reached then complexity science can be easily understood as the study of such systems. However, given the lack of a concrete definition of complex systems, many of my contemporaries (if concerned with the question at all) prefer a more *pragmatic approach* to the question - what is complexity science? A fair summary of the consensus view is that complexity science represents a broad set of mathematical and computational tools which find application in increasingly diverse areas of the natural and social sciences. This expansion is driven by the ever increasing availability of data; advances in high performance computing; and the development of the tools themselves. For some the tools used and the fields of study are too disjoint to call complexity science a unified field. It is my view that this pragmatic approach is not dissimilar to the *instrumentalist*, or "shut-up-and-calculate", interpretation of quantum mechanics[85], which finds popular support in the face of the difficult and unresolved philosophical questions posed by that theory. Further discussion on the nature of complexity is not relevant to this thesis other than to say that, if an accepted definition of a complex system did exist, then its criteria would surely be met by the *ecological community*<sup>1</sup>.

We now introduce key areas of theory??

## 1.2 Literature review for immigration

(From chapter 5) Possibly include here a summary of recent work on immigration, including IBT and meta-community theory. (Alternative is that this goes in introductory chapter.)

IBT -> species area relationship (SAR), does high immigration reduce this effect. [114] - Search for immigration.

[59] Many studies of dispersal/immigration take a metacommunity approach. In general higher dispersal between communities is found to promote better species richness, and lower variation between communities. However we are interested in...

---

<sup>1</sup>Actually this may not be quite right! E.g. neutral theory. We treat it as a complex systems though..

[67] shows, using a metapopulation model, that competitive plant communities benefit from high immigration. In these communities competition for space would lead to the extinction of all but one species. Therefore some immigration is required for diversity to exist. (Similar to our case of zero IR. Is there evidence for spatial competition in our simulations?) Some metrics change with IR some do not (check which ones, do they agree with their experiment, and ours?) They have species specific immigration rates, and an intensity parameter (otherwise same as ours with effective rate proportional to number of vacant sites...nice). Our simple case is simpler than their simple case 1 (IBT -size of regional pool and degree of isolation). They call the immigration a 'propagule rain'.

Their model displays classical competitive exclusion: "This result is different from that usually found in metapopulation competition models (Levins and Culver 1971; Horn and MacArthur 1972; Slatkin 1974; Hastings 1980; Nee and May 1992; Tilman 1994). These models implicitly or explicitly allow interference competition between species within sites, and coexistence is obtained when there is a trade-off between competitive superiority and colonization ability. Because we do not consider interference, there is no possibility for such a trade-off in our model and, hence, coexistence in a closed community is impossible. As mentioned earlier, we deliberately ignore trade-offs associated with interference competition because we wish to explore the effects of immigration from an external source on their own."

They never get stochastic extinctions because their model is continuous, therefore use an extinction threshold.

"Thus, there is always stable coexistence in a community with a propagule rain." (Interesting discussion at end : where does the propagule rain come from?)

As immigration goes to zero they find that most species have a density close to zero (therefore likely extinct) and that dominance is determined by basic reproductive rate ( $r = c/m$ ). Also space is more fully occupied as IR increases. (Figure 1 nice plot of expected number of species versus immigration intensity). Interesting change in dominance relationships (relative abundances) with IR.

Experimental results by Mouquet and Loreau [82] suggest that immigration has a positive impact on plant community diversity. (Manipulation of seed rain.) However it also shows that certain community-level properties do not depend on IR e.g. total biomass (check others). Also local competition effects.<sup>2</sup>

[38] importance of the propagule pool in determining local diversity, and shifting limitations hypothesis (SLH) [17] dispersal as a community structuring mechanism, immigration increasing local diversity, effect differs between plants and animals, intermediate dispersal rates best?

[121] Importance of dispersal in maintaining diversity in fragmented habitat patches (small tropical forest fragments). Dispersal limitation versus edges effects. Must be other REFS for this as well?

---

<sup>2</sup>Nice structure here - theoretical results, followed up a few years later by experimental study. Talk about this somewhere.

[19] Obviously proponents of neutral theory, but suggests the importance of immigration, and measures immigration rate close to ours..

[28] mathematical treatment of 3D prediction model, effect of immigration rates.

[93] "Although previous studies as far as we are aware did not investigate the effects of landscape context on immigration rates, immigration has consistently been shown to increase population size in patches both in field studies and simulation models [44], [47]." Importance of landscape context.

[62] Changes in demographic rates due to HFRAG, in addition to population sizes and extinctions.

[44] 70 per cent of remaining forest within 1km of forest edges. Fragmentation effects. Smallest and most isolated patches most vulnerable. Reduce biodiversity by 13 to 75 per cent (metrics?) and reduce biomass, affecting ecosystem functions.

[106] Testing different HL scenarios on neutral and non-neutral communities. Habitat alteration reduces average level of specialisation - 'functional homogenisation'.

[1] Communities are more repeatable (similar between repeats) for high immigration rates. Use IBM model. Immigration from regional species pool. Zero sum natural theory dynamics.

[46] SAR updated to account for habitat fragmentation i.e. habitat loss that is not contiguous. Very relevant for discussion somewhere.

[39] spatial competition.

[32] food web plasticity, stabilising and predictable.

### 1.3 Community ecology

An ecological community may be broadly defined as a collection of species that coexist in time and space. The study of these collections, *community ecology*, attempts to understand their structure, dynamics and function. The field is too large to give a coherent overview. In this section we introduce some key themes from community ecology that are relevant to the current project. Key to community ecology is the understanding that coexisting species interact with one another. These interactions generate a *tangled web* of inter-dependence between species that has been discussed at least since Darwin [23] (see quote in front matter), and probably long before. This inter-dependence between species can make communities sensitive to perturbation - if a species, upon which another species is strongly dependent, goes extinct it is likely that the dependant species will be lost also. These *co-extinctions* and other knock-on effects, such as *trophic cascades*, have been empirically observed [60, 96], giving evidence to the importance of species interactions in shaping communities. However the tangled web may also generate communities which are remarkably robust to perturbation, and which persist through time despite varying environmental conditions. The story of community ecology has been one of trying to understand the general mechanisms and factors that shape communities, generating the

observed patterns of biodiversity. This task is far from complete, and urgently required given the crisis facing life on this planet.

Mention metacommunities? - and IBT (referred to here from conclusion in chapter 3) And summary on persistence - put more refs here on stabilising role of dispersal..

(Mention language of community ecology - versus e.g. network science. Robustness, modularity - see next section.)

(neutral versus niche?) (figure - trophic cascade? Yellowstone?)

### 1.3.1 Species interactions

Individuals interact with individuals belonging to either the same or different species, as well as with their abiotic environment. The interactions with other individuals are classed as intra- and inter-specific respectively. Inter-specific interactions broadly fall into three groups based on the effect that each species has on the other: antagonism, mutualism and competition (strictly this list also includes commensalism and amensalism). Antagonisms being interactions where there is benefit to one species and detriment to the other, whereas mutualisms and antagonisms give benefit and detriment to both parties respectively. Such interactions may easily be represented as a network, which codifies the pattern of interactions between the species in a community. This is useful because tools taken from network science, and developed specifically for application in community ecology, may be used to analyse the community (see section [REFS] for more details on these methods). There are various conventions used for the network representation, but all of them use some kind of interaction or *community matrix*. We will call this matrix **A** such that the elements  $a_{ij}$  give some measure of the effect that species  $j$  has on species  $i$ . Therefore a two species predator-prey system may be represented as:

$$(1.1) \quad A = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix},$$

where it is clear that species 1 is the predator since it has a negative impact on species 0 ( $a_{01} < 0$ ), whilst receiving benefit itself ( $a_{10} > 0$ ). Notably here  $a_{00}$  and  $a_{11}$  represent intra-specific interactions, or self-loops in the network, which may be non-zero due to mechanisms such as competition for space, predator-interference, or Allee effects [REFS]. In this representation the matrix is binary, and therefore the network is directed but unweighted. Conventionally ecologists have used unweighted networks since it is much easier to identify the presence or absence of an interaction than it is to quantify its strength. Therefore the ‘first wave’ of ecological network metrics were developed without reference to interaction strengths, whereas subsequent quantitative metrics took interaction strength into account [12]. Several studies highlighted the importance of quantitative metrics over qualitative ones. For example.. Interactions may also be classified as *trophic* or non-trophic. Trophic interactions, such as predation, represent a directional flow of energy from one species to another. Therefore a food web can be thought of as

a map of trophic energy or biomass flows through a community. In such case the link weight can take the natural units of energy flow rate, although this is not necessarily easy to measure. It is common in empirical studies to use the observed frequency of interaction as a proxy for the interaction strength. This is reasonable - if lynx are observed to prey on hares more frequently than on lemmings, then we may conclude that there is more biomass flow along one pathway than the other. Similar procedure are applied to pollination a mutualistic interaction which is also trophic..However the measurement of interaction strength is not simple, and as discussed by Berlow [11] there is no single definition or method. (return to this issue)

In nature all the types of interaction mentioned above have the potential to coexist. In fact it is not necessarily clear which interaction belong to which type. For example the interaction between a bee and a flower is trophic - the bee takes nectar from the plant. This has an energetic cost to the plant. There is also the potential for benefit to the plant since it may be pollinated, and also gives up pollen which will later go on to pollinate a mate. Therefore we expect the species in a single community to be engaged in multiple interaction types. Until recently most community ecology has focused on single interaction types in isolation. - that is they have studies subsets of the full community. So, for example, pollination studies have sampled interactions between plants and pollinators to construct mutualistic networks [? ], whilst parasitism studies have sampled rates of parasitism on each host to construct antagonistic networks [111]. It is also worth noting here that such focus on single interaction types has often meant that the network being studied is *bipartite* - i.e. one type of species linked to another via the interaction in question (host-parasitoid, plant-pollinator etc.) The single interaction focus has generated interesting results on the structure and functioning of such networks and sub-communities (see section 1.3.2). However more recently there has a move towards integration of multiple interaction types as they are found in nature.

(Worth noting that May's classic study involves multiple interaction types, other theoretical studies have..in stability section?)

(Include figure!)-Pocock or Kefi?

### 1.3.2 The role of structure

SEE DISCUSSION ON STABILITY IN (NOW DEPRICATED) CHAPTER 5!

Structure in this context refers to the topology of a community's network of interactions. The question we ask, as have many before, is to what extent are patterns observed in the community dependent on this structure? These patterns may be temporal (e.g. population dynamics) or spatial or may relate to biodiversity properties such as the distribution of biomass across species. This is a key goal of community ecology. But in attempting to answer this question (and, in fact, many of those posed in this introduction/thesis) there is often a gulf between theory and empirical observation, which is largely due to relative ease and frequency with which mathematical models have been applied compared with the relative difficulty of designing and

carrying out experiments to answer such questions. To give a simple, but well known, example - *population dynamics* models such as the Lotka-Volterra model use ordinary differential equations (ODE) to describe how species abundances change over time. The theoretical field itself is well advanced and the properties of such models have been studied as mathematical objects in their own right. Yet this has almost always been done with the belief that these models are ecologically meaningful and useful. Many would argue that this field in particular is divorced from reality because it has continued to develop these models with disregard for empirical validation (REF:rebutatlpaper). One consensus reached from this type of modelling is that predator-prey systems may exhibit stable oscillatory dynamics in which there is a phase lag between the prey and predator populations. Eventually this effect was able to be demonstrated in laboratory experiments conducted by Luckinbill [REF], although his systems required constant inflow of individuals from an external source for the oscillations to persist. Although the laboratory demonstration made it possible that such a mechanism exists in nature it has not been conclusively demonstrated [REF]. The best known example of possibly predator-prey cycles is the hare and lynx dataset from the Hudson Bay Company. However it has not been demonstrated conclusively that this represents a genuine predator-prey dynamic. For example it has been demonstrated that hare population may oscillate in the absence of predation [REF], lending credibility to the hypothesis that the hare oscillations at least may be intrinsic (self-governed)[100], and that the lynx population oscillate passively in response. An alternative hypothesis involves a third species, implicit in the data - the hunters that caught the animals. This hypothesis states that the hunter preferred hunting the larger and lynx, but would switch to hare when they were very abundant, thus leading to oscillations in the number of lynx pelts that are synchronised with the hare oscillations, but phase shifted. Finally another study concludes that hare eat lynx based on the phase relationship between the two populations at certain points in the time series<sup>3</sup>. This is not an altogether series conclusion, rather it serves to remind us that the wrong results can be obtained by an apparently sensible analysis, demonstrates that extreme care must be taken , especially given the propensity for human confirmation-bias [REF] and the tendency for large bodies of theory to be proven incorrect or incomplete.

The hare-lynx example shows us that even the simplest structure - a single prey connected to a single predator - can pose problems when we try to relate it to observed patterns (in this case population dynamics). One reason for this is that nature is messy and complex. In fact the hare and lynx populations are embedded in a larger community [104] and the observed dynamics is likely to be the results of a combinations of extrinsic and intrinsic factors [5]. Fairly conclusive that predation plays a role [63]. The point is that a vast amount of research has gone into understanding just two species and whether the interaction between plays an important role in the observed dynamics. We can understand then the challenges faced when attempting to answer similar questions of a whole community.

---

<sup>3</sup>We will actually return to this in the final chapter. Phase difference appears to be very relevant, and may shift of be in apparently the wrong order due to interactions with other species.

(Need to work space and climate into the above discussion. Also include picture of hare-lynx food web? Use hare-lynx model fit plot from first year report?)

Structure-stability : antagonism, mutualism, combined - general May, Thilo - does appear that structure plays a role, but this role is not yet understood - what is clear is that stability seems unlikely, and yet ecosystems appear stable - something fishy! The push to find structural properties that confer stability.. [54, 113]. Nestedness, mutualism etc.

This may be a suitable place to touch on confusion regarding the term stability, and clear up some confusion...(robustness, different types of stability, persistence) - no refer forwards to section 2.7.2

Structure vs function: ecosystems services and functions, pollination as the perfect example. Tylianakis involved in: [108]

### 1.3.3 Network generation

Empirical methods, frequency as a proxy for interaction strength. What is meant by interaction strength? What is relevant? (Refer forwards to final chapter, more detail in intro there).

Possible section: how are networks created - refers to the problem of estimating species interaction strengths and relating these to dynamics and function.

Network inference - brief reference to methods for doing this, and why you would want to...

## 1.4 Ecology *in silico*

Define: IBM and CA and spatially explicit modelling.

As discussed in section 1.4 such models have become increasingly popular in ecology [56], but have rarely been used to model multi-trophic communities with many species [43, 68].

realistic than previous models of complex food web dynamics (e.g. (Pimm 1979; McCann et al. 2005; Brose et al. 2006)) in the following aspects: (i) individuals within species have different extinction rates, which are not dependant on stochastic events, thus eliminating the need to define fixed extinction probabilities for all species in the community (e.g. (Sol and Montoya 2006; Fortuna et al. 2013)); (ii) more complex demographic processes such as reproductive ability and immigration based on available space are taken into account; and (iii) bioenergetic constraints such as energy gathering efficiency and energy l

In section 1.3.2 we touched on the gap between theory and experiment in the field of population dynamics. Barraquand [9] provides a useful discussion of some pressing issues regarding this problem. One issue cited is poor feedback between studies conducted by theoreticians and empiricists. The fault does not lie with either group but rather in the nature of their subject matter. Traditionally physical matter has proven more amenable to the rigorous application of mathematical theory because of the relative ease with which theoretical predictions can be tested with controlled experiments. As such applied and theoretical physics have been able to

proceed, more or less, hand in hand. Numerous problems faced by field ecologists have hampered such smooth progress in the field of ecology. As we saw previously, the result is a tremendously advanced field of theoretical population dynamics, which is not closely tied to reality and certainly not useful as predictive models for natural systems. Perhaps this in itself is not a problem, it is simply that in some cases theory has advanced beyond the point at which it can be properly tested with data. To draw another analogy with physics, this is not dissimilar to the development of string theory [REF]. Coming from a background in physics, but with a desire to work on theoretical problems that are relevant to the application area, and testable, this is an important point to note.

Promote ties and discussion between theorists and empiricists! Invariability is a good example. Do no be too critical of the field!

However, this is not to say that theoretical work is not important. It is just to realise the challenges faced in this field. In fact, given the difficulty and expense of ecological field work, theoretical studies are perhaps even more important.

Justify modelling - bottom-up modelling of complex systems, what it can and cannot help us with.

BUT - must make a very clear distinction between modelling and empirical work, one which is not clear enough in Ecology.(find examples) Rift between theoreticians and empiricists, must make clear what each can reasonably ask of the other and how best to advance the field as a whole (how has this been done so successfully in physics).

Difficulty of obtaining data in ecology, and also experimental design e.g. control of extrinsic factors! - Means that computational experiments are particularly useful. In fact this is perhaps one of the most useful functions of this thesis - advancing our ability to model ecosystems represents.

Hypothesis generation, mechanistic modelling - if behaviour in nature is different then mechanism is wrong or missing..this advances understanding.

Perhaps one good definition of complex system - one for which we do not have a concise and unifying theory/model - once this are explained they appear simple. Some things give you reason to think they will never be made simple. Kolmogorov....???

IBM models - popularity in ecology...get references on history, development, refer to common ways they are used.. whole community? And of course, habitat loss - refer forwards to section [1.6.1](#).

The importance of using theoretical models to generate testable predictions...

## 1.5 Habitat loss

More facts and figures - if not included above?

Importance of considering full multi-trophic strucutre: [\[101\]](#)

In nature habitat tends to be destroyed in a spatially-autocorrelated manner - for example urban development, agriculture and logging all occur in concentrations rather than being distributed totally at random throughout space. Therefore the result of human activity is often a patchy and fragmented landscape [REFS]<sup>4</sup>. The study of our simulated communities under contiguous HL represents the study of communities in single such fragment, with immigration from an external source. In reality we know that such fragments support a lower richness of species, beyond a certain size. In this chapter we saw that a high immigration rate prevented a loss of species richness. In the next chapter we begin to look at how communities respond to changes in the immigration rate.

### 1.5.1 Modelling HL

Refer to previous studies. How has HL been modelled? Gives details of the most interesting/important findings (see Dani's list of refs.)

Especially random versus contiguous...

### 1.5.2 Beyond species richness

Many empirical studies, and most of the theoretical studies cited above focus on richness and extinctions. Are there exceptions (in the theory ones)?

The loss of interactions...

However certain studies have stressed to importance of other effects - loss of interactions, which as we have seen are important for ecosystem functioning...Tylianakis. Why is this interesting?

## 1.6 Outline of thesis

What this thesis is and is not.. investigation and development of a new model. Hypothesis generation. Test of model fitting procedure. Not a replacement for field studies.

Where it starts from.. model has been used to look at the effect of including mutualisms into an antagonistic system. We now intend to explore how these simulations are affected by HL. Again hypothesis generation and mechanism.

What we hope to achieve..?

- First: preliminary investigation of how simulated communities respond to HL
- Second: further analysis of the model, especially under varying immigration rate. "Stress test"
- Third: HL under varying IR

---

<sup>4</sup>Reference in word final sent by Dani - email with correction to chapter 3.

- Fourth: Estimating species interactions

A habitat may be defined as the environment containing an organism, or collection of organisms. It has both biotic and abiotic components. Therefore habitats are constantly changing due to ongoing environmental processes. These changes may make the habitat more or less hospitable to different organisms, generating emergent effects at the species and community levels. Human activity in particular creates pronounced and significant changes in habitat. There is good evidence [90] that anthropogenic climate change has affected living systems by changing regional habitat suitability. An example of this is the northward shift in butterfly species ranges attributed to rising temperatures [89]. Other activities such as agriculture, deforestation and urbanisation interfere directly with physical habitat components and with local flora. This alters the type of species and the community that can be supported [15, 64]. Globally the scale of these man-made effects is huge. Various studies have suggested that habitat modification is the leading cause of global species extinctions [34, 111]. Therefore an understanding of how ecological communities respond to changes in habitat is essential in order to mediate the destructive effects of human activity, and to create beneficial conservation, land management and restoration strategies. The subject has received much attention in the ecological literature, and this project is a continuation of that dialogue.

The destruction of habitats due to human activity has also received much attention in the media. This has done a lot to raise public awareness, and to fuel a growing number of campaign groups, charities and conservation organisations. In most cases the focus is on *single species effects*, especially on those threatened with extinction. The most notorious example of this may be the polar bear as the media face of global warming (see figure 1.1). Similarly the habitat loss literature has largely focused on the loss of species [34, 109], and has reinforced the notion of *species richness*<sup>5</sup> as a measure of biodiversity and ecosystem health. This is perhaps because species level effects are the most visible results of ecosystem damage, and the easiest to study empirically. However they are symptomatic of underlying system processes. At least since Darwin's marvel at the complexity of the "Tangled Bank" [23] ecologists have understood that species exist in highly interdependent communities. Therefore the ecological impacts of habitat destruction, and other human activities, must be approached from a systems perspective.

In community ecology the system of study is the ecological community - a local collection of co-existing species. The focus is on the structure, patterns and processes within the community. A key aspect of this is the pattern of *interactions between species*, which underlies many of the processes that shape the community (for more detail refer Chapter 2). Recently the habitat loss literature has begun to move away from species level effects, towards community wide effects and especially inter-specific interactions [112]. This has been facilitated by the wider availability of ecological network data, improved methods for data collection, and the ability to simulate large ecological networks and communities. Advances in ecological network theory have also provided

---

<sup>5</sup>Simply defined as the number of different species present in a community.



Figure 1.1: Stranded polar bears on Cross Island outside Prudhoe Bay, Alaska. The plight of the polar bear has received much attention in the media. The habitat loss it suffers from is very visible. However the focus of conservation strategies must be on the ecological communities, of which it is one member species. (Source: [www.greenpeace.org.uk](http://www.greenpeace.org.uk))

many new metrics for community stability, biodiversity and for analysis of network structure (section 2.7). Our approach to the study of habitat loss is situated in this context.

There is now a growing consensus that ecological interactions are the key to understanding the effects of habitat loss on ecological communities [42, 45, 77]. In addition to the loss species, it has long been known that habitat loss also leads to the important loss of inter-specific interactions. As Janzen remarked [55] in 1974: “what escapes the eye, however, is a much more insidious kind of extinction: the extinction of ecological interactions”. It has since been demonstrated that ecosystems experiencing habitat alteration often suffer loss of interactions *before* loss of species [2, 37, 112]. This can result in detectable changes in community structure, without any detectable change in species richness [111]. These structural changes have consequences for community stability, robustness and population dynamics. A significant part of the ongoing challenge is to identify meaningful measures for the structural (network) changes, and to generalise the ways in which they impact on the community. The bulk of the recent literature supports the belief of Valiente et al. [112] in “the importance of focusing on species interactions as the major biodiversity component on which the ‘health’ of ecosystems depends.”

### 1.6.1 Habitat loss

### 1.6.2 Communities of single and multiple interaction types

In the habitat loss literature most studies have looked at communities with a single type of interaction. The same has been true for network ecology in general, with the bulk of the literature focused on either antagonistic or mutualistic networks. In these networks a node represents a species, and a directed link represents a certain type of interaction (for example predation). Such networks represent the interaction structure of an idealised and closed community. For example it is common to study mutualistic communities, such as plants and their pollinators, in isolation. This is represented as a bipartite network of plant and pollinator species, with mutualistic interactions between them. Both empirical and *in silico* studies have derived some apparently general results on the response of such single-interaction communities to habitat loss. We discuss some of these findings here. However in nature a single-interaction community is a subset of a larger group of species with multiple types of interaction (predation, mutualism, competition, parasitism). There has been a recent move towards studies of communities with multiple types of interaction [58], which are less simplistic models of natural systems. These hybrid communities are represented as networks with more than one type of link. We also discuss this body of work, some of which challenges previous finding based on single-interaction communities.

Perhaps most the general result, already discussed, is that habitat destruction leads to a loss of inter-specific interactions. This may be accompanied by lower interaction frequencies, changes in interaction strength, reduced connectivity, or other structural changes in the network due to rewiring. Tylianakis et al. [111] showed that empirical antagonistic communities (host-parasitoid) responded to habitat degradation with reduced evenness in interaction frequencies. This means that certain interactions became relatively more frequent, so that energy flow through the community became concentrated along certain pathways. Also, importantly, the quantitative changes in network structure that they observed were not detectable by equivalent qualitative metrics. Neither were conventional diversity metrics, based on species abundance or richness, able to distinguish between habitats at different levels of degradation. Similarly Albrecht et al. [2] showed that insect food webs in a grassland system lost interaction diversity faster than species diversity, when subjected to habitat alteration. This suggests a biodiversity reduction in the interaction structure that is not measurable by metrics based on species abundance. Both of these examples highlight the sensitivity of results to the metrics used, when studying community response to habitat loss. Hence the large suite of metrics introduced and discussed in section 2.7.

An issue of particular interest is community stability, its response to habitat loss and its relationship to network structure. Mutualistic networks tend to have a highly nested structure and low modularity [10]. These properties are believed to improve the stability of the community [107]. It has been shown that habitat destruction can push mutualistic networks towards higher modularity, higher connectivity, and lower nestedness, thereby reducing stability [45, 102].

Conversely antagonistic networks tend to be modular in structure, which is believed to promote stability and robustness in these communities [107]. Habitat loss has been shown to destabilise antagonistic communities by lowering modularity and increasing interaction strengths [45]. Generally the literature suggests, as expected, that habitat loss reduces community stability, irrespective of the interaction type. However the underlying changes driving this loss in stability appears to differ between mutualistic and antagonistic communities. It should also be noted here that the definition and measurement of stability is non-trivial. Lurgi et al. [68] have shown that certain stability metrics may respond differently to a changing control variable, meaning that a combined, or multi-stability approach is required.

The above examples represent attempts to understand the structural changes that occur due to habitat loss, prior to the occurrence of species extinctions. From a conservation perspective this highlights the importance of targeting inter-specific interactions and the maintenance of network structure and function, rather than focusing on species level effects [77]. Fortuna and Bascompte [36] have demonstrated that real-world networks have better persistence against habitat loss than random networks assembled using null-models. This suggests that artificially managed ecosystems may be more vulnerable to perturbations than their ‘wild-type’ equivalents, unless careful attention is paid to those properties that promote stability and robustness. In food webs there appear to be certain simple properties that mediate the impacts of habitat destruction [76]. For example omnivory is shown to increase extinction thresholds, as is a reduction in top-down control by predators. However these numerical results are for small model networks and remain to be demonstrated empirically.

Recently ecologists have realised the importance of studying ecological networks that contain multiple types of inter-specific interaction [35, 58, 78]. It is known that mutualistic communities have knock on effects on food webs, and vice versa. Indeed certain species are simultaneously involved in more than one type of network or community. A powerful example of this phenomenon was demonstrated empirically by Knight et al. [60]. They showed the presence of a trophic cascade, crossing ecosystem and habitat boundaries, by which freshwater fish were able to facilitate terrestrial plant reproduction. The inclusion of such indirect and cascading effects is one of the many strengths of the network paradigm in ecology. However this study highlights the limitations of focusing on localised community subsets and single-interaction types.

A large scale study by Pocock et al. [91] was one of the first to combine networks of different types into a network of ecological networks. They used empirical networks constructed over different habitats on a farm, to construct a whole farm network. This included host-parasitoid, seed-dispersal, plant-pollinator and predator-prey networks. Using quantitative robustness analysis (section 2.7), they were able to identify keystone plant species which generated significant cascading effects across networks, and also determined the most fragile components of the meta-network. This type of integrated analysis has different implications for conservation and restoration than an approach which looks at the individual networks in isolation.

The integration of multiple interaction types has begun to shed new light on the stability of ecological communities. This is because the conventional understanding is based on studies of communities with single-interaction types. In general complex antagonistic networks with strong interactions are thought to be unstable [86]. This presents a problem for ecological theory since natural food webs, which are inherently complex, appear to be stable. The problem may lie in the fact that antagonistic networks have been studied in isolation. It has been shown theoretically that introducing mutualistic interactions into the network can be stabilising [68, 81]. Specifically Lurgi et al. [68] propose that increasing the proportion of mutualistic interactions at the base of a food web reduces the overall strength of species interactions. They found that this improved the stability of their model communities, according to a spatial aggregation metric (section 2.7).

Recently Sauve et al. [98] have brought into question the established wisdom on the relationship between network structure and stability. As discussed previously, the structural properties believed to promote stability differ between antagonistic and mutualistic communities. High modularity and high nestedness are thought to promote stability in antagonistic and mutualistic networks respectively. However Sauve's work suggests that, for a combined network of mutualisms and antagonisms, modularity and nestedness do not strongly affect stability. The results of Lurgi et al. also support this finding [68]. Therefore new metrics, accounting for diversity in interaction type, may be required in order to understand community structure and stability in hybrid networks<sup>6</sup>.

Since hybrid networks of multiple interaction type are relatively new, there are few studies relating them to habitat loss. One study, by Evans et al. [31], uses the same empirical network of networks as [91]. They employed a robustness algorithm to determine how vulnerable the hybrid network is to the loss of different habitats from the farm<sup>7</sup>. Aside from this study there is a lack of empirical and theoretical results on the response of hybrid networks to habitat loss. This project aims to make a contribution towards this area. We will extend on the work of Lurgi et al. [68] to simulate multi-trophic communities with mutualistic and antagonistic interactions. By investigating the response of these communities to simulated habitat destruction we will be generating novel results and predictions which can be tested empirically in the future. To do this we will employ a range of metrics to quantify structural changes and community stability. We will focus on the regime before species are lost from the community, with an interest in the underlying changes that occur as a result of habitat destruction.

### 1.6.3 Spatially explicit model and metrics

Another novel aspect of this work is the spatially explicit modelling approach... And some of the spatial analysis employed...

[101] - spatially explicit analysis.

---

<sup>6</sup>See suggestions in the text of [98] and talk to Alix about possibly including these in our analysis?

<sup>7</sup>Interestingly they reported that two of the most important habitats, relative to their sizes, were hedgerow and wasteland.

[37] mutualistic interactions decrease non-linearely. Connectance increases? Abrupt change in number of interactions, spatial skewness in number of interactions.

[57] - quantitative food web metrics did not vary between fragmented habitat patches in different landscape contexts.

[86] - interaction strengths is focus, but also spatial stability. c.f. a,b,g stability and Lurgi et al.

#### **1.6.4 Modelling Habitat Loss**

The importance of the spatial pattern of HL! - "As we saw in section 1.6.4 numerous studies have found that the spatial pattern of habitat destruction plays an important role in mediating the effect on the community or meta-community (for example [53, 88?])"

Habitat loss has been modelled in various ways..Spatial auto-correlation..how does our approach fit in with the literature..

[33] - controlled habitat destruction, large empirical project

#### **1.6.5 The effects of habitat loss**

Tylianakis - go into detail of the changes and why this is important (see intro of chapter 3)

CHAPTER



## METHODOLOGY

### 2.1 Introduction

The research presented in this thesis represents an *in silico* investigation of ecological community dynamics; how they respond to habitat loss and the role of species interactions. All of the experimental results, except for those at the beginning of chapter 6, are generated using the same modelling framework. In this chapter we provide the details of that framework; including the procedure for generation of interaction network topologies (section 2.2), the specification of the individual-based model (IBM) used for simulations (section 2.3), and the two algorithms used to model habitat loss (section 2.4). We also outline the implementation of this framework in code (section 2.5) and give examples of the dynamics generated by the IBM model (section 2.6). At the end of the chapter (section 2.7) we define the suite of ecological metrics and analysis methods that are used to characterise simulated communities throughout the thesis. This suite is not exhaustive. In particular additional analytic techniques are introduced in chapters 4 and 6. However the methods defined in this chapter have all been used previously in empirical or theoretical ecology studies - together they represent a *community ecology* toolbox.

### 2.2 The interaction network

Before we can simulate community dynamics we need to define the network of interactions between the species. We refer to this as the *underlying interaction network* because it defines the potential for interactions in the IBM. If two individuals meet, and they belong to species that share a link in the underlying network, they may interact. This is in contrast to the *realised interaction network*, which is simply the network of all interactions that actually occur between members of species during a given period of time. As motivated in section 1.3.1, the inclusion of

multiple interaction types in the model is a key feature of this research. Therefore we require underlying interaction networks that contain both *antagonistic* (predator-prey) and *mutualistic* (plant-pollinator) links. This allows individuals to interact either via predation or mutualism depending on the type of link that they share. Additionally the networks must be multi-trophic since we wish to study whole community patterns and responses. It would be possible to use empirically derived topologies for the underlying network. Indeed there is some precedent for this in simulation studies [37]. However we use a method, pioneered by Lurgi et al. [68], to generate artificial network topologies with desired properties. The method first creates an artificial food web (section 2.2.1), and then replaces some of the links to introduce mutualism between some species in the first two trophic levels (section 2.2.3). The strength of this approach is that it allows us to specify the ratio of mutualistic to antagonistic interactions between species in the first two trophic levels, whilst controlling the total number of species ( $S$ ) and the connectance ( $C$ ) of the network. Therefore we are able to vary the level of mutualism and determine the effect that this has on community responses to habitat loss.

### 2.2.1 The niche model

The niche model (NM) of Williams & Martinez [117] is used in the first stage of network generation to create artificial food webs. The model was first published in 2000 and it was shown to produce networks with statistical properties similar to empirical food webs. Since that date numerous competing models have been proposed [105] to explain the structure of complex food webs. However the NM has been shown to outperform the other methods in the generation of *realistic* food web topologies [118], and has proven a useful tool for the creation of artificial networks in theoretical studies such as our own [16, 29, 68, 103]. Part of the attraction of the model is its simplicity - species are randomly distributed along a one-dimensional trophic niche, and are then assigned to consume all other species that fall within a certain region of the niche space. The details of the procedure are as follows.

The model has two parameters: the number of species  $S$ , and the desired connectance  $C$ . The model output is a binary adjacency matrix  $\mathbf{a}$  that defines the presence/absence of links between species. The element  $a_{ij} = 1$  implies that species  $i$  consumes species  $j$ , and  $a_{ij} = 0$  implies the absence of an interaction. Connectance is defined as the proportion of the maximum possible number of links that are realised i.e.  $C = L/S^2$ , where  $L$  is the number of links in the network.

Figure 2.1 illustrates the concept of niche space, and shows the niche value  $n_i$  and feeding range  $r_i$  for a particular species  $i$ . Niche space is the 1-dimensional range of real numbers  $[0, 1]$ . Each of the  $S$  species is assigned a niche value  $n_i$ , drawn uniformly at random from the niche space. The species is then assigned a feeding range with a central value  $c_i$  and a width  $r_i$ . Species  $i$  consumes all species, including itself, whose niche values fall within its feeding range. To determine the width of the feeding range, a beta function with expectation  $2C$  is used to draw a

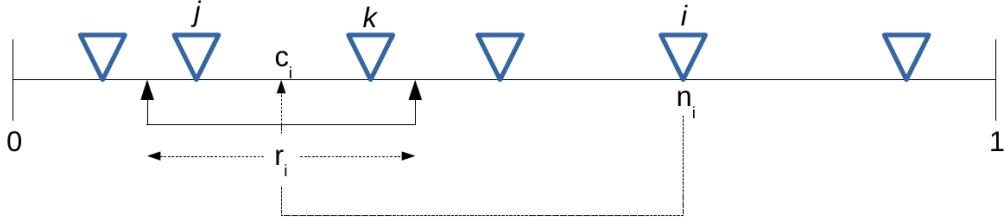


Figure 2.1: A representation of 1-dimensional *niche space* as visualised in the original publication [117], for number of species  $S = 6$ . The blue triangles represent the placement of species in niche space. The *niche value* of species  $i$  is given by  $n_i$ . The width and centre of the *feeding range* for species  $i$  are denoted by  $r_i$  and  $c_i$  respectively. Species  $i$  consumes all species whose niche values fall within its feeding range, here  $j$  and  $k$ .

number  $x_i$  from the range  $[0, 1]$ . This number is then multiplied by the species niche value  $n_i$  to give the feeding range width:  $r_i = x_i \times n_i$ . Since  $n_i \sim U(0, 1)$ , we know that the expectation value  $E(n_i) = 0.5$ , and so  $E(r_i) = C$ . Therefore on average each species consumes a fraction  $C$  of the total number of species, resulting in a network with close to the desired connectance (an approximation that improves for larger number of species). The only departure from the above is for the species with the smallest niche value  $n_i$ , which is assigned a zero-width feeding range  $r_i = 0$ . Therefore this species consumes no others, and there is guaranteed to be at least one basal species in the web.

A beta function has two parameters:  $\alpha, \beta \in \mathbb{R}^+$ . The choice of  $\alpha = 1$  simplifies the probability density function to:

$$f(x; 1, \beta) = \begin{cases} \beta(1-x)^{\beta-1} & \text{if } 0 < x < 1, \\ 0 & \text{otherwise.} \end{cases}$$

The cumulative distribution function is derived by:

$$\begin{aligned} P(x) &= \int_0^x \beta(1-x')^{\beta-1} dx' \\ &= 1 - (1-x)^\beta. \end{aligned}$$

Therefore, by choosing a probability value  $y$  uniformly at random from the interval  $[0, 1]$ , we can draw an  $x$  value from our beta distribution:

$$y = 1 - (1-x)^\beta, \quad \text{such that}$$

$$x = 1 - (1-y)^{1/\beta}.$$

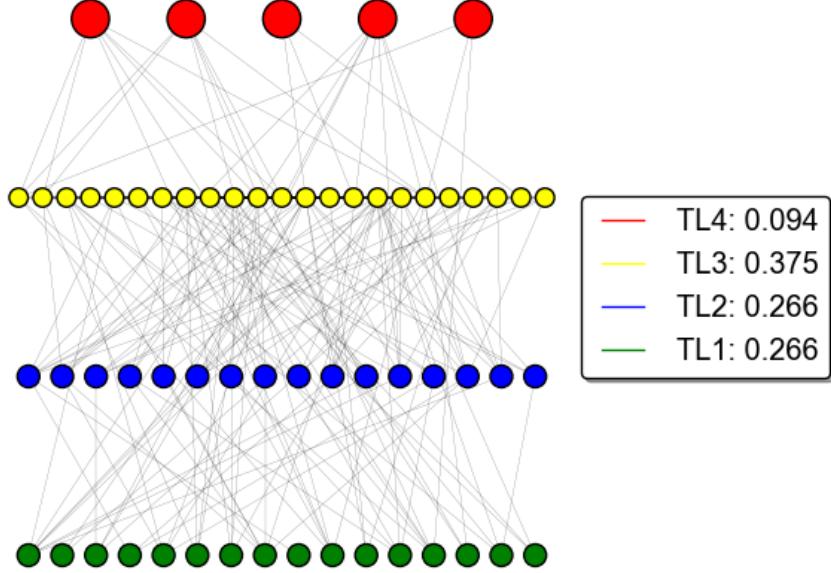


Figure 2.2: An example food web of 60 species, generated using the niche model with trophic constraints (described in the text). Numbers given in the legend are the fraction of species belonging to each trophic level.

The expectation value of this beta distribution is given by  $E(x) = \frac{1}{1+\beta}$ , therefore we choose

$$\beta = \frac{1}{2C} - 1$$

to give the desired expectation of  $E(x) = 2C$ .

Once the width  $r_i$  has been chosen, the feeding range is placed in niche space by drawing the range centre  $c_i$  uniformly at random from the interval  $[r_i/2, n_i]$ . Therefore cannibalism and looping are possible because up to half of the feeding range may contain niche values  $\geq n_i$ . In some cases the generated network may not be connected (i.e. contains one or more disconnected components), or two species may be trophically identical (i.e. have the exact same links as another species). In these cases the guilty species are deleted and reassigned new niche values  $(n_i, r_i, c_i)$  until the resulting network is connected and without identical species.

### 2.2.2 Trophic constraints

As alluded to previously the niche model produces multi-trophic food webs. Specifically the resulting networks have four trophic levels (see figures 2.2,2.3). The first trophic level consists of basal species, which have no prey and therefore represent plants. The second trophic level comprises species which feed only on plants. Therefore these species represent animals which are either strictly herbivorous, or may in fact be mutualists (see section 2.2.3). The third trophic level contains species which feed on other animal species, but which are also predated upon by other animals. These species may also feed on basal species, in which case they represent omnivores. The fourth trophic level contains species which feed on animal species, but which have no predators of their own. Therefore these species represent top-predators. It is worth noting here that the niche model does not require top-predators to be strictly carnivorous. Therefore top-predators may happen to be omnivorous, feeding both on basal and non-basal species. We address this artefact of the niche model later in chapter 4.

The niche model gives us control over the number of species and the connectance. However the proportion of species belonging to each trophic level cannot be specified. Williams and Martinez [117] showed that on average the proportion of species belonging to basal, intermediate (levels two and three) and top trophic levels closely match those proportions found in empirical webs. However this is an ensemble statistic and so does not guarantee proportions for an individual web. Furthermore it is known that the niche model, and other food web assembly models, significantly underestimate the number of herbivore species[118]. That is, although the number of intermediate species may be ‘correct’, there are often too many species in the third trophic level and not enough in the second. To ensure that all the networks we generate contain a reasonable distribution of species across trophic levels we impose *trophic constraints*. We stipulate that at least 25%, 25% and 5% of species must belong to the first, second and fourth trophic levels respectively. If the niche model output does not meet these constraints the network is rejected and we generate another. The percentages used were determined heuristically in the development of the model by Lurgi et al. [68]. They ensure that there is always sufficient species richness at each level, especially at the base of the web, and that the community is not dominated by the third trophic level. An example of a niche model network which meets the trophic constraints is shown in figure 2.2.

### 2.2.3 Link replacement

The second stage in network creation, having obtained a food web from the niche model, is to introduce mutualistic interactions. This is done by replacing some of the links between species in the first two trophic levels i.e. between plants and herbivores. This changes the way that some species interact from antagonism to mutualism (see section 2.3.1). The fraction of these links switched is defined as the *mutualistic vs. antagonistic interaction ratio* (hereafter MAI ratio). Figure 2.3 is a schematic of a possible interaction network generated by this procedure, for a

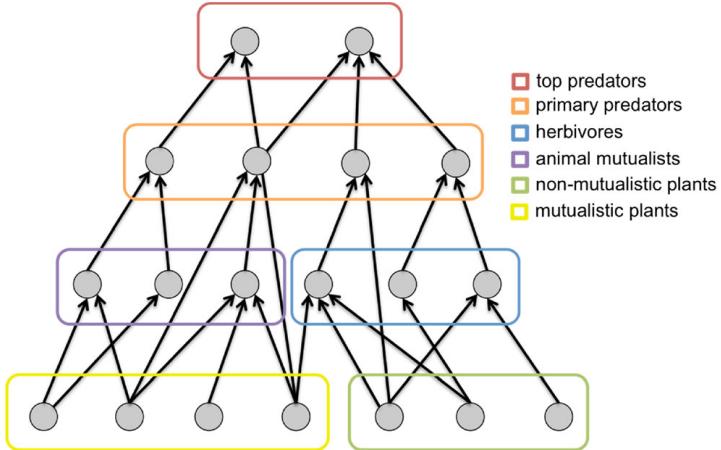


Figure 2.3: Schematic of an underlying interaction network (reproduced from [68]). Nodes correspond to species, and arrows to trophic links (antagonistic or mutualistic) from resource to consumer. The six *functional groups* of species are colour coded, and named in the legend. In this case there are twelve links between the first two trophic levels, six of these have been replaced by mutualistic links giving a MAI ratio of 0.5. Mutualistic plants and animal mutualists are defined by any species that has *at least* one mutualistic link. However species in both these groups may also have antagonistic links.

nineteen species community. In this case there are twelve links between the first two trophic levels, and six of these have been replaced by mutualistic links. The other six links remain antagonistic. Since half of the basal links have been replaced, the MAI ratio for this community is 0.5.

The result of link replacement is a hybrid network that defines two types of interaction between species (although only one between any given pair). We can identify two functional groups in each of the first two trophic levels, based on the way species interact. In the first trophic level *non-mutualistic plants* are basal species which do not have any mutualistic links. This group represents wind-dispersed plants which only have antagonistic interactions with the trophic level above. *Mutualistic plants* are any basal species with at least one mutualistic link. This group are dispersed by species from the second trophic level via their mutualistic interactions and can no longer be wind dispersed. They may also be predated upon by herbivores, if they have such links. Similarly *herbivores* are members of the second trophic level which only predate on basal species, whereas *animal mutualists* are any species in the second trophic level with at least one mutualistic link. The top two trophic levels represent distinct functional groups, which we refer to as *primary predators* and *top predators*. See figure 2.3 for a visualisation of these functional groups.

For most of the results presented we generated networks with eleven different MAI ratios in the range:  $[0, 0.1, 0.2, \dots, 1.0]$ . Communities with  $\text{MAI} = 0.0$  contain no mutualism, whereas those with  $\text{MAI} = 1.0$  contain only mutualistic interactions between the first two trophic levels. This

is in accordance with the previous study [68], and allows us to look at how communities with different MAI ratios respond to habitat loss.

## 2.3 Individual-based model

Community dynamics is simulated using a spatially explicit, individual-based model (IBM) that was developed by Lurgi et al. [68]. The landscape consists of a homogeneous two-dimensional lattice ( $200 \times 200$  cells) on which individuals move around and interact subject to bio-energetic constraints. The lattice has periodic boundary conditions such that the topology of the landscape is toroidal. Each lattice cell has a space for an *inhabitant* and a *visitor*, such that a cell may contain at most two species. Basal species may only occupy the inhabitant space, whilst all other species may occupy either or both spaces. Distance on the lattice is defined as follows. The immediate neighbours of any given cell are the eight adjacent cells, including diagonals (i.e. a Moore neighbourhood). These eight neighbours are distance-1 from the central cell, whilst the sixteen cells surrounding them are distance-2, and so on (see figure 2.4). This distance metric is used in the rules for movement and reproduction (section 2.3.1), the habitat loss algorithms (section 2.4), and also in the calculation of the spatial metrics (section 2.7.3).

The model has a large parameter space - there are seventeen free parameters, which are defined in table 2.1. A discussion of the values chosen for these parameters can be found in section 2.3.2. Initial conditions are defined randomly by the following procedure. For each cell in the landscape an individual belonging to a randomly selected basal species is placed in the inhabitant space, so that all cells contain a plant individual. Then individuals from randomly selected non-basal species are placed in the visitor space of randomly selected cells, until the desired fraction of the landscape (given by parameter *OCCUPIED\_CELLS*) is filled with animal individuals. The simulation is then run for a given number of time steps following the local rules described in section 2.3.1 below.

### 2.3.1 Local rules

The following local rules define the behaviour of individuals, which together generate the global dynamics of the IBM. In what follows capitalised-italicised words refer to model parameters, which are defined in table 2.1. Each individual stores energy (or resource), which it expends to perform actions. Initially all individuals are given a random amount of energy between *MIN\_RESOURCE* and *MAX\_RESOURCE*. If the energy of an individual drops below *MIN\_RESOURCE* it dies and is removed from the landscape. On each time step an initial cell is randomly selected and all cells are updated sequentially, starting at the initial cell. Cell update consists of the following ordered processes which occur first for the visitor individual and then for the inhabitant:

1. Immigration

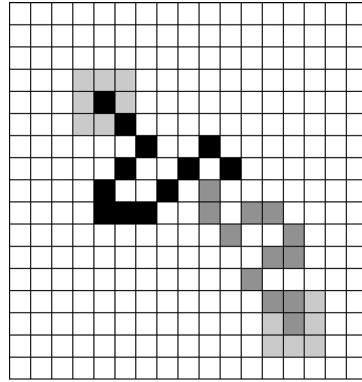


Figure 2.4: The trajectories of two individuals over 12 time steps are shown in *black* and *dark grey*. The distance-1 neighbourhoods of the two individuals on the first time step are shown in *light grey*. Figure reproduced from [68].

2. Death
3. Movement
4. Reproduction
5. Feeding
6. Metabolic loss

### 1) Immigration

An immigrant individual is created with probability given by *IMMIGRATION*. The species of the immigrant is selected uniformly at random from the original species pool. There must be space in the cell for the immigrant to be placed, or the immigrant must be able to feed upon the species present in the cell (in which case it does so and replaces it). Otherwise the immigrant is discarded. If placed, the immigrant is given a random starting energy.

### 2) Death

If the energy of an individual in the cell has fallen below *MIN\_RESOURCE*, it is removed from the landscape.

### 3) Reproduction

An individual may only reproduce if its stored energy is greater than *MATING\_RESOURCE*. This is true for all species. Animals reproduce sexually, plants reproduce asexually.

- **Sexual reproduction:** If an individual's energy exceeds *MATING\_RESOURCE* it searches its distance-3 neighbourhood. If it finds an individual of the same species, with sufficient

energy to mate, and it finds a destination cell with space for an animal (inhabitant or visitor space), then mating occurs. Both parents give a fraction of their stored energy (*MATING\_ENERGY*) to the offspring, which is placed in the destination cell. If an individual has reproduced it carries out no further actions on that time step.

- **Asexual reproduction:** This occurs for plants via two possible mechanisms.

1. If the individual is a non-mutualistic plant it reproduces with a probability equal to *REPRODUCTION\_RATE*. If reproduction occurs the offspring is placed in a randomly selected available cell in the distance-3 neighbourhood. For plants, available means empty or only occupied by an animal individual. If no cells are available the plant cannot reproduce. Again a fraction of the parent plant's stored energy (*MATING\_ENERGY*) is given to the offspring.
2. Mutualistic dispersal occurs for mutualistic plants. This action is carried out by the animal partner, and is done in the 'feeding' phase (see 5). The 'seed' of the parent plant is carried by the animal partner, so it may be placed beyond the distance-3 neighbourhood.

#### 4) Movement

If the individual is a plant it does not move. Otherwise a neighbouring cell (distance-1) is selected uniformly at random. If the selected cell contains a prey species, feeding occurs (see 5). Otherwise, if there is an available space in the selected cell, the individual moves there. The motion is therefore a two-dimensional random walk, as represented in figure 2.4.

#### 5) Feeding

Having selected (in 4) to move into a cell containing prey, there are three possible trophic interactions:

1. **Predation:** If neither individual belongs to a basal species a predation event occurs with probability *CAPTURE\_PROB*. The prey species dies and a fraction of its energy *EFFICIENCY\_TRANS* is given to the predator. The predator moves into the new cell.
2. **Herbivory:** If one individual is a non-mutualistic animal, the other is a plant, and there is space to move into the selected cell, they interact. A fraction of the plant's energy *HERB\_FRACTION* is lost, and a fraction (*HERB\_EFFICIENCY*) of this energy is given to the herbivore. Both individuals continue living and the herbivore moves into the new cell. If the animal is an omnivore an additional trade-off (*OMNI\_TRADEOFF*) is applied to its energy gained, since omnivores are less efficient at digesting plant matter than straight herbivores.

3. **Mutualism:** If the individuals share a mutualistic link, and there is space for the animal to move, they interact. A fraction of the plant's energy (*MUT\_FRACTION*) is transferred to the animal. The animal also keeps track of which plant it interacted with. If it later reaches an available cell in the landscape it creates an offspring of this plant with probability *MUT\_EFFICIENCY*. On each time step that an offspring is not produced, the mutualistic efficiency is reduced by a fraction *MUT\_COOLING*.

### 6) Metabolic loss

If the individual is an animal it reduces its stored energy by a fraction *LIVING\_EXPEND*, to account for metabolic losses. If the individual is a plant it auto-trophically increases its energy by a fraction *SYNTHESIS\_ABILITY*. This, along with the randomly generated immigrants, are the only energy input to the system.

#### 2.3.2 Model Parameters

A complex model such as this may display great variation in output depending on parameter values. During the model development by Lurgi et al. [?] a set of parameter values were selected that produced *realistic community patterns* and stable dynamics (see section 2.6). In particular the rank-abundance (see section 2.7.1.1) and degree-distributions were shown to be well fitted by log-normal and exponential functions, which is a pattern that has been observed in natural communities [80]. Where possible these parameters are based on *ecological realism*; the main example being trophic assimilation efficiency. It is well known that energy is lost when transferred between trophic levels, and that transfer rates are different depending on the type of resource consumed (plant vs. animal biomass) [52]. As such the assimilation rate is higher for plant biomass than animal biomass (*HERB\_EFFICIENCY* > *EFFICIENCY\_TRANS*). The extra reduction in transfer efficiency *OMNI\_TRADEOFF* models the fact that omnivores are less well adapted to consume plant material because they also consume meat. Other than the omnivory trade-off all species within a functional group have identical parameters, and therefore differences between species are defined only by feeding relationships<sup>1</sup>

A key mechanism, and novel feature of the model, is mutualism. Mutualistic interactions are trophic, so energy is transferred from plant to consumer, but less than in a herbivorous interaction (*MUT\_FRACTION* < *HERB\_FRACTION* × *HERB\_EFFICIENCY*). Therefore a mutualistic animal benefits energetically from the interaction, but less so than if it were herbivorous. A mutualistic plant benefits significantly by having less of its resource consumed, and receiving improved dispersal ability. There is a potential disadvantage to the plant that it must wait for a partner to reproduce. However the combined effect is that mutualism shifts some of the benefit of interaction in favour of the plant, whereas herbivory only benefits the consumer and harms the plant.

---

<sup>1</sup>Does this mean that the model is neutral within functional groups? Clairfy what is meant be neutral.

Lurgi et al. conducted a sensitivity analysis, which showed that their results were not significantly affected by a  $\pm 10\%$  variation in the value of all parameters (see S.I. for [68]). Given the extensive effort that went into finding a stable and interesting region of parameter space, we choose to begin the investigation of habitat loss by using the same parameter values. These values, hereafter referred to as the *default values*, are given in table 2.1. In chapters 4 and 5 we explore the effect of varying certain parameters, with particular focus on the immigration rate.

## 2.4 Modelling habitat loss

In order to study the effect of habitat loss (HL) on simulated communities we extend the IBM of Lurgi et al. [68] by implementing two habitat loss algorithms. Simulations are set up and run as detailed in the previous sections but on the  $1000^{th}$  time step, after the initial transience (see section 2.6), a given fraction of the lattice cells are destroyed. The individuals inhabiting the destroyed cells are removed. Subsequently individuals may select a destroyed cell to move into (see section 2.3.1), in which case it is unable to move and remains in place. In the reproduction phase destroyed cells are counted as unavailable for the placement of offspring. Throughout the thesis results are presented for incrementally affected landscapes, representing a gradient of habitat loss. The levels of destruction are referred to by the percentage of destroyed cells:  $HL = [0, 10, 20, \dots, 90]\%$ . The cells to destroy are chosen by two simple algorithms, giving two habitat loss scenarios: 1) Random and 2) Contiguous. These scenarios represent two extremes of the spatial pattern in which we may expect habitat to be destroyed in Nature (see section 1.6.1).

**1) Random habitat loss** proceeds by selecting lattice cells uniformly at random from the set of non-destroyed cells. This is repeated until the desired percentage HL is achieved. The result is a patchy and fragmented landscape.

**2) Contiguous habitat loss** proceeds by selecting a ‘seed cell’ uniformly at random from the pristine landscape. Destruction then spreads radially outwards from the seed cell, according to the distance metric defined in section 2.3 and the toroidal boundary conditions of the lattice. This results in contiguous regions destroyed and pristine habitat.

## 2.5 Implementation

The code for the simulation model was originally written by Miguel Lurgi for research leading to the publication [68]. He and Daniel Montoya were responsible for the bulk of the model development, testing and parameter selection - a considerable task. My task was to take this ‘legacy code’ and apply it to a study of habitat loss. The model is implemented in *Python* [4], with several switches that ensure portability between different versions of the language. The programme makes extensive use of *numpy* and *networkx*, amongst other *Python* libraries. A

## CHAPTER 2. METHODOLOGY

---

Parameter name	Value	Description
OCCUPIED_CELLS	0.4	Fraction of the grid initially occupied by individuals randomly placed on it.
MAX_RESOURCE	20	Maximum amount of resource an individual may possess at any given time.
MIN_RESOURCE	3	Death threshold: minimum amount of resource an individual may possess. Any individual possessing less than this amount at any given iteration will die (see text).
LIVING_EXPEND	0.01	Fraction of resource an individual spends in living every iteration of the model. Metabolic rate.
MATING_RESOURCE	0.5	Fraction of MAX_RESOURCE that is required for an individual to be able to reproduce.
MATING_ENERGY	0.2	Fraction of resource given to the offspring by the parent during reproduction. Each parent gives the same fraction. The total amount depends on how much resource the parent possesses at the time of reproduction.
IMMIGRATION	0.005	Probability that a new individual will appear in a cell of the grid each iteration. The species this individual belongs to is randomly chosen from the original species pool.
SYNTHESIS_ABILITY	0.1	Fraction of resource that is autotrophically created by each individual from the basal species every iteration. This is the only energy input to the system.
HERB_FRACTION	0.7	Fraction of resource lost to herbivores by individuals belonging to a basal species during a trophic event, i.e. a species in the first trophic level feeding on a species in the basal level.
OMNI_TRADEOFF	0.4	Fraction of resource that omnivores are effectively able to gather when feeding on a species from the basal level (a plant).
MUT_FRACTION	0.25	Fraction of resource of a primary producer (basal species individual) that a mutualistic partner obtains when an interaction of this type occurs.
CAPTURE_PROB	0.4	Probability that a predator individual embark upon a trophic relationship with one of its prey individuals when it encounters it.
EFFICIENCY_TRANS	0.2	Fraction of the resource the prey that is assimilated by the predator in a carnivorous interaction, i.e. trophic interaction not involving individuals from the basal species.
HERB EFFICIENCY	0.8	Fraction of the resource of the prey assimilated by the herbivore in an herbivorous interaction.
MUT_EFFICIENCY	0.8	Efficiency of an individual mutualist when dispersing a plant partner. In other words, the probability with which a mutualistic individual will facilitate the creation of a new individual of the last species of plant it visited when it is positioned on an empty cell immediately after it interacted with a mutualistic plant partner.
MUT_COOLING	0.9	Cooling factor for the mutualistic efficiency of plant dispersers (mutualists). This is the fraction of mutualistic efficiency that remains after each iteration.
REPROD_RATE	0.01	Reproduction rate of non-mutualistic plant species. Probability with which an individual belonging to a plant species that does not possess mutualistic partners for dispersal will create an offspring in any given iteration of the simulation run.

Table 2.1: Definitions of model parameters, and *default values* used. Reproduced from [68].

change in the return value of some methods in *networkx* (specifically some methods from v1.9 return *generator* objects, whereas in v1.7 they return *lists*) forced me to make some amendments to the code. During this process I profiled the code using *cProfile* and *pstats*, because I noticed it was running very slowly. I noticed that the cause was repeated calls to methods of the *networkx* library, which were made every time an interaction needed to be checked in the network (for example to see if two individuals are able to interact). By storing two representations of the network - one as a *networkx* object and one as a simple *numpy* array - I was able to speed up the simulations by almost an order of magnitude. The *networkx* methods need only be called when some network metric needs to be calculated. Having made the above changes I ran an ensemble of test simulations. Cross-checking the results with those in [68] confirmed that the simulations were behaving as expected.

The original code was well written to be extendible, and there was a prototype algorithm for contiguous habitat loss. Therefore my job of implementing and testing the habitat loss algorithms (section 2.4) was made easy. I also added methods to save the spatial state of the system. This allows for *post-hoc* spatial analysis, and also facilitated the creation of spatial animations of the evolution of the system (see animations at [75], and figure 2.7). All results presented in this thesis are from simulation ensembles run on *Blue Crystal Phase 3* [40], the university's high performance computer cluster. Ensembles were run as parallel job arrays, and output save in pre-constructed directory trees. I conducted post-simulation analyses in the statistical package *R* [94] and in *Python* making extensive use of the *open source* libraries available.

## 2.6 Dynamics of the model

In the absence of habitat loss the dynamics of the model is characterised by an initial period of transient dynamics, with high amplitude oscillations, after which the populations settle into a *relatively steady state*. A typical example of the population dynamics, aggregated by trophic level (TL), is shown in figure 2.5. The simulation for that produce this figure used default parameters, and no habitat loss. In general we observed that the main transience, characterise by relaxation from the initial conditions, is contained within the first 1000 time steps. Therefore we use the *heuristic* that habitat loss is applied on the 1000<sup>th</sup> time step. Figure 2.6 shows the trophic dynamics of a mutualistic community ( $MAI = 1.0$ ), with 40% HL applied. The drop in abundance on the 1000<sup>th</sup> time step is clearly visible. The figure also appears to show that the system has not reached steady state before HL is applied - the abundance of TL1 in particular is still increasing. However there is no requirement that the system should be at steady state before habitat is destroyed. More important is for the system to be a steady state when we sample from it to conduct analysis - we want to characterise the system reliably, and not unwittingly perform calculations on some section of transience. This issue turns out to be non-trivial, and we return to it in chapter 4. However, for the initial analysis in chapter 3, we make the same assumption as

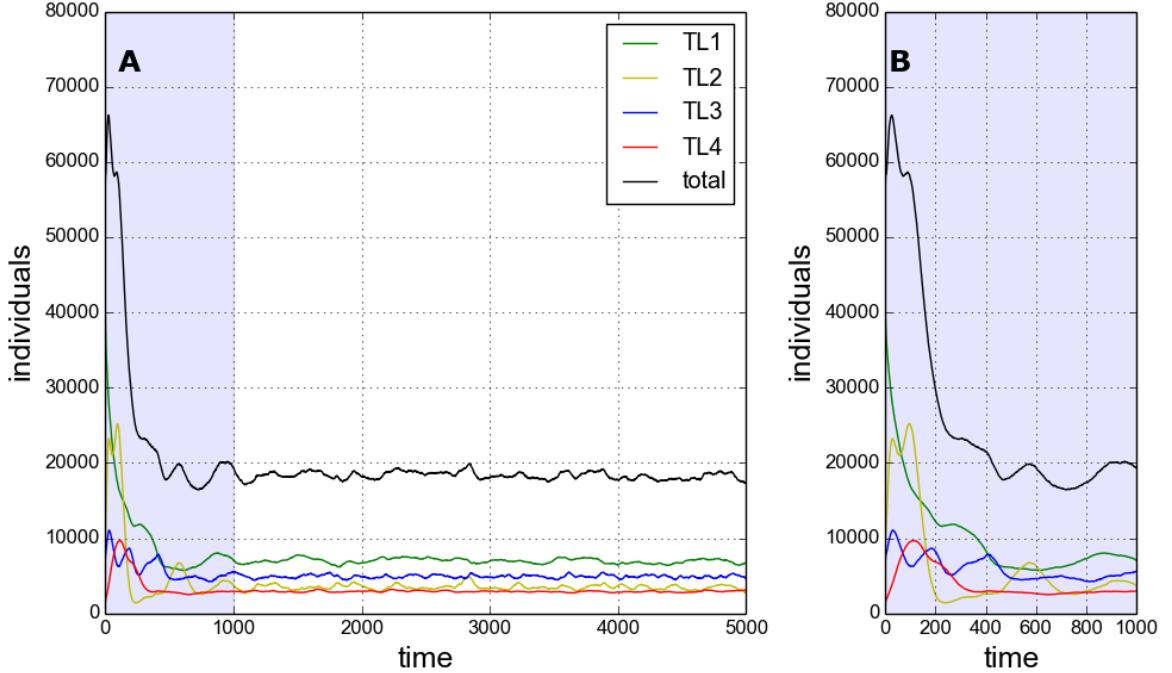


Figure 2.5: Example of community by trophic level, simulated using the IBM with *default parameters*. No habitat loss. Antagonistic community:  $MAI = 0.0$ . (A) The full dynamics for 5000 time steps. (B) Magnification of the first 1000 time steps, containing initial transience.

in [68]: that the system has reached steady state after 5000 time steps, even with the addition of HL.

Figure 2.7 shows the spatial state of the system on the 5000<sup>th</sup> time step, for a community with  $MAI = 0.5$ . There is no HL. Obviously we cannot generalise from a single image. However observing the communities in space does provide intuition about the dynamics and other properties of interest. For example the community shown displays visibly higher aggregations in the lower two trophic levels (panels A-D) than in the top two (panels E-F). It also appears that the more aggregated species are the most abundant. Such insights, gained from *watching* the system, feed into the analysis and discussion (at least implicitly) throughout the thesis. I have also produced *animations* of the spatial dynamics under various conditions which can be viewed at [75], and are referenced at relevant points in the thesis.

## 2.7 Ecological metrics and analysis methods

In this section we introduce the metrics and methods used to analyse simulation output, which are standard tools of community ecology. These methods fall nicely into four categories. *Biodiversity* metrics (section 2.7.1) capture some aspect of the diversity of species. *Stability* metrics (section 2.7.2) try to quantify stability - a task that is fraught with confusion and therefore given

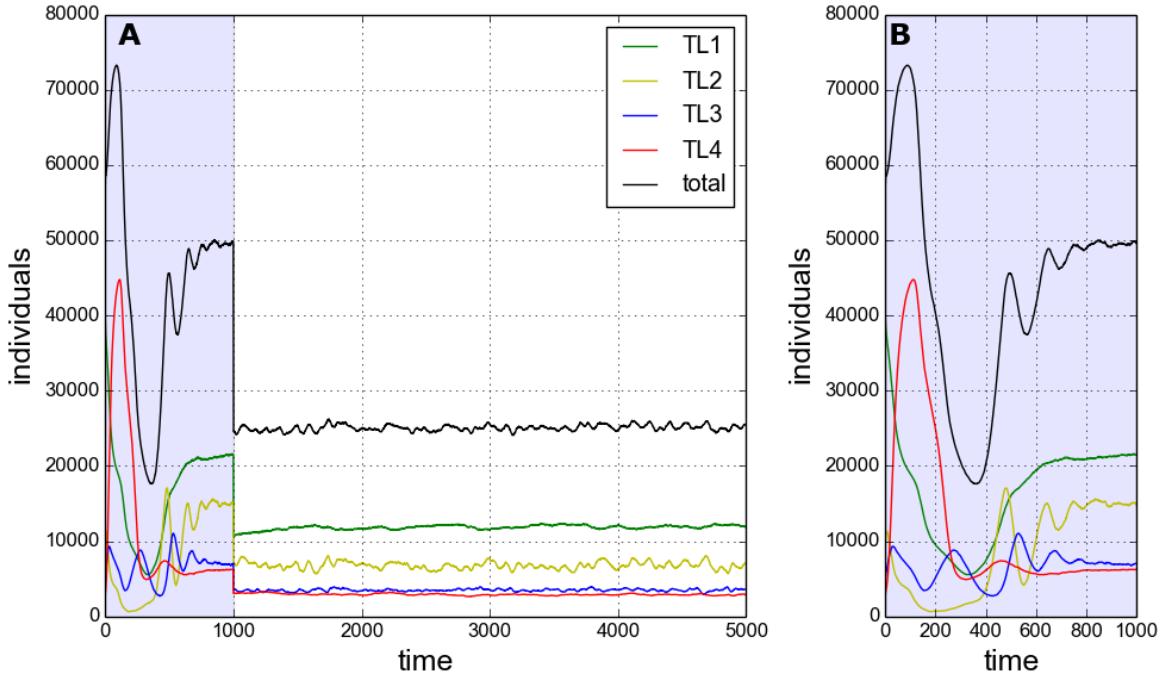


Figure 2.6: Similar to figure 2.5, but with  $MAI = 1.0$ , and with 40% contiguous habitat loss applied on the 1000th time step.

some thought here. *Spatial metrics* characterise the properties of the distribution of species in space. *Network metrics* use properties of the *realised* interaction network to provide insight about community structure. The choice of metrics provides a balance across the aforementioned categories, and largely is determined by the analysis previously conducted by Lurgi et al. [68]. Additionally we present a way to use rank-abundance distributions to measure *evenness* in species abundances (section 2.7.1.1), and implement several new stability metrics based on *invariability* (section 2.7.2.1). In all that follows we use the number of individuals belonging to a species to measure *abundance*. An alternative would be to use *biomass*, which is potentially more informative due to allometric scaling [115]. However biomass information is not included in the model, and number of individuals is a commonly used alternative.

### 2.7.1 Biodiversity metrics

**Species richness** is very simply defined as the number of different species present (in a sample). As discussed in section 1.6.1, many HL studies have focused in species richness. However this metric only changes when species become locally extinct, or sufficiently rare that they are not present in samples. Therefore this metric is insensitive to many ways that a community may

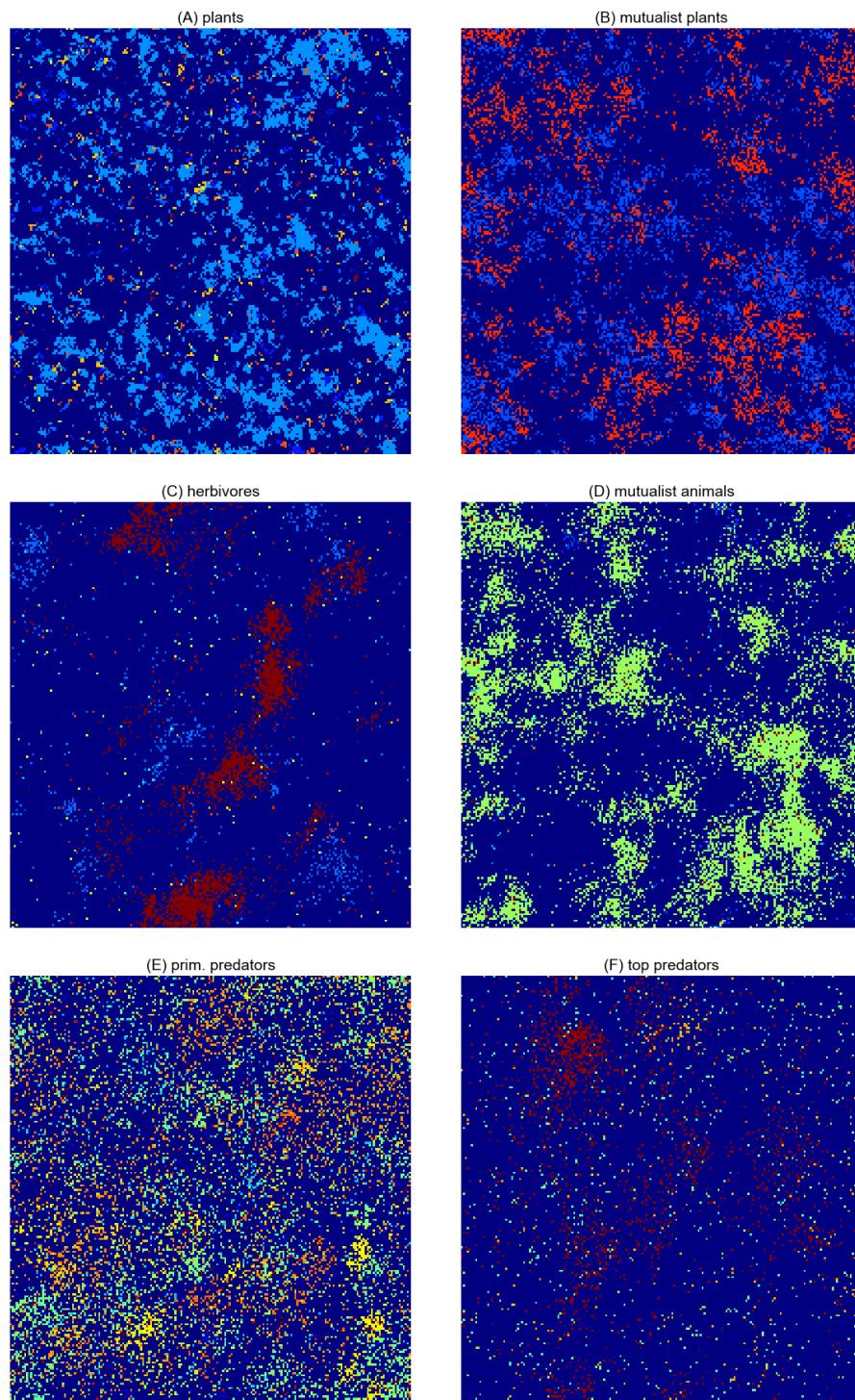


Figure 2.7: A spatial representation of each functional group. Each panel (A-F) shows the full landscape ( $200 \times 200$  cells). Dark blue represents empty cells. Other colours indicate presence of an individual, coloured by species.  $MAI = 0.5$ ,  $HL = 0\%$ .

respond to perturbation. We define the **relative abundance** of a species  $i$  by:

$$(2.1) \quad r_i = \frac{N_i}{\sum_{j=1}^S N_j},$$

where  $N_i$  is the number of individuals belonging to species  $i$ ,  $S$  is the number of species. Unless otherwise stated the sum in (2.1) is taken over all species, such that the relative abundance is the proportion of individuals in the whole community that belong to species  $i$ . The relative abundance of species within a subset of the whole community (such as a functional group) may also be calculated. In this case the sum is only over the subset of species, giving the abundance relative to the other species in the group. The relative abundance is used to calculate the **Shannon diversity index**:

$$(2.2) \quad D_{Sh} = -\sum_{i=1}^S r_i \log(r_i),$$

which is simply the Shannon entropy of species relative abundances. Therefore diversity, by this metric is maximal when all species have the same relative abundance, and approaches zero if one species out competes all others. There is no standard convention for base of the logarithm. Throughout this thesis we use the natural logarithm (base  $e$ ) for all information theoretic measures. Normalising  $D_{Sh}$  by the maximum diversity gives us the **Shannon equitability index**:

$$(2.3) \quad E_{Sh} = \frac{D_{Sh}}{\log(S)},$$

since the maximum diversity is equal to the logarithm of the number of species  $S$ . The metric is constrained between 0 and 1, and is maximal when all species have the same relative abundance. Both the Shannon metrics can be computed for a subsets of the community (for example a functional group) by using the relative abundance within the subset (2.1), and reducing  $S$  accordingly. Throughout the thesis we use only the Shannon equitability index because it controls for the number of species present, and therefore changes in this metric are not driven by changes in species richness that may be induced by HL. Hereafter *we use the terms equitability and diversity interchangeably* to refer to the normalised metric given in (2.3).

### 2.7.1.1 Rank abundance distributions

Rank abundance distributions (RADs) provide another tool to study the patterns in species abundances. They are sometimes referred to as Whittaker plots, after his 1965 paper on species abundances in plant communities [116]. To construct the RAD, species are simply ranked from most to least abundant, such that the distribution is monotonically decreasing. It is conventional to use the logarithm of the abundance measure when plotting. We use RADs to *visually inspect* the distribution of species abundances, and to develop *alternative measures of evenness*. Example RADs for simulated communities can be seen in figure 3.4.

In natural communities it has been observed that RADs tend to be *long-tailed* - with relatively few species of high abundance, and relatively many species of low abundance. As such RADs have often been found to fit well to a *log-normal* distribution [68, 116], or modifications thereof [74]. However numerous alternative models have been proposed [119], and are implemented in the *vegan* package [87] of the programming language *R*. **We propose two measures of evenness**, calculated by fitting two alternative models to a RAD. We use two models from those available in *vegan*: the *Zipf* and the *preemption* model, because both have a parameter which is easily interpreted as a measure of the evenness of the distribution. The **Zipf model** is given by the power law:

$$(2.4) \quad \hat{a}_r = N \hat{p}_1 r^\gamma$$

where  $\hat{a}_r$  is the predicted abundance of the species rank  $r$ ;  $N$  is the total number of individuals;  $\hat{p}_1$  is the estimated proportion of the most abundant species (rank 1); and  $\gamma \in \mathbf{R}^-$  is the estimated exponent of the power law. So  $\gamma$  gives the gradient of the line defined by (2.4) in log-log space. Therefore a smaller value of  $|\gamma|$  indicates a shallower line and more even distribution of abundances (see for example figure 3.4). The **preemption model** is given by a geometric sequence:

$$(2.5) \quad \hat{a}_r = N \alpha (1 - \alpha)^{r-1}$$

where  $\alpha \in [0, 1]$  is the single model parameter, and other symbols have the same meaning as in (2.4). Therefore the estimated abundance decreases by a fraction  $(1 - \alpha)$  for each rank, and the choice of  $\alpha$  is constrained such that the estimated abundances sum to  $N$ . In semi-log space, as is used to plot the RADs, the preemption model gives a straight line, since (2.5) implies:

$$(2.6) \quad \log(\hat{a}_r) = \log(1 - \alpha)r + C,$$

where  $C$  is constant. Therefore the smaller the value of  $\alpha$ , the closer the gradient of the line is to zero, and the more even the distribution of abundances. In our case, as is common [87], we use relative abundances to allow comparison of the RADs between communities with a different total number of individuals. Therefore  $N = 1$ , and  $\hat{a}_r$  are fractions, in (2.4) and (2.5). The models are fitted using *R*, and the parameters  $\gamma$  and  $\alpha$  used as complementary metrics for evenness - in both cases *the smaller the absolute value of the parameter, the more even the distribution of abundances in the community*. As with the Shannon metrics, RADs may be produced for subsets of the community.

### 2.7.2 Stability metrics

There are various different theoretical concepts relating to stability in ecosystems, which can lead to confusion [7, 83]. It is remarkable that complex ecosystems are able to function with relative

consistency in the face of extrinsic variability and perturbations (e.g. due to environmental factors and human activity). Ecological theory does not yet fully explain this phenomenon. Donohue et al. [26] explain that stability is a multi-dimensional concept involving related but distinct components such as variability, resilience, robustness and persistence. Communities may appear stable in some respects, but unstable in others, especially under perturbations such as habitat loss. Our main approach to stability focuses on features of temporal population dynamics and is outlined below. However we first introduce two key stability concepts for clarification. **Persistence** simply means lack of extinction i.e. a population of the species continues to exist through time [30]. In the current context persistence of a species is defined by the presence of individuals belonging to that species in the artificial landscape. This concept plays a key role in chapter 4. **Robustness** is a measure of how resistant a community is to perturbation. It is often used in the context of species extinctions, with robustness metrics measuring the expected extent of cascading extinctions after removal of some species [31]. **Dynamic stability**, in the current context, refers to properties of the dynamics of the system. Its definition requires further clarification<sup>2</sup>.

A seminal work on ecosystem stability is May's 1972 paper [71]. Here he showed, using random matrix theory, that high connectance and strong interactions lead to instability. This lead to the conclusion that ecosystem *complexity* reduces stability, which contradicts observations of Nature and has been the subject of ongoing debate ever since (see discussion in section 1.3.2). The measure of stability used in this body of literature is derived from the mathematics of dynamical systems. The assumption is that the dynamics of an ecosystem can be modelled as a system of non-linear ordinary differential equations (ODEs). It is possible to calculate the local stability of the equilibrium state of this system - it is stable if all eigenvalues of the Jacobian have negative real parts. Therefore this commonly used theoretical concept of stability relates to the stability of ecosystem models. It is often referred to as *asymptotic resilience* [7] because the magnitude of the real parts of the eigenvalues determine the rate of return to equilibrium after perturbation.

Asymptotic resilience is used to measure stability in chapter 6. However, in practical applications, this measure cannot be used because of the difficulty in estimating the elements of the Jacobian from empirical data (see the aforementioned chapter). Therefore empiricists use an alternative measure that can be computed directly from population dynamic time-series [18, 26]. **Temporal variability** is measured as the coefficient of variation in the abundance time-series of a species. So for species  $i$ :

$$(2.7) \quad CV_i = \frac{\sigma_i}{\mu_i},$$

where  $\sigma_i$  and  $\mu_i$  are the standard deviation and mean of the species abundance respectively, over the time period in question. To derive a community level measure we average  $CV_i$  over all species, and in all subsequent analysis refer to the resulting metric as **mean\_CV**. The metric

<sup>2</sup>Could give more examples of how different stability concepts may relate to each other.

captures the variability of population dynamics, with the assumption that *higher variability implies lower stability*. However the relationship to asymptotic resilience (AR) is not clear. For example a system may exhibit a limit-cycle which is globally stable such that all trajectories approach it asymptotically (an attractor). Therefore AR classifies the system as stable. However the *mean\_CV* may be high, depending on the amplitude of the limit cycle, suggesting instability. It appears that the two stability concepts are incompatible.

### 2.7.2.1 Invariability

A group of theoreticians at the *Station d'Ecologie Théorique et Expérimentale* in Moulis are currently attempting to clarify the issues surrounding stability, and characterise the relationship between AR and stability metrics based on variability [7, 79]. Their work is currently restricted to the case of systems with a stable equilibrium point, rather than an attractor. They introduce four new stability metrics based on *invariability*, which are named: ecosystem ( $I_{eco}$ ), population ( $I_{pop}$ ) and minimum invariability ( $I_{min}$ ), and ecosystem synchrony ( $Sync$ ). These are defined, respectively, as follows:

$$(2.8) \quad I_{eco} = \frac{1}{CV(X_{tot})^2},$$

$$(2.9) \quad I_{pop} = \frac{(X_{tot}^*)^2}{(\sum_i \sqrt{Var(X_i)})^2},$$

$$(2.10) \quad I_{min} = \min_i \frac{(X_i^*)^2}{Var(X_i)},$$

$$(2.11) \quad Sync = \frac{Var(X_{tot})}{(\sum_i \sqrt{Var(X_i)})^2},$$

where  $X_i$  is the abundance time-series vector of species  $i$ ;  $X_{tot}$  is the total abundance time-series (i.e.  $\sum_i X_i$ );  $CV$  is the coefficient of variation (as defined in (2.7)); and  $X_i^*$ ,  $X_{tot}^*$  are the species level and total abundance values of the stable equilibrium point respectively. Therefore  $I_{eco}$  is the squared inverse of the metric *mean\_CV*, whilst  $I_{min}$  is the squared inverse of the coefficient of variation of the *most variable species*.  $I_{pop}$  is a weighted sum of the species level invariabilities, and  $Sync$  is given by the ratio  $I_{pop}/I_{eco}$ . In [79] it is shown that AR is strongly dependent on species composition, and is closely related to  $I_{min}$  - it is often dominated by the least abundant and most variable species. Therefore they conclude that AR is not a reliable metric, and advocate some combination of invariability metrics to characterise stability. We implement and test these metrics in our analysis (chapter 3<sup>3</sup>), and note here the following relevant properties:

- $I_{eco}$  is a measure of system level stability, and should behave similarly to *mean\_CV*.
- $I_{min}$  is the stability of the most variable species, which is often the least abundant [79]

---

<sup>3</sup>And chapter 5?

- $I_{pop}$  is weighted in favour of the more abundant species, to reduce sensitivity to very rare species.
- $Sync$  is always less than or equal to one. If all species abundances vary in perfect synchrony  $Sync = 1$ . However any deviation from synchrony results in some cancellation in variability when abundance is aggregated over species, meaning that invariability is greater at the ecosystem level, than when summed over species i.e.  $Sync < 1$ .

### 2.7.3 Spatial metrics

The spatially explicit nature of the IBM allows us to study the spatial structure of the communities. We use the spatial-autocorrelation metrics Moran's I and Geary's C to measure aggregation in species spatial distributions [22, 68]. **Moran's I** is a measure of global correlation, and for a single species  $s$  is given by:

$$(2.12) \quad MI_s = \frac{N}{\sum_{i=1}^N \sum_{j=1}^N w_{ij}} \frac{\sum_{i=1}^N \sum_{j=1}^N w_{ij}(X_i - \bar{X})(X_j - \bar{X})}{\sum_{i=1}^N (X_i - \bar{X})^2},$$

where  $N$  is the total number of cells;  $X_i$  is equal to 1 if the species is present in cells  $i$  and equal to 0 if not;  $\bar{X}$  is the mean of  $X_i$  over all cells; and  $w_{ij}$  is spatial weight between cells  $i$  and  $j$ . We take  $w_{ij}$  equal to 1 if the cells are neighbours (distance-1), and 0 otherwise. In the extreme of total aggregation, when every cell containing the species is surrounded by other cells containing the species,  $MI_s = 1$ . In the opposite extreme, when no cells containing the species are adjacent to each other,  $MI_s = -1$ . For a random distribution of individuals the expected value  $E(MI_s) = -1/(N - 1)$ , which is close to zero for large  $N$ . For a single species  $s$ , **Geary's C** is given by:

$$(2.13) \quad GC_s = \frac{(N - 1)\sum_{i=1}^N \sum_{j=1}^N w_{ij}(X_i - X_j)^2}{2W\sum_{i=1}^N (X_i - \bar{X})^2},$$

where  $W$  is the sum over all  $w_{ij}$ , and all other symbols have the same meaning as in (2.12). GC is a measure of local aggregation and is constrained between maximum aggregation at  $GC_s = 0$ , and minimum aggregation at  $GC_s = 2$ . The expected value for a random distribution is  $E(GC_s) = 1$ , indicating no local correlation structure. To obtain community level metrics, the species level aggregation values are averaged over all species. In [68] the aggregation metrics were associated with stability, since higher aggregation may confer species with greater reproductive stability and reduced vulnerability to predation. However this relationship is not concrete.

Three further spatial metrics are used to characterise species *spatial ranges*. The **species range centroid** is calculated as the two-dimensional centre of mass of all individuals on the lattice. The **species range area** is calculated as the area of the circle, centred on the centroid, that contains 97% of the individuals of that species. **Species density** is then defined as the number of individuals divided by the area of the range. These three metrics are used to calculate the

temporal variability of species spatial ranges, by taking the coefficient of variation in each metric (as defined in (2.7)) over the time period of interest. The coefficients of variation are averaged over all species to obtain **community level metrics for spatial variability**: *mean\_CV\_centroid*, *mean\_CV\_area*, and *mean\_CV\_density*. These metrics provide an interesting comparison with variability in species abundance, and represent another alternative component of system stability.

## 2.7.4 Network metrics

The importance of the structure of the interaction network in determining the high level patterns and functions of an ecological community was discussed in section 1.3.2. A wide range of *qualitative* and *quantitative* descriptors are available to characterise the structure of the interaction network [12, 13, 68, 111]. Qualitative metrics refer to those that depend only on the presence/absence of links i.e they are used on unweighted networks. Quantitative metrics refer to those that use some measure of interaction strength i.e. a weighted network. It has been shown that the quantitative metrics are less sensitive to sampling intensity and generally more reliable than their qualitative counterparts [8, 14]. Also Tylianakis and collaborators have shown [111] that quantitative metrics may be required in order to detect the effects of habitat degradation on ecological communities (host-parasitoid networks in their study). In our analysis we use the same metrics as in [68], which are defined below.

### 2.7.4.1 Qualitative network descriptors

The unweighted interaction network is defined by the asymmetric adjacency matrix **a**, where element  $a_{ij}$  is 1 if there exists a trophic link from prey  $i$  to predator  $j$ , and 0 otherwise. The **connectance** is defined as:

$$(2.14) \quad C = L/S^2,$$

where  $L$  is the number of links (non-zero elements in **a**), and  $S$  is the number of species. Qualitative **generality** ( $G$ ) and **vulnerability** ( $V$ ) are defined as the mean number of prey per predator, and the mean number of predator per prey respectively. These are given by

$$(2.15) \quad G = \frac{L}{N_B + N_I},$$

$$(2.16) \quad V = \frac{L}{N_I + N_T},$$

where  $N_B$ ,  $N_I$  and  $N_T$  are the number of species belonging to basal, intermediate and top trophic levels respectively. **Compartmentalisation** is a measure of how frequently species share common neighbours, and is given by:

$$(2.17) \quad P = \frac{1}{S(S-1)} \sum_{i=1}^S \sum_{j \neq i} c_{ij},$$

where  $c_{ij}$  is the number of species with which both  $i$  and  $j$  interact divided by the number of species with which either  $i$  or  $j$  interact [68]. Finally **nestedness** is a measure of the extent to which the diets of specialist species are subsets of the diets of more generalist species [68]. As discussed in section 1.3.2 this is may be an important feature of mutualistic networks. Therefore nestedness is calculated for the mutualistic sub-network only, using the *NODF* algorithm of Almeida-Neto et al. [3].

#### 2.7.4.2 Quantitative network descriptors

The quantitative network metrics used are based on the *Shannon entropy* of link weights - analogous to the way it is used to measure diversity of species abundances. It is common practice to use interaction frequency to define link weights [14, 68, 111], in part because this is easier to measure empirically than, for example, biomass flow. We present here the standard definitions of the Shannon network metrics using the notation of Bersier et al. [12]. As previously, the natural logarithm is used for all metrics. It is necessary in what follows to define  $\log(0) = 0$ , which is equivalent to excluding zero elements in the interaction matrix from the calculations. Also rows or columns with sum equal to zero are removed, to avoid division by zero.

The asymmetric weighted matrix of interaction frequencies is denoted by  $\mathbf{b}$ , where element  $b_{ij}$  is the number of individuals of species  $i$  consumed by species  $j$  during the sampling period. Therefore we can define interaction diversities for the links coming into a species  $k$  from its prey, and the links going from  $k$  to its predators:

$$(2.18) \quad H_{N,k} = -\sum_{i=1}^s \left( \frac{b_{ik}}{b_{\cdot k}} \log \left( \frac{b_{ik}}{b_{\cdot k}} \right) \right)$$

$$(2.19) \quad H_{P,k} = -\sum_{j=1}^s \left( \frac{b_{kj}}{b_{k\cdot}} \log \left( \frac{b_{kj}}{b_{k\cdot}} \right) \right)$$

where  $H_{N,k}$  is the diversity of inflows from prey;  $H_{P,k}$  is the diversity of outflows to predators;  $s$  is the total number of species; and  $b_{\cdot k}$ ,  $b_{k\cdot}$  are column and row sums giving the total number of interactions that species  $k$  has with its prey and predators respectively. The interaction diversity metrics behave just as the Shannon entropy - the higher the number of interaction partners and the more even the interaction frequencies across these partners, the higher the interaction diversities. The exponents of (2.18) and (2.19) are used to calculate the *effective number* of prey and predators respectively:

$$(2.20) \quad n_{N,k} = \begin{cases} e^{H_{N,k}} \\ 0 \text{ if } b_{\cdot k} = 0 \end{cases}$$

$$(2.21) \quad n_{P,k} = \begin{cases} e^{H_{P,k}} \\ 0 \text{ if } b_{k\cdot} = 0 \end{cases}$$

where the symbols have the same meaning previously. These metrics have the property that if the interaction frequencies of species  $k$  are distributed equally amongst its interaction partners, then the effective number of prey/predators is equal to the actual number. However, if the interaction frequencies are not equal between partners, the effective number of prey/predators is reduced, since there is some preferential interaction. These effective number of species are used to calculate **weighted quantitative generality and vulnerability**, which are defined as:

$$(2.22) \quad G_q = \sum_{k=1}^s \left( \frac{b_{\bullet k}}{b_{\bullet\bullet}} n_{N,k} \right)$$

$$(2.23) \quad V_q = \sum_{k=1}^s \left( \frac{b_{k\bullet}}{b_{\bullet\bullet}} n_{P,k} \right)$$

where  $b_{\bullet\bullet}$  is the total number of interactions. So the metrics in (2.22) and (2.23) give a weighted average of the effective numbers of prey and predators respectively. They are weighted by the fraction of the total interactions the species  $k$  is involved, such that species with more interactions contribute most to the average.

We also calculate metrics for **interaction diversity** ( $H2$ ) and **specialisation** ( $H2'$ ). These are applied to the mutualistic sub-network only because the specialisation metric  $H2'$  is implemented in *R* in the *bipartite* package [27]. (It may be possible that a multi-trophic version of the specialisation metric can be derived.) We define:

$$(2.24) \quad p_{ij} = \frac{\sum_{j=1}^S b_{ij}}{\sum_{i=1}^S \sum_{j=1}^S b_{ij}},$$

as the proportion of the total number of interactions which occur between prey  $i$  and predator  $j$ . Then the interaction diversity is given by:

$$(2.25) \quad H2 = -\sum_{i=1}^S \sum_{j=1}^S p_{ij} \log(p_{ij}),$$

which is maximal when all interactions between prey and predator have the same frequency i.e. the distribution of interaction frequencies is most even. The specialisation metric ( $H2'$ ) is obtained by standardising  $H2$  with respect to the theoretical maximum and minimum diversity:

$$(2.26) \quad H2' = \frac{H2_{max} - H2}{H2_{max} - H2_{min}},$$

where  $H2_{max}$  and  $H2_{min}$  are the extremes of the diversity values that can be achieved if the total number of interactions of each species is held constant. As such  $H2'$  is equal to 0 and 1 for maximum generalisation and specialisation respectively. We calculate  $H2'$  using the *bipartite* package in *R*<sup>4</sup>.

---

<sup>4</sup>Really? - check this. I thought it is calculated in the simulations.

### 2.7.4.3 Interaction strength metrics

A key feature of ecological communities, and a central theme of this thesis, is the strength of interactions between species. As with stability (section 2.7.2), there are numerous ways to define the strength on an interaction [11]. We have already seen that interaction frequency ( $b_{ij}$ ) serves as one proxy for interaction strength. In the IBM we may expect that the frequency of an interaction will be largely determined by the abundances of the interacting species. If two species are very abundant then the individuals of these two species are more likely to encounter each other in space. However other factors may contribute. The distribution of species in space may affect how often they interact. For example two species may be abundant but highly aggregated in different regions of the landscape, and so are unable to interact. Factors such as competition and stochasticity may also influence interaction frequencies.

In addition to frequency we define an interaction strength metric  $IS$ , which for the interaction between species  $i$  and  $j$  takes the value:

$$(2.27) \quad IS_{ij} = \frac{b_{ij}}{N_i N_j},$$

and as such represents the per-capita effect of a predator, per-capita of prey. In subsequent analysis  $IS$  refers to the mean value of (2.27) over all interactions in the network. This metric is useful because it is analogous to the Lotka-Volterra interaction coefficients and their generalisation to other ODE population dynamic models [11]. We return to look in more detail at this metric and its interpretation in chapter 6.



CHAPTER



## HABITAT LOSS WITH HIGH IMMIGRATION

### 3.1 Introduction

In this chapter we conduct a preliminary investigation of how simulated communities respond to habitat loss. Using the individual based model (IBM) developed in section 2.2 we study the response of communities with *varying levels of mutualism (MAI ratio) to the destruction of different percentages of the landscape*. As discussed in section 1.6.2, a solid theoretical understanding of communities comprising multiple types of interaction is lacking. Studies in this area are relatively few and recent [31, 35, 58, 81, 91, 98], compared with studies of single interaction communities, which have been ongoing for several decades. Therefore the inclusion of both mutualistic and antagonistic interactions in the model represents a key novel feature of this investigation.

To destroy habitat we use the two algorithms described in section 2.4. Therefore there are two habitat loss (HL) scenarios to study, which we refer to as the *random* and *contiguous* scenarios according the way in which habitat is destroyed. As we saw in section 1.6.4 numerous studies have found that the spatial pattern of habitat destruction plays an important role in determining the effect on the community or meta-community (for example [53, 88, 101]). The two HL scenarios represent two extremes of the way in which a natural habitat may be destroyed, ranging from almost complete spatial auto-correlation to complete randomness. Therefore we are able to study the differences between community responses at these two extremes, and drawn comparisons with the results of previous studies.

To determine the effect of HL on the simulated communities we employ the suite of metrics defined in section 2.7, and study how they change along the gradient of increasing HL. These metrics capture aspects of diversity, stability, spatial organisation, and properties of the network of interactions between species. We predict that both diversity and stability will be negatively

impacted by HL, and that beyond a threshold of destruction some species will become extinct from the landscape. Based on the findings of previous HL studies (discussed in section 1.6.1) we also expect that random HL will impact communities more than contiguous HL, at least in terms of the number of extinctions. However, following the work of Tylianakis et al. [111], we may also expect to see underlying changes in the community structure and network properties prior to the loss of any species from the landscape. This region of the HL gradient, prior to species extinctions, will be of particular interest since such underlying changes are not well understood, and may not be detected by conventional studies and metrics (see section 1.6.5 for more discussion on this). Therefore we expect changes in network properties, which will be detectable in particular by the quantitative metrics introduced in section 2.7.4.2. The role of MAI ratio in mediating the effects of HL is also of interest. Based on the findings of Lurgi et al. [68], which used the same IBM, we may expect communities with high MAI ratio to be more robust to the HL. In that their paper it was shown that communities with higher MAI ratio tended to have stronger interactions and be more highly aggregated in space, and were also suggested to be more stable. We hypothesise here that mutualism may confer stability in the face of habitat loss by improving the reproductive success of basal species, making more energy available to species in higher trophic levels. If this hypothesis holds we may also expect to see dominance of mutualistic plants and animals over their mutualistic counterparts.

## 3.2 Methods

Here we outline the experimental procedure used to generate the results presented in section 3.3 below. Full details of the methodology are given in chapter 2, including the analysis metrics which are defined in section 2.7.

### 3.2.1 Simulation procedure

For this chapter we ran two ensembles of simulations using the IBM model. One ensemble employed the random habitat loss algorithm, and the other employed the contiguous algorithm. In both ensembles the level of habitat loss (HL) is varied between 0% and 90% in steps of 10%. At each value of HL there are 25 repeat simulations at each of the 11 MAI ratios ( $MAI = [0.0, 0.1, \dots, 0.9, 1.0]$ ). Therefore there are a total of 2750 ( $= 25 \times 11 \times 10$ ) simulations in each ensemble. Each simulation uses a unique interaction network, generated using the method described in section 2.2. Each ensemble contains the same simulations at  $HL = 0\%$ , since no habitat is destroyed and all else is held constant between HL scenarios. All model parameters take the *default values* as given in table 2.1 and published in [68]. The model, implemented in *Python* [4], was simulated on the Blue Crystal computer cluster [40] (details in section 2.5).



### 3.2.2 Sampling and analysis

To calculate the analysis metrics from the simulation output we follow the precedent set in [68]. Metrics based only on species abundances are calculated from a ‘snapshot’ of the simulation state on the final time step. Other metrics (for example temporal variability and network metrics) are calculated from samples aggregated over the final 200 time steps of the simulation. For such metrics we use the mean species abundance over the 200 time steps, where a measure of abundance is required. The *realised network* of interactions is constructed by counting the total number of interaction events between each pair of species during the 200 time steps. The use of snapshots and short time windows taken from the end of the simulations assumes that the system has reached steady state by then. For now we make this assumption, but return to assess its validity and consequences in chapter 4.

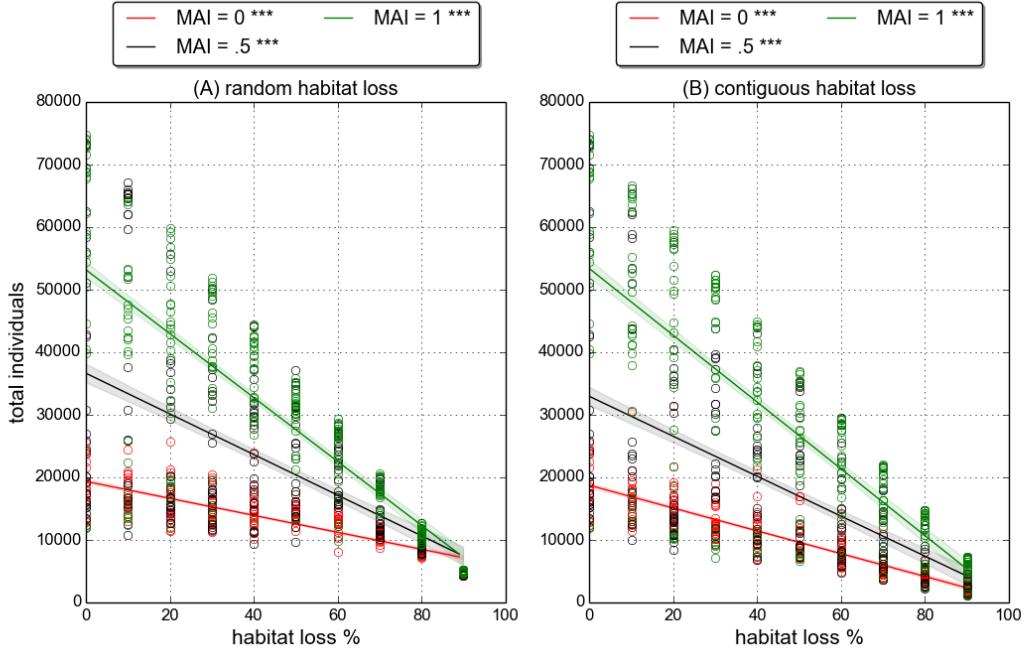
Throughout the analysis we look for robust changes of the metrics evaluated in response to HL. Therefore, as in [68], we fit linear models to identify statistically significant trends, and where necessary we log-transform the data to attempt to linearise the response (see for example figure 3.16)<sup>1</sup>. The linear models are fitted using the *ordinary least squares* method of the *statsmodels* package in *Python*. The *p-value* of the *f-statistic* of the fits are given, which is a significance value for the test of the null hypothesis that the slope of the linear trend is equal to zero. Therefore a low p-value is evidence to reject the null hypothesis and conclude that the slope of the trend is non-zero. Different levels of significance are indicated in the plots throughout this chapter. However we always follow the convention that  $p < 0.05$  are statistically significant, whilst  $0.05 < p < 0.1$  indicates marginal significance.

## 3.3 Results

We find that simulated communities respond differently under the two HL scenarios. Therefore throughout this section we draw comparisons between the results for *random* and *contiguous* HL, and seek to understand the differences. In general we find that there is little qualitative difference in the results for communities with different MAI ratios. Certain quantitative differences, reported in [68], hold true across the HL gradients. For example high MAI communities are more aggregated in space, have more biomass in the lower trophic levels, and support a greater total number of individuals. Despite these quantitative differences, we find that communities with different MAI ratios respond in qualitatively the same way to both types of HL. Therefore, to simplify matters, we present results for three MAI ratios only ( $MAI = 0.0, 0.5, 1.0$ ). In general we refer only to qualitative trends in the response metrics which hold true across MAI ratios, but make it clear where this is not the case.

For clarity we separate the results into subsections according the type of metric analysed: *Diversity* (section 3.3.1); *Network properties* (section 3.3.2); *Stability and space* (section 3.3.3); and

<sup>1</sup>Check this! And refer forwards to SEM modelling?



**Figure 3.1: Total number of individuals** against percentage habitat loss, for both scenarios: (A) Random HL, and (B) Contiguous HL. Circles represent the number of individuals for a single community; lines represent a linear fit to the data and the shaded regions indicate the standard error of the mean.  $p$ -value  $< 0.001$  for all linear model fits (indicated by \*\*\*).

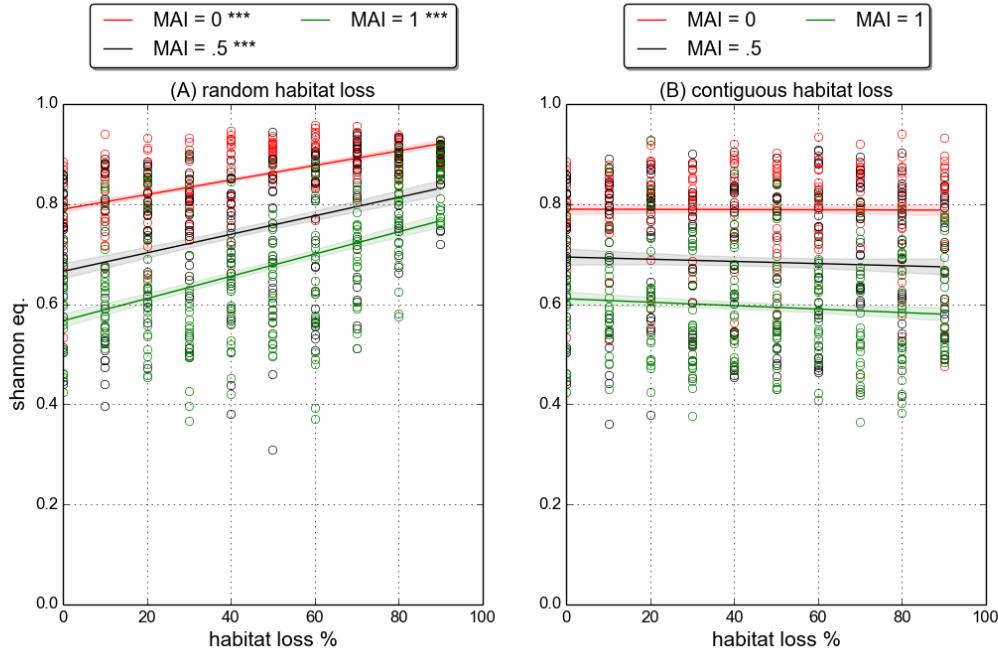
*Invariability* (section 3.3.4). In general we find significant changes in the metrics associated with diversity, stability and space, whereas we find fewer significant changes in network properties.

### 3.3.1 Diversity

Figure 3.1 shows that the total number of individuals decreases in response to HL, in both the random and contiguous scenarios. This is expected because there are fewer available cells, so the landscape is able to support fewer individuals. The linear fits to the data suggest that the average number of individuals is similar at each point in the HL gradient in the two HL scenarios. Therefore there is little to distinguish between the two scenarios based on the number of individuals. The figure also shows that communities with higher MAI ratio support more individuals than those with lower MAI ratio, as was previously reported [68].

Despite the decrease in the total number of individuals, species do not go extinct (results not shown). At 90% HL, when averaged over all MAI ratios and replicates, we observed a mean of 0.0 and 0.23 species extinctions per community for random and contiguous HL respectively. These numbers are before the extinction threshold. Therefore we effectively see no change in species richness in response to habitat loss. The lack of extinctions is due to a relatively high immigration rate (IR). The IR can be thought of as the

### 3.3. RESULTS



**Figure 3.2: Shannon equitability** against percentage habitat loss, for both scenarios: (A) Random HL, and (B) Contiguous HL. Circles represent the Shannon index value for a single community; lines represent a linear fit to the data and the shaded regions indicate the standard error of the mean. \*\*\* indicates  $p$ -value  $< 0.001$  for linear model fit, whilst no marker indicates  $p$ -value  $> 0.1$ .

probability per time step that an empty cell receives an immigrant (see section 2.3.1). Therefore at the default value ( $IR = 0.005$ ), if the landscape were empty, we would expect on each time step an average of 200 ( $= 0.005 \times 200 \times 200$ ) immigrant individuals drawn uniformly at random from the pool of 60 species. Therefore species can recover from extinction via a *rescue effect* that is common to all species. The immigration mechanism also allows for the maintenance of very low species abundances, which are sustained by immigration rather than feeding and reproduction. The absence of extinctions means that all changes observed across the HL gradient are not associated with changes in species richness. Therefore all of the results presented in this chapter represent the sort of underlying structural changes reported by Tylianakis et al. [111].

So, although there is no change in species richness under either HL scenario, there are changes in community structure. Figure 3.2 shows that the *Shannon equitability* (equation 2.3) increases for random HL, but does not change significantly for contiguous HL. Figure 3.3 shows that the same trends hold when the Shannon equitability is calculated separately for each trophic level. Since there is no change in species richness we can interpret this metric in terms of both diversity and evenness. Therefore random HL seem to make communities more diverse by increasing the evenness in the distribution of species abundances. On the other hand, contiguous HL appears to

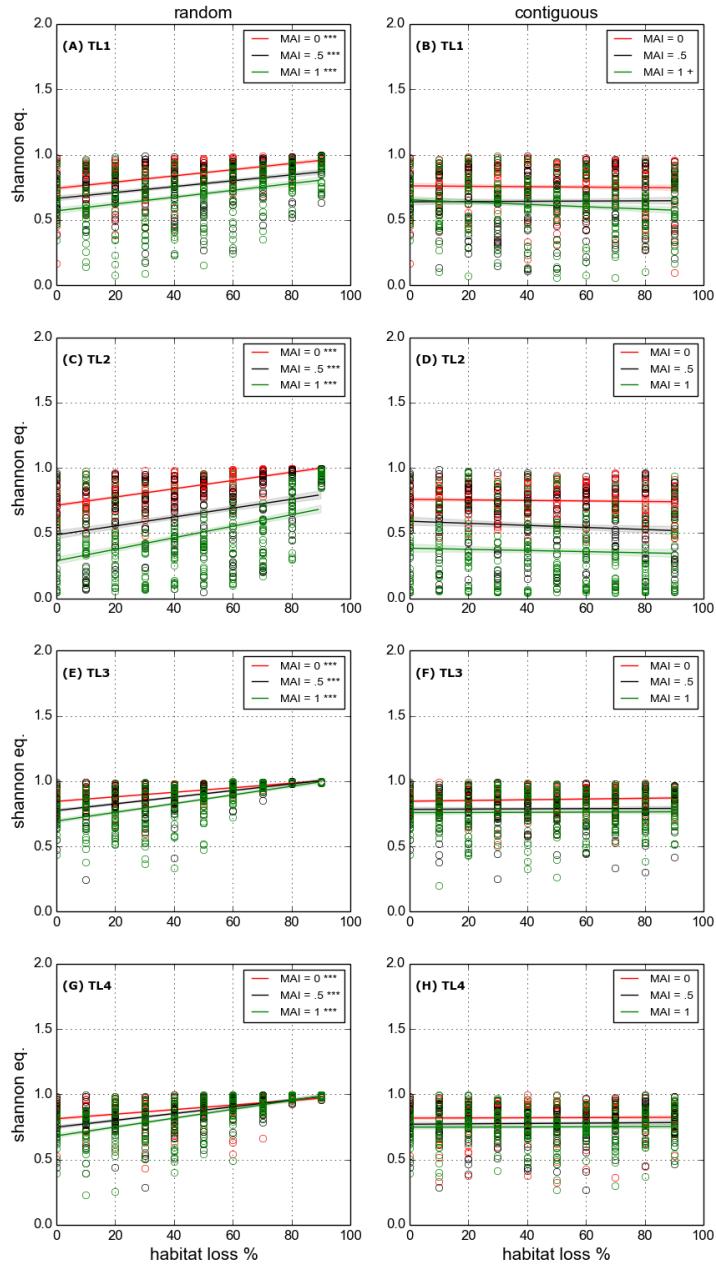
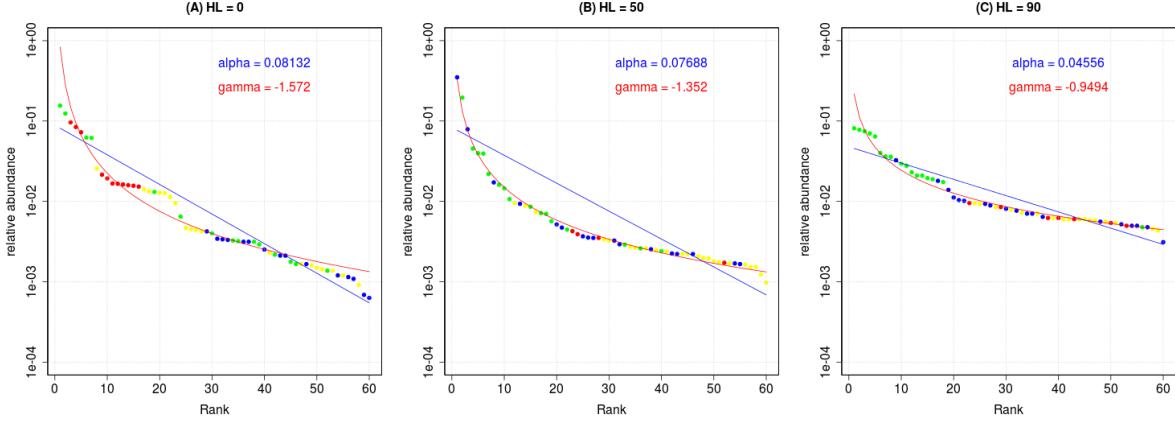


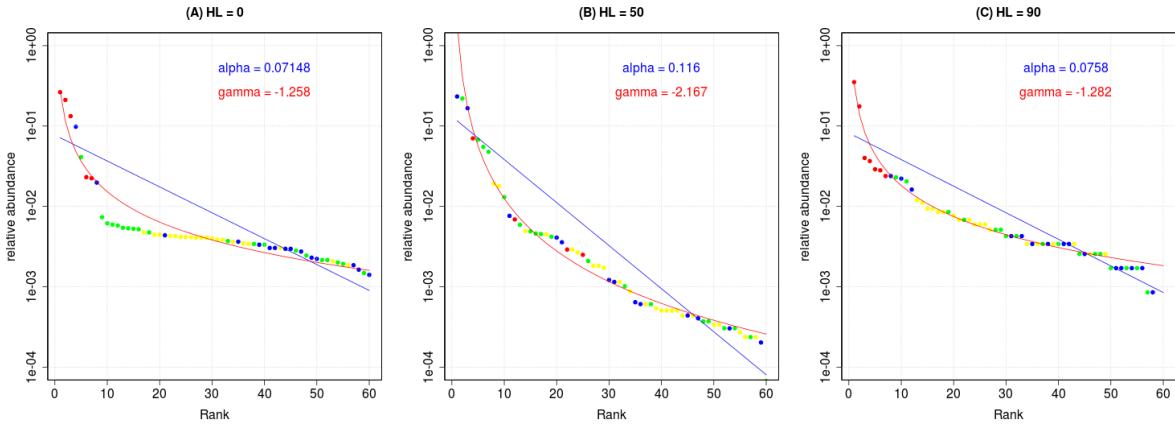
Figure 3.3: **Shannon equitability** against percentage habitat loss, for each trophic level. Left column: random HL. Right column: contiguous HL. Circles represent the Shannon index value for a single community; lines represent a linear fit to the data and the shaded regions indicate the standard error of the mean. \*\*\* indicates  $p\text{-value} < 0.001$  for linear model fit, + indicates  $p\text{-value} < 0.1$ , and no marker indicates  $p\text{-value} > 0.1$ .

### 3.3. RESULTS

have no significant affect on the diversity or evenness of species abundances.



**Figure 3.4: Example rank abundance distributions (RADs)** for three communities with **MAI=0.5** at different levels of **random HL**: (A) HL= 0%; (B) HL= 50%; (C) 90%. Species abundances are relative to the total number of individuals in the community, and plotted on a logarithmic scale. Points represent species, coloured according to trophic level: green=basal; blue=herbivore/animal-mutualist; yellow=primary predator; red=top predator. Blue and red lines give the pre-emption and Zipf model fits respectively (see text in section 2.7.1.1 for definitions), best fit parameter value for each model given as annotations on plot.



**Figure 3.5:** Similar to figure 3.4 but for **contiguous HL**.

To look explicitly at changes in evenness we construct *rank abundance distributions* (RADs) for each community, and fit standard models to these distributions (defined in section 2.7.1.1). Figures 3.4 and 3.5 show example RADs for three communities at different levels of HL, all with MAI= 0.5 (qualitatively the RADs and how they change under HL are similar across all MAI ratios). Since these plots are for single communities we cannot draw general conclusions from them. However, they serve to visualise the model fits and their interpretation. The solid blue

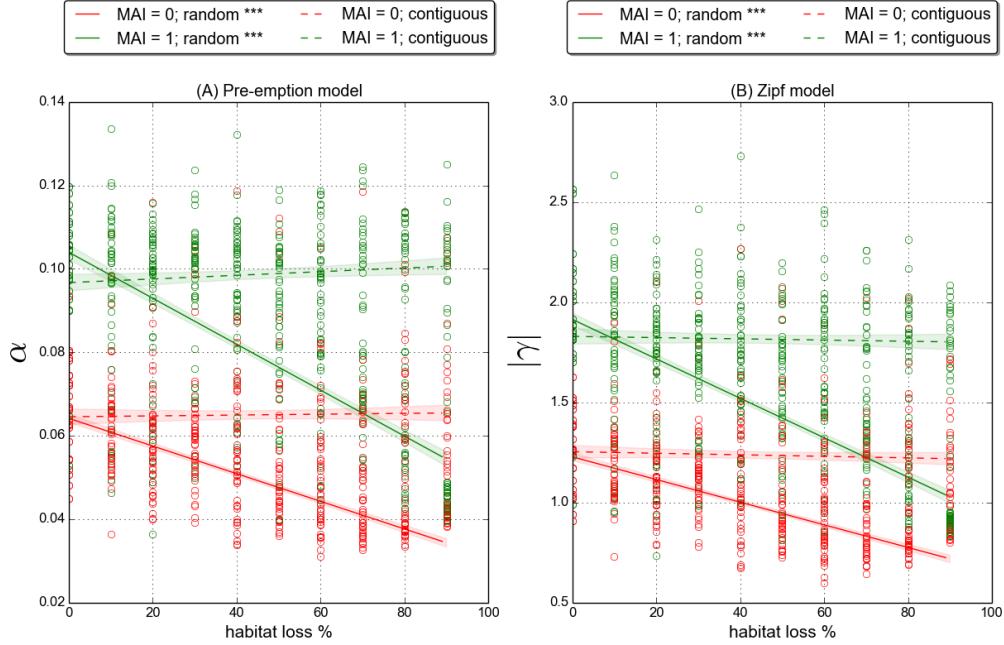


Figure 3.6: **Rank abundance model fit** parameters against HL. Panel A: Pre-emption model parameter  $\alpha$  is smaller for more even distributions. Panel B: Absolute value of Zipf model parameter  $|\gamma|$  is smaller for more even distributions. (See model definitions in section 2.7.1.1). Solid lines represent linear fits to the random HL data, dashed lines indicate linear fits to the contiguous HL data, and error bars give  $\pm 1$  standard-deviation. \*\*\* indicates  $p$ -value  $< 0.001$  for linear model fit, whilst no marker indicates  $p$ -value  $> 0.1$ .

and red lines in these plots show the *preemption* and *Zipf* model fits respectively. Each model has an evenness parameter and, as discussed in section 2.7.1.1, the lower the magnitude of the parameter the more even the distribution. From visual inspection panel C in figure 3.4 is the most even RAD of the six displayed. Correspondingly the model fits to this RADs have the lowest magnitude values for  $\alpha$  and  $\gamma$ . Although the Zipf model appears to give a qualitatively better fits to the data, we use both models the test for evenness in our simulated communities. In this way we can check for consistency in the conclusions.

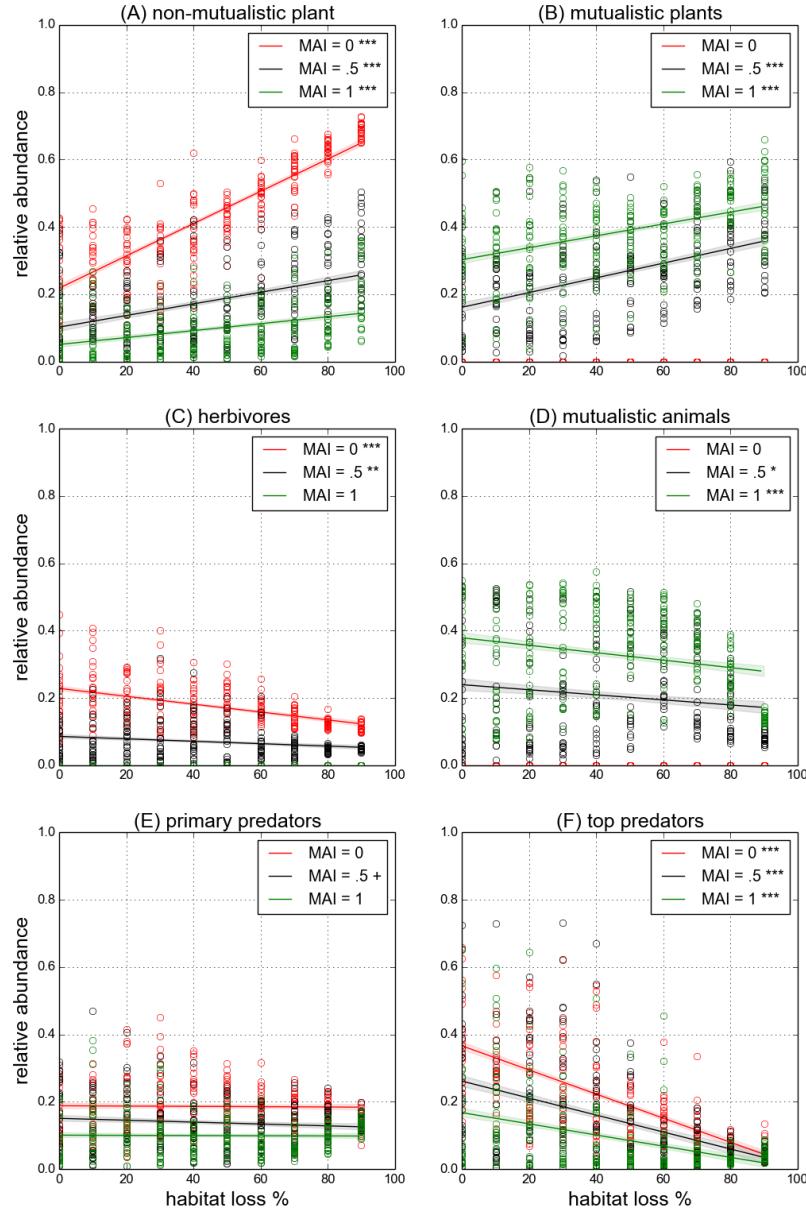
Figure 3.6 uses all replicates at a given MAI ratio to show how the evenness parameters,  $\alpha$  and  $\gamma$ , change in response to HL. For clarity the results are shown for two MAI ratios only (0.0, 1.0), but are consistent across all MAI ratios. The modelling suggests that under contiguous HL the RADs show no significant change in evenness. However both the preemption and Zipf models indicate that communities under random HL show a significant increase in evenness. These findings are consistent with the changes in Shannon diversity discussed above.

The RADs in figure 3.4 suggest another trend in response to random HL - there appears to be a systematic shift in the relative abundances according to trophic level. This is most visible for the

### 3.3. RESULTS

---

basal and top trophic levels, shown in green and red respectively. For the example communities shown, the top predator species have high relative abundance in pristine landscape (panel A), but reduced abundances in damaged landscapes (panels B and C). At 90% HL all top predators in the depicted community have a relative abundance of less than 0.01. On the other hand basal species have a wide range of relative abundances in pristine landscape, but come to dominate the community at 90% HL with all but one basal species in the top 18 ranks. This is consistent with empirical observations, since species in higher trophic levels tend to be more sensitive to perturbations [25, 95]. To investigate this effect further we look at the relative abundances of all six functional groups in response to habitat loss. These are plotted in figures 3.7 and 3.8 for the random and contiguous scenarios respectively. For the contiguous scenario community structure is remarkably constant across the habitat loss gradient, according to the relative abundances of the functional groups. The only statistically significant changes are in the non-mutualistic plant and top predator species at MAI= 0, where there is a slight decrease in top predator abundance relative to plant abundance (figure 3.8, panels A and F). In the random scenario there are clear systematic shifts in the distribution of abundance across the functional groups (figure 3.7). In particular, and in agreement with the RADs in figure 3.4, there is a relative increase in plant abundance and decrease in top predator abundance, which is statistically significant across all MAI ratios. There is also a slight decrease in the relative abundance of species in the second trophic level (panels C and D). Overall there is a shift in relative abundance towards the basal level, but interestingly there is no significant change in the abundance of primary predator species (panel E). This suggests that there is some benefit to being a primary predator in this context (see discussion in 3.4).



**Figure 3.7: Relative abundance by functional group for random HL.** Abundance relative to total number of individuals in the community. Circles represent the value for a single community; lines represent a linear fit to the data and the shaded regions indicate the standard error of the mean. The markers \*\*\*; \*\*; \* and + corresponds to linear model fit p-values of < 0.001, < 0.01, < 0.05 and < 0.1 (marginal significance) respectively.

### 3.3. RESULTS

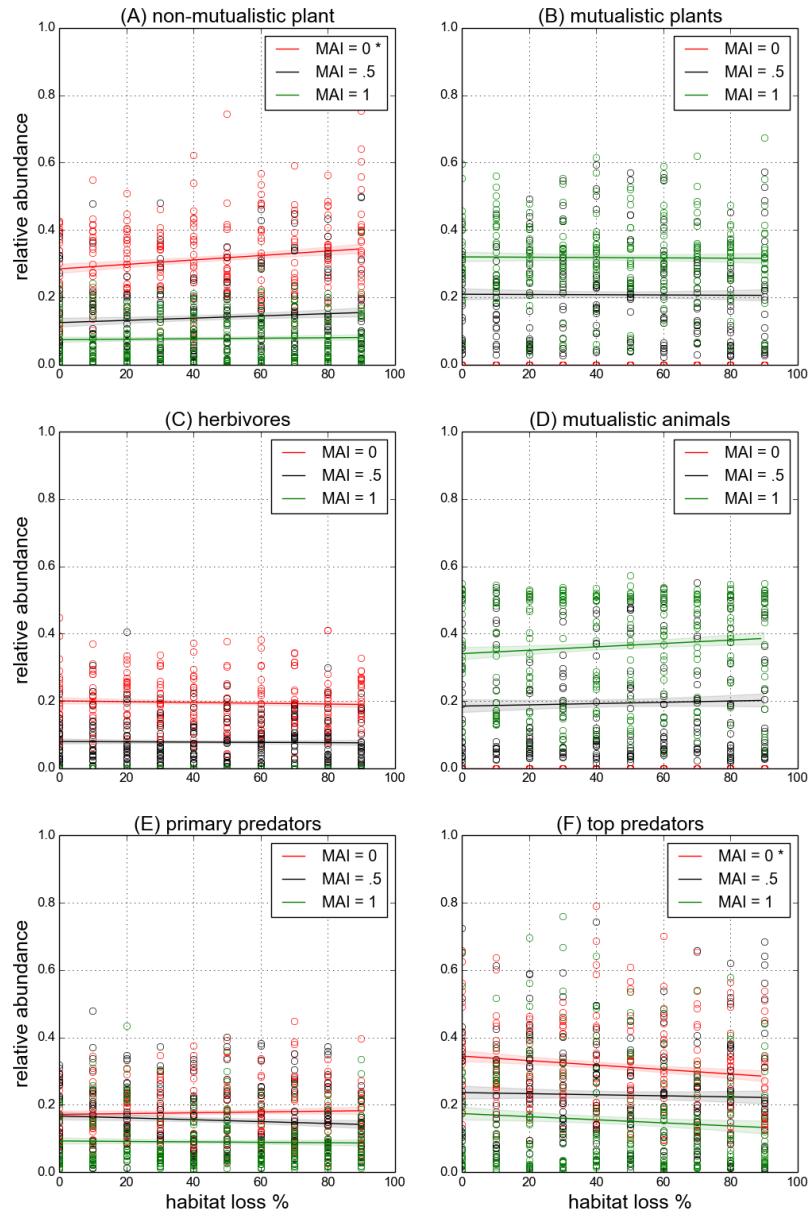


Figure 3.8: Similar to figure 3.7, but for **contiguous HL**.

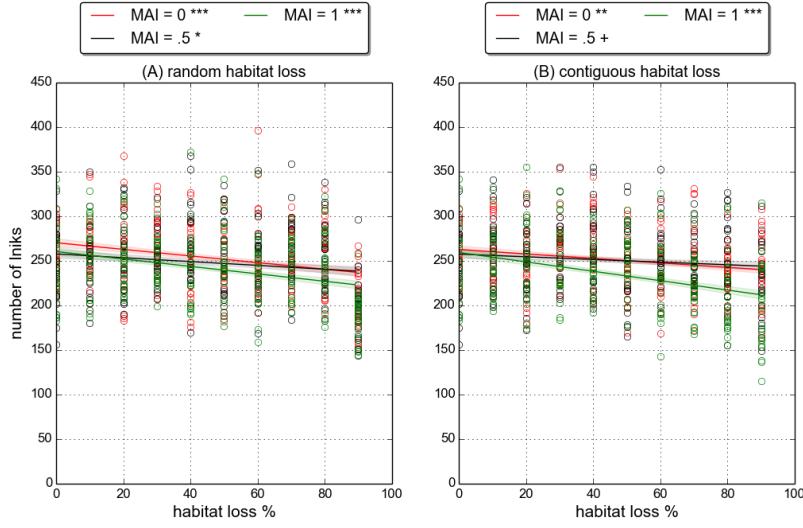


Figure 3.9: **Number of links** of links in the realised network (see text in section 3.2.2 for definition). Linear fits and p-value markers as in previous plots (see caption of figure 3.7).

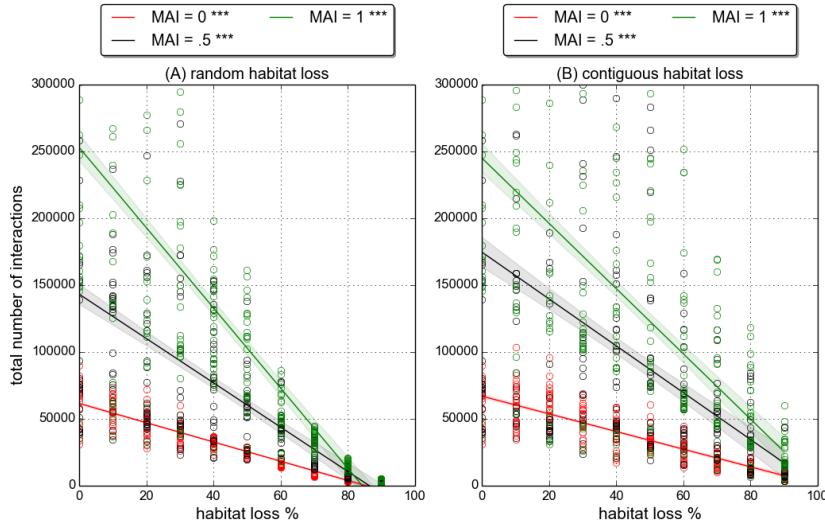
### 3.3.2 Network properties

Figure 3.9 shows that, in both HL scenarios, there is a slight decrease in the number of links in the realised network. This result indicates that some species, which would be able to interact, do not encounter each other in space as a result of HL. The number of links lost, on average is few - the greatest loss according to the linear models is under contiguous HL for MAI= 1.0 where the number of links falls from  $\approx 250$  in pristine landscape to  $\approx 200$  at  $HL = 90\%$ <sup>2</sup>. However loosing any number of links is enough to create changes in network topology that can be detected by the qualitative network metrics (section 2.7.4.1). In addition to the loss of links, figure 3.10 show that there is a decrease in the total number of interactions in both scenarios. This overall decrease in the frequency of interactions is not a surprising result of HL. As we argued in section 2.7.4.3, interaction frequencies are largely determined by the abundances of the interacting species. So when there are fewer individuals in the landscape (figure 3.1) we should expect fewer interactions. However it appears from figure 3.10 that more interactions are lost due to random HL than contiguous. For example at  $HL = 70\%$  all communities in the random scenario have fewer than 50,000 interactions in total, whereas at the same level of contiguous destruction many communities have more than 50,000 interaction and several even exceed 100,000. Also the slopes of the linear fits in panel A are steeper than in panel B. These results suggest a mechanism that produces fewer interactions under random HL than under contiguous, despite similar numbers of individuals. We propose such a mechanism in section 3.4.

Together the loss of links, and the change in interaction frequencies indicate changes in

<sup>2</sup>Why is the greatest change here?

### 3.3. RESULTS



**Figure 3.10: Total number of interactions** between all species during final 200 time steps of a simulation. Linear fits and p-value markers as in previous plots (see caption of figure 3.7).

realised network of interactions. Therefore we seek to characterise these changes, using the network metrics defined in section 2.7.4. The nestedness and compartmentalisation of communities showed no significant change under either HL scenario. Therefore these plots are not shown. The response of connectance, and qualitative generality and vulnerability, follows directly from the loss of links since there is no change in species richness. Therefore connectance, generality and vulnerability all decrease in both scenarios, as a direct result of the loss of links shown in figure 3.9 (results not shown). The response of the quantitative network metrics is more subtle because these metrics are based not only on the presence/absence of links, but on the frequency of each interaction.

The results for the **re-iterate intuitive meaning of concepts as they are re-introduce** (hereafter referred to as generality and vulnerability) are shown in figures 3.11 and 3.12. For contiguous HL neither metric shows significant change. This tells us that the loss of links and decrease in interaction frequencies occur in a such a way that the *average number of effective prey and predators* (equations (2.20) and (2.21)) per species is constant across the contiguous HL gradient. For this to occur the changes must be distributed homogeneously across the network such that there is no change, on average, in the number of interactions per species and the relative frequencies of those interactions. This finding relates to the previous observation that diversity and rank-abundance patterns do not change under contiguous HL (section 3.3.1). If the relative abundance of all species is constant across the HL gradient, and interaction frequencies are mainly driven by species absolute abundances, then the lack of change in quantitative network metrics follows directly. This appears to support the conclusion that interaction frequencies are driven by species abundance patterns.

In the random scenario generality and vulnerability decrease and increase respectively (figures 3.11 and 3.12). This represents an asymmetry between the responses of predator interactions compared to prey interactions. The increase in vulnerability means that the average *effective* number of predators per prey increases. It is unlikely that the *actual* number of predators per prey increases, since links are lost from the network. Therefore the only way that vulnerability could increase is if the interaction frequencies of prey with their predators became more even. We know from section 3.3.1 that species abundances become more even in response to random HL, and that diversity increases both at the community level and within functional groups. Therefore the change in vulnerability appears to be explained by the increased evenness of prey and predator species, leading to more even interaction frequencies between the two.

The decrease in generality, shown in figure 3.11(A), tells us that the average *effective* number of prey per predator decreases under random HL. Since we have just concluded that interaction frequencies must become more even, we must attribute this to a decrease in the *actual number of prey* per predator. Since we know that the relative abundance of top predators is greatly reduced (figure 3.7), it is reasonable to conclude that some struggle to find prey. Interestingly the change in generality at MAI= 1.0 is not significant (although it was significant at all ten other MAI ratios) suggesting that either prey are not lost, or more likely that the loss of prey to top predators is offset by the increased evenness of interactions. We know that high MAI communities have fewer individuals in the top trophic level (figures 3.7 and 3.8) and so the top predator contribution to generality is likely to be lower. Also high MAI communities contain more individuals than low MAI communities across the HL gradient, meaning a greater absolute number of prey. Together these two mechanisms explain why generality does not change significantly at MAI= 1.0<sup>3</sup>.

The results for the bipartite network metrics, interaction diversity and H2', are shown in figures 3.13 and 3.14. They support the conclusion that the changes in quantitative network metrics are driven by changes in species relative abundances. In the contiguous scenario neither metric shows significant change, in agreement with the generality and vulnerability results. In the random scenario interaction diversity increases, whereas the specialisation metric H2' shows no significant change. The increase in interaction diversity means that the interactions frequencies between plant and animal mutualists become more even. Again we assert that this is due to the observed increase in evenness of species relative abundances in these two groups (figure 3.3).

The fact that H2' does not change significantly means that interactions in the mutualistic sub-network become neither more even nor less diverse relative to the minimum and maximum interaction diversity,  $H2_{min}$  and  $H2_{max}$  (see definition in section 2.7.4).  $H2_{min}$  and  $H2_{max}$  are theoretical values calculated with the constraint that the total number of interactions of each species is fixed. If we assume, as we have argued, that the interaction frequencies are determined by species abundances, then the constraint is equivalent to holding all species abundances

---

<sup>3</sup>This is still not quite right.

### 3.3. RESULTS

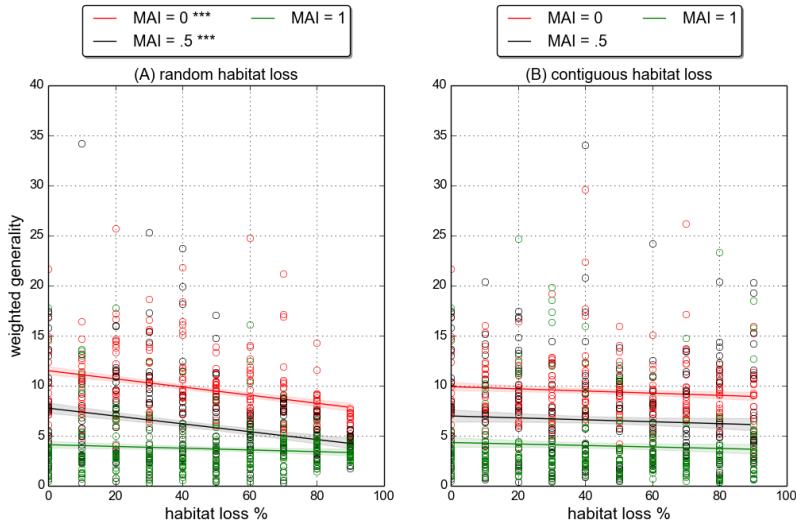


Figure 3.11: **Weighted quantitative generality**, of whole network (see section 2.7.4 for definition). Linear fits and p-value markers as in previous plots (see caption of figure 3.7).

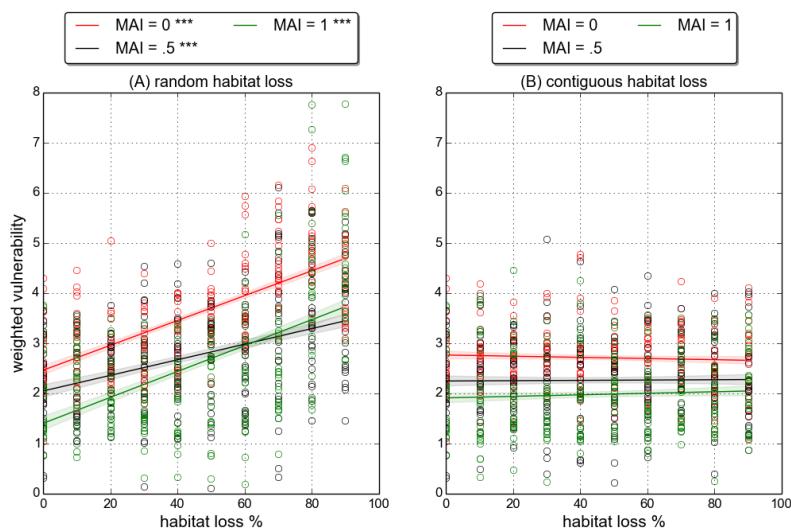


Figure 3.12: **Weighted quantitative vulnerability**, of the whole network (see section 2.7.4 for definition). Linear fits and p-value markers as in previous plots (see caption of figure 3.7).

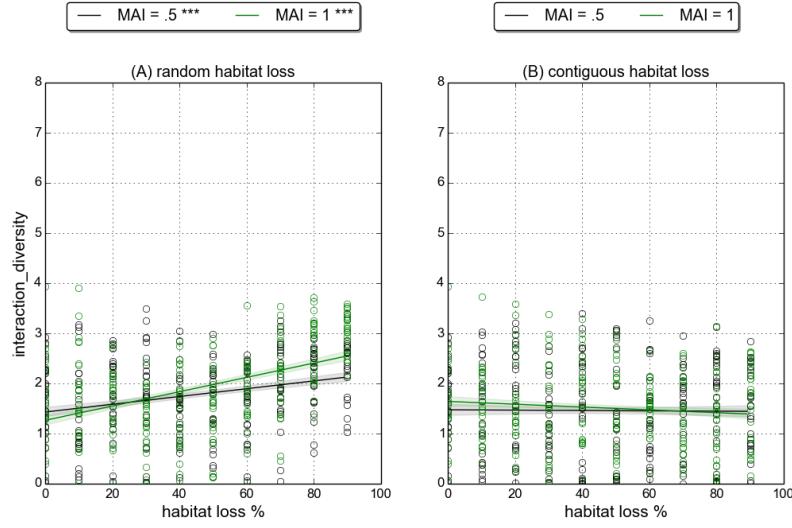


Figure 3.13: **Interaction diversity** in the mutualistic sub-network (see section 2.7.4 for definition). Linear fits and p-value markers as in previous plots (see caption of figure 3.7).

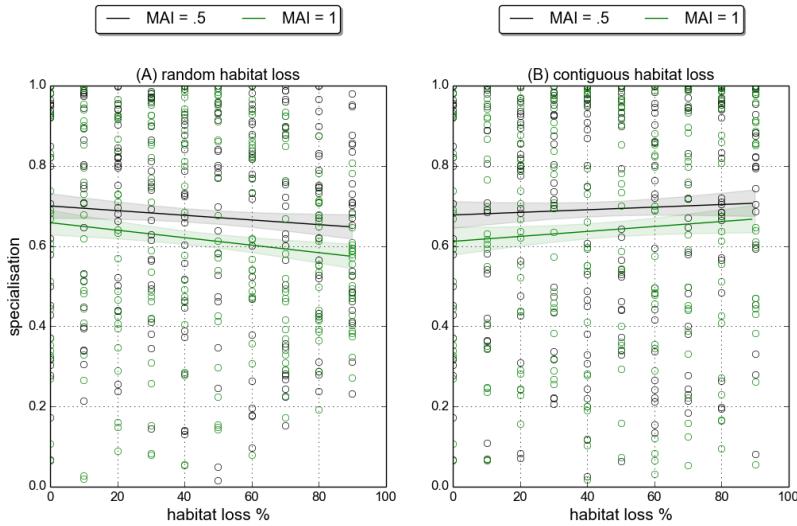


Figure 3.14: **Specialisation H2'** in the mutualistic sub-network (see section 2.7.4 for definition). Linear fits and p-value markers as in previous plots (see caption of figure 3.7).

constant. The H2' result then implies that interactions do not become more diverse than expected, given the abundance of each species. In other words, the change in the diversity of interactions is due to changes in species relative abundances<sup>4</sup>.

<sup>4</sup>This is still not clear!

### 3.3.3 Stability and spatial metrics

In this section we measure temporal stability using four metrics for temporal and spatial variability defined in sections 2.7.2 and 2.7.3: the temporal coefficients of variation (CV) in species abundances (*mean CV*), species range area (*mean CV area*), species range centroid (*mean CV centroid*), and species spatial density (*mean CV density*). At the end of the section we also look at the spatial aggregation metrics, as defined in section 2.7.3. All metrics are calculated at the species level, and then averaged over all species in the community. In the following section (3.3.4) we show that the stability results are consistent with the alternative *invariability* metrics, which are defined in section 2.7.2.1.

The stability results are consistent across the four variability metrics - *communities become more variable under random HL, and less variable under contiguous HL*. The raw data for these metrics show trends that are clearly non-linear, and appear approximately exponential (responses change at an increasing rate). Therefore we log-transform the data prior to fitting linear models. The log-linear trends are illustrated for species area variability in figure 3.15, and for species abundance variability in the middle row of figure 3.16. The same trends are observed for the other two variability metrics (results not shown), and they are statistically significant across all MAI ratios. Therefore we conclude that the afore

**Explain better**

The variability response also represents a striking difference between the two HL scenarios. In all other metrics presented communities either respond in the same way (e.g. number of

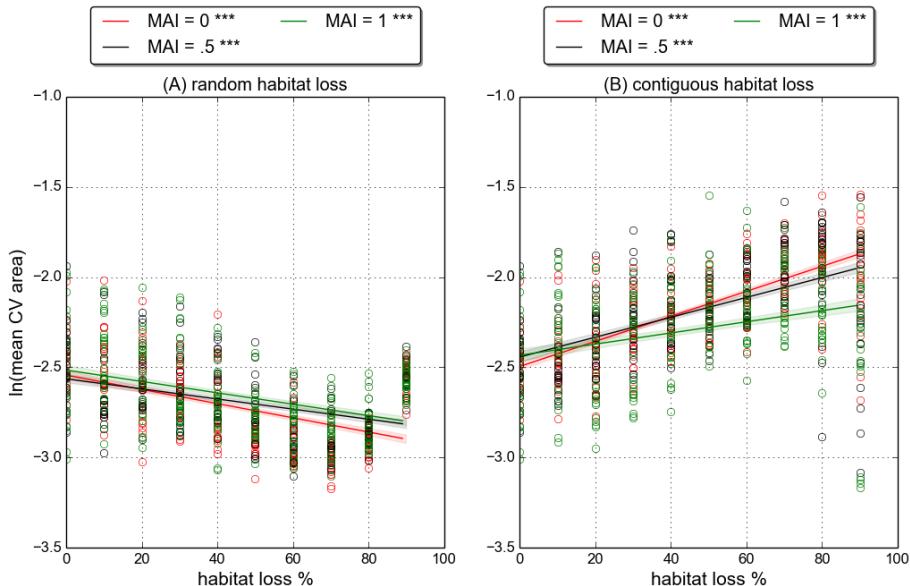
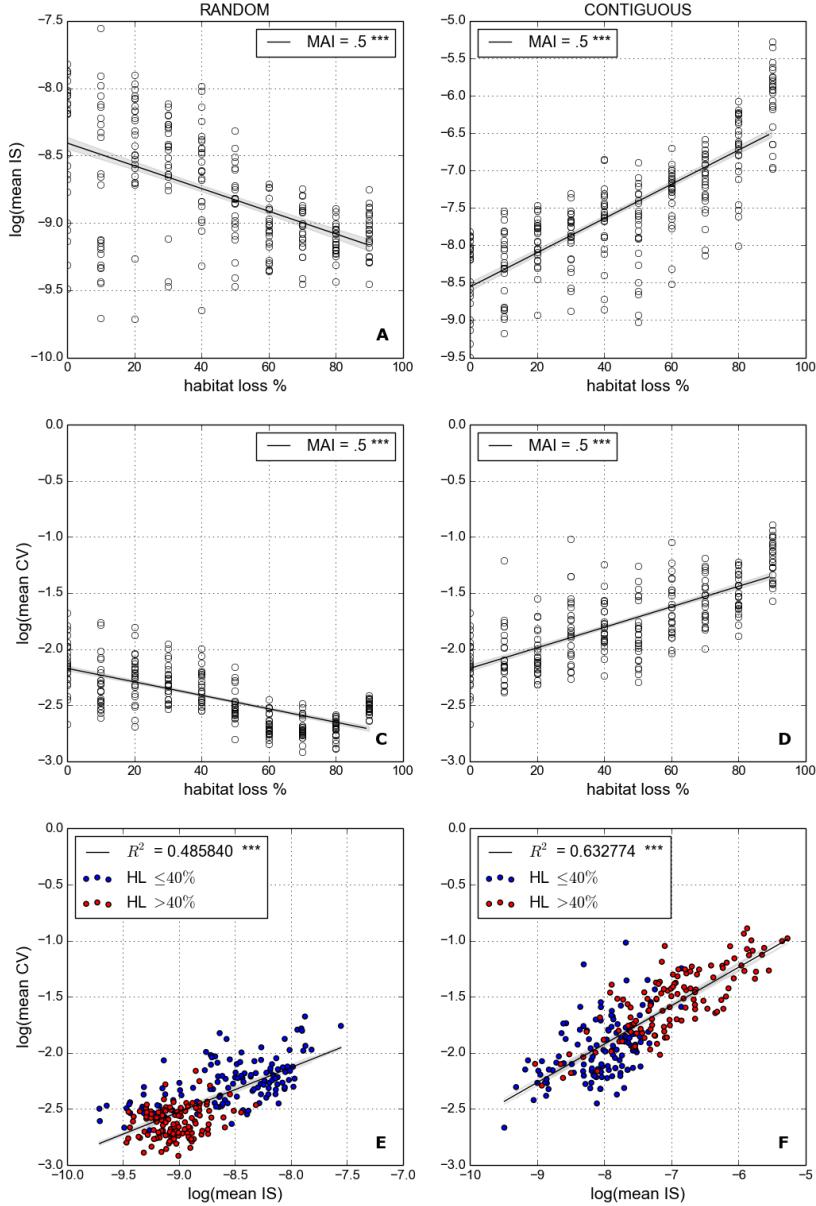


Figure 3.15: **Species area variability:** coefficient of variation in area of species range, averaged over all species (definition in section 2.7.2). Linear fits and p-value markers as in previous plots (see caption of figure 3.7).



**Figure 3.16: Interaction strengths and temporal variability.** Both natural log-transformed to linearise trends. Panels A-B: Interaction strength metric IS (defined in section 2.7.2) averaged over all interactions in realised network. Panels C-D: mean CV, coefficient of variation in species abundances (defined in section 2.7.2) averaged over all species. Panels E-F: IS as a linear predictor for mean CV, with low and high HL communities indicated by blue and red circle respectively. All communities for MAI= 0.0, 0.5, 1.0 shown. Linear fits and p-value markers as in previous plots (see caption of figure 3.7).

### 3.3. RESULTS

---

individuals, number of links), or display significant changes under one scenario but not the other (e.g. population evenness). However the trends observed in variability are qualitatively opposite for the two HL scenarios. This leads us to question the mechanism behind the changes in variability.

The only metric, aside from variability, which shows trends in opposite directions under different types of HL is *interaction strength* (IS). These trends are illustrated in the top row of figure 3.16, where we again use the natural log-transform of the data. Under random HL the average interaction strength decreases, whereas under contiguous HL it increases. We also observe a dependence of interaction strength on MAI ratio, as reported in [68]. In pristine landscape higher MAI communities have weaker interaction strengths. Under contiguous HL this ordering of IS according to MAI is conserved across the HL gradient. However under random habitat loss communities with a high MAI ratio do not lose interaction strength as much as low MAI communities. The result is that beyond about 70% HL high MAI communities tend to have greater interaction strength than low MAI communities. Although only shown for three MAI ratios, the pattern described is consistent across all eleven MAI ratios. The possible explanations for the dependence of the IS response on MAI ratio are discussed in section 3.5.

The bottom row (panels E and F) of figure 3.16 shows the log-transformed values of IS plotted against the log-transformed abundance variability (mean\_CV). These figures show of all the simulation repeats with MAI= 0.0, 0.5, 1.0. In both the random and the contiguous scenario there is a significant linear trend between the log-transformed interaction strength and variability. The coefficient of determination  $R^2$  values of these linear models are, as given on the plot,  $\approx 0.5$  and  $\approx 0.7$  in the random and contiguous cases respectively. This means that, on aggregate, interaction strengths can explain at least half of the variance in temporal variability, and more than this in the contiguous case. The dependence on habitat loss is also striking. In the random case, high HL ( $> 40\%$ ) shifts communities towards lower IS and lower variability, whilst the converse is true for the contiguous case. In both cases the linear relationship between IS and variability is consistent at low and high HL. Therefore we conclude that there is a strong correlation between IS and temporal variability, which is mediated by habitat loss. In section 3.4 we present the evidence that this correlation represents is a causal relationship, and that *interaction strengths are key to understanding how simulated communities respond to HL*.

We observe changes in spatial aggregation under random HL, but not under contiguous HL. Spatial aggregation is measured at the local and global scales using the metrics Moran's  $I$  (MI) and GEAR. Can be phrased much nicer, and why using the metrics are averaged over all species in the community. Figures 3.17 and 3.18 show that, across the board, species are not highly aggregated on average. The observed values of MI and GC are close to 0 and 1 respectively, which are in space. Also higher MAI communities are found to be more What is aggregation? Why interesting? As argued in [68], this is caused by an increase in the aggregation or plant species due to reduced herbivory pressure and increased local reproduction.

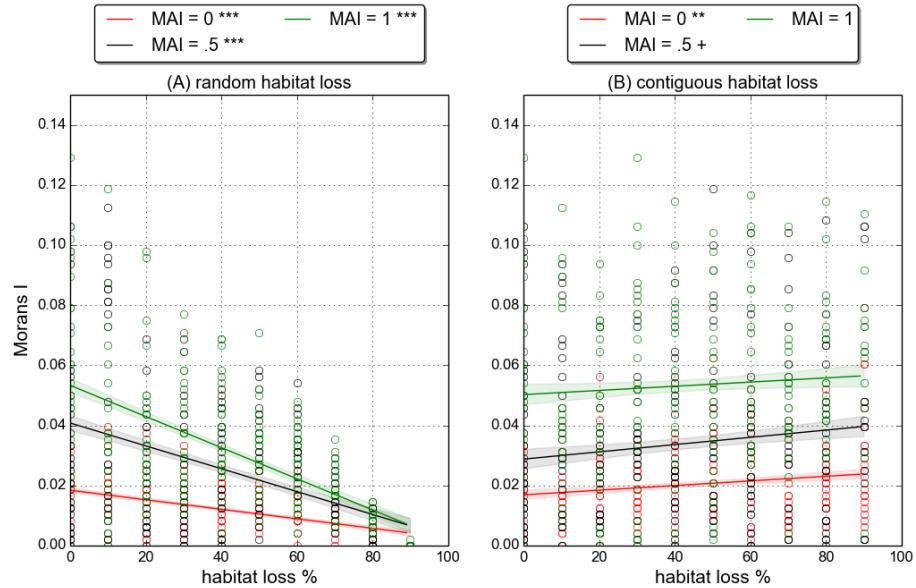


Figure 3.17: **Moran's I** metric for global aggregation (defined in section 2.7.3, averaged over all species. Linear fits and p-value markers as in previous plots (see caption of figure 3.7).

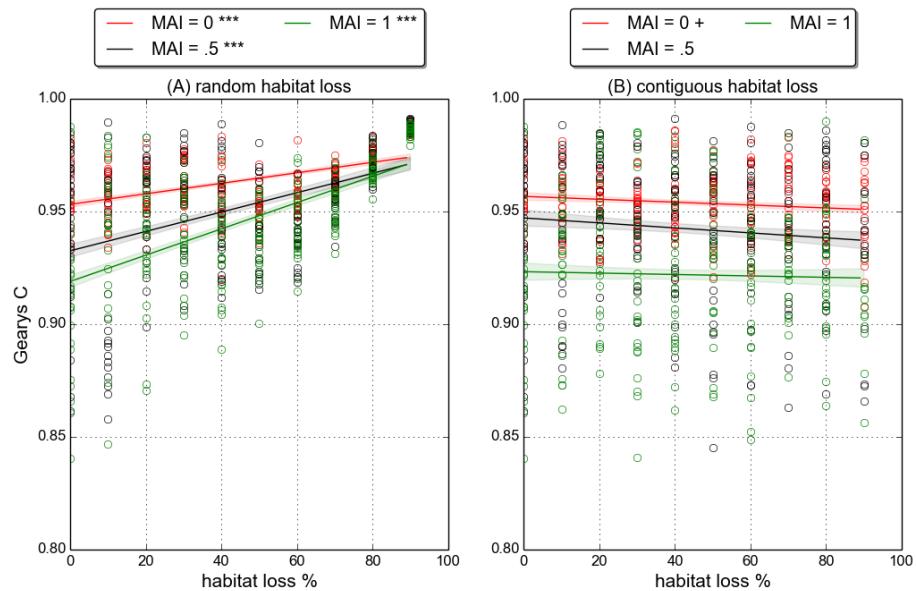


Figure 3.18: **Geary's C** metric for local aggregation (defined in section 2.7.3, averaged over all species. Linear fits and p-value markers as in previous plots (see caption of figure 3.7).

### 3.3. RESULTS

---

The increase in plant aggregation cascades up to higher trophic levels since species thrive close to aggregations of food resource. (This effect can be seen in some of the animations at [75]<sup>5</sup>.) In response to random HL, species on average become less aggregated in space at both the local and global scales. This is expected because the way in which habitat is destroyed creates a patchy landscape which acts against aggregation. Conversely there is some evidence that contiguous HL leads to a slight increase in aggregation. However the only significant linear trend is in MI at MAI= 0.0 (figure 3.17B). Therefore we conclude that contiguous HL does not create significant and robust changes in aggregation. This may be linked to the reduction in stability, since dynamics become more variable it may be harder for species to form local aggregations in space.

---

<sup>5</sup>Specifically which ones?

### 3.3.4 Invariability

The alternative stability metrics defined in section 2.7.2.1 are based on *invariability*. These include minimum invariability, population invariability, ecosystem invariability and ecosystem synchrony. Here we demonstrate that the first three of these metrics respond in the same way to HL as the *variability metrics*, giving more insight into the dynamics of community invariability. Sec ref, and what conclusions? The fourth metric, ecosystem synchrony, gives us a new piece of evidence about how the dynamics of the communities change in response to HL.

Figure 3.19 shows the response of minimum invariability to HL. We use the natural log-transform of the data, as in section 3.3.3, because of the apparent exponential nature of the trends in the raw data. From the figure we see that minimum invariability increases in response to random HL, whereas it decreases in response to contiguous HL. The trends are the same for population and ecosystem invariability, and statistically significant at all MAI ratios (results not shown). These changes in invariability are in agreement with the results of the previous section - community dynamics becomes less variable under random HL, and more variable under contiguous HL. The interpretation of these metrics in [79] lends support to the conclusion that the observed changes in temporal variability are also associated with changes in the stability of the community in a dynamical systems sense (see section 2.7.2.1). Specifically the changes in minimum invariability likely to correspond to similar changes in the asymptotic resilience (derived from the system Jacobian) that governs the rate of return to equilibrium.

What is Jacobian?

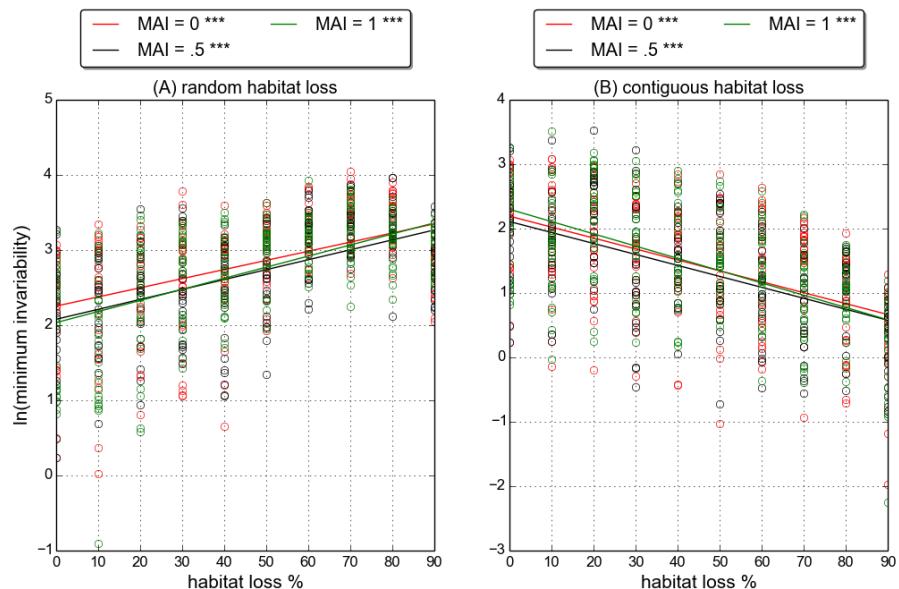


Figure 3.19: **Minimum invariability** as defined in section 2.7.2. Linear fits and p-value markers as in previous plots (see caption of figure 3.7).

### 3.3. RESULTS

---

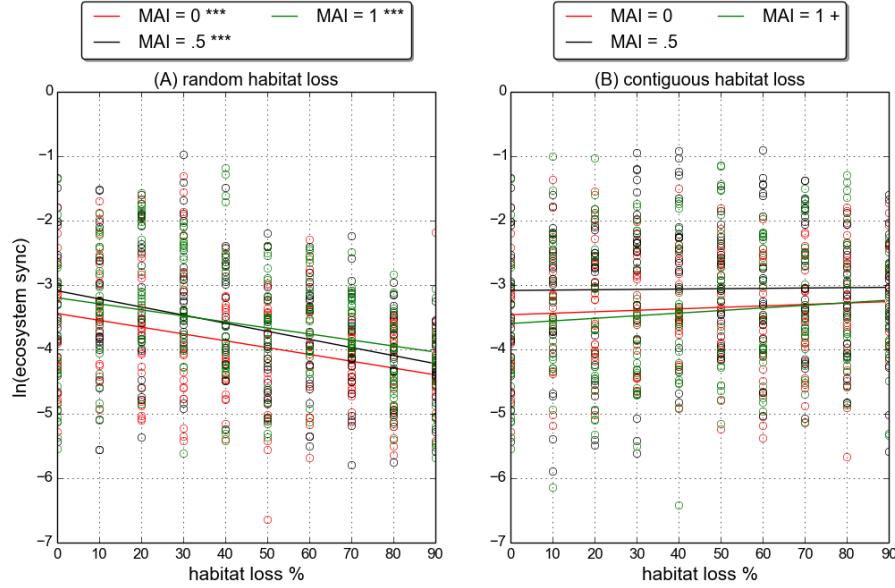


Figure 3.20: **Ecosystem synchrony** as defined in section 2.7.2. Linear fits and p-value markers as in previous plots (see caption of figure 3.7).

Figure 3.20 shows linear fits to the ecosystem level synchrony for the two HL scenarios. In the random case there is a significant decrease in synchrony, whereas in the contiguous case there is no significant change. Also in all cases the synchrony is relatively low. Perfectly synchronised dynamics for all species would give a log-synchrony value of 0. Therefore all communities are well below perfect synchrony. This is expected from trophic dynamics, since it is well known the predator-prey dynamics leads to a phase lag between the population of predator and prey. However it is important to note that trophic dynamics should also lead to some level of synchrony, as species tend to respond in the same way to fluctuations in a shared resource or predator. Therefore it may be that the change in synchrony under random HL is a signature that the trophic component of the population dynamics is reduced, relative to the stochastic component. This would agree with the observed reduction in interaction strengths shown in figure 3.16.

### 3.4 Discussion

Coupling?

In this section we summarise the results presented in section 3.3, and attempt to synthesise them into a coherent explanation of mechanisms driving community responses to HL. Under contiguous HL communities displayed fewer changes than under random HL. *In the contiguous scenario* we saw that the number of individuals, the frequency of interactions, and to a lesser extent, the number of links, all decreased with HL. However the mean interaction strength and temporal variability increased, representing a reduction in dynamic stability. There were no robust trends in network properties, diversity, evenness, relative abundance by functional group or spatial aggregation, under contiguous HL. *In the random scenario*, as in the contiguous, the number of individuals, the frequency of interactions, and the number of links, all decreased. Unlike the contiguous case, the mean interaction strength and temporal variability decreased, representing an increase in dynamic stability. Under random HL communities became more diverse and more even; displayed a shift in relative abundance towards basal species; and became less aggregated in space. There was also an increase in quantitative vulnerability; a decrease in quantitative generality; and an increase in the interaction diversity of the mutualistic sub-network.

In section 3.3.3 we demonstrated that there is a significant correlation between mean interaction strength and temporal variability. Theoretical population dynamics, in general, suggests that strong inter-specific interactions are destabilising in antagonistic systems [20, 61, 71, 73, 84], and there have been some empirical observations of this effect [86]. In a simple predator-prey model, such as the Lotka-Volterra model, it can be shown that increasing the strength of the coupling between species leads to larger amplitude oscillations and may, in certain models may lead to the extinction of species due to over-predation. This consequence of strong interactions is intuitive since the stronger the coupling, or interaction, between species, the greater the effect that one has on the other. In 1972 May showed [71] that for large assemblages of species, with random interactions, the probability of stability is reduced by strong interactions. Together the body of work cited represents a general consensus that antagonistic dynamics are destabilised by strong interactions. Less is known about the dynamics of mutualistic systems, especially when they are embedded in a larger trophic system [68, 98]. However, in some studies mutualisms have been found to play a destabilising role [20, 72]. Again this consequence of mutualisms is intuitive, since an interaction which provides mutual benefit to both parties can easily lead to a destabilising positive-feedback. In the IBM we model mutualisms as trophic interactions. The animal mutualist consumes resource from its plant partner in addition to providing the reproduction service. Therefore it is not clear whether the net result of mutualistic interactions on the plant population as a whole is positive or negative. However in either case it can be argued that stronger mutualistic interactions will be destabilising. *Therefore we assert that the observed changes in interaction strengths are driving the changes in temporal variability under HL* (figure 3.16).

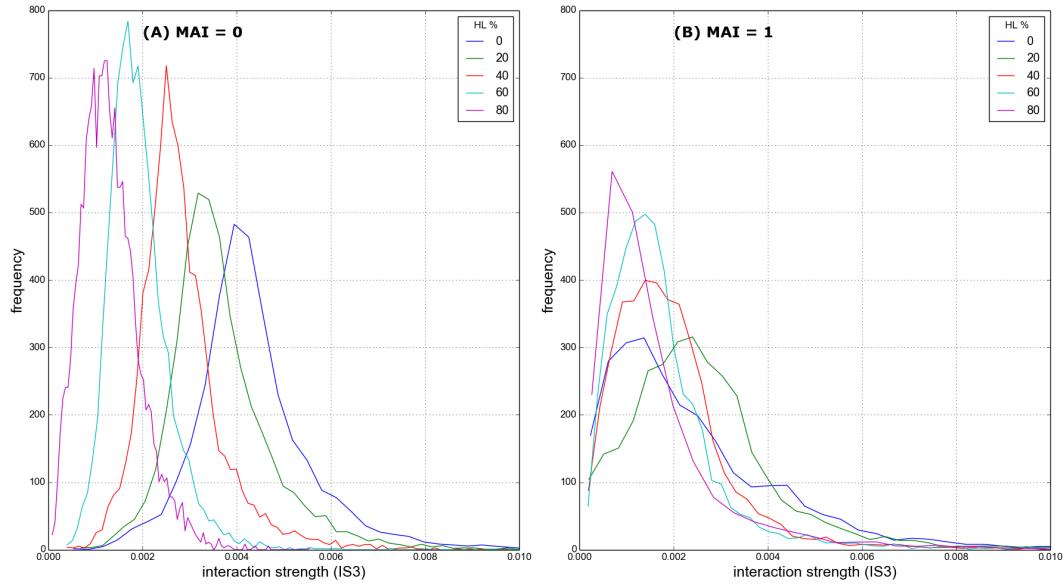
In section 3.3.2 we argued that *changes in the distribution of species abundances can explain how the quantitative network properties respond* under the two HL scenarios. In particular we explained how the evenness of species abundances affects interaction frequencies. Since evenness changes in the random but not in the contiguous scenario, this is able to explain why quantitative network properties are observed to respond to one type of HL but not the other. However we have yet to explain the differential response of species relative abundances under the two scenarios. Therefore we are left with two main questions: what is driving the change (or lack therefore) in the distribution of species abundances? And, what is driving the change in interaction strengths? In what follows we show that the two are closely related.

We first address the question of interaction strengths. In both HL scenarios the total number of individuals and the total number of interactions decreases. However in the random and contiguous scenarios mean interaction strengths decrease and increase respectively. Figures 3.21 and 3.22 show that these changes in mean interaction strength are due to the entire distribution of IS shifting in opposite directions. Distributions are plotted for MAI= 0.0 and MAI= 1.0, showing the same shifts in both cases. We focus on the MAI= 0.0 case because the pattern is clearer (panel A in these two figures). Under random habitat loss the distribution of IS shifts left, towards weaker interactions. The spread also decreases slightly, suggesting less variance in interaction strengths. Under contiguous habitat loss the distribution shifts towards higher interaction strengths, and also becomes much flatter, such that there is a greater spread in interaction strengths. Importantly we see that conclusions based on the mean value of IS were not misleading, for example due to a highly skewed distribution. In fact the majority of interactions in a contiguous landscape are stronger than the the majority of interactions in a randomly destroyed landscape, for a given level of HL. However we know, from figure 3.1, that the total number of individuals is similar in either landscape. Since IS is calculated by dividing the interaction frequency by the abundance of the two interacting species, it must be the case that interaction frequency is lower for a given number of individuals in a randomly destroyed landscape than a contiguous one. Indeed we saw evidence in figure 3.10 that species interact less frequently in the random scenario than the contiguous, despite comparable total numbers of individuals.

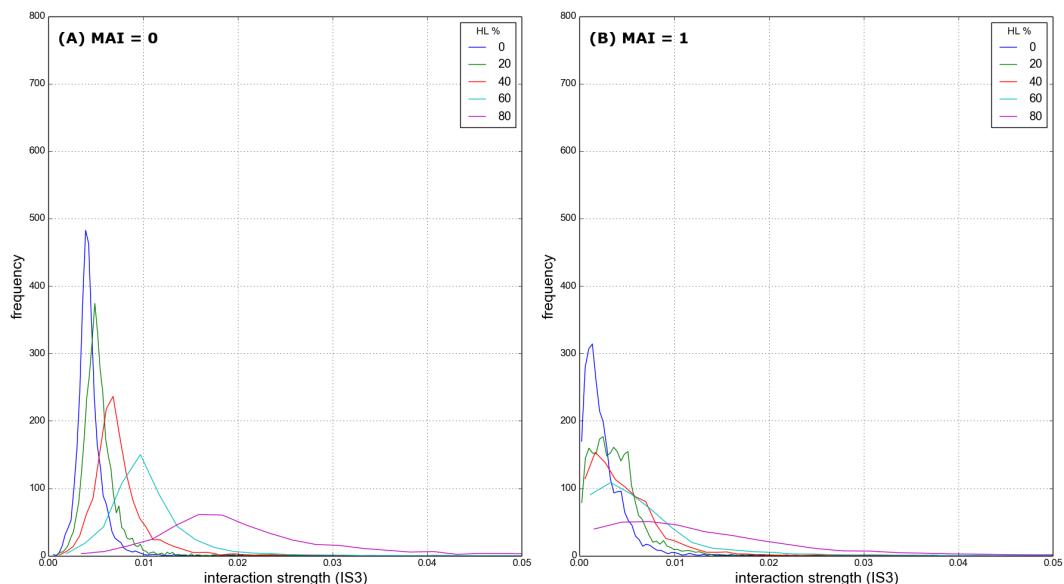
Why would species with the same abundance interact more frequently<sup>6</sup> in a contiguous landscape compared to a landscape with random HL? The answer is simply that randomly destroyed cells present a barrier to the motion of individuals. To demonstrate this we conduct a series of simulation experiments, using the following procedure. We place a single individual randomly in a  $200 \times 200$  landscape. The individual moves according to the same rules that the animal individuals follow in the IBM (defined in section 2.3), but without the bioenergetic constraints. We record what fraction of the available (i.e. not destroyed) landscape cells the individual visits during 5000 time steps. The experiment is repeated 100 times for each level of

---

<sup>6</sup>Make clear statement here, or somewhere else that IS can be thought of as a probability of encounter.



**Figure 3.21: Interaction strength distributions (IS) under random HL.** Panel A: MAI = 0.0. Panel B: MAI = 1.0. IS values for all interactions in each of 25 replicate simulations at the given MAI and HL value, frequency in 100 bins of equal width.



**Figure 3.22:** Similar to figure 3.21, but for **contiguous HL**.

HL, and for both HL scenarios, to obtain the *expected range of motion* of an individual in each type of landscape. The results are shown in figure 3.23. Panel A shows an example trajectory of an individual over 5000 time steps in a pristine landscape. Panel B shows the same, but for a landscape at 40% random HL. It is clear that the range of motion is severely restricted by the destroyed cells. In the contiguous case at 40% HL, depicted in panel C, we see that an individual does not experience such barriers to motion (except at the edges of the available habitat). Therefore an individual has the same dispersal ability, but a smaller space to search. The result is that the percentage of available landscape that an individual can explore during 5000 time steps increases under contiguous HL, but decreases under random HL. This explains the observed changes in interaction strength. If an individual is less mobile it is harder for it to find interaction partners, even if the potential partners are present in the landscape at the same abundance. The converse is true for individuals with increased mobility.

Another consequence of the reduced mobility of individuals is that, *ceteris paribus*, it decreases the probability of intra-specific interactions. These interactions are required for the sexual reproduction of non-basal species since individuals must encounter a member of the same species in space in order to create offspring. Intra-specific interactions are not recorded during the simulations, but the total number of births and total number of immigrations for each species is recorded. In figure 3.24 we plot the proportion of total births during a simulation that are due to immigration. Here total births includes all new individuals that are created due to immigration, sexual reproduction, mutualistic reproduction, and wind dispersal of plants. From the figure it is clear that immigration is the main source of new individuals contributing over 50% of new individuals in almost all simulations. We also see that the relative contribution of immigration is roughly constant under contiguous HL, whereas immigration becomes more important under random HL. We can attribute the differing contribution of immigration, in part, to the changes in mobility illustrated in figure 3.23. In the random scenario it becomes harder for individuals to find a mate and therefore reproduce, shifting the balance in favour of immigration. In the contiguous case we may expect the opposite i.e. a reduced contribution from immigration due to individuals' increased ability to find a mate. However such an effect is not present and the contribution of immigration is constant under contiguous HL. **Too strong** production due to increased mobility is offset by some other mechanism. A candidate mechanism is increased predator mobility, meaning that prey species are more likely to be consumed before they can find a mate. This effect may be compounded by the high relative abundance of predator species in the contiguous scenario, even at high levels of HL. Indeed relative abundance and mobility of predators here suggests a strong predation pressure, which is supported by the high inter-specific interaction strengths.

This leads us to the second question posed above, regarding the distribution of species abundances. We propose that the contribution of immigration to total births explains the observed changes in species relative-abundance and spatial distributions. The immigration mechanism is a

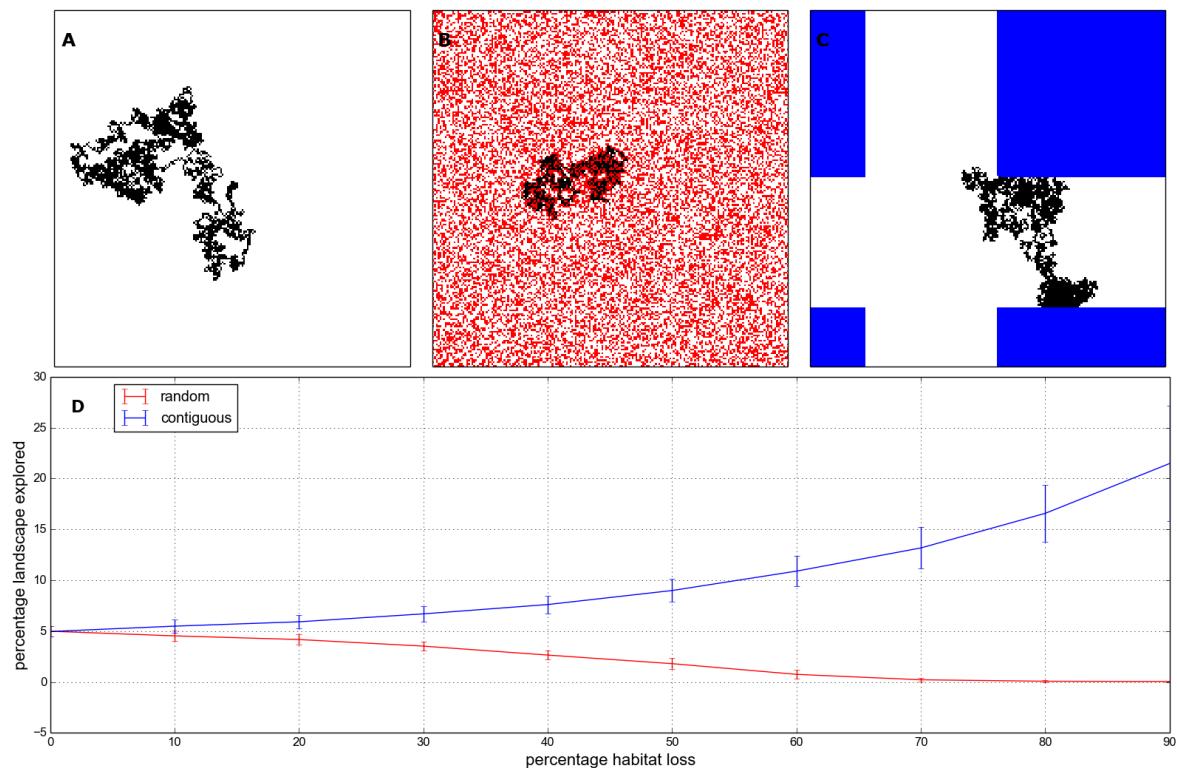
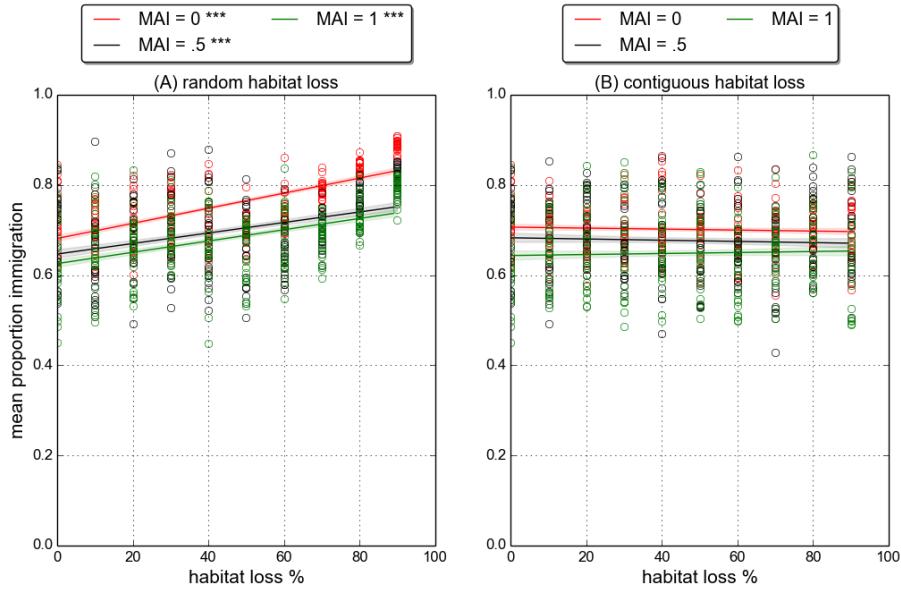


Figure 3.23: **An individual's range of motion** in different habitat conditions. Top row: example trajectory for a single individual over 5000 time steps in (A) pristine landscape; (B) 40% random HL; (C) 40% contiguous HL. Pristine landscape cells shown in white, destroyed cells in red and blue for random and contiguous destruction respectively. Bottom row (D): Percentage of the pristine landscape cells explored by an individual during 5000 time steps. Solid lines indicate mean over 100 repeat runs; errorbars indicate  $\pm 1$  standard deviation.

*levelling influence* on the communities, both spatially and between species. All species are equally likely to immigrate, with no dependence on their abundance or distribution in the landscape. Also all empty cells are equally likely to receive an immigrant at each time step. Therefore, although there is no spatial preference built into the immigration mechanism, areas of space with a high density of individuals will be locally less likely to receive immigrants than those areas with a low density, simply due to the number of available cells. Therefore immigration, in isolation, acts to make the distribution of individuals more even between species and throughout space. These are both changes that we observe in the random scenario and not the contiguous, and therefore we attribute them, at least in part, to the increased dependence on immigration to supply new individuals.

Another important change, observed in the random but not the contiguous scenario, is the shift in relative abundances of the functional groups. The greatest changes are the increase and decrease in relative abundance of basal and top-predator species respectively. However there is also a decrease in the relative abundance of species in the second trophic level (herbivores and

### 3.4. DISCUSSION



**Figure 3.24: Number of immigrants** as a fraction of the total number of new individuals created over the course of a simulation. Linear fits and p-value markers as in previous plots (see caption of figure 3.7).

mutualistic-animals). Only primary predator species display no change in relative abundance. We assert that all of these changes in relative abundance can be explained by the change in mobility and its knock on effect on interaction strengths, reproduction and dependence on immigration.

As previously discussed, the reduction in mobility makes it harder to find prey and therefore reduces interaction strengths. The result is an overall decrease in predation. So species lower in the food chains benefit from reduced predation pressure, whilst species higher in the food chains suffer because less energy is being transferred up the trophic pathway. In addition reduced mobility makes it harder for animal species to find partners with which to reproduce. This negatively affects all animal species and mutualistic plants. However mutualistic plants, although they may be reproducing less frequently, receive the incidental benefit of losing less energy to their mutualistic partners. Therefore we see a shift in relative abundance towards basal species, which benefit from reduced consumption whilst all other species suffer from a reduced ability to find food and reproduction partners. Animal species in the lower trophic levels receive a compensatory benefit of lower predation pressure, which partly explains why species in the second trophic level are hit less hard than top predators. The fact that the relative abundance of **emph, predation pressure from their predators.** be a combination of reduced predation pressure, but also a slight quirk of the modelling framework. Due to the way the networks are constructed, there are significantly more primary predator species than any other functional group (this an known artefact of the niche model - see section 2.2.2 and figure ??). Therefore, since immigrant

individuals are drawn uniformly at random from all species. This is not clear e  
primary predators. As we have seen (figure 3.23) random HL increases the contribution of immigration. We now understand that this benefits primary predators disproportionately. Conversely the functional group with fewest species is the top-predators. Therefore this group receives fewest immigrants, compounding the other effects detailed above. Therefore it appears that mobility, along with immigration, accounts for all observed changes in species relative abundances.

It is worth here reiterating the point about ecosystem synchrony made in section 3.3.4, where we attributed the reduced synchrony under random HL to a reduction in the trophic component of the dynamics. In this section we have shown that random HL increases the contribution of immigration to the overall number of births (figure 3.24). The creation of individuals due to immigration is a random processes. Births due to reproduction, although stochastic, are dependent on trophic and non-trophic interactions. Therefore the increased contribution of immigration represents a reduction in trophic dynamics and an increase in stochasticity under random HL.

Indeed! We conclude that it is indeed this shift driving the decrease in synchrony. We will return to the issue to deterministic versus random dynamics in chapter 4.

### 3.5 Conclusion

A key feature of the results presented in this chapter is that the community response to HL is dependent on the spatial pattern of the perturbation. This spatial dependence has been reported previously in numerous studies [30, 48, 53, 88, 101, 110, 120]. In general the cited studies support the conclusion that spatially-correlated HL is less detrimental to communities than random HL because the former leaves larger areas of landscape unaffected, while the latter results in patchy and fragmented landscapes. Our findings broadly agree. However, while previous theoretical studies have focused largely on extinction thresholds, we have conducted a comprehensive analysis of community responses in the absence of extinctions.

Contiguous habitat loss was mainly found to reduce temporal stability due to increased interaction strengths. Other aspects of the communities (diversity, network properties, spatial aggregation) were generally constant across the contiguous HL gradient. As such contiguous HL resulted in communities with lower absolute abundances, confined to a smaller region of space, and displaying greater temporal variability but with the same biodiversity patterns. This leads us to speculate that there may exist a scaling of time and space under which the systems at low and high HL are equivalent. However we do not explore this possibility further.

Random habitat loss was found to induce a breakdown in the trophic structure of communities, resulting from weaker species interactions and characterised by the collapse of top-predator populations. Primary predator populations were only spared, and extinctions prevented, by the immigration mechanism. Somewhat surprisingly communities under random HL became more even in terms of species relative abundances, and the temporal dynamics became less variable. Taken in isolation these results suggest that random HL is beneficial for diversity and stability. This mistaken conclusion highlights the importance of considering multiple aspects of diversity and stability when analysing multi-trophic communities (see section 2.7.2). In other aspects random HL is clearly damaging to diversity, mainly in terms of top-predator abundance. Also the low abundance of top-predator species makes them more vulnerable to stochastic extinction, thus reducing community stability in this sense<sup>7</sup>.

We predicted that MAI ratio would mediate community response to HL. In fact, in almost all cases, mutualistic communities responded in qualitatively the same way as antagonistic communities. Therefore we find no significant evidence that mutualism makes communities either more or less robust to HL. One notable difference in response was IS under random HL (figure 3.16). High MAI communities did not lose interaction strength to the same extent as low MAI communities. This effect has not been explained. We suggest that it may be due to spatial aggregation, which was higher for mutualistic communities across the HL gradient (figure 3.17). Spatial aggregation means that individuals are not homogeneously distributed through the landscape which could cause spatial effects that skew the distribution of interaction strengths. Indeed this may explain why there is more spread in the IS distributions at MAI= 1.0 than at

---

<sup>7</sup>What about the network properties - vulnerability and generality.

$MAI = 0.0$  (figures 3.18 and 3.17). Interestingly the fact that IS is less sensitive to random HL in mutualistic communities does not translate into different variability responses.

Nestedness and compartmentalisation showed no significant trends in response to either HL scenario. These metrics have often been associated with stability in mutualistic and antagonistic communities respectively [107]. It is interesting to note that neither metric changed, despite significant and opposite trends in temporal stability under random and contiguous HL. However we did not explicitly test for the impact of these metrics on stability. It would be possible, and potentially informative, to study whether communities with higher nestedness/compartmentalisation responded differently to other communities.

Crucially we have found that, in the default parameter regime, the effects of HL are governed largely by interaction strengths and immigration. In particular interaction strengths are responsible for changes in temporal variability, and to some extent biodiversity patterns. In subsequent chapters (especially chapter ??) we will return to the role of interaction strengths. Immigration has been shown to be a stochastic influence on the dynamics, and a levelling influence between species and throughout space. Also immigration was found to prevent species extinction, allowing communities to persist in the face of high levels of HL. In this sense the communities modelled in this chapter represent *open systems* with a high rate of influx from an external source. In the ecological context, although the local landscape is almost entirely destroyed there is sufficient immigration from surrounding sources (see IBT discussion in section 1.3)<sup>8</sup> to maintain diversity. In what follows we will consider what happens when the immigration rate is varied. We begin chapter 4 with a study of *closed-communities* ( $IR = 0$ ). In chapter ?? we investigate how communities respond to HL over a range of immigration rates.

---

<sup>8</sup>Remove this ref?

CHAPTER



## PERSISTENCE AND STATIONARITY

### 4.1 Introduction

In the chapter 3 we analysed in detail how simulated communities responded to two habitat loss scenarios. These simulations used the *default parameter values* for the IBM model, as published in [68]. An important feature of these simulations was that they exhibited almost no species extinctions, even up to 90% habitat loss (HL). This due to a high immigration rate ( $IR = 0.005$ ), providing a rescue effect for all species. In nature the destruction or degradation of habitat often leads to a loss of species [34? ]. Therefore we also intend to use the IBM to investigate the regime where habitat loss causes extinctions. The obvious way to do this is to reduce the IR. Since immigration proved to be a key driver of the dynamics in chapter 3, in this chapter we conduct a detailed analysis into the consequences of varying the IR parameter. This is largely in preparation for chapter 5 where we will study, from a more ecological perspective, how changing IR affects the response of communities to HL. We restrict the scope of this chapter to pristine landscapes ( $HL = 0$ ). In the first part of this chapter (section 4.2) we study *closed communities* by removing immigration from the model altogether ( $IR = 0.0$ ). In the second part (sections 4.3 to 4.6) we address an issue that is found to arise from reduced IR, namely *increased temporal variability*. In section 4.4 we revisit the assumption made in the previous analysis, that simulations reach steady-state after 5000 time steps (see section 3.2.2). In section 4.5 we test for determinism in the simulated population dynamics, and in section 4.6 we apply different sampling regimes to compare the convergence and repeatability of our experimental results.

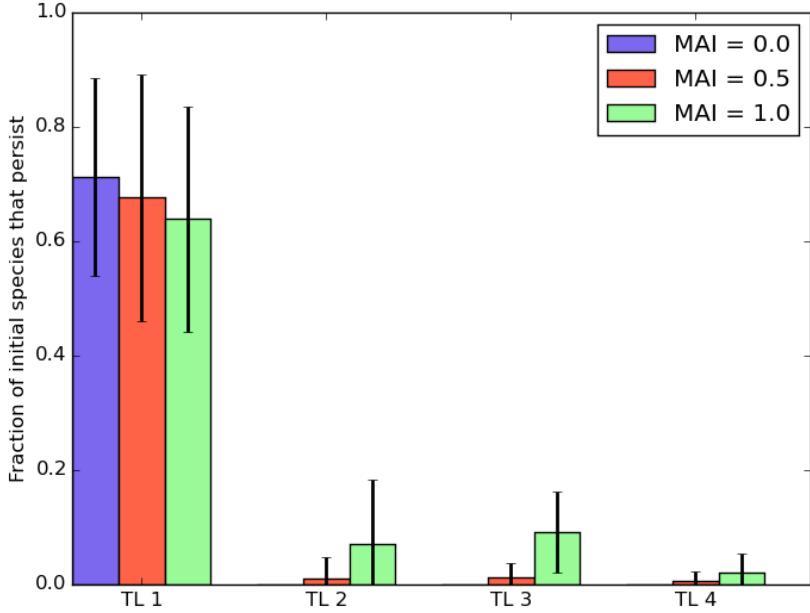


Figure 4.1: **Fractional persistence** by trophic level for three different MAI ratios, with **zero immigration** ( $IR=0$ ). All other parameters set to default values. Fractional persistence measured by the fraction of species initially belonging to a trophic level which have not gone extinct by the end of a simulation (5000 time steps). Solid bars give the mean value over 22 repeat simulations. Error bars show  $\pm 1$  standard deviation.

## 4.2 Persistence in closed communities

We define a *closed-community* as one in which there is no inflow of individuals from an external source. In the model this is achieved by setting the immigration rate to zero ( $IR=0$ ). Here we study such communities in a pristine landscape i.e. without habitat loss ( $HL=0$ ). The simulation procedure is the same as in chapter 3. All parameters, except  $IR$ , take default values unless otherwise stated, and each simulation uses a distinct network topology generated with the method described in section 2.2. A species is said to *persist* if there is at least one individual belonging to that species present in the landscape on the final time step of the simulation. Therefore *fractional persistence* is defined as the fraction of the initial pool of species (or subset of that pool) that persist. We also use the *number of persistent species*, where absolute values are desirable. All abundance measurements given are calculated from the number of individuals belonging to each species on the final (5000th) time step of the simulation. Simulations were run with three MAI ratios ( $MAI = 0.0, 0.5, 1.0$ ) to explore the full range between antagonism and mutualism. To reduce the number of figures we sometimes present results for only one or two MAI ratios only, but make it clear where results differ across MAI ratios.

An initial set of simulations with zero IR shows that species persistence is low (Figure 4.1). We see that for an antagonistic community ( $MAI = 0$ ) all non-basal species go extinct, as do around 30% of basal species. Introducing mutualism into the communities slightly improves fractional persistence in the higher trophic levels. We see some persistence of non-basal species, but no more than 10% of the initial pool on average survive. Therefore, even in the *best case* at  $MAI = 1.0$ , we expect around 90% of non-basal species to go extinct. It is clear that, given the default parameters, immigration is a requirement for the maintenance of species richness. This is a troubling feature of the model. In nature a true closed-community may not exist, but there are systems which come close (for example remote island ecosystems). It is desirable for our model to be able to describe such systems, and for us to investigate the effects of habitat loss in this extreme case. It may also be informative to discover which factors in the IBM model contribute to persistence in closed-communities. Here we conduct a systematic review of certain model parameters and their impact on persistence. Due to the large parameter space of the model we study these parameters individually, with other parameter values held constant. We restrict the analysis to four parameters, namely: MAI ratio (section 4.2.1); reproduction rate (section 4.2.2); landscape size (section 4.2.3); and number of initial species (section 4.2.4). In addition we look at the effect of the interaction network structure (section 4.2.5). The choice of each parameter is justified in the corresponding subsections. As in chapter 3 we use linear models to detect trends in persistence in response to the parameter values varied. Again the *p-value* of the fit is used to test for significance in any trend identified (see section 3.2.2).

### 4.2.1 Mutualistic-to-antagonistic interaction ratio

In [68] it was shown that increasing the levels of mutualism (MAI ratio) in the IBM can have a stabilising effect. However in the previous chapter we saw that this stability did not translate into a greater robustness of mutualistic communities to habitat loss. From figure 4.1 we see that mutualism has a small but positive effect on fractional persistence at zero IR. An interesting feature of this effect is that it *cascades* to higher trophic levels, benefiting species other than mutualists. Here we explore the effect of mutualism on persistence in more detail.

Figure 4.2 shows the average abundance dynamics, by functional group (FG), for four different MAI ratios. These plots are produced by taking the mean abundance of each trophic level, at every time step, over 22 repeat simulations. In panel A we see that the abundance of producers rises to fill the whole landscape ( $200 \times 200 = 40,000$ ), while the abundance of all other FGs is at or near zero. From the other panels (B-D) we see that increasing the MAI ratio particularly benefits the mutualistic-plants and mutualistic-animals. However this also appears to also confer some small benefit to primary-predators and top-predators, as their abundances are small but non-zero. Ecologically this makes sense. If mutualism strongly benefits mutualistic species, it will also benefit those species that feed on them. However the benefit may only be a transient effect as the abundance of producers in panel D is still rising, on average, at the end of the simulations.

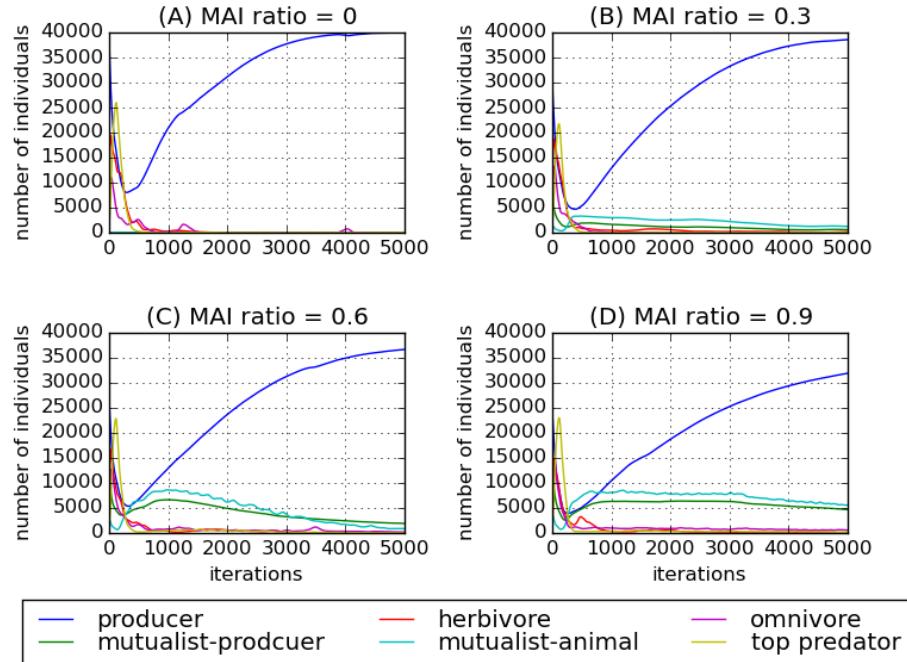


Figure 4.2: **Mean dynamics by functional group** for four different MAI ratios (A-D), without immigration. Lines given by number of individuals belonging to a functional group, averaged over 22 simulations. Colours indicate functional group (see legend).

Figure 4.3 shows the number of persistent species by trophic level for a range of MAI ratios. We see that increasing MAI ratio in fact has very little effect on the absolute number of persistent species in each trophic level. For example at MAI= 1.0 there are only two persistent species on average in the second trophic level. Therefore the increase in abundance seen in figure 4.2 must be due to a small number of dominant species. We conclude that mutualism has a negligible effect on overall persistence in terms of species richness.

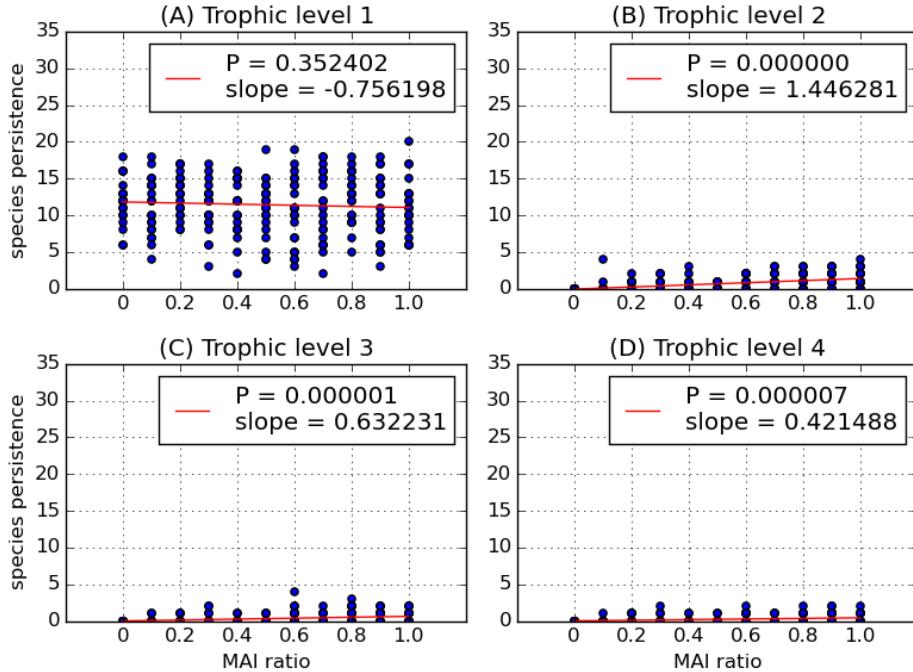


Figure 4.3: **Number of persistent species** plotted against MAI ratio, for each trophic level (A-D). Blue circles give number of persistent species, for a single simulation, on the 5000<sup>th</sup> time step. 22 repeat simulations at each MAI ratio. Red lines show linear regression fits, with slope and p-value of fit given.

#### 4.2.2 Reproduction rate

The initial transience in the abundance dynamics at zero IR (figure 4.2) is characterised by a sharp decline in plant abundance (mutualist and non-mutualist), which reaches a minimum and then rises again. We hypothesise that this overconsumption of producers, and therefore limited availability of food for animal species, causes many of the extinctions. Indeed in these simulations ~85% of the extinctions occur during the first 500 time steps. Therefore we look at the possibility of improving persistence by increasing the reproduction rate (RR). The RR parameter defines that rate at which non-mutualist plants reproduce (via the wind-dispersal mechanism, see section 2.3). Therefore increasing this mechanism should improve the availability of plant biomass in the system, with potentially cascading effects. The RR parameter does not affect mutualistic-plants, which only reproduce via their interactions with mutualistic-animals and not via wind-dispersal. Here we vary the RR between 0.01 (default value) and 0.2 and look at the effect on persistence.

Increasing the RR improves the fractional persistence of all trophic levels, as shown in figure 4.4. This effect is greatest for the two lowest trophic levels, but has a small cascading effect on the upper two trophic levels. However, as we saw with MAI ratio, the improvement in persistence due to RR is small. Even for high RRs we still find that many species go extinct. Looking at the

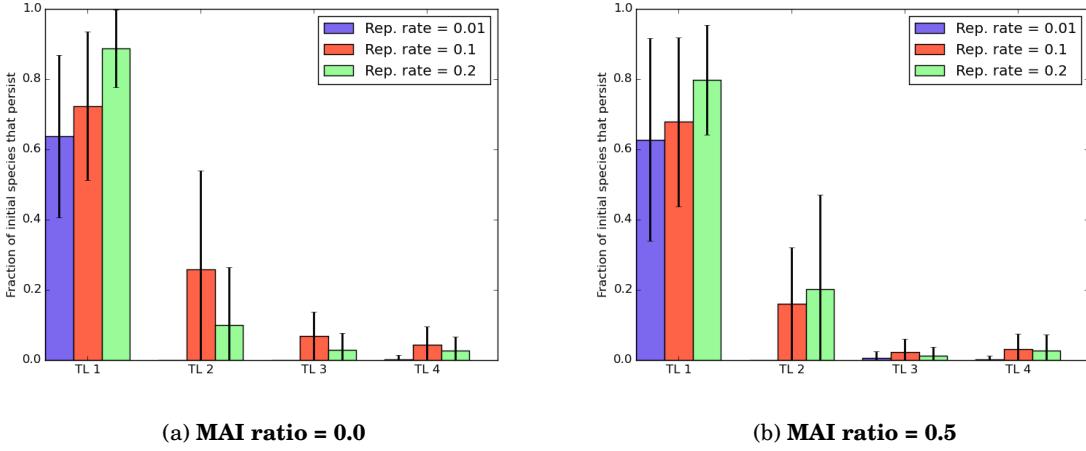


Figure 4.4: **Fractional persistence** by trophic level for three different reproduction rates (RR), at two MAI ratios (A-B). Fractional persistence measured by the fraction of species initially belonging to a trophic level which have not gone extinct by the end of a simulation (5000 time steps). Solid bars give mean value over 22 repeat simulations. Error bars show  $\pm 1$  standard deviation.

abundance dynamics in figures 4.5 and 4.6 we see that increasing the RR reduces the severity of the decline in plant abundance during transience. The resulting increase in the availability of plants does benefit all trophic levels by increasing the number of individuals present. However, because of the low fractional persistence, it must again be the case that these communities are dominated by just a couple of species in each trophic level, with all other species going extinct. The results for MAI= 1.0 (not shown) are qualitatively similar, although these communities contain fewer plants that reproduce via wind-dispersal and can therefore benefit from increased RR. In the remaining simulations for this section, as we continue to search for improvements in persistence, we choose to use RR= 0.1 . This choice reflects the resulting improvement in abundances across trophic levels compared with the default value. It was decided that this effect, combined with changes in other parameter values may help to improve persistence further.

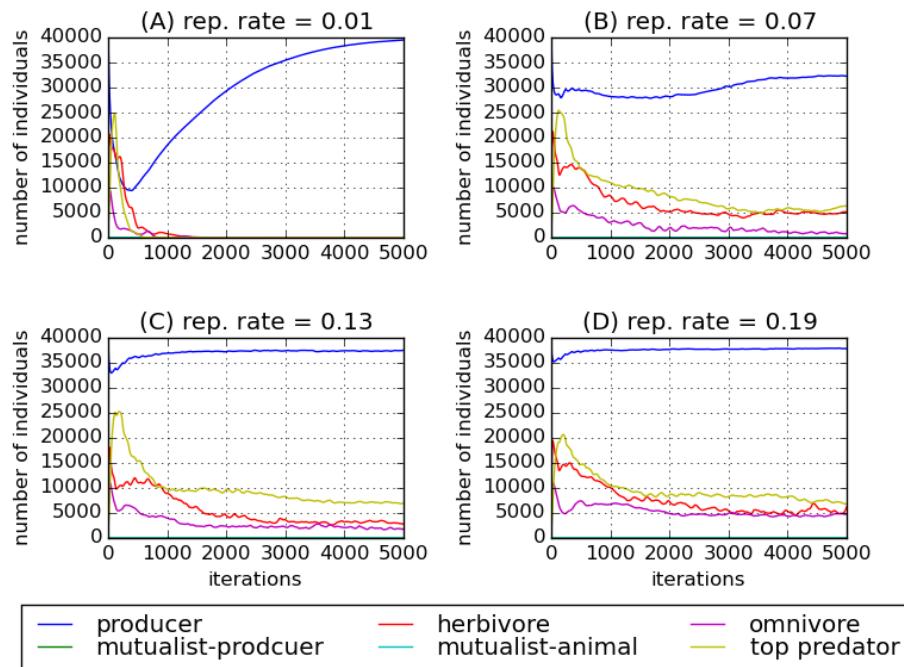


Figure 4.5: **Mean dynamics by functional group** for four different reproduction rates (A-D), at **MAI=0**. Lines given by number of individuals belonging to a functional group, averaged over 22 simulations. Colours indicate functional group (see legend).

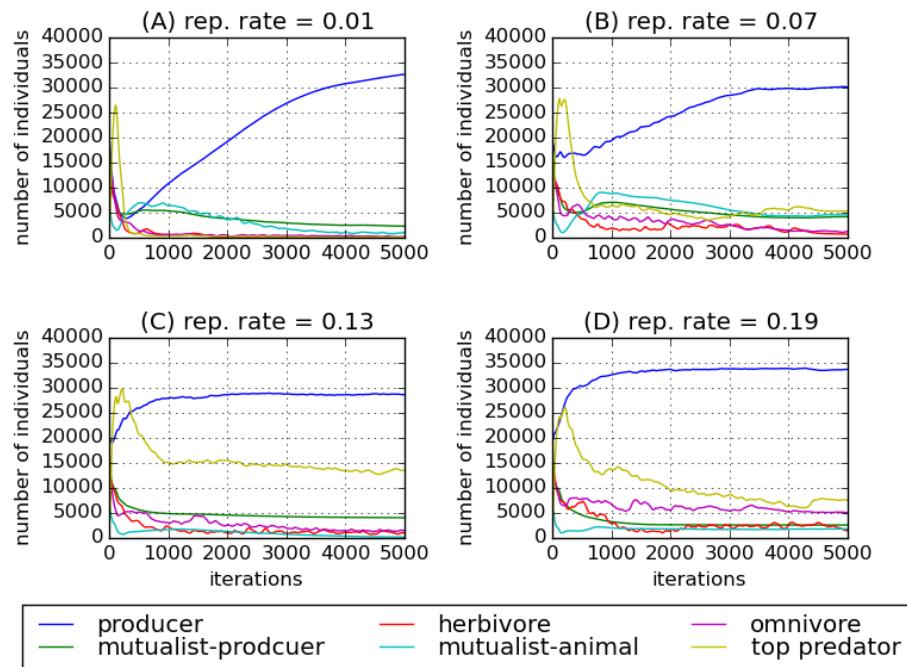


Figure 4.6: **Mean dynamics by functional group** for four different reproduction rates (A-D), at **MAI=0.5**. Lines given by number of individuals belonging to a functional group, averaged over 22 simulations. Colours indicate functional group (see legend).

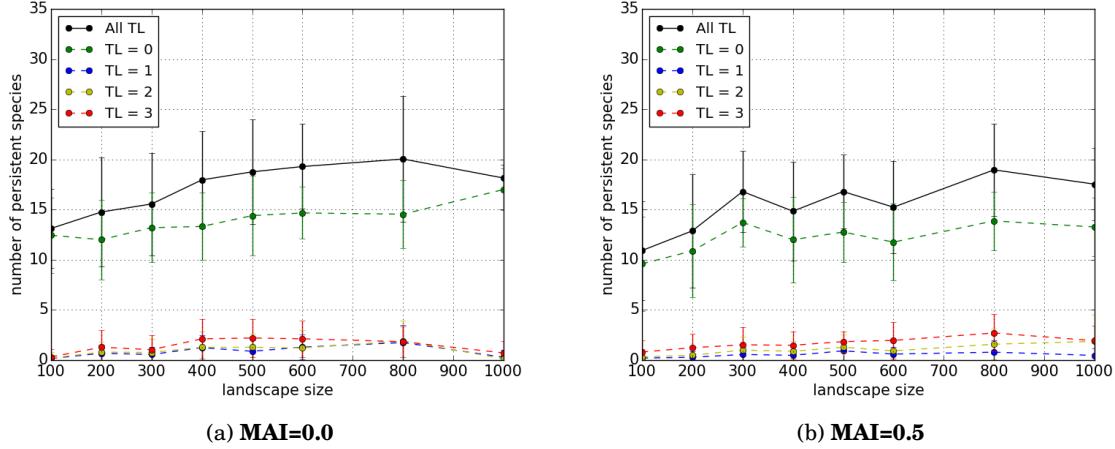


Figure 4.7: **Landscape size** against species persistence (on 5000<sup>th</sup> time step), for two MAI ratios (A,B). Total persistence, and persistence by trophic level. Points show mean value over 25 replicates. Error bars show  $\pm 1$  standard deviation.

### 4.2.3 Landscape size

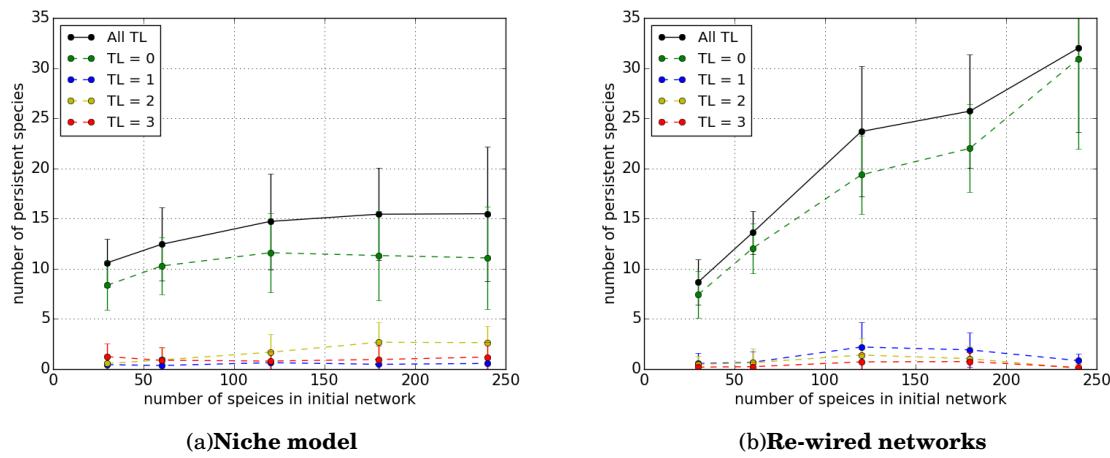
One might hypothesise that competition for space is causing species extinctions, in which case increasing the size of the landscape may improve persistence. We varied the width of the square landscape between 100 and 1000 cells, running 25 repeat simulations at each width. These simulations were run for MAI= 0.0, 0.5, 1.0, and used the increased reproduction rate RR= 0.1. Figure 4.7 shows how the number of persistent species changes in response to landscape size for MAI= 0.0 and 0.5 (results qualitatively similar for MAI= 1.0). For both MAI ratios depicted there is an overall increase in species persistence up to a landscape width of 800. This increase in persistence is driven by small increases in the species richness of each trophic level. Beyond a width of 800 cells persistence appears to decline, although this may be a statistical anomaly and we do not explore this result further. Even with a landscape of 800  $\times$  800 the results show that we should not expect more than 20 persistent species, which represents the extinction of two thirds of the initial species pool. Therefore we conclude that landscape size does not resolve the problem of extinctions. Further simulations are therefore run using the default landscape width of 200, which requires significantly less CPU time than for a width of 800.

#### 4.2.4 Number of initial species

All previous simulations have been run with an interaction network consisting of 60 species. We consider the possibility that beginning the simulation with a larger network may result in a greater number of persistent species. In this way we may hope to evolve to a stable network structure as extinctions kill off species that are not viable. Here we vary the number of species in the network between 30 and 240, and look at the effect this has on persistence. As in the previous section, 25 repeat simulations are run for each number of species at all three MAI ratios (0.0, 0.5, 1.0).

Using a network of more than around 70 species causes a problem with our network generation procedure. As described in section 2.2 we use the niche model to create interaction networks, but reject those that do not satisfy the trophic constraints. With a large number of species the probability of the niche model generating a network that meets the trophic constraints becomes low enough that the procedure cannot be run within a reasonable time (*time* > 2 days on Blue Crystal for 100 species). In particular, as discussed in section 2.2.2, the niche model overestimates the number of primary-predator species, and underestimates the number of herbivores. To overcome this problem we here construct networks in two different ways: 1) we use the standard niche model with no trophic constraints; and 2) we *rewire* networks generated using the niche model iteratively updating the network until it meets the trophic constraints. The procedure is as follows. A species is selected uniformly at random from the trophic levels that contain more than the desired number of species. This species is then moved to a trophic level that contains too few species, and is linked to other species selected at random from a pool of possible candidates. Possible candidates are defined by the new trophic level of the species that has been moved e.g. a herbivore can only eat plants, and only be eaten by species in higher levels. The number of new links created is chosen to preserve the degree of the moved species, and therefore the connectance of the network (in/out degree cannot necessarily be preserved because of the change in trophic level). The procedure is repeated, moving randomly selected species until all trophic constraints are satisfied. Mutualistic links are then introduced using the standard *link replacement* method (section 2.2.3). Here we use the same constraints as were applied in the original network generation procedure. That is the proportion of species belonging to trophic levels one, two and four must be at least 25%, 25% and 5% respectively. We refer to the networks generated using procedures 1 and 2 as *niche* and *rewired*, respectively. Simulations are run for both types of network.

Figures 4.8 and 4.9 show how the number of persistent species depends on the number of species in the initial network. For the *niche* networks we see that there is little change in species persistence. Whereas for the *rewired* networks we see a large increase in total persistence. However this increase is mainly due to plant species, with only a small change in persistence for trophic levels two, three and four. The difference between the *niche* and *rewired* network results is due to the fact that large niche model networks are dominated by omnivores. It appears



**Figure 4.8: Number of initial species** against species persistence at **MAI=0**, for two types of initial network: (A) Interaction network generated using standard niche model, and (B) generated using re-rewired niche model topology (see text). Total persistence, and persistence by trophic level. Points show mean value over 25 replicates. Error bars show  $\pm 1$  standard deviation.

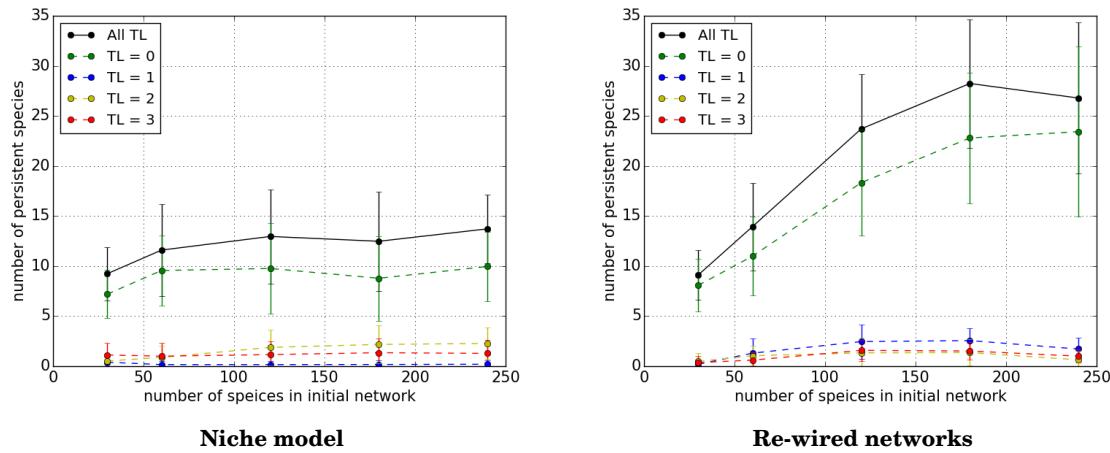


Figure 4.9: Similar to figure 4.8, but with **MAI=0.5**.

that the omnivore species out-compete each other and therefore we do not see a large change in persistence. Conversely, for the *rewired* networks the number of species in each trophic level grows proportionally with the total number of species. The only significant benefit is to plant species, which seem to be able to coexist in larger numbers. We conclude that increasing the number of initial species is not able to generate the diverse communities that we had hoped for. However the *rewiring algorithm* developed for this analysis represents a novel modification to the niche model, which may prove useful for the generation of realistic interaction networks.

### 4.2.5 Network structure

A notable feature of the results presented in this section is that there is significant variation in persistence between replicate simulations. Presumably some of this variation is due to noise. However some of this variation may be due to the structure of the interaction network used. As we know (section 1.3.2) the structure of an interaction is believed to impact on various aspects of community stability. Here we look for systematic differences in persistence that are due to network structure. To do this we choose one ‘good’ and one ‘bad’ network from the simulations used in the section 4.2.4. From the 25 repeat simulations using a *rewired* network of 120 species we select the two communities that display the best and worst persistence profiles. The best case, giving the ‘good’ network, had persistent species in all trophic levels and two persistent species in each of trophic levels three and four. In contrast the worst case, giving the ‘bad’ network, had no species persistence in any trophic level above one. With each of these two networks we ran 100 repeat simulations ( $MAI = 0.0$ ) to test if the difference in persistence was repeatable and therefore due to the network structure.

Panel A in figures 4.10 and 4.11 shows the structure of the 120 species network, for the good and bad cases respectively. Visually there is little to distinguish the two networks. However we find that they have very different persistence properties. Panel C of these figures shows the average number of persistent species in each trophic level. It is clear that the good network performs better on average, in terms of persistence, than the bad network. We do not analyse the network structures here, however this could be an interesting avenue of further work. We simply conclude that the network structure is an important determinant of community dynamics in our simulation, with an effect on persistence that is repeatable.

### 4.2.6 Summary

<sup>1</sup>.

From the results presented in this section it appears that immigration is required in order for the IBM to generate communities with more than a handful of persistent species in the higher trophic levels. Having not exhaustively searched the parameter space we cannot be certain of this fact. However this finding is not in disagreement with our observations of the natural world. Indeed the first laboratory experiments to produce persistent predator and prey populations in a microcosm, conducted by Gause [41], required either spatial heterogeneity (to provide refuge for the prey) or immigration in order to prevent extinctions [24]. Also, in larger systems the principle of competitive exclusion is expected to cause extinctions. The principle states that the number of coexistent species cannot be greater than the number of niches (or alternatively, the number of limiting resources). The reasoning is that two species competing for the same resource is unstable, such that if either species happens to gain an advantage it drives the other

<sup>1</sup>Note that this does not rule out competition as a cause of extinctions? Just that increasing the available space does not remove it.

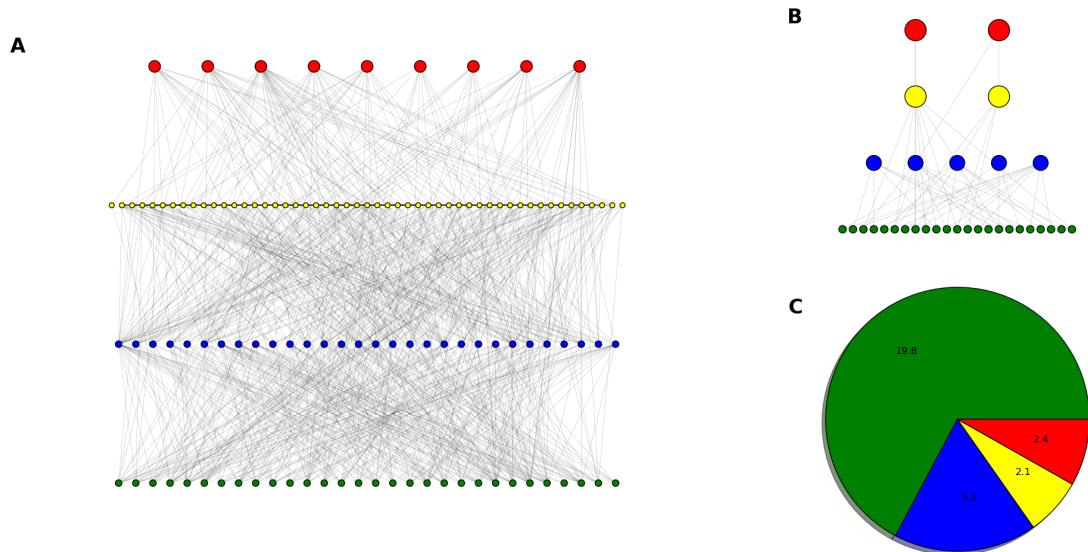


Figure 4.10: Example of a ‘**good network structure**’. Colours represent trophic level. (A) Antagonistic network of 120 species, rewired from the niche model (see text), used as input for simulations. (B) Example of ‘pruned’ network consisting of the species that persist after 5000 time steps. (C) Mean number of species belonging to each trophic level in the pruned network, averaged over 100 replicate simulations (all using A as the initial network).

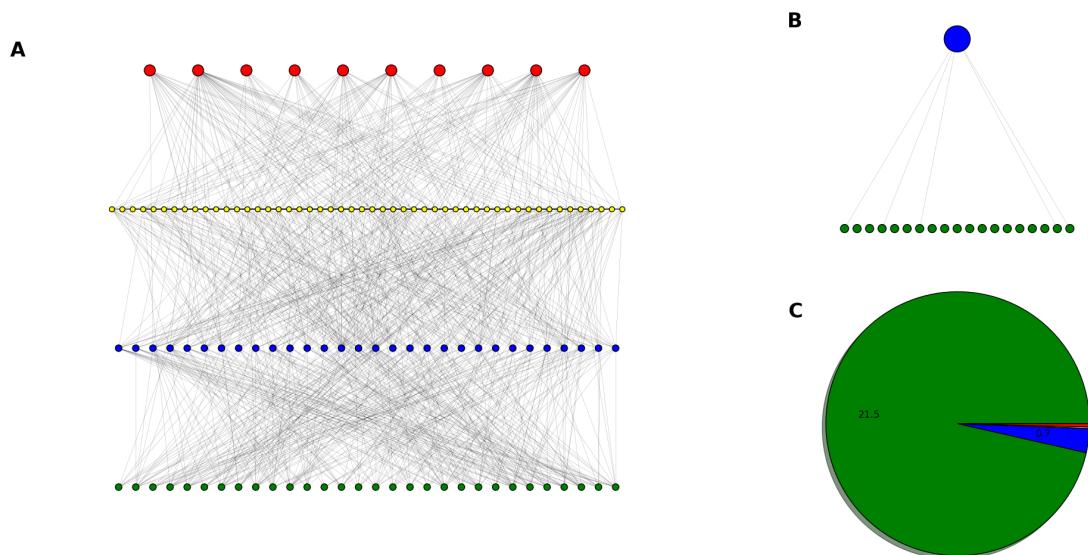


Figure 4.11: Example of a ‘**bad network structure**’. Similar to figure 4.10, but with a different initial network of 120 species used as input for the simulations.

to extinction. The principle of competitive exclusion presents an apparent paradox in Nature because many communities exhibit species richness much greater than the apparent number of limiting resources. A prime example of such communities are oceanic plankton, hence the *paradox of the plankton* [51]. One proposed solution to these paradoxes comes from meta-community theory (see section 1.3). That is, although local populations may be unstable in isolation, they exist as part of a meta-community. Various models predict that meta-community dynamics, which involves dispersal between local patches or communities, can stabilise local populations [49]. The key feature here is *dispersal*, which is not dissimilar to our immigration mechanism because it provides an external source of individuals. Therefore we propose that the extinctions we are seeing in the IBM at zero IR are due to competitive exclusion, an effect which is reduced by the introduction of immigration<sup>2</sup>.

We demonstrated in section 4.2.5 that network structure *can* have an effect on persistence, and therefore on community dynamics also. This results implies that the IBM could be used to study the effect of various network properties on different aspects of stability (see sectionref:WHERE). We also presented a novel algorithm to modify niche model networks according to desired trophic constraints. In order to justify the use of this method further investigation would be required to compare the realism of the derived topologies to empirical webs.

---

<sup>2</sup>Also discuss differential traits?

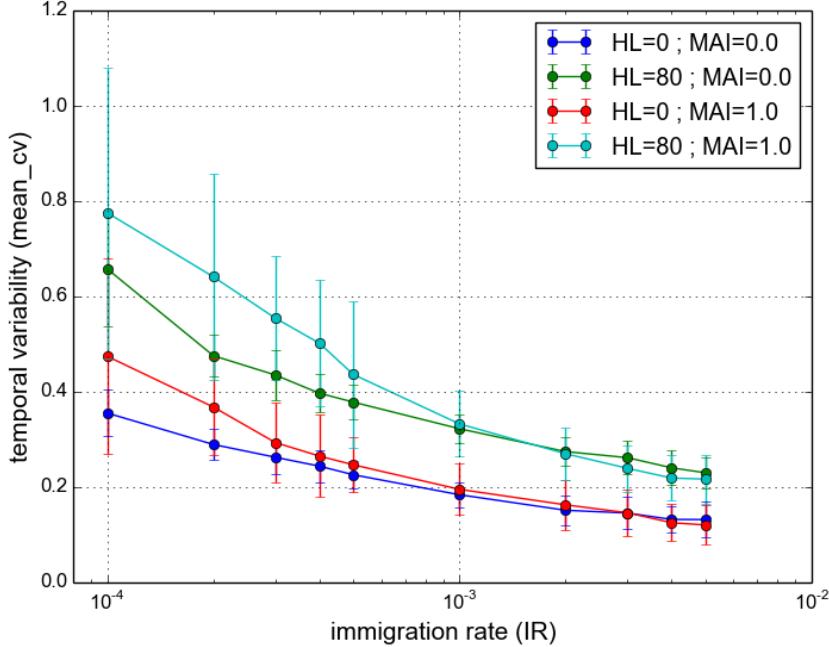


Figure 4.12: **Temporal variability** against immigration rate (IR). Variability calculated as *mean\_CV* (defined in section 2.7.2). Points show mean value over 50 repeat simulations. Error bars show  $\pm 1$  standard deviation.  $HL = 0$ , all other parameters take *default values*.

### 4.3 Temporal variability and immigration

In section 4.2 we concluded that immigration is required for our model to produce persistent communities. Therefore to study communities that are persistent in pristine landscape, but exhibit HL induce extinctions, we must investigate the region between zero immigration and the *default rate* ( $IR \sim 0.005$ ). As we began to investigate this region we found a robust feature of the model: reducing immigration rate *increases the temporal variability* of the dynamics. We again measure temporal variability using *mean\_CV* (section 2.7.2) i.e. the coefficient of variation in species population dynamics, averaged over all species in the community. In the previous chapter *contiguous* habitat loss was shown to increase temporal variability, up to a average value of  $\sim 0.3$  at  $HL = 90\%$  (figure 3.16). Figure 4.12 illustrates that reducing the immigration rate can push the temporal variability well above this level, even in pristine landscape.

This increased variability may bring into question the validity of results derived from the simulation output. Implicit in our analysis so far is the assumption that the system reaches *steady-state* by the end of the simulation (see section 3.2.2). If the system is not at steady-state then our results, especially those calculated from snapshots or short samples, may not be reliable. In the high immigration regime (HIR) the steady-state assumption may be reasonable. From inspection, the HIR simulations contain transient dynamics within the first 1000 time steps,

followed by a period of relatively constant abundance (see for example figure 2.5). However the high level of temporal variability illustrated in figure 4.12 now motivates us to question whether the system does reach steady-state, especially at low IR. Therefore, in section 4.4, we conduct a detailed analysis of the *stationarity* of population dynamics generated by the IBM model. We then consider whether the increased variability is due to stochastic or deterministic effects. In section 4.5, we use *recurrence quantification analysis* to look for signatures of determinism in the simulation output. Finally we close the chapter (section 4.6) by looking at how the accuracy and reliability of our numerical results are affected by increased variability, and how sensitive the results are to the sampling regime used.

## 4.4 Second-order stationarity

We introduce here three tests for second-order (or ‘weak’) stationarity in time series. Second-order stationarity may be defined as the time invariance of the first and second moments of the data. Specifically Hsu [50] states that a random process  $X(t), t \in \mathbf{Z}^+$  is second-order stationary if:

$$(4.1) \quad \mathbf{E}[X(t)] = \mu \text{ (constant)},$$

$$(4.2) \quad R_X(t,s) = \mathbf{E}[X(t)X(s)] = R_X(|s-t|),$$

where  $R_X(t,s)$  is the *autocorrelation* function of the process. Conceptually these conditions state that a second-order stationary time series has constant mean, and an autocorrelation dependent only on time separation. From now on we refer to a time series satisfying 4.1 and 4.2 as *stationary*. If the conditions are not met we call the time series *non-stationary*. Non-stationary time series cannot be fit to a constant distribution. Non-stationarity may be due, for example, to a trend in the data or a change in the parameters of the data generator.

In our case the data generator is the IBM model and there are several possible causes of non-stationarity. It may be that there is no steady-state equilibrium in the model. For example the number of individuals may undergo a random-walk. From previous analysis this situation seems unlikely, since we have observed what appear to be deterministic population cycles. However randomness has not been explicitly tested for. Another possibility is that the dynamics is characterised by large amplitude oscillations in abundance, which produce non-stationarity. Again such oscillations have not been tested for. Alternatively a steady-state equilibrium may exist, but a long period of transience means it is not reached during the time frame of our simulations (5000 time steps previously).

### 4.4.1 Tests for stationarity

We compare three different tests of stationarity: the Kwiatkowski-Phillips-Schmidt-Shin (KPSS) [65]; the Augmented Dickey-Fuller (ADF) [97]; and the the Priestley-Subba Rao (PSR) [92]

tests. These tests were chosen for their popularity in the time series literature. All three are implemented in the programming language *R* [94]. The KPSS and ADF are provided in the *tseries* package, and the PSR is in the *fractal* package. All three tests are applicable to univariate time series.

The ADF test has a null hypothesis that the time series is non-stationary. The test models the data as an auto-regressive process (see section ??), and the null hypothesis is that this process has a *unit root*. The test produces a statistic that is negative. The greater the magnitude of the test statistic the more evidence there is to reject the null hypothesis in favour of stationarity.

The KPSS test complements the ADF test in that the null hypothesis is stationarity. The data is modelled as the sum of a random-walk and an error component, and tests the hypothesis that the variance of the random walk is zero. The test statistic is always positive, and the greater its magnitude the more evidence there is to reject the null hypothesis in favour of non-stationarity.

The null hypothesis of the PSR test is also that the series is stationary. The test is based on the idea that non-stationary processes have power spectra that change over time [92]. These are called *evolutionary spectra*. The test, as implemented in *R*, returns several statistics. We quote the ‘p-value for T’ which can be thought of as the confidence that the estimated spectral density functions are constant in time.

#### 4.4.2 Characterising the tests

To understand the performance of the stationarity tests (section 4.4.1) we apply them to three example time series, which we refer to as HI, RW and NS. The first series, HI, is taken from a single IBM simulation run with IR= 0.001, zero mutualism (MAI= 0.0) and otherwise default parameters (table 2.1). The series represents the total number of individuals of all species at each time step. The simulation was run for 50,000 time steps, compared with the 5000 used in previous chapters. The increased length allows more time for the simulation to reach steady-state, and allows comparison of tests applied to different sections of the series. The first 1000 time steps were discarded, since these contain clearly transient dynamics (see figure 4.15B), leaving a time series of 49,000 points. The IR used is lower than the *default value*, but an order or magnitude higher than the lowest IR used in this chapter (see figure 4.12). The rate  $IR = 0.001$  was chosen because it gives slightly increased temporal variability over the default value and therefore increased chance of non-stationarity. In the context of what follows we refer to IR as high immigration (HI).

The series RW and NS are chosen as a negative and a positive control respectively. Both have the same length as HI. RW is a non-stationary series generated by a one-dimensional *random-walk*, defined as:

$$(4.3) \quad x(t) = \sum_t^{i=1} Z_i,$$

where  $Z_i$  are independent random variables that may take a value of either  $-10$  or  $+10$ , both with probability half. An ensemble of such random walks was generated and a single instance was chosen with mean and variance closest to those of the HI series. RW has mean and standard deviation of 15525.2 and 1549.8 respectively, compared to 15915.8 and 1545.6 for HI. For comparison, NS is a stationary series generated by drawing each value independently from a normal distribution with mean and variance equal to that of HI. The three series are plotted in figure 4.13.

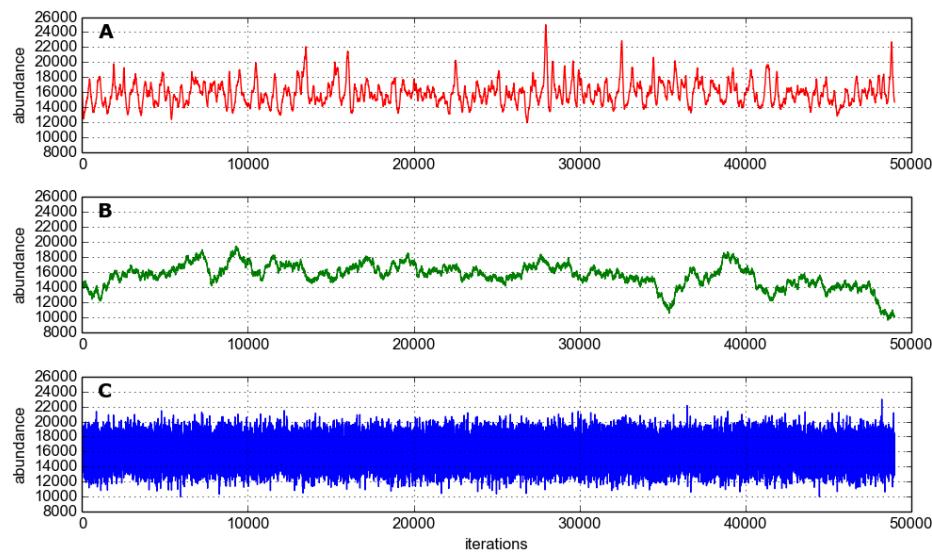


Figure 4.13: The three time series used to characterise the performance of the stationarity tests (see text for details of how they are generated). The initial 1000 points are removed such that all are 49,000 points long. (A) HI: total abundance dynamics of a simulation with high immigration rate; (B) RW: a random walk without drift; and (C) NS: a series generated by independent sampling from a normal distribution.

	A.D.F.		P.S.R.		K.P.S.S.	
	stat	p-value	stat	p-value	stat	p-value
HI	-15.401	<0.01	-	0.0004782808	0.5395	0.03277
RW	-4.0386	<0.01	-	0.9929773	18.7453	<0.01
NS	-37.5348	<0.01	-	0.811097	0.0466	>0.1

Table 4.1: Results of applying the three stationarity tests to the three time series shown in figure 4.13. P-values that indicate evidence for stationarity at 95% confidence are highlighted in green. The test statistics are also given for the ADF and KPSS tests.

Initially we apply the three stationarity tests to the entire length of the time series. The results are shown in table 4.1. ADF finds significant evidence that all three series are stationary, at 99% confidence. We may be suspicious of this result since we know that RW is generated by

a non-stationary process. However RW is a particular instance of a random walk, chosen from several thousand to closely match the mean and variance of HI. Therefore this random walk is likely to appear more stationary than others in the ensemble generated. The test statistic for ADF indicates that there is most evidence for NS to be stationary, followed by HI, then RW. The KPSS test ranks the series in the same order, based on the magnitude of the test statistic. According to this test NS is clearly stationary (accept  $H_0$ ), and RW is clearly non-stationary (reject  $H_0$  at 99% confidence), whilst HI is borderline. For HI we would accept the null-hypothesis of stationarity at 95% confidence, but reject it at 99%.

The PSR test gives unexpected results. It concludes that RW and NS are both stationary, whilst HI is non-stationary with a high degree of confidence ( $p\text{-value} < 0.001$ ). In fact, according to PSR, RW is more likely to be stationary than NS. This result contradicts what we know about the data generators that produced the time series. Therefore we do not use this test in the analysis that follows. Nevertheless the apparently erroneous result may contain interesting information about the HI series and the process that generated it (see the discussion in section ??).

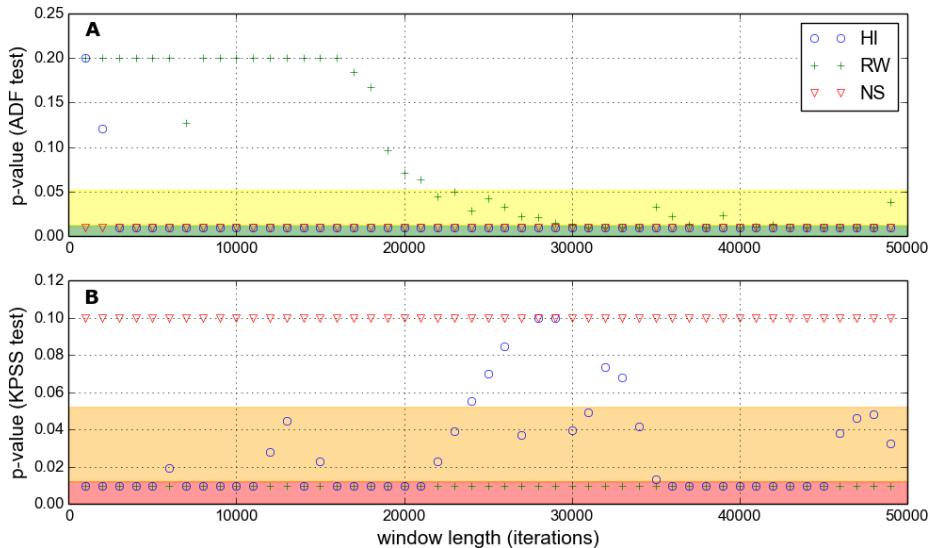


Figure 4.14: Two tests for stationarity applied to samples of varying size (window length). Samples are taken from the three time series (HI,RW,NS) shown in figure 4.13. All three time series contain 49,000 points. Sample windows begin at the first point and increase in length from 1000 to 49,000 points. Minimum p-value plotted is 0.01, actual values may be lower. (A) ADF test, with p-values capped at 0.20. The 95th and 99th percentile shaded yellow and green respectively, indicate significant evidence for stationarity. (B) KPSS test, with p-values capped at 0.1. The 95th and 99th percentile shaded orange and red respectively, indicate significant evidence for non-stationarity.

Having discarded the PSR test, we now apply ADF and KPSS to samples of varying sizes, taken from the three series (HI,RW,NS). Sampling begins at the first point of the series and takes consecutive points up to the desired sample length. Sample lengths range from 1000 to

49,000 data points. Again, as we saw in table 4.1, the two tests perform differently. The KPSS test correctly identifies RW and NS as non-stationary and stationary respectively, for all sample sizes. This is shown in figure 4.14B. The ADF test (figure 4.14A) correctly identifies NS as stationary for all sample sizes. For short sample sizes it also correctly identifies RW as non-stationary. However, for sample sizes much above 20,000, ADF finds significant evidence that RW is stationary at 95% confidence. This is an interesting result. Although RW is generated by a non-stationary process, it appears to fool the ADF test by staying ‘stationary enough’ over many time points.

There is mixed evidence for the stationarity of HI, as shown in figure 4.14. ADF, for all sample sizes above 2000, finds significant evidence that HI is stationary. Whereas KPSS, on the whole, gives significant evidence that HI is non-stationary - There are only seven cases where there is insufficient evidence to reject the null hypothesis that the HI series is stationary, and these occur at sample sizes between 24,000 and 34,000. *From these results it appears that the KPSS test is a stricter test of stationarity, and is less sensitive to the size of the sample.* Although it appears that the ADF test is biased in favour of stationarity, the test statistic does order the series correctly in the above examples (table 4.1) and is a useful complement to KPSS. Also it may be that the sensitivity of ADF to sample length is useful, since processes may appear stationary/non-stationary at different scales.

We consider the possibility that the method of sampling from the time series affects the results of the stationarity tests. For example samples taken near the beginning of an IBM simulation run may be more likely to give the non-stationary series because of transient dynamics. Alternatively a non-stationary data generator may produce sections of time series that appear stationary purely by chance. This sensitivity to sampling is investigated by *reversing* the time series and repeating the above analysis (results not shown). For HI, RW and NS we see no qualitative change in the results presented above. We also scan sampling windows of fixed length along the series to look for time dependence in the test results. The time at which samples are taken appears to make no qualitative difference, and there is no systematic change in the results that would suggest the simulation becomes more stationary the longer it is run.

**HI simulation.** <sup>3</sup> We now focus on the simulation data used to generate HI and look in more depth at whether this dataset can be considered stationary. We use the same two tests, ADF and KPSS, for the stationarity of univariate time series. Since our abundance vector is 60-dimensional ( $N = 60$  species), it is necessary to perform some manipulation to get a time series we can test. Previously we used the total number of individuals as our time series. However simply summing over species (L1-norm) is not necessarily the most informative metric to use. One possible issue is due to the phase differences between species oscillations that we might expect due to trophic interactions. Such oscillations may mean that temporal variability is cancelled out when aggregating abundances in this way (see discussion on ecosystem synchrony, section 3.3.4). It

---

<sup>3</sup>Trophic dynamics figure: refer to comparison with figures from methods chapter..

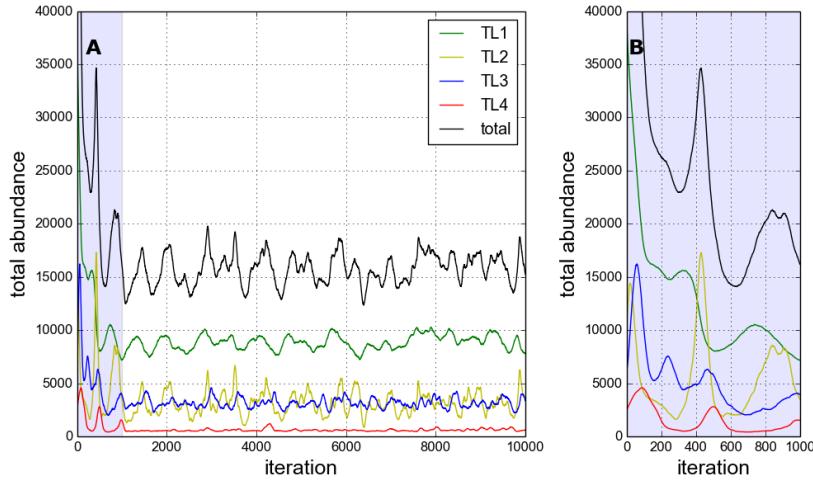


Figure 4.15: Population dynamics from HI ( $Ir = 0.001$ ) simulation: total number of individuals, and by trophic level. (A) First 10,000 time steps of simulation run (total = 50,000 time steps). (B) Enlargement of first 1000 tim steps, showing transience.

is possible that simulations which appear stationary according to some aggregate metric (e.g. total number of species) may have non-stationary underlying dynamics. This suggests that it is most informative to consider stationarity at the species level. We also consider the stationarity of abundances by trophic level, as an alternative aggregate metric.

The dynamics of the HI simulation are aggregated by trophic level to create four new time series TL1 – 4. These *trophic dynamics* are plotted in figure 4.15, and display visibly more variability than the equivalent plot produced using the default IR (figure 2.5). The initial period of transience is expanded in panel B. As in the previous analysis this part of the time series (first 1000 time steps) is removed. The ADF and KPSS tests are applied to the four trophic series separately and the results are shown in figure 4.16. According to ADF all trophic levels are stationary for samples sizes greater than 4000. TL1 appears to be least stationary according to ADF, requiring a sample size of at least 4000 to reject the null hypothesis at 95% confidence. According to KPSS TL1 is non-stationary for all sample sizes, whilst TL2 and 3 are stationary for samples sizes above 6000 and 1000 respectively. KPSS gives mixed results for TL4, with no clear dependence on sample size. It is hard to reconcile these results with an observation of the dynamics in figure 4.15, indicating the usefulness of the statistical tests.

The dynamics of all the species belonging to each trophic level are plotted in figure 4.17. It is clear here that the community is dominated by a few abundant species, mainly in the lower trophic levels, with a large number of relatively scarce species. This agrees with the rank abundance plots from chapter 3. It also appears from this figure that the more abundant species exhibit large amplitude oscillations in their dynamics. This leads us to hypothesise that the most abundant species may be non-stationary, whereas the least abundant species may be stationary.

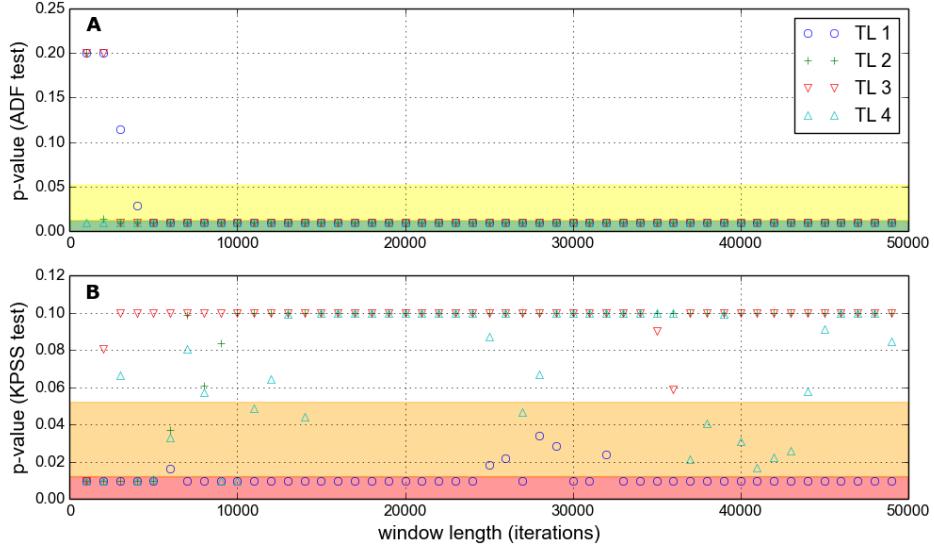


Figure 4.16: Similar to figure 4.14, but here the tests are applied separately to each trophic level of the HI simulation shown in figure 4.15.

We test this hypothesis by applying the ADF and KPSS tests to the three most abundant and three least abundant species in the HI simulation. Species are selected based on their average abundances over the whole simulation (minus the intial transience).

We see from figure 4.18 that all six species are stationary according to ADF, given a sufficiently large sample size. However the sample size for all three of the abundant species to be stationary is greater (panel A:  $\geq 9,000$  points) than for the three least abundant species (panel C:  $\geq 2,000$  points). This suggests that the most abundant species are indeed less stationary than the least abundant species. The KPSS test supports this conclusion. KPSS finds that the three least abundant species are stationary above samples sizes of  $\sim 18,000$ , whereas two of the most abundant species are non-stationary for almost all sample sizes. Inspecting the dynamics in figure 4.17 we see that these non-stationary species are those with largest amplitude fluctuations in their abundances.

In general we conclude that the choice of metric used to generate the time series does affect the conclusions about stationarity. Overall we cannot be confident that the HI simulation is stationary, based on the results presented above for species, trophic and total abundances. This lack of stationarity is largely due to the apparent strictness of the KPSS test. We also assert that considering species dynamics individually is the most informative. It allows for the possibility that some species abundances may be more variable than others, and information is not lost due to aggregation. In the following analysis we apply stationarity tests to species dynamics individually, and calculate the number of stationary species (NSSP) as an aggregate statistic. If NSSP equals the total number of species, then the community dynamics is fully stationary

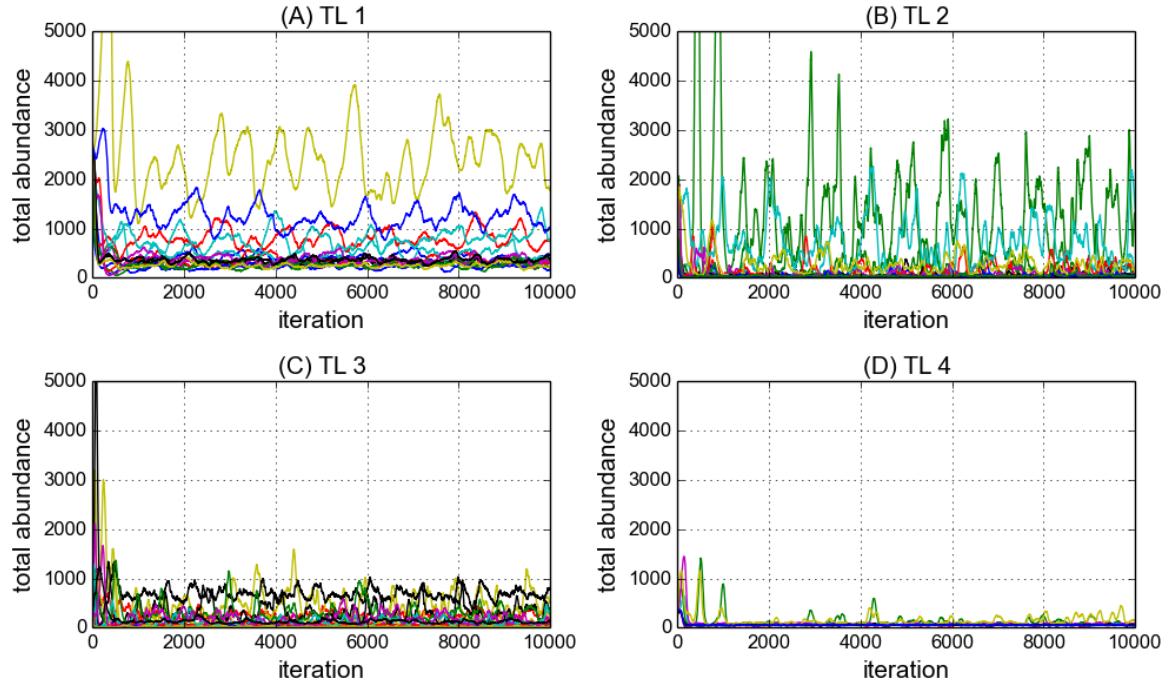
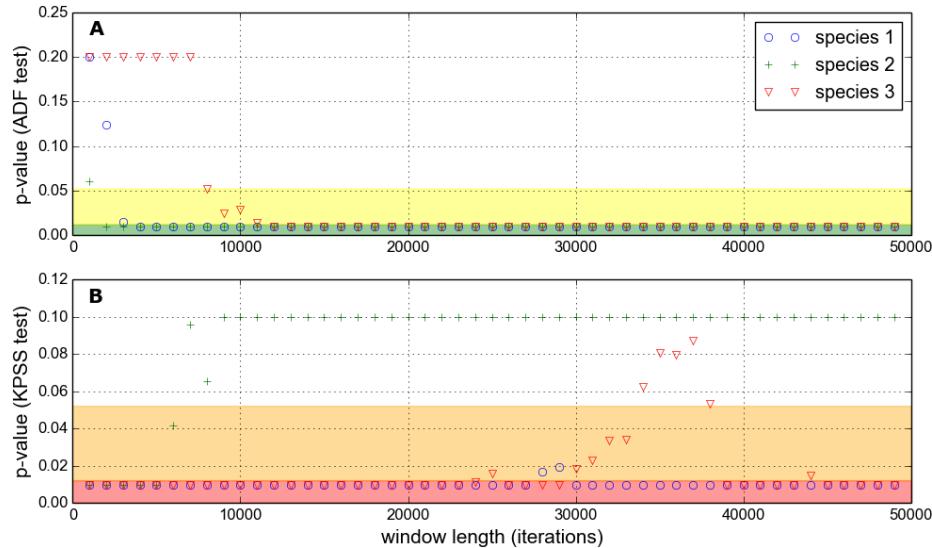
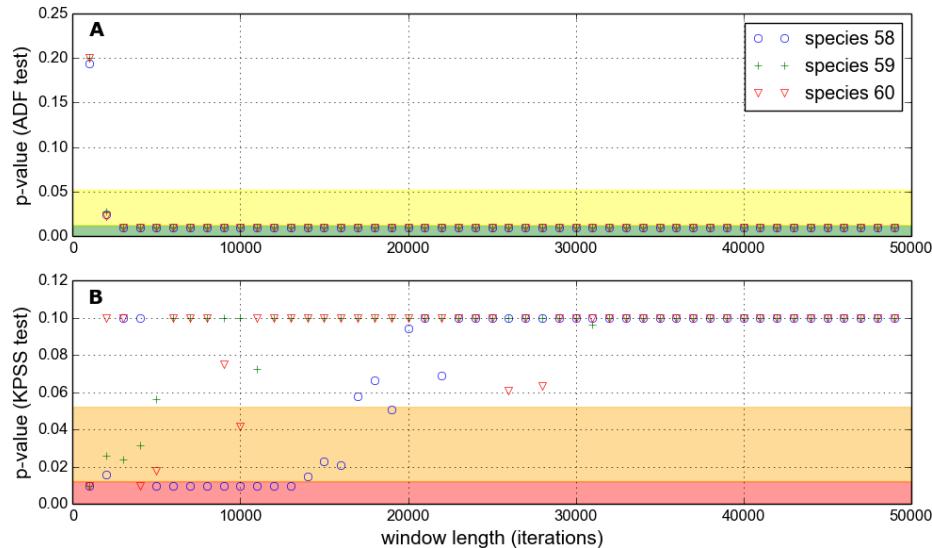


Figure 4.17: Dynamics of all species in the first 10,000 time steps of the HI simulation shown in figure 4.15. Each colour represents of different species, each panel (A-D) shows a different trophic level (TL1 – 4).

(according to the test used).



(a) Three most abundant species



(b) Three least abundant species

Figure 4.18: Similar to figure 4.14, but here the tests are applied separately to individual species from the HI simulation shown in figure 4.15. (A) The three species with highest long-term average abundance. (B) The three species with lowest long-term average abundance.

#### 4.4.3 General stationarity results

We now conduct a general investigation of the stationarity of communities simulated with the IBM model. First we test the stationarity of three ensembles of new simulation runs (figure 4.19). The new simulations use default parameters, unless otherwise specified, and are run for 50,000 time steps. Two ensembles of 100 simulations each are run for high ( $IR= 0.001$ ) and low ( $IR= 0.0001$ ) immigration rates, which we refer to here as the HI and LI ensembles respectively. All networks are generated using the niche model as described in section 2.2, with zero mutualism ( $MAI=0.0$ ), and each simulation uses a uniquely generated network. A third ensemble of 50 shorter simulations (10,000 time steps) is run, at high immigration rate, using a fixed interaction network selected at random from the HI ensemble. This third ensemble we call NM1. Stationarity testing is done using the ADF and KPSS tests characterised in section 4.4.2 above. As standard the initial 1000 time steps are discarded in an attempt to remove transience. The tests are then applied to a sample taken from the abundance time series of each species. The results presented in this section give the number of stationary species (NSSP) in the community, at the 95% confidence level.

As the length of the samples taken from the abundance time series increases, the average NSSP also increases. This is true of both tests and for all three ensembles, as we can see from figure 4.19A. According to ADF all species are stationary, on average, for sufficiently large sample length. The required length of sample is larger for the LI ensemble than for HI. For KPSS, although the NSSP does increase with sample length, it is not clear that it will asymptotically approach 60 species in the limit of many time steps. The average NSSP at 49,000 sample points is just under 40 and just over 20 for HI and LI ensembles respectively.

To check the time dependence of stationarity (i.e. are species more likely to be stationary towards the end of a simulation) samples of length 3000 were taken from different points along the time series. From figure 4.19B we can see that there is no clear trend in stationarity over 49,000 time steps. The average NSSP is almost the same whether the sample is taken from time steps 1000-3000 or 46,000-49,000. This lack of time dependence in NSSP also holds for windows of different length (results not shown).

On average we see that the LI ensemble is less stationary than the HI ensemble. This we expected based on the previous observation that reducing IR increases temporal variability. However we cannot be confident that either ensemble contains communities in which all species are stationary. This observation may be problematic for the interpretation of our results, and we discuss this further below. Interestingly the NM1 ensemble gives very similar results to the HI ensemble. It may be that stationarity of the simulation output is not dependent on the interaction network structure. Alternatively the observation may be anomalous, owing to the fact that we have accidentally chosen NM1 to closely resemble the average of this ensemble (see both panels of figure 4.19).

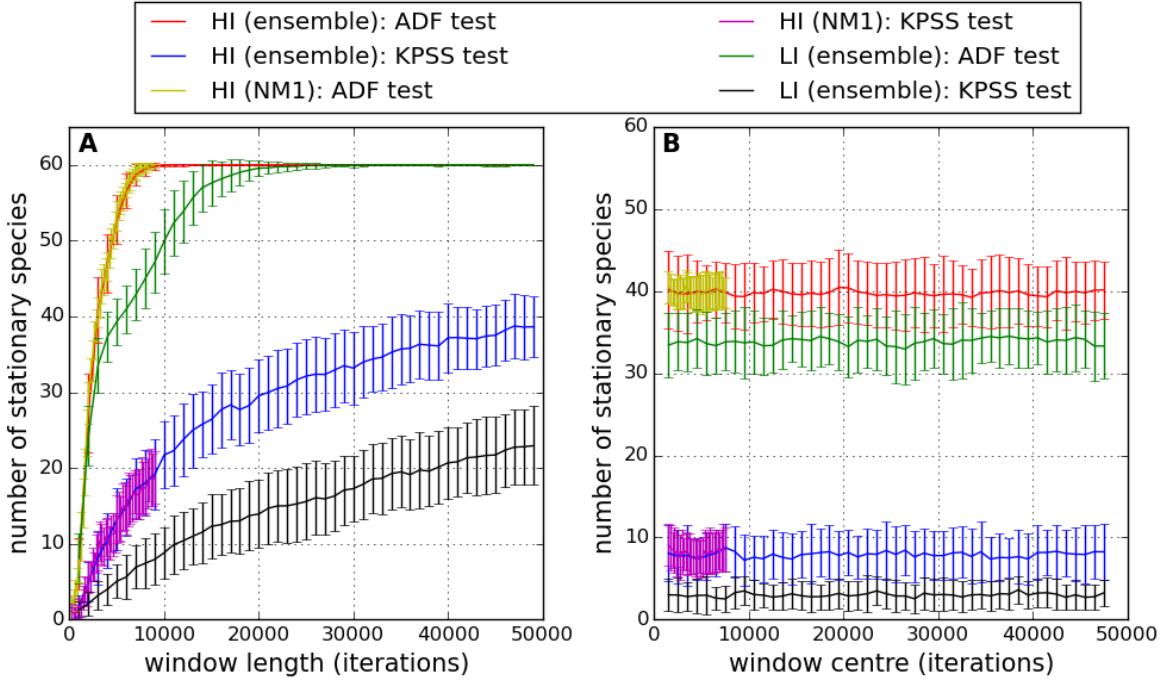


Figure 4.19: The number of stationary species (NSSP) according to the two stationarity tests (ADF and KPSS) at the 95% confidence level. Results averaged over three different ensembles of simulations: HI; NM1 (high immigration) and LI as described in the text. The first two are high immigration ensembles, whilst the latter is low immigration. Solid lines indicate mean results for the ensemble. Error bars indicate  $\pm 1$  standard deviation. (A) Each species abundance time series is sampled with a window of increasing length, as in figure 4.18. (B) Each species time series is sampled with a window of length 3000 time steps, which is scanned along the series to check for changes in stationarity over time.

**In chapter ?? we will** explore a slice of the IBM parameter space by varying both immigration rate and habitat loss. Here we study the stationarity of the simulations that we will use in that chapter. All simulations are 5000 time steps long. The initial 1000 time steps are discarded and the ADF and KPSS tests applied, species by species, to the remaining 4000. Figure 4.20 shows the average NSSP across the region of parameter space investigated, for three MAI ratios (MAI= 0.0, 0.5, 1.0). The results are qualitatively the same for both tests, although NSSP is higher for ADF than for KPSS as expected. On average reducing IR reduces the NSSP. A weaker, but still visible, effect is that increasing HL reduces the NSSP. Most striking is the effect of MAI ratio on stationarity. The average NSSP is greater across the whole parameter region at MAI= 0.0 than at MAI= 1.0, with MAI=0.5 in between the two. Increasing mutualism also appears to reduce the dependence of NSSP on IR. Figure 4.21 summarises these trends using cross sections taken from the ADF heat maps (panels A,C,E) in figure 4.20, with error bars added. It is clear that there is high variability across replicate simulations, and this variability appears to be greatest for high

mutualism (MAI=1.0).

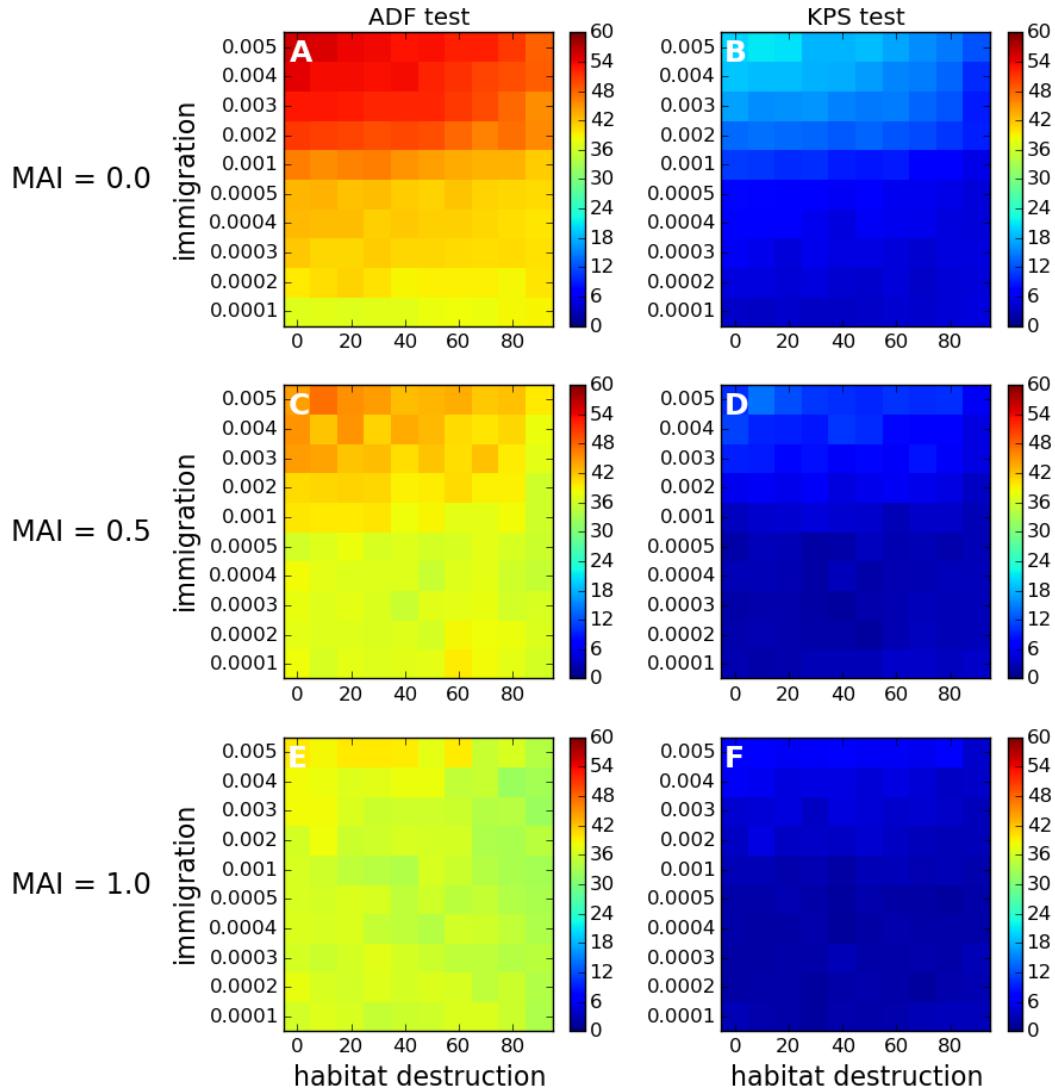


Figure 4.20: The average number of stationary species (NSSP) according to the two stationarity tests (ADF and KPSS) at the 95% confidence level. Each cell on the grid represents the mean value over 50 repeat simulations. All simulations are 5000 time steps with default parameter values. Tests are applied to the final 4000 time steps of each species abundance time series. These correspond to the simulations presented in chapter ??.

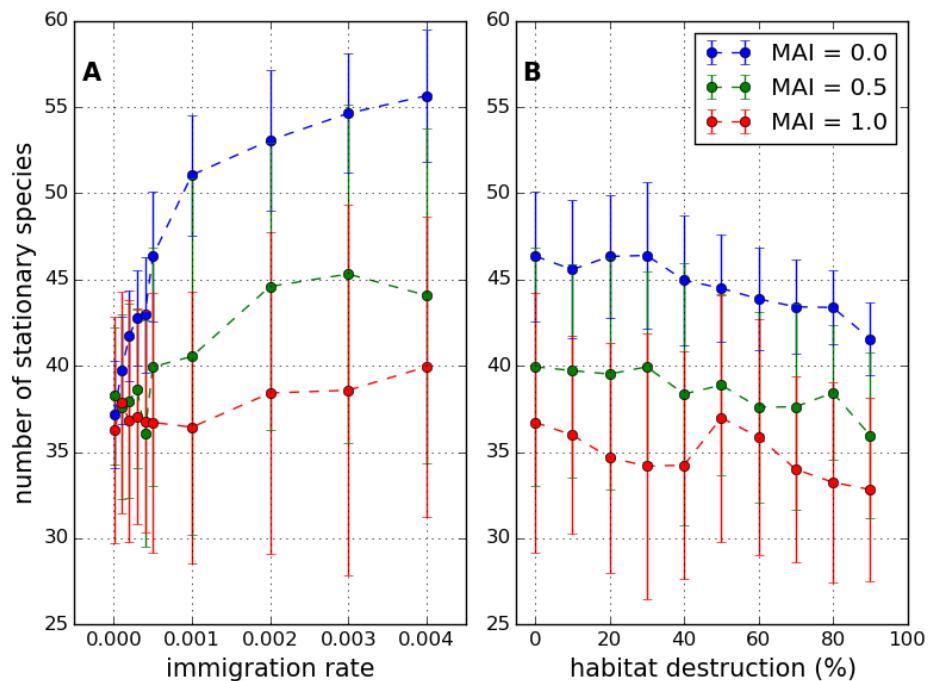


Figure 4.21: The number of stationary species (NSSP) according to the ADF test at the 95% confidence level. Points show mean value over 50 replicates. Error bars show  $\pm 1$  standard deviation. Tests are performed on the same simulations depicted in figure 4.20. (A) Plotted against immigration rate (IR), with zero habitat destruction. (B) Plotted against habitat destruction, with IR= 0.001 (high immigration regime).

## 4.5 Determinism tests

We have seen that population dynamics becomes highly variable, especially at low IR, and indeed may not be stationary. This leads us to speculate that in some instances the simulation output may become dominated by stochastic effects. This randomness is an interesting feature of the model, and indeed noise is ubiquitous in natural systems. However much of our analysis (chapters 3 and 5) involves interpreting the simulation results in terms of the ecological mechanisms built into the model, in particular the patterns and structures generated by species interactions. If the model output lacks determinism it may not be meaningful to conduct such analyses. Therefore we present here a test for determinism based on *recurrence quantification analysis* (section 4.5.1), which we then apply to the IBM output (section 4.5.1.1).

### 4.5.1 Recurrence quantification analysis

Recurrence quantification analysis (RQA) maybe be used to detect signatures of determinism [6, 47, 70]. The analysis is based on the idea of *recurrence* - deterministic dynamical systems tend to return to similar regions of phase space. Moreover, when they do, their trajectories tend to remain close in phase space for some amount of time. In the case of chaotic systems the trajectories will diverge, at a rate which is broadly determined by the maximal Lyapunov exponent [47]. However, even for chaotic systems, there is some tendency for neighbouring trajectories to remain close. RQA aims to detect the presence of this feature in the dynamics. The analysis is often used with univariate time series (such as in [47]), in which case time-delay embedding must be used to increase the dimensionality of the phase space. However, since we have a time series for each of the 60 species, our phase space is already high dimensional. Therefore our first step is to construct a recurrence matrix (RM). The RM is a binary matrix whose elements, for a system  $x(t) \in \mathbb{R}^N$ , are given by the function:

$$(4.4) \quad d_{ij} = \begin{cases} 1 & \text{if } \|x(i) - x(j)\| < r \\ 0 & \text{if } \|x(i) - x(j)\| \geq r \end{cases},$$

where  $r$  is a threshold distance that defines the measure of ‘closeness’ in phase space. Various methods have been proposed to choose the value of  $r$ . These methods are discussed in [6], but we use the method which they adopt for the results in that paper:  $r$  is chosen such that  $\sim 10\%$  of the elements of the RM are equal to one.

Having constructed the RM it can be visualised as a *recurrence plot* (Figure 4.22). Such plots can be visually striking (search recurrence plots for examples), but are difficult to interpret without experience. An important feature, in the search for determinism, are *diagonal lines* - lines parallel to the leading diagonal. Such lines indicate that trajectories which find themselves ‘close’ in phase space remain close, for a period of time given by the length of the line. Visual

inspection can detect the presence of diagonal lines but it is better to use a quantitative pattern detection method. A common metric that quantifies the relative abundance of diagonal lines is the *percentage of determinism* (%DET) [6, 70], and is given by:

$$(4.5) \quad \%DET = \frac{NPD}{NREC} \times 100,$$

where NREC is the number of entries in the RM equal to one; and NPD is the total number of points found on diagonal lines of length greater than or equal to two. The %DET allows quantitative comparison of the level of determinism between different RMs. In [6] they develop three statistical tests, based on %DET and two similar metrics, for the null hypothesis of pure randomness in the time series. However in the current analysis we make do with a comparison of the value of %DET between different test cases.

#### 4.5.1.1 Results

We test for determinism in the simulation ensembles HI and LI used in section 4.4 (see for example Figure 4.19). Each ensemble consists of 100 repeat simulations with different network structures. These simulations are used because one ensemble is from the high immigration regime (HI: IR= 0.001) and one is from the low immigration regime (LI: IR= 0.0001). Therefore the temporal variability is higher on average in the LI ensemble than the HI ensemble. The simulations in these ensembles are also longer than other simulations - being run for 50,000 time steps, they allow for more chance of recurrent dynamics and therefore provide more information from which to detect a signature of determinism. In both ensembles the communities are antagonistic (MAI= 0) and the landscape is pristine (HL= 0). Therefore we do not cover the full range of parameters explored in other analyses, but we at least consider one highly variable and one less variable scenario.

We construct randomised control data in two different ways. The control data is used to compare %DET results from the IBM with %DET values generated by randomness. In [6] they state a random process would generate a RM with NREC points distributed uniformly at random. We construct such matrices by randomly permuting the elements of another RM, to obtain a randomised RM with the same number (NREC) of points. The second method is to construct RMs by randomising the order of the time series of each species. This preserves the mean and variance of the population dynamics, but removes any determinism. The randomised time series are then used to construct an RM.

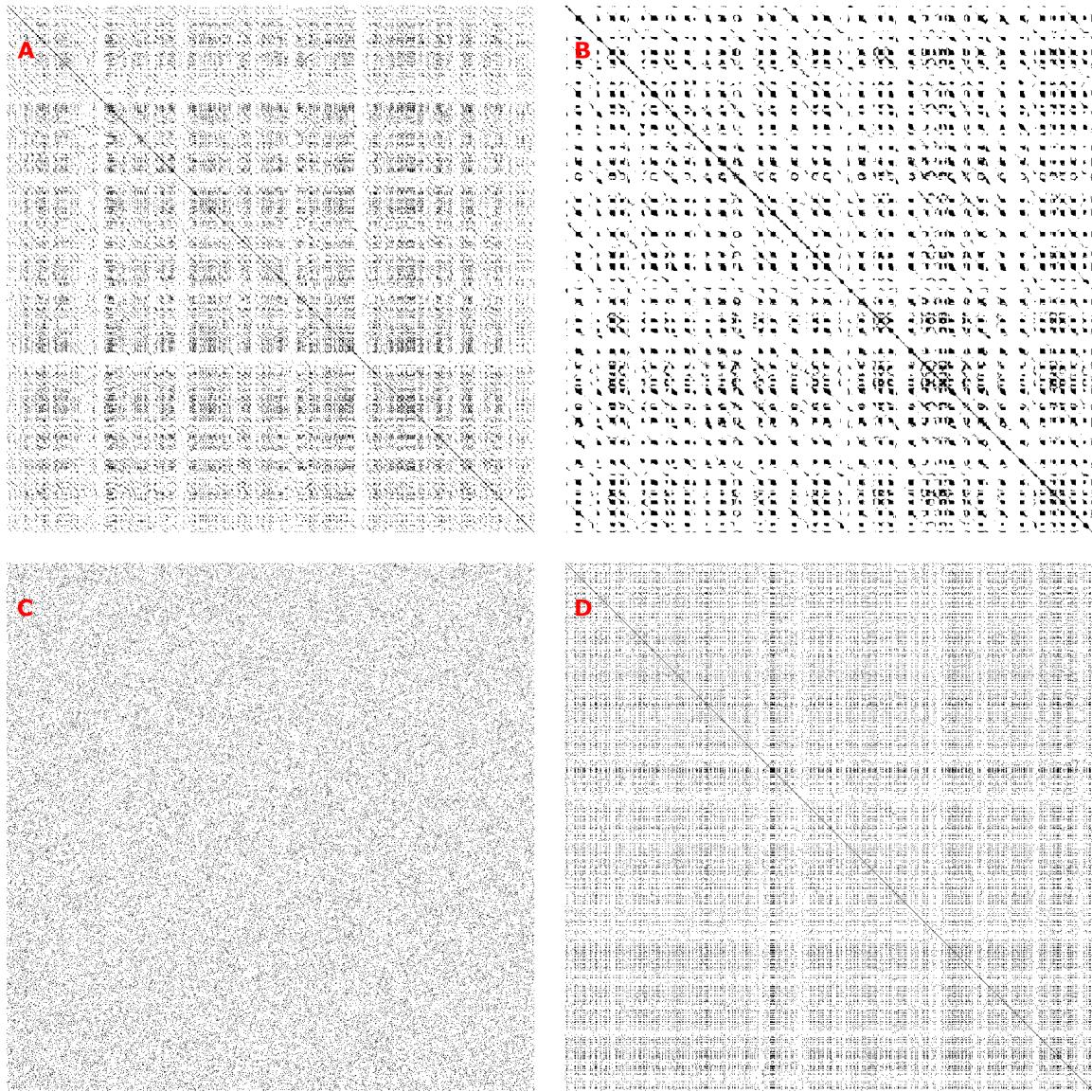
Figure 4.22 shows four example recurrence plots. Panels A and B show the RMs of single simulations from ensembles HI and LI respectively. Panels C and D show randomised versions of panel B generated by randomising the matrix, and by randomising the time series respectively. Panels A and B clearly have some structure, whilst the structure is lost in panel C as expected. Panel D retains some structure. In particular it contains the leading diagonal, since even the

randomised time series is identically equal to itself. It also contains horizontal and vertical bands, created by points in the time series that are unusually distant (or close) to a large number of other points in the phase space. A subtle but detectable feature of panels A and B is the presence of diagonal lines (parallel to the leading diagonal) suggesting determinism. This feature is lost in panel D, as we would expect due to the randomisation of the time series. Interestingly panel A looks very similar to figure 1(c) in [6]: an RM generated from the dynamics of a noisy chaotic Hénon attractor. This leads us to speculate that the deterministic component of the dynamics may be chaotic. We do not search for signatures of chaos, partly because the presence of noise severely complicates the available methods for doing so (such as estimation of the Lyapunov exponent)[47].

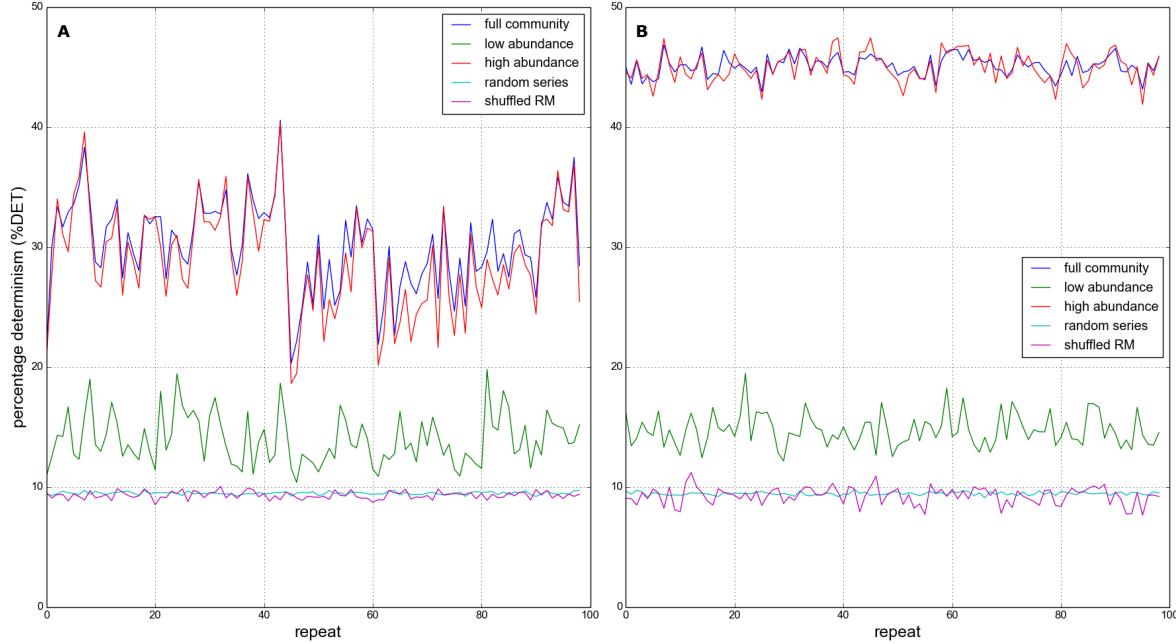
We calculate the %DET for each simulation in the HI and LI ensemble. In each case, for comparison, we also calculate the %DET for the dynamics of the ten most and ten least abundant species, and for the two randomised versions of the data. These results are shown in figure 4.23. The randomised data shows little variation, with a  $\%DET \approx 10$  in all cases. The %DET for the least abundant species is slightly greater than for random data, suggesting some evidence of determinism but also a strong stochastic component. In both the HI and LI ensembles there is good evidence for determinism at the whole community level, with %DET consistently  $> 30$  and  $> 40$  respectively. It is interesting that there is more evidence of determinism in the LI ensemble than in the HI ensemble, as we were previously concerned that the high temporal variability of the LI dynamics could mean that it lacked determinism. It is also clear for these results that the deterministic component of the dynamics is dominated by the most abundant species. This, together with the results from previous sections tells us a lot about the dynamics of the model (see discussion in section 4.7). We conclude here that the analysis so far suggests that non-stationarity may be largely due to deterministic population dynamics/oscillations of the most abundant species, while the dynamics of the less abundant species are more random and more stationary although they do have a deterministic component.

## 4.6 Convergence and repeatability

In chapter 3, and at the beginning of this chapter, we used ‘snapshots’ of the simulation state to calculate species abundances. The snapshot method was justified by the assumption that simulations reached stationarity and therefore we were sampling from a steady-state distribution. However, as we saw in section 4.3, stationarity cannot be guaranteed. The results of section 4.5 suggest that this lack of stationarity is due to deterministic population dynamics, especially those of the most abundant species. With oscillatory dynamics it is clear that snapshot sampling will yield different results depending on when the measurement is taken. We make the assumption that the best way to characterise the system is to take the long-term average of the metric in question. This approach is also justified by the observation that stationarity does not increase



**Figure 4.22: Recurrence plots** defined by equation 4.4. (A) High immigration simulation ( $IR=0.001$ ). (B) Low immigration simulation ( $IR=0.0001$ ). (C) Randomised recurrence matrix (permutation of the elements of the matrix in B). (D) Randomised time series (permutation of species population time series from B, such that mean and variance for each species are preserved). In simulations for both A and B:  $HL=0$  and  $MAI=0.0$ ; number of time steps = 50,000; sampled every 50 time steps to construct plots.



**Figure 4.23: Percentage determinism (%DET)** defined by equation 4.5. Value calculated for whole community, ten least and ten most abundant species, and for two randomised versions of the data (see text). For 100 repeat simulations at: (A) High immigration rate ( $IR=0.001$ ); and (B) Low immigration rate ( $IR=0.001$ ). In simulations for both A and B:  $HL=0$  and  $MAI=0.0$

of decrease over the course of a simulation, but remains relatively constant<sup>4</sup> (figure 4.19B). In section 4.6.1 we look at how temporal variability affects the *convergence* of our results on the long term average. In section 4.6.2 we consider the *repeatability* of our results by running replicate simulations with the same network structure and parameter values.

### 4.6.1 Convergence

In this section we compare the performance of different estimators for long-term mean species abundance. In all that follows the initial 1000 time steps are removed from the simulated time series, as has been done throughout the chapter. Therefore when we refer to a time series below, we refer to this truncated version of the time series. We define the long-term average abundance of species  $i$  as:

$$(4.6) \quad \bar{X}_i = \frac{1}{T} \sum_{t=1}^T x_{i,t},$$

where  $T$  is the number of time steps in the time series, and  $x_{i,t}$  is the abundance of species  $i$  at time point  $t$ . We denote by  $e_i$  an estimator of the long-term average abundance  $\bar{X}_i$ . The snapshot estimator  $e_{1i}$  for species  $i$  is defined as the abundance of species  $i$  on the 5000<sup>th</sup> time step. Other

<sup>4</sup>Really we should also check for trends.

estimators are calculated by taking the mean species abundance over different sample lengths. Sample are always taken from the beginning of the time series.

Figure 4.24 shows three different estimators plotted against long-term average abundances. Two simulations are shown, one with high IR (0.001) and one with low IR (0.0001). Also plotted are linear regression fits to the data, which for a perfect estimator would lie on the line  $y = x$  and therefore have a regression coefficient ( $R^2$ ) equal to one. In this way  $R^2$  can be used to assess the quality of any given estimator as a measure of the long-term abundances of all species in a community. From the plots it is clear that the longer samples produce better estimates than the snapshot. In particular at low IR, where there is increased temporal variability, the performance of the snapshot estimator is poor. We define another metric for estimator quality, which we call *mean relative error* MRE:

$$(4.7) \quad MRE = \frac{1}{N} \sum_{i=1}^N |\bar{X}_i - e_i|,$$

where  $N$  is the number of species, and  $e_i$  is the value of the estimator in question for species  $i$ . This metric is equal to zero for a perfect estimator. Figure 4.25 shows the performance of estimators, calculated from increasing sample lengths, applied to the HI and the LI ensembles. Performance is measured by both the  $R^2$  value and by MRE. In all cases the estimators converge on the long term average as expected. However for the LI ensemble the MRE converges very slowly. Even with sample lengths of 30,000 time steps the MRE is still around 0.1. Furthermore the performance of the snapshot estimators is very poor, with MREs of around 0.3 and 0.8 for the HI and LI ensembles respectively. From these results it appears that the snapshot estimator would introduce large errors into our results in cases where there is high temporal variability. However it is also clear that some trade-off must be made, because the length of sample required to guarantee convergence is very large.

FINISH: The estimator convergence depicted in figure 4.25 suggests that very long samples are required in order to reduce the expected deviation in abundance estimates from the long term average (measured by MRE (4.7)). However, based on the regression analysis, much shorter samples...suggesting that samples of length 4000 should be sufficient to characterise the system with a high level of accuracy.

### 4.6.2 Repeatability

Here we briefly consider the issue of repeatability in our experiments. I have heard IBMs (or *agent-based model*) referred to as ‘*very complicated random number generators*’. Although I can find no reference to such a critique in the literature, it is certainly possible that sufficient stochasticity could result in replicates of the same model producing wildly different outputs. Various results presented in this chapter suggest that this is not the case here. In section 4.2.5 we saw that replicate simulations with two different network structures generated consistently different persistence profiles. Also the RQA analysis (section 4.5) revealed a strong deterministic

signature in species population dynamics. The detection of determinism suggests that the local rules of the IBM successfully define mechanisms responsible for the observed patterns and dynamics, rather than simply producing randomness. As a final demonstration of repeatability we selected three replicate simulations at random from the NM1 ensemble (section 4.4.3). These three simulations use the same interaction network structure. We compare the abundance distributions of the three replicates using *rank-abundance spectra* (RAS) [69]. RAS are a useful tool for the comparison of abundances in communities that contain the same species. Species are ranked according to their abundance in one community (as in RADs - section 2.7.1.1). The rank of each species is then fixed, giving its spectral location, which allows the comparison of the initial community with several others. RAS are plotted for the three NM1 replicates in figure 4.26, using the long-term average abundance for each species (4.7). We provide no more than a qualitative description of the RAS, which clearly illustrate a high level of repeatability in the abundance distribution. There are only a few significant deviations from the rank ordering in the initial community (panel A) in the replicate communities (panels B,C). In particular the species at rank 8 has a much lower abundance in panel C than in A and B. It appears that this is due to competition from species in the same trophic level (ranks 38 and 54).

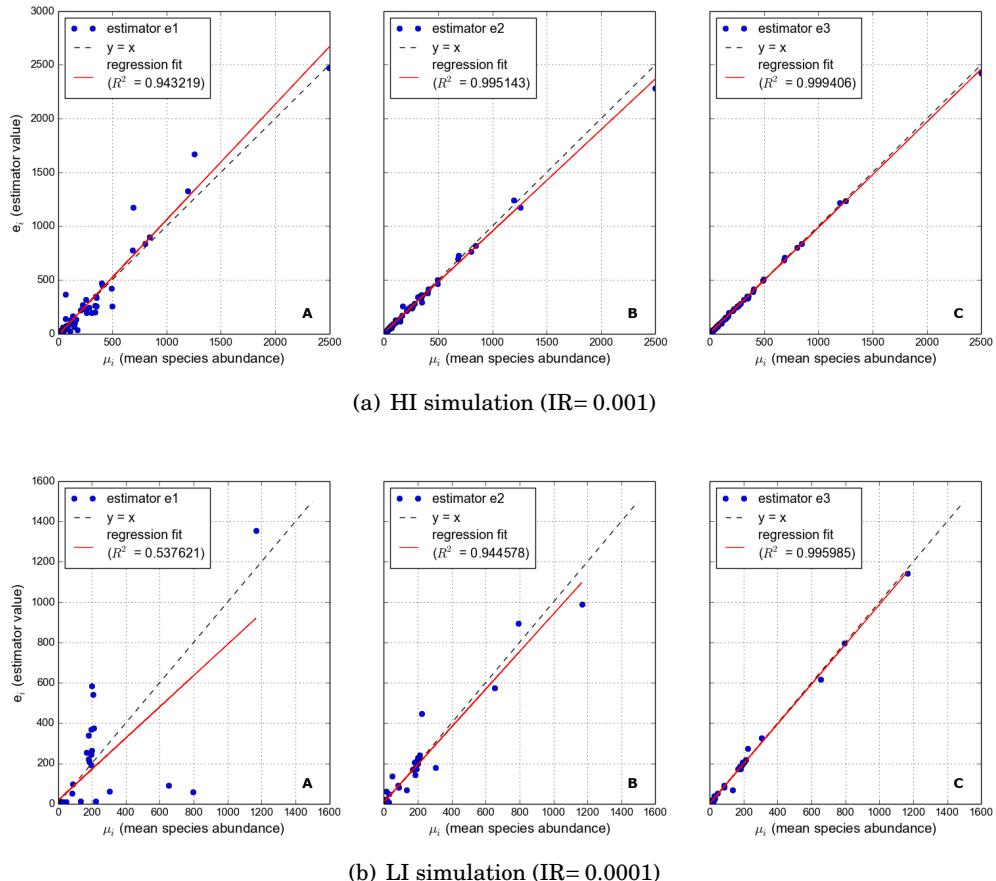


Figure 4.24: Performance of three different estimators of species abundance, applied to a (A) single HI simulation and (B) a single LI simulation. Estimator ‘e1’ is a snapshot of abundance on 5000th time step; ‘e2’ is an average over 4000 time steps; and ‘e3’ is an average over 29,000 time steps. Points are the long-term mean abundance of a species, plotted against the estimator value for that species. Red lines show linear regression fits for the estimator, and how close the estimator is to modelling the ‘true value’ of species abundances (dashed blue line).

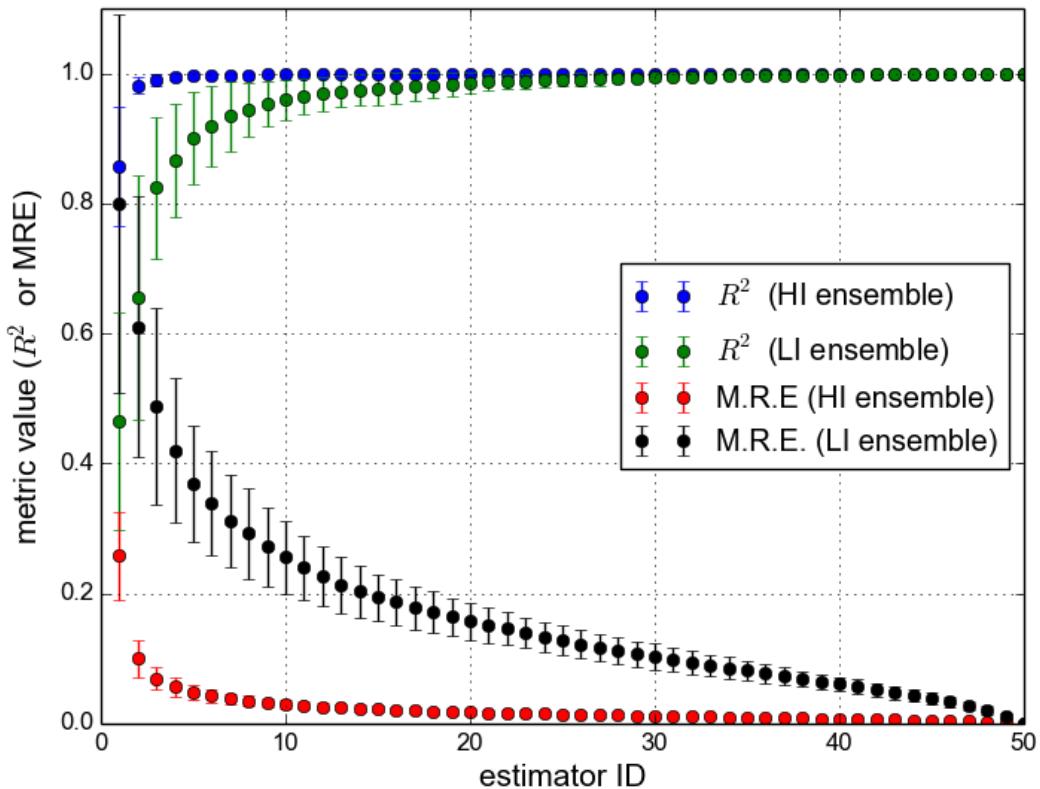
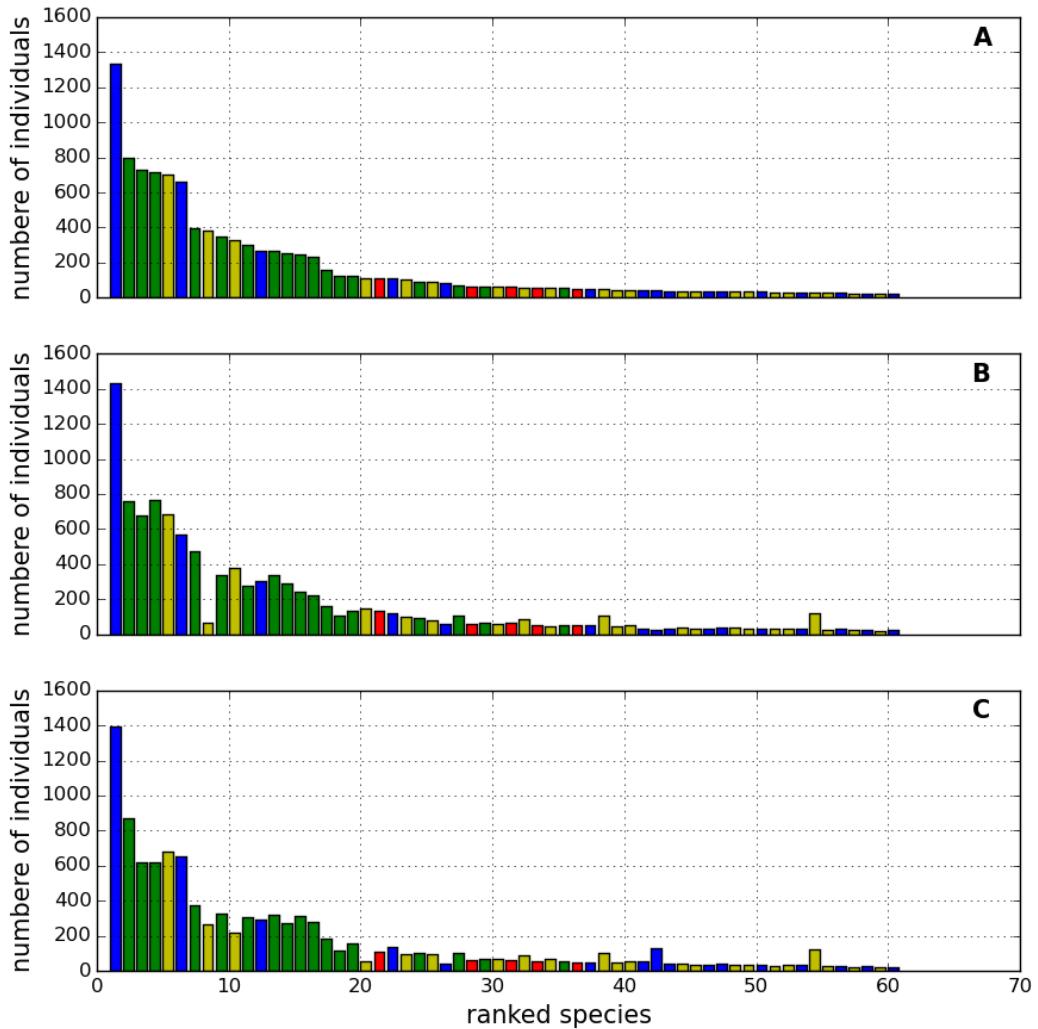


Figure 4.25: Performance of 50 different estimators for species abundance, measured by  $R^2$  value and mean relative error MRE (see text for definitions). The first estimator (estimator ID=1) is uses ‘snapshot’ of simulation state at 5000th time step (as used in previous chapter), the remaining estimators (IDs= 2,..,50) use average abundances over sample windows ranging from length 1000 to 50,000 in steps of 1000. Points indicate mean value of metric over ensemble of simulations. Error bars indicate  $\pm 1$  standard deviation.



**Figure 4.26: Rank abundance spectra (RAS)** for three simulations from the NM1 ensemble (see text). Species abundances measured as long term average (over time steps 1000-9,000). Species are ranked according to their abundances in the first simulation (panel (A)). Ranking is retained in panels (B) and (C), which show abundances from two different simulations. Colouring of species by trophic level is consistent with previous figures.

## 4.7 Conclusion

In this chapter we have discovered that immigration is required for the IBM to generate persistent communities. In the absence of immigration the majority of species in higher trophic levels go extinct, and we have argued that this is due to competition effects. We have also demonstrated the population dynamics becomes more variable as the immigration rate is reduced. This may be a feature of the system passing between a regime of stable persistence at high IR into an unstable regime at low IR. Furthermore we have demonstrated that species population dynamics may be non-stationary, especially at low IR. However results from the RQA suggest that this non-stationarity is due to deterministic population dynamics rather than stochastic effects. We have also shown that the dynamics of species with low abundance tend to be more stationary and less deterministic, whereas species with high abundance tend to be less stationary and more deterministic. Together these results suggest that the abundant species are undergoing larger amplitude trophic dynamics, generated by their interactions with other species. The least abundant species, we propose, are those maintained mostly by immigration (even at lower IR values) because their relative scarcity makes them less likely to interact with other species. As we argued in chapter 3, immigration is a random process. So the dependence of low abundance species on immigration would explain their lower %DET values and, together with their reduced ability to interact with other species explains their higher stationarity.

We have addressed the impact of increased variability on our experimental results. We showed that, especially at low IR, snapshot sampling can produce abundances measures with large deviation from the long term average. However it is also clear that, for highly variable communities, the length of sample required for the measured abundance to converge on the long term average is infeasible. Therefore we accept that increase variability may introduce random error into our results. In chapter ?? we address this issue by comparing two different sampling regimes in the analysis. We have also briefly considered the issue of repeatability in our experimental results. In particular we saw that replicate simulations using the same interaction network produced very similar rank-abundance spectra (RAS). The reproducible RAS, together with the observed signature of determinism in the IBM output, given us confidence in the repeatability of our results. Although clearly the strong stochastic component requires that replicates are part of the experimental procedure.



CHAPTER



## VARYING IMMIGRATION RATE

[96]

### 5.1 Motivation

In chapter 3 we discovered that the immigration mechanism provides a *rescue effect* for all species, preventing extinctions even at high levels of habitat loss (HL= 90%). This allowed us to study community responses to HL in the absence of extinctions. However in Nature it is common that HL does lead to the local loss of species [34]. Therefore we are motivated to reduce the immigration rate (IR), weakening the rescue effect, to study communities in which HL results in extinctions. In chapter 4 we found that communities at zero IR - *closed communities* - displayed many extinctions, even in pristine landscape. We concluded that this was due to competition between species. Therefore in this chapter we study the response of communities to HL at IRs between zero and the *default value* ( $IR = 0.005$ ). We may expect such communities to respond differently to HL than those studied in chapter 3.

In the previous analysis we have demonstrated that immigration is a key determinant of the dynamics and structure of simulated communities. In chapter 4 we saw that high IR reduces both the temporal variability and the determinism of the population dynamics. In chapter 3 we demonstrated that immigration acts to increase the evenness in the distribution of species abundances and that this can affect network properties, for example by making interaction frequencies for even. In particular we saw that the *dependence of a community on immigration* (measured by the relative contribution of immigration to total births) could account for the different structural responses of communities to the two HL scenarios. The contiguous HL scenario did not produce significant changes in many of the metrics analysed, and it was argued that this was due to the constant dependence of communities on immigration across the HL

gradient. In contrast random HL increased community dependence on immigration, and produced significant changes in most of the metrics analysed. In this chapter we employ the same two HL algorithms (section 2.4), *random* and *contiguous*, at different IRs. We are particularly interested in finding an IR at which HL *reduces community dependence on immigration*. Reduced dependence on immigration is not an effect we have seen previously. In such a case we hypothesise that communities will become less even, and that there will be corresponding changes in network properties. It is most likely that such an effect will be found in the contiguous scenario, due to the role of species interactions.

Species interactions have proven to be another key driver of population dynamics and community structure, as expected. In chapter 3 strong inter-specific interactions, as measured by IS, were shown to produce high temporal variability. In chapter 4 it was demonstrated that this variability was due to deterministic oscillations, characteristic of predator-prey interactions. We have also found that the two HL scenarios have different effects on species interactions due to changes in the mobility of individuals (figure 3.23). Random HL presents a barrier to motion making both inter- and intra-specific interactions less likely. This accounts for the increased dependence on immigration in the random scenario, since sexual and mutualistic reproduction is hindered by the physical barriers of destroyed cells that make it difficult to find mates. Contiguous HL makes interactions more likely because all individuals are contained within a smaller region of space and maintain the same dispersal ability. Therefore in the contiguous scenario at low IR we may expect to see a reduced dependence on immigration, and increase in the effects associated with strong species interactions. Conversely we may expect low IR to weaken the effects of random HL, associated with increased dependence on immigration, that were observed at the default IR.

In general we anticipate that reducing IR will strengthen the effects associated with species interactions, and weaken those associated with immigration. In particular we expect communities to become less even and more temporally variable as IR is reduced. There may also be an increase in *ecosystem synchrony* as the stochastic component of the dynamics is reduced compared to the deterministic component. Such effects should be visible at any given HL value under both scenarios. Finally we expect that at lower IRs, HL will induce species extinctions under both scenarios. In section 5.2 we discuss alternative ways to define extinction in the current model.

AFTER PREDICTIONS! - OUTLINE OF CHAPTER...

## 5.2 Experimental approach

As in chapter 3 we run two ensembles of simulation, one for random and one for contiguous HL. Here we vary HL between 0% and 90% in steps of 10%, as before. At each value of HL we run replicates at 10 different values:  $1 \times 10^{-4}, 2 \times 10^{-4}, 3 \times 10^{-4}, 4 \times 10^{-4}, 5 \times 10^{-4}, 1 \times 10^{-3}, 2 \times 10^{-3}, 3 \times 10^{-3}, 4 \times 10^{-3}, 5 \times 10^{-3}$ . Therefore each ensemble explores a two-dimensional slice of the parameter space. This same region of parameter space was visualised in figure 4.20 during the

stationarity testing. The inclusion of variable IR increases that number of required simulations by a factor of ten, over the ensembles run in chapter 3. To reduce to computational cost we make two savings. Simulations are only run for three MAI ratios (0.0, 0.5, 1.0), giving the full range between antagonism and mutualism but with lower resolution. Additionally we restrict the number of metrics that are calculated during simulations. The spatial metrics in particular are computationally expensive. Therefore all of the metrics defined in section 2.7.3, that characterise spatial aggregation and variability, are discarded. This speeds up the simulation run times by about a factor of five.

From section 4.6 we understand that the increased temporal variability resulting from reduced IR can make our results more sensitive to sampling procedure. It may be desirable to run the simulations for an increased number of time steps, allowing for longer samples to be used. However it was decided that this is an unnecessary computational expense. The regression analysis of estimator quality suggested that a sample length of 4000 characterises species abundances with good accuracy ( $E(R^2) > 0.9$ ) in pristine landscape at low IR ( $1 \times 10^{-4}$ ). Therefore all simulations in this chapter were run for 5000 time steps, with the first 1000 time steps discarded before analysis (as in chapters 3, 4). In the initial analysis in section 5.3 we further address the issue of sampling by statistically comparing results obtained using two different sampling methods. We also increase the number of replicate simulations from 25 to 50 to account for the increased variability. Therefore each ensemble contains 15,000 ( $= 50 \times 10 \times 3$ ) simulations. Again these were run on Blue Crystal [40].

We have seen that immigration provides a rescue effect that is common to all species. A species may be extinct at some time in the simulation, but later recover due to immigration. Especially at low IR we expect that some species abundances will hover close to zero. We make the assertion that such species are effectively locally extinct because they fail to maintain a viable local population and are only sustained by immigration. Indeed in empirical studies sufficiently rare species are unlikely to be detected and therefore do not contribute to estimates of species richness. Therefore our previous definition of extinction, as the absence of any individuals belonging to a given species, proves problematic. Snapshot measurements of species on the border of extinction will be particularly sensitive to the time at which the measurement is taken. Measurements of abundance made by averaging over a number of time steps will result in small but non-zero population sizes for borderline species. Requiring that a species population size is at zero will therefore mean we again detect no extinctions. In the initial analysis (section 5.3) we set an *arbitrary threshold* of three individuals, defining any species with population sizes smaller than this as extinct. In subsequent analysis<sup>1</sup> we revisit this definition of extinction with more rigour.

NOTE: on vegetarian predators! < 5% of links on average, so should not make too much difference<sup>2</sup>. ALSO: will be used RADs, or just Shannon? Why?

---

<sup>1</sup>Where?! And what exactly do we do?!

<sup>2</sup>Need to recalculate proportions.

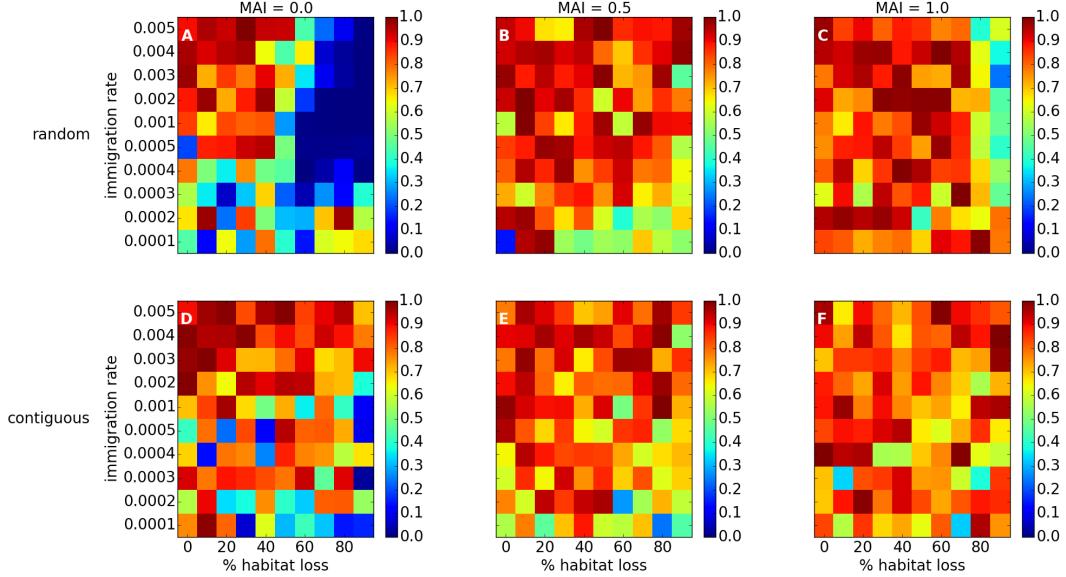


Figure 5.1: P-values for t-tests to compare the *Shannon equitability* calculated by two different sampling methods: *snapshot* and *averaged* sampling (see text for definitions). Each point in the plot represents the p-value of the test comparing the *snapshot* and *averaged* Shannon equitability results for the 50 replicate simulations at the corresponding HL and IR value. A p-value < 0.05 (i.e. dark blue) represents 95% confidence that the two sampling methods produce different average equitability results.

### 5.3 Initial analysis

In this section we provide an overview of the results for both the random and contiguous scenarios, over the region of parameter space explored. Results are presented as heat-maps over parameter space. Each pixel corresponds to a unique pair of HL and IR values, with the temperature (colour) given by the corresponding value of the metric in question. In this way it is possible to gain a qualitative impression of how the various metrics respond as HL and IR are varied. We begin in section 5.3.1 with a comparison of results obtained using two different sampling methods. In sections 5.3.2 and 5.3.3 we then provide results for selected metrics associated with diversity, and variability respectively.

#### 5.3.1 Estimator comparison

The *Shannon equitability* metric (equation (2.3)) is calculated for all simulations using two different sampling methods. The first method uses snapshot sampling (as in chapter ??) i.e. species abundances are measured on the last time step of the simulation. The second method takes the mean species abundance over the final 4000 time steps of the simulation. Results obtained using

the two sampling methods are referred to as *snapshot* and *averaged*, correspondingly. We compare the results obtained using a *two-sided t-test*, which is implemented in the *Python* package *scipy*. The test is used to compare two datasets of independent samples, testing the null hypothesis that the expected value of the two datasets are equal. If the *p-value* of the test is smaller than the confidence threshold then there is sufficient evidence to reject the null hypothesis and conclude that the means of the two datasets are significantly different. For each test we are comparing the *snapshot* and *averaged* equitability results, calculated from the 50 replicate simulations at a given HL and IR value. If the test is significant then we conclude that the two sampling methods give significantly different results in calculation of the average Shannon equitability at that value of HL and IR. We conduct tests for all HL and IR values, and all three MAI ratios, under random and contiguous HL. The *p-values* of these tests are depicted in figure 5.1.

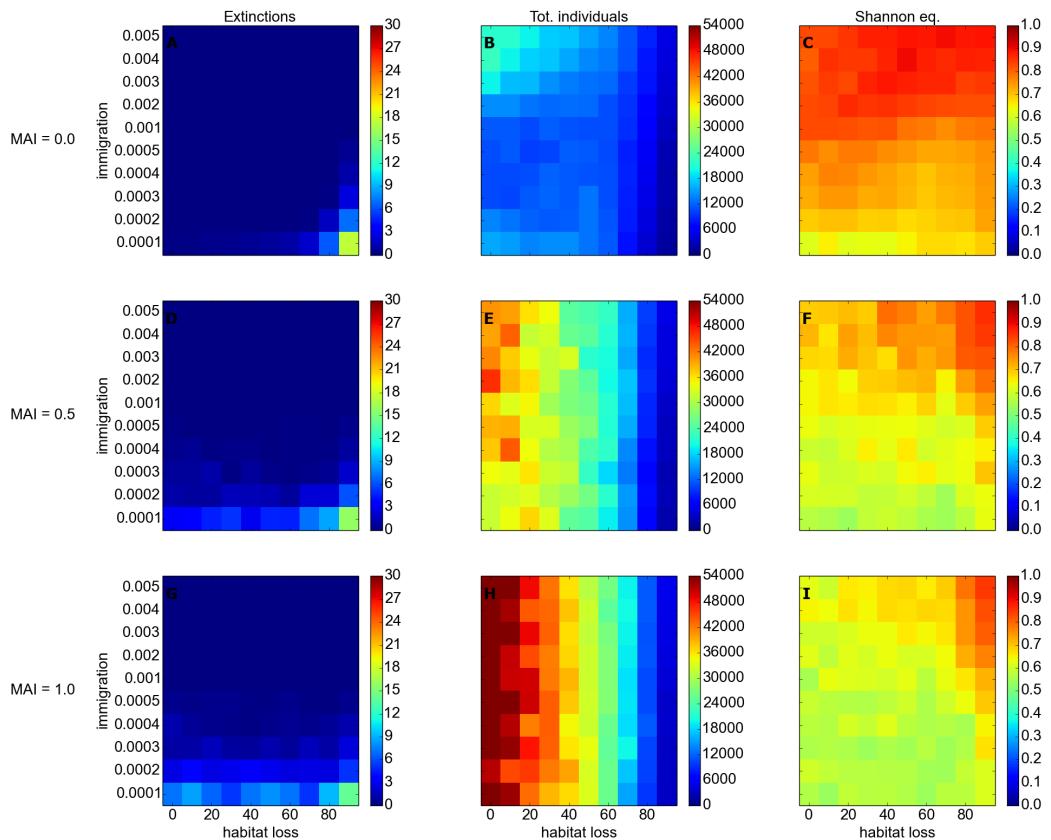
In general figure 5.1 shows that there is strong support for the conclusion that the two sampling methods produce the same average equitability results. The worst case is random HL at MAI= 0.0 (panel A). In this case there is a region of parameter space, above HL= 60%, where the p-values are significant. Therefore in this region the methods appear to give statistically different results. However, as stated, most of the tests suggest that the two methods produce statistically similar results. The similarity between the two methods is surprising given the results of the estimator analysis in section 4.6.1. Presumably averaging over 50 replicate simulations provides some reduction in noise. It may also be that a high level of precision in the estimates of species abundances is not required to calculate community level metrics. Based on the comparison presented here we conclude that snapshot sampling is sufficient to draw general conclusions about community structure over the ensemble of simulations. This allows consistency with the analysis in chapter 3. However we acknowledge that the use of snapshot samples may introduce some error, and increased variability, into our calculations. We treat the region of parameter space in which the equitability results proved dissimilar (panel A: random HL. MAI= 0.0) with particular caution.

NOTE: on network and diversity metrics - sample length of 200.

### 5.3.2 Diversity

Figures 5.2 and 5.3 illustrate the mean value of three metrics over the region of parameter space, *number of extinctions*, *total number of individuals* and *Shannon equitability*, for the random and contiguous scenarios respectively. All results shown are calculated from snapshot samples. As stated in section 5.2, an extinction is defined as the presence of less than three individuals in the landscape. The three metrics show broadly similar pattern across the region of parameter space in both HL scenarios. We highlight to following general observations:

- Reducing IR increases the number of extinctions (panels A,D,G, both figures). In the contiguous case there are, on average, more extinctions than the random case. These extinction are clearly induced by habitat loss, appearing to increase monotonically along



**Figure 5.2: Random HL:** Mean values of diversity metrics at each combination of HL and IR (Average over 50 replicate simulations). All metrics use snapshot sampling (see text). Each row corresponds to a different MAI ratio, as labelled. Panels A,D,G: Number of extinctions, defined as number of species with less than 3 individuals at end of simulation. Panels B,E,H: Total number of individuals in the community. Panels C,F,I: Shannon equitability metric.

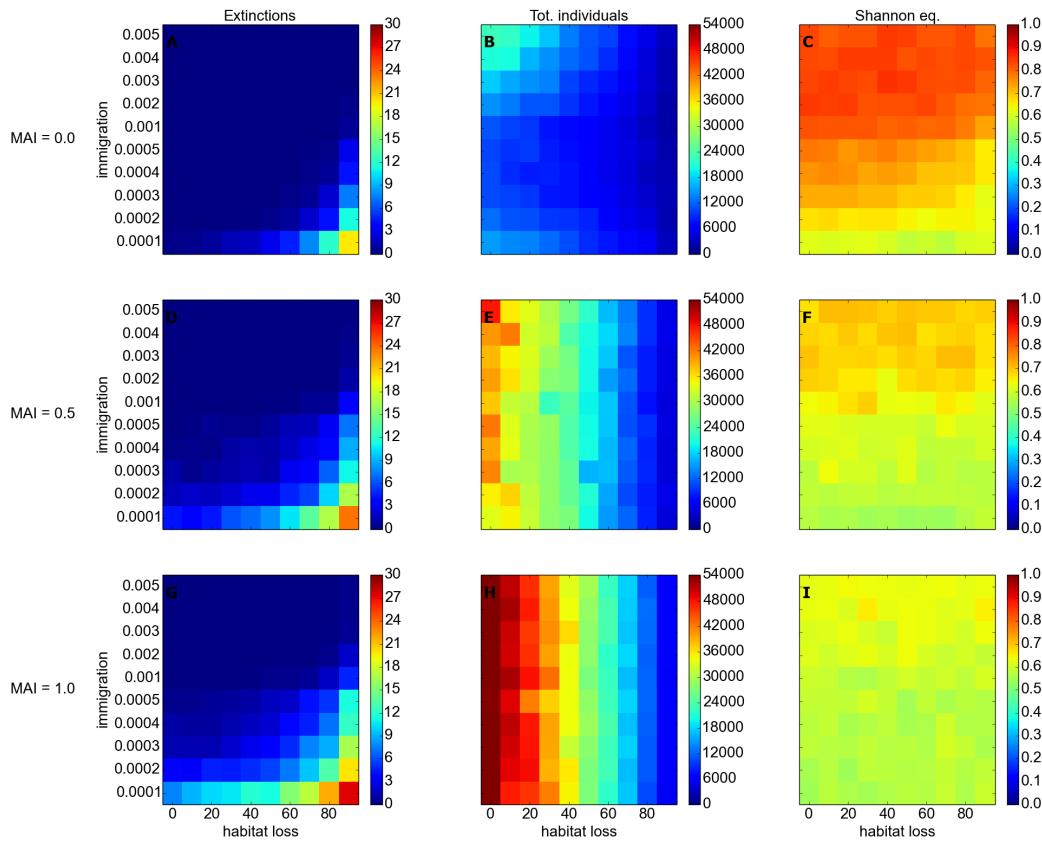


Figure 5.3: Similar to figure 5.2, but for **contiguous HL**.

the HL gradient. Communities under random HL exhibit fewer extinctions. And, especially in mutualistic communities the number of extinctions appears to be less sensitive to HL.

- In antagonistic communities the total number of individuals varies with IR (panel B, both figures). Initially reducing IR reduces the number of individuals, but at the lowest IR values the number of individuals increases again. In mutualistic communities the dependence of total individuals on IR is greatly reduced, if not removed altogether (panels E,H, both figures). In all cases HL reduces the number of individuals, as expected and in agreement with the results from chapter 3.
- In general mutualistic communities have higher total abundance than antagonistic ones. Under contiguous HL they exhibit more extinctions.
- In all cases reducing the IR reduces the Shannon equitability i.e. communities become less even. On the whole communities become more even under random HL (panels C,F,I),

figure 5.2), as observed in chapter 3. However at some IRs random HL appears to make antagonistic communities more even ( $IR= 0.0003$  to  $0.002$ , panel C, figure 5.2). Also in the contiguous scenario, at certain IRs, antagonistic communities become less even along the HL gradient ( $IR= 0.0002$  to  $0.003$ , panel C, figure 5.3). This effect is not clear in the case of mutualistic communities, where there is little visible dependence of equitability on HL (panels F,I figures 5.2).

The observations listed above confirm some of our predictions from section 5.1. The role of immigration in driving evenness is clear, since reducing IR reduces the Shannon equitability in all cases. In the random scenario there is some evidence that lower IRs reduces the magnitude of the change in evenness, as predicted. And in both scenarios there appear to be some cases where the evenness of the communities decrease with HL. This requires further analysis. Also the predicted extinctions are observed at low IR, due to a smaller *rescue effect*. However in the random scenario the number of extinctions does appear very sensitive to HL, which is unexpected. This, and the fact that the random scenario produces fewer extinctions than the contiguous, suggest that the mechanisms behind the species extinctions differ between the two scenarios. Based on what we saw in chapter 3 we propose that extinctions in the contiguous scenario are due to strong predation driven by high IS, whereas in the random scenario they are likely due to a collapse in the trophic structure of the community due to low IS.

The diversity results also highlight certain other new and unexpected features. In particular we find that MAI ratio play an important role in mediating how communities respond to HL across the range of IR values. This role is most clearly visible in the total number of individuals, which becomes insensitive to IR at high MAI ratios. In this sense at least mutualism can be said to confer robustness on communities in the face of variable IR. However, based on the persistence analysis in section 4.2.1, we may expect that the mutualistic communities at low IR are dominated by a small number of species in the non-basal trophic levels. It is perhaps the increased dominance of these species which accounts for the constant total abundance across the range of IR values. From the equitability results we know that, although abundance may not change, mutualistic communities do become less even at low IR. It will be informative to study community composition in more detail for such cases.

### 5.3.3 Variability and interaction strengths

Figures 5.4 and 5.5 show the mean value of three variability and interaction strength metrics for the random and contiguous scenarios respectively. Panels A,D, and G show the natural-logarithm of the mean interaction strength (IS, defined in section ??). Panels B,E, and H show the total number of (inter-specific) interaction events between all individuals, and the remaining panels show the natural-logarithm of the temporal variability metric (mean CV) (defined in section 2.7.2). The results for the contiguous scenario are simplest (figure 5.4), and are summarised as follows:

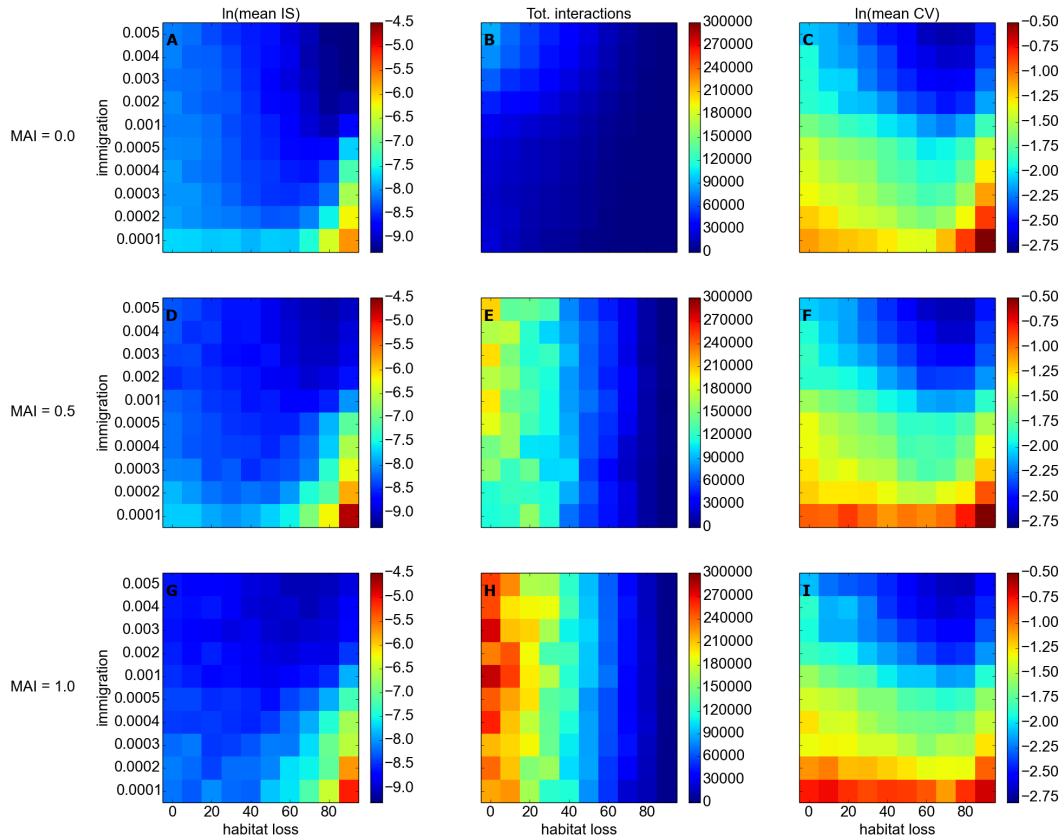


Figure 5.4: Similar to figure 5.2, but showing three different metrics: the natural logarithm of the mean interaction strength (ln(mean IS)); the total number of interactions between all species; and the natural logarithm of the mean temporal variability (ln(mean CV)). **Random HL**.

- Interaction strength and temporal variability increase with contiguous HL in all cases. So varying IR does not change the response of these metrics that was observed in chapter 3.
- Reducing IR increases temporal variability, as predicted. The increase in variability is associated with a slight increase in interaction strengths, which is weaker at high MAI.
- For antagonistic communities the number of interactions decreases with both IR and HL. For mutualistic communities the the number of interactions decreases with HL, but is less sensitive to IR.

In the random scenario there is a more subtle interaction between HL and IR. We observe the following from figure 5.5:

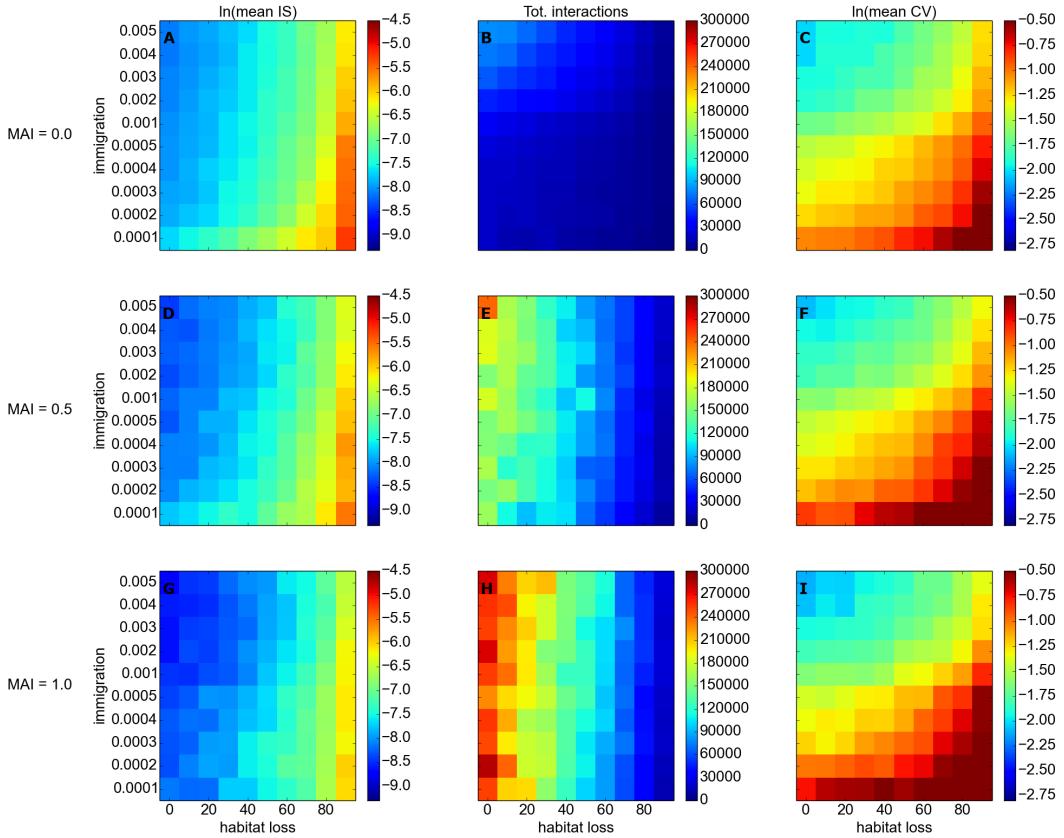


Figure 5.5: Similar to figure 5.4, but for **contiguous HL**.

- The response of the total number of interactions is qualitatively the same as in the contiguous case. However slightly more interactions are lost due to random HL than contiguous HL, in agreement with the results from chapter 3
- As in the contiguous case, reducing IR increases variability, with an associated increase in IS. However, especially at intermediate IR values, increasing HL initially reduces variability but causes variability to increase at high levels of destruction. Visually the profile of changing variability along the HL gradient appears to be approximately but not exactly correlated with changes in IS.

The results can largely be extrapolated from those in the previous two chapters and confirm our predictions. There are no surprises in the contiguous case. Although we note the variability and IS responses are consistent with, and therefore support conclusions drawn from, previous analysis. Also the similarity between the results for number of interactions and number of individ-

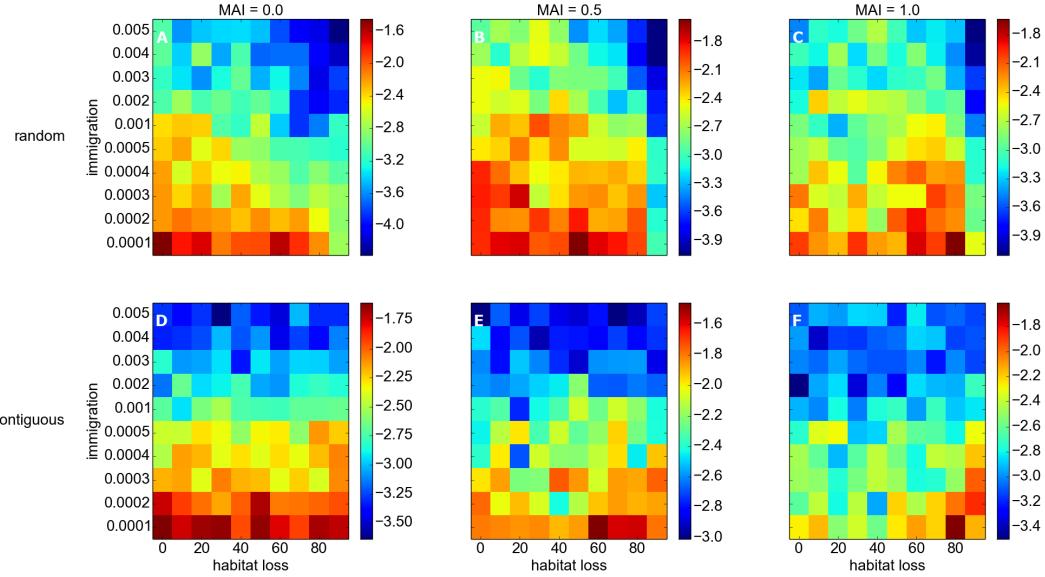


Figure 5.6: Natural logarithm of *ecosystem synchrony* ( $\ln(\text{Sync})$ ), defined in section 2.7.2.1 for both HL scenarios, and all three MAI ratios.

uals, in both HL scenarios, supports our assertion from chapter 3 that interaction frequencies are largely driven by abundance. We did not expect to see an increase in IS as IR is reduced, an effect which is present in all cases. However we did predict that effects due to species interaction would come to play a more important role as immigration was reduced. It is not clear why reducing IR increases the mean probability of trophic interactions between species (see discussion in section 3.4 for interpretation of IS in terms of probability). We suggest that it may be due to the dominance of communities by abundant and strongly interacting species. This hypothesis is investigated in section ??.

The variability response under random HL is another unexpected feature. In many instances random HL causes an initial decline, followed by a subsequent increase in temporal variability. This pattern appears robust across IRs and MAI ratios (panels C,F,I, figure 5.4). In fact, returning to figure 3.16 the same effect is present in our original results. The temporal variability increases at 90% HL. We explore this feature further in section ?? below.

### 5.3.4 Dependence on immigration

## 5.4 Bivariate Analysis

In this section we will present results for a fixed, low IR value against HL. And for a fixed intermediate HL value against IR.

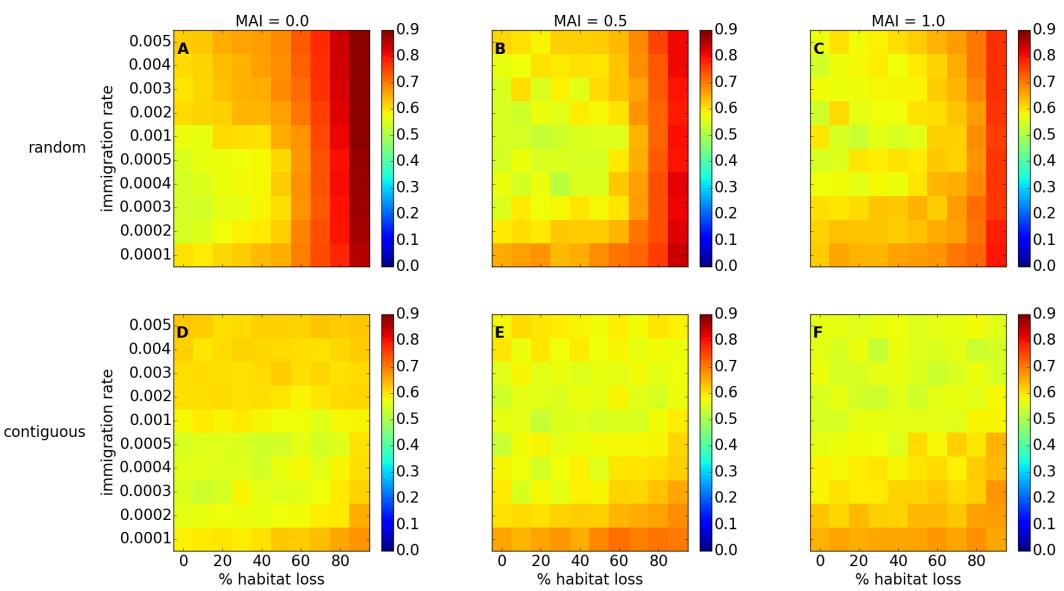


Figure 5.7: Similar to figure 5.6, but displaying the proportion of the total number of individuals born that come from immigration, calculated as in figure 3.24.

CHAPTER



## INFERRING SPECIES INTERACTIONS

### 6.1 Motivation

In this chapter we investigate the possibility of quantifying species interaction strengths from observed population dynamics. This was discussed at length in the introduction (section ??), however we now reiterate some of the important points here and to motivate our approach to the problem.

- metrics for interaction strength (choice of IM metric)
- hare-lynx dataset (ongoing debate!)
- Availability of time-series data. Plankton system (complications! e.g. seasonality)
- Motivate simulation approach (known interactions)
- Timme approach to fit GLV (use of GLV in general - constant interaction strengths)
- Simplify the problem (two species predator prey, but see extension)

We implement and test a novel method for quantifying species interactions from population dynamics. Initially we test this method using two distinct ODE simulation models to generate the population dynamics. These two models are referred to as the *linear* and *holling* models. The GLV can fit the *linear* model exactly. Therefore the results presented in section 6.3.1 serve as a test of our numerical method. We characterise the robustness of the method to the addition of noise, and under sparsity of sampling. The *holling* model cannot be fitted exactly by the GLV because it uses a different form of *functional response*. Therefore applying our method to population

dynamics simulated using the *holing* model provides a test of its ability to give approximate estimates of interaction strength, when underlying structure of the system is not of GLV type. This is also tested in the presence noise and under sparse sampling, and the results are presented in section 6.3.2. In section 6.3.3 we provide a preliminary look at how the method could be used to determine the type of functional response present, by observing the dynamics. Having characterised the performance of our method using ODE models, we then apply it, in section 6.5, to dynamics generated using the IBM model of previous chapters. This represents a first step towards applying the method to empirical data, since it involves a spatial system and a larger number of species. We conclude the chapter with a discussion of how the method could be developed further towards empirical application.

## 6.2 Methodology

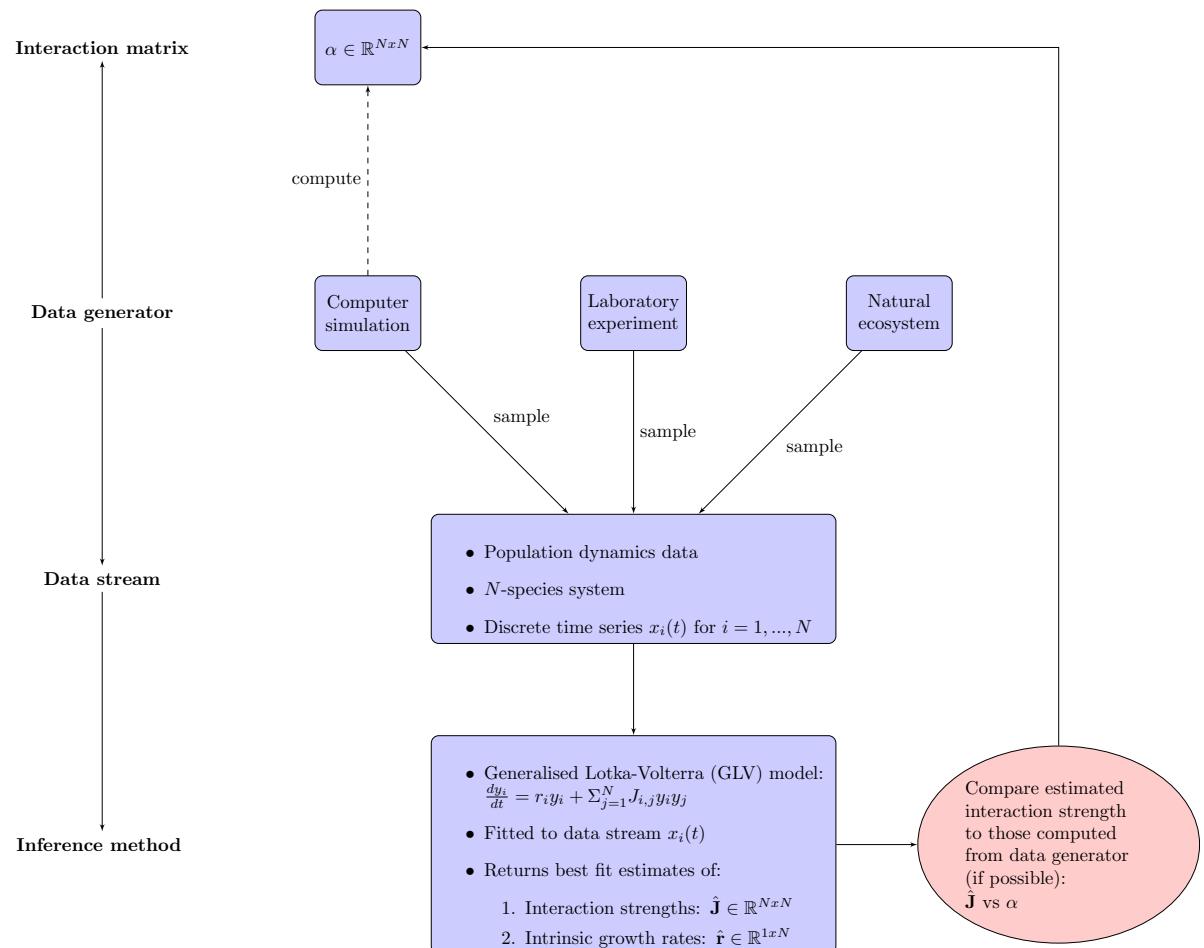


Figure 6.1: This is what we do.

To summarise our methodology, we simulate population dynamics then sample these dynamics

and fit a *generalised Lotka-Volterra* (GLV) model. The fitted GLV parameters give us estimates of the species interaction strengths (and other parameters), which we then compare to those used in the original simulation. The details of all the stages are given below. In section 6.2.1 the *interaction matrix* (IM) is introduced. The IM is the metric used to quantify the strength of species interactions and is key to this chapter. We also introduce the *generalised Lotka-Volterra* (GLV) model, and show that this model has constant interaction strengths, given by the coupling matrix  $J$ . In section 6.2.2 we give a general framework for ODE predator-prey modelling, and derive the two models that we use to simulate population dynamics. We then discuss, in section 6.2.3 the details of how these models are simulated *in silico*, including the selection of model parameters. Section 6.2.4 gives the details of the numerical method we use for fitting the GLV model to sampled population dynamics. In section 6.2.5 we give an example of the full methodology in action.

### 6.2.1 Interaction strength

The metric used for species interaction strength is key to this chapter. As discussed in section 6.1 there are several metrics available [REF]. However there is one that is a natural choice, given our methodology. This metric allows us to calculate the interaction strengths from our simulation models, and to directly compare them to the estimates obtained by fitting the GLV model. The metric is called the *interaction matrix* (IM). The elements of the IM,  $\alpha_{ij}$ , quantify the effect of a small change in the population density of species  $j$  on the per capita growth rate of species  $i$  [REF]. Therefore the IM elements are given by:

$$(6.1) \quad \alpha_{ij} = \frac{\partial}{\partial x_j} \left( \frac{1}{x_i} \frac{dx_i}{dt} \right),$$

where  $x_i$  and  $x_j$  are the population densities of species  $i$  and  $j$  respectively. In the case of our two species systems the IM is a  $2 \times 2$  matrix, but trivially extends to quantify all  $N^2$  pair-wise interactions between species in a  $N$ -species system (including self-interactions).

Using the IM we are able to calculate the interaction strengths exactly from the models that we use for simulation. This is because they are ODE models with explicit expressions for  $dx_i/dt$ , so we can evaluate the partial derivative in equation 6.1 to obtain analytic forms for all the IM elements ( $\alpha_{00}, \alpha_{01}, \alpha_{10}, \alpha_{11}$ ). Depending on the model used the IM elements are either constants, or are functions of prey density. The interaction strengths for our simulation models are given at the end of section 6.2.2, and are illustrated in figure 6.2.

**Generalised Lotka-Volterra.** The GLV model is the extension of the Lotka-Volterra equations to  $N$  species. We use this model to fit to simulated population dynamics and obtain estimates of the underlying interaction strengths. The GLV model is given by:

$$(6.2) \quad \frac{dx_i}{dt} = r_i x_i + \sum_{j=1}^N J_{ij} x_i x_j,$$

where  $x_i$  is the population density of species  $i$ ;  $r_i$  is the intrinsic growth rate;  $N$  is the number of species and  $J_{ij}$  is the coupling between species  $i$  and  $j$ . Applying equation 6.1 to equation 6.2 we see that  $\alpha_{ij} = J_{ij}$ , such that the IM for the GLV model is equal to the coupling matrix. Therefore, by fitting the GLV model to population dynamics, the fitted parameters  $J_{ij}$  give us numeric estimates of interaction strength. To perform the fit we use the method detailed in section 6.2.4.

### 6.2.2 Population models

We use ordinary differential equation (ODE) models to simulate two species predator-prey dynamics. Therefore the mathematics presented in this section focuses on this case. However the framework may be easily extended to include other interaction types (competition and mutualism), and to model larger systems in a similar way. For a two species system the dynamics are governed by two coupled equations, one for each species. Importantly it is possible, using the IM introduced in section 6.2.1, to calculate the species interaction strengths from the model. The model equations take the general form:

$$(6.3) \quad \begin{aligned} \frac{dx_0}{dt} &= G_0(x_0) + \alpha_{01} x_1 H(x_0, x_1), \\ \frac{dx_1}{dt} &= G_1(x_1) + \alpha_{10} x_0 H(x_0, x_1) \end{aligned}$$

where species  $x_0$  and  $x_1$  are the population densities of the prey and the predator species respectively; the  $G_i(x_i)$  are the intrinsic growth functions of each species; the  $\alpha_{ij}$  are constant coefficients and  $H(x_0, x_1)$  is the functional response (FR) of the predator. This form is standard in the literature [REFS] and many models may be expressed in this way by choosing different functional forms for  $G$  and  $H$ . The coefficients  $\alpha_{01}$  and  $\alpha_{10}$  are negative and positive respectively, such that the prey losses biomass, and the predator gains biomass as a result of the interaction. These coefficients may be used to introduce asymmetry into the interaction terms. For example it is common to choose  $|\alpha_{01}| > |\alpha_{10}|$ , to model the inefficiency of the predator in the conversion of biomass from the prey. For the intrinsic growth functions we use the functional forms:

$$(6.4) \quad G_0(x_0) = r_0 x_0 \left(1 - \frac{x_0}{K_c}\right)$$

$$(6.5) \quad G_1(x_1) = -r_1 x_1,$$

where  $r_i \in \mathbb{R}^+$  is the intrinsic growth rate of species  $i$ , and  $K_c$  is the carrying capacity of the prey species. Therefore the predator has an exponential intrinsic mortality, whereas the

prey species has logistic intrinsic growth. These use of these functional forms in predator-prey modelling was made popular by Rosenzweig and MacArthur [REF]. They are now widely used [REFS].

The FR defines the per-predator rate of consumption of prey. We focus on the forms proposed by Holling in the 1950s [REFS], which are also widely used [REFS]. However it is worth noting that various other forms have been proposed and there is an ongoing debate about which to use [REFS] (see discussion in section 6.6). There are three types of Holling FR, referred to as types I, II, and III. These can be expressed as:

$$(6.6) \quad H_I(x_0, x_1) = x_0,$$

$$(6.7) \quad H_{II}(x_0, x_1) = \frac{x_0}{x_0 + K_s},$$

$$(6.8) \quad H_{III}(x_0, x_1) = \frac{x_0^2}{x_0^2 + K_s^2},$$

where  $x_0$  is the prey density, and  $K_s$  is the saturation constant for the predator, giving the prey density at which the per-predator consumption rate reaches half-maximum. We choose to narrow our investigation by focusing here on the first two forms: Holling type I and type II. Therefore we have two distinct simulation models, which we refer to as the *linear* and *holling* models.

**Linear model.** This model uses the type I FR. As is clear from equation 6.6, type I is the simplest of the Holling functions: the per-predator predation rate increases linearly as the abundance of available prey increases. The slope of this linear relationship given by the  $a_{ij}$  parameters in equations 6.5. The linear FR is the same as is used in the famous Lotka-Volterra equations [REF], meaning that our linear model may be expressed in GLV form (equation 6.2). We may rescale the equations of the linear model in order to reduce the number of parameters. This makes the local stability analysis simpler, and reduces the dimension of the search space when probing the equations numerically via simulation. The re-scaled equations are given by:

$$(6.9) \quad \frac{d\chi_0}{dt} = A\chi_0(1 - \chi_0) - B\chi_0\chi_1$$

$$(6.10) \quad \frac{d\chi_1}{dt} = -\chi_1 + C\chi_0\chi_1,$$

where  $\chi_0$  and  $\chi_1$  are the re-scaled prey and predator population densities respectively; the parameters  $A, B, C \in \mathbb{R}^+$ . The equilibrium population densities are given by:

$$(6.11) \quad \chi_0^* = \frac{1}{C}$$

$$(6.12) \quad \chi_1^* = \frac{A}{B} \left(1 - \frac{1}{C}\right),$$

Therefore  $\chi_0^*$  is always positive, and  $\chi_1^*$  is positive if  $c > 0$ . This is a requirement for physical realism, since it is not possible to have negative populations of species. In most applications it is also required that this equilibrium is stable, to allow for the coexistence of species. We use these conditions on the equilibrium for parameter selection, which is discussed in section 6.2.3. By applying equation 6.1 to equations 6.9 and 6.10 we can evaluate the elements of the IM for the linear model. This gives:

$$(6.13) \quad \alpha_{linear} = \begin{bmatrix} -A & -B \\ C & 0 \end{bmatrix},$$

such that all the interaction strengths are constants, which is illustrated in figure 6.3.

**Holling model.** This model uses the type II FR, which is a non-linear function of prey density. As we can see from equation 6.7 and figure 6.2, this FR models predator saturation - individuals take a certain amount of time to process and digest prey - such that the response curve flattens out at high prey densities. The difference between the type I and II functions is illustrated in figure 6.2. We may perform a similar rescaling as we did with the linear model to reduce the number of parameters. The resulting equations for the holling model are given by:

$$(6.14) \quad \frac{d\chi_0}{dt} = A\chi_0(1-\chi_0) - \frac{B\chi_0\chi_1}{\chi_0+D}$$

$$(6.15) \quad \frac{d\chi_1}{dt} = -\chi_1 + \frac{C\chi_0\chi_1}{\chi_0+D},$$

where the saturation constant  $D \in \mathbb{R}^+$ , and the other symbols are the same as in equations 6.9 and 6.10. The equilibrium populations for this model are given by:

$$(6.16) \quad \chi_0^* = \frac{D}{C-1}$$

$$(6.17) \quad \chi_1^* = \frac{ACD(C-1-D)}{B(C-1)^2},$$

Therefore  $\chi_0^*$  is positive if  $C > 1$ ,  $\chi_1^*$  is positive if  $C - D > 1$ . Again we use these conditions, and the requirement of stability, to constrain our choice of parameters (6.2.3). As we did for the linear model, we can evaluate the IM for the holling model, giving:

$$(6.18) \quad \alpha_{holling} = \begin{bmatrix} -A + \frac{B\chi_1}{(\chi_0+D)^2} & \frac{-B}{\chi_0+D} \\ \frac{CD}{(\chi_0+D)^2} & 0 \end{bmatrix},$$

such that the three non-zero elements of the IM are functions of prey density  $\chi_0$ , instead of constants. The shape of these interaction functions is shown in figure 6.2, and we will return to them in section 6.2.5.

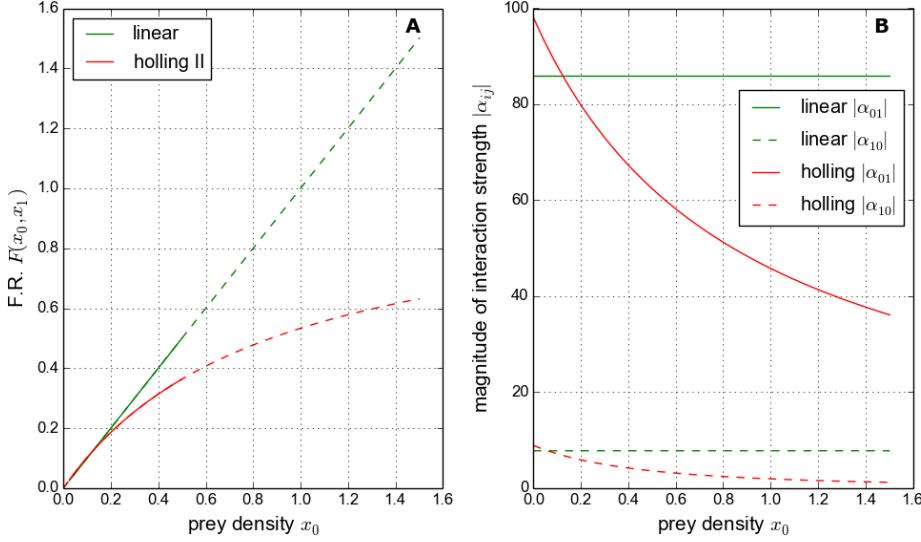


Figure 6.2: Example of (A) the functional response curve, and (B) the corresponding inter-specific interaction strengths for one parameter set of the *linear* model, and one of the *holling* model.

	A	B	C	D
linear	0.1 - 100	0.1 - 100	1 - 100	N/A
holling	0.1 - 100	0.1 - 100	1.1 - 100	0.1-99

Table 6.1: Ranges from which parameters were selected uniformly at random for the two ODE simulation models. The parameters are all allowed to vary over at least three orders of magnitude, to ensure that our investigation covers a large region of parameter space. The restrictions on parameters C and D ensure that it is always possible to achieve an equilibrium population of both species that is strictly positive (see equations 6.11, 6.12, 6.16, 6.17)

### 6.2.3 Simulation procedure

We apply a strict recipe when running simulations in order to ensure consistency and to allow comparison of our numerical results. Key to this is the control of certain variables across simulations, and also our method for parameter selection, both of which are discussed below. All simulations are run using the first-order forward Euler approximation to the ODE model. We use additive Gaussian noise to simulate *process error*. Therefore we implement the stochastic difference equation :

$$(6.19) \quad \chi_{i,t+1} = \chi_{i,t} + \Delta t \Delta \chi_{i,t} + \xi_{i,t},$$

where  $\Delta t$  is the integration time step,  $\Delta \chi_i$  is given by the right hand side of ODE model being simulated (e.g. equations 6.9, 6.10 for the linear model); and the additive noise term

$\xi_{i,t} \sim \mathcal{N}(0, \sigma_{noise} \chi_{i,t} \Delta t)$ . The value of  $\sigma_{noise}$  is quoted as *noise intensity* in what follows. In the event of stochastic extinction of either species, both population densities are reset to their initial conditions. The case where  $\sigma_{noise} = 0$  is referred to as the *deterministic* model. In all the results presented the simulations were run with a time step  $\Delta t = 10^{-4}$ . All code was implemented in the language *Python*, and large computations were performed on the UoB HPC cluster *Blue Crystal* [REF].

The goal of fitting the GLV model to simulated population dynamics requires that the dynamics contain enough information to perform the fit - it is not possible to fit the a model if species populations are sitting at equilibrium. Therefore we follow the precedent set in [REF], such that all simulated dynamics of the *deterministic models* exhibit two ‘large amplitude’ oscillations about a stable equilibrium (see condition 2 below). Every simulation is run with the initial population densities set to half of their equilibrium value. This ensures that all systems start consistently away from equilibrium.

**Parameter selection.** We select an ensemble of 100 parameter sets for both simulation models (*linear* and *holling*). Parameters are selected uniformly at random from predefined ranges, which are given in table 6.1. This range ensures that a positive equilibrium population is possible (see equations 6.11, 6.12, 6.16, 6.17), but also allows for parameters to vary over at least three orders of magnitude so that our investigation covers a large region of parameter space. The selected parameters are then accepted if they meet the following conditions:

1. The equilibrium population is positive, and is locally a stable spiral (eigenvalues of the Jacobian have negative real part and complex conjugate imaginary part).
2. The deterministic dynamics exhibit at least two full rotations in the phase plane before relaxing to within 5% of the equilibrium (Euclidean distance in the phase plane).
3. The population densities do not differ by more than an order of magnitude, in the deterministic case.

The two parameter sets generated by the above procedure are used for the simulation results presented in section 6.3. All simulations , including those with  $\sigma_{noise} \neq 0$ , are run for the length of time  $T_{2P}$  required to achieve two full oscillations in the deterministic case, for that parameter set.

#### 6.2.4 Numerical estimation method

To estimate the inter-specific interaction strengths we use a numerical method, adapted from [99], to fit the GLV model to the population dynamics. The method gives ‘best fit’ estimates of the GLV parameters, which include constant coefficients for the interaction strengths as we saw in section 6.2.1. We include here a derivation of the method, slightly adapted and simplified for

our purposes. Say the dynamics of the population density of a species  $x_i$  is governed by coupled differential equations of the form:

$$(6.20) \quad \dot{x}_i = r_i f_i(x_i) + \sum_{j=1}^N J_{i,j} g_{ij}(x_i, x_j),$$

where  $\dot{x}_i = \frac{d}{dt}x_i$ ;  $N$  is the number of species in the system and  $i, j$  index the species. The  $r_i, J_{i,j} \in \mathbb{R}^+$  are constants, and the functions  $f_i$  and  $g_{ij}$  are known. This form looks familiar, indeed all of the ODE models discussed so far may be expressed in this form - there is an intrinsic growth term, and a linear sum of interaction terms. To express the GLV model (equation 6.2) in this form, we have:

$$(6.21) \quad f_i(x_i) = x_i$$

$$(6.22) \quad g_{ij}(x_i, x_j) = x_i x_j,$$

It would be possible to use this method to fit models other than the GLV, so long as the functions  $f_i$  and  $g_{ij}$  are *known and parametrised*. Since the functions  $f_i$  and  $g_{ij}$  are known there are  $N+1$  unknowns in equation 6.20:  $r_i$  and  $J_{i,j}$  for  $j = 1, \dots, N$ . Therefore, if we knew the exact values of  $\dot{x}_i, x_i$  and the  $x_j$ 's, at  $N+1$  time points, then we could solve the equation for  $r_i$  and the  $J_{i,j}$ 's. However in any practical application our knowledge of the system is not *exact*; the system is subject to noise; and the model may be an imperfect description of the dynamics. So the equation cannot be solved exactly. We must look for an approximate solution. To do this the full state of the system is sampled at  $M+1$  time points  $t_m$  for  $m \in 1, \dots, M, M+1$ . These samples are used to construct estimates for the states  $x_i$  and their time-derivatives  $\dot{x}_i$  at  $M$  intermediate time points, for every species  $i$ .

The simplest way to estimate the time-derivatives is to take the finite difference between observations at two consecutive time points, giving estimates:

$$(6.23) \quad \hat{x}_i(\tau_m) := \frac{x_i(t_m) - x_i(t_{m-1})}{t_m - t_{m-1}},$$

where  $\tau_m \in \mathbb{R}, m \in \{1, \dots, M\}$  is the midpoint of the two time-points:

$$(6.24) \quad \tau_m := \frac{t_{m-1} + t_m}{2}.$$

To evaluate the functions  $f_i, g_{ij}$  at these new time-points we must estimate the states  $x_i(\tau_m)$  from our observations. We use the linear interpolation:

$$(6.25) \quad \hat{x}_i(\tau_m) := \frac{x_i(t_{m-1}) + x_i(t_m)}{2}.$$

So from equation 6.20 we can now construct  $M$  equations using our  $M + 1$  samples:

$$(6.26) \quad \hat{x}_i(\tau_m) = r_i f_i(\hat{x}_i(\tau_m)) + \sum_{j=1}^N J_{i,j} g_{ij}(\hat{x}_i(\tau_m), \hat{x}_j(\tau_m)).$$

We now simplify the notation such that equation 6.26 may be written

$$(6.27) \quad \hat{x}_{i,m} = r_i f_{i,m} + \sum_{j=1}^N J_{i,j} g_{ij,m},$$

where the subscripts  $i, j$  indicate the species, and  $m$  indicates the time-point  $\tau_m$  for which the equation holds. This system of  $M$  equations can be expressed in matrix form:

$$(6.28) \quad X_i = J_i G_i,$$

where we have

$$(6.29) \quad X_i = \begin{pmatrix} \hat{x}_{i,1} & \hat{x}_{i,2} & \cdots & \hat{x}_{i,M} \end{pmatrix} \in \mathbb{R}^{1 \times M},$$

$$(6.30) \quad J_i = \begin{pmatrix} r_i & J_{i,1} & J_{i,2} & \cdots & J_{i,N} \end{pmatrix} \in \mathbb{R}^{1 \times (N+1)},$$

$$(6.31) \quad G_i = \begin{pmatrix} f_{i,1} & f_{i,2} & \cdots & f_{i,M} \\ g_{i,1,1} & g_{i,1,2} & \cdots & g_{i,1,M} \\ g_{i,2,1} & g_{i,2,2} & \cdots & g_{i,2,M} \\ \vdots & \vdots & \ddots & \vdots \\ g_{i,N,1} & g_{i,N,2} & \cdots & g_{i,N,M} \end{pmatrix} \in \mathbb{R}^{(N+1) \times M}.$$

The system 6.28 has  $N + 1$  unknowns ( $J_{i,k}$  for  $k = 1, \dots, N + 1$ ) and  $M$  equations. In the case when  $M > N + 1$  the system is over constrained and there is no exact solution in general. We look for an approximate solution  $\hat{J}_i$  that minimises the error between the LHS and RHS of equation 6.28. We take the error function:

$$(6.32) \quad E_i(\hat{J}_i) = \sum_{m=1}^M (X_{i,m} - \sum_{k=1}^{N+1} \hat{J}_{i,k} G_{i,k,m})^2,$$

which we want to minimise with respect to the matrix elements  $\hat{J}_{i,k}$ . That is

$$(6.33) \quad \frac{\partial}{\partial \hat{J}_{i,k}} E_i(\hat{J}_i) \stackrel{!}{=} 0.$$

By taking the derivative of the RHS of equation 6.32 we have that:

$$\begin{aligned}\frac{\partial}{\partial \hat{J}_{i,k'}} E_i(\hat{J}_i) &= \frac{\partial}{\partial \hat{J}_{i,k'}} [\sum_{m=1}^M (X_{i,m} - \sum_{k=1}^N \hat{J}_{i,k} G_{i,k,m})^2] \\ &= -2 \sum_{m=1}^M [(X_{i,m} - \sum_{k=1}^N \hat{J}_{i,k} G_{i,k,m}) G_{i,k',m}]\end{aligned}$$

To find the minimum of the error function we equate this to zero, giving:

$$\begin{aligned}0 &= \sum_{m=1}^M (-X_{i,m} G_{i,k',m} + G_{i,k',m} \sum_{k=1}^N \hat{J}_{i,k} G_{i,k,m}) \\ &= (-X_i G_i^T)_{k'} + \sum_{m=1}^M G_{i,k',m} (\hat{J}_i G_i)_m \\ &= (-X_i G_i^T)_{k'} + \sum_{m=1}^M (\hat{J}_i G_i)_m G_{i,m,k'}^T \\ (6.34) \quad &= -X_i G_i^T + \hat{J}_i G_i G_i^T\end{aligned}$$

Therefore we conclude that:

$$(6.35) \quad \hat{J}_i = X G_i^T (G_i G_i^T)^{-1},$$

which, in our case, is the analytic form for the best estimate of the row corresponding to species  $i$  in parameter matrix of the GLV model. For a two species system, by applying equation 6.35 to each species, we can obtain the full set of GLV parameter estimates:

$$(6.36) \quad \hat{J} = \begin{pmatrix} \hat{J}_{0,0} & \hat{J}_{0,1} \\ \hat{J}_{1,0} & \hat{J}_{1,1} \end{pmatrix},$$

and

$$(6.37) \quad \hat{r} = \begin{pmatrix} \hat{r}_0 & \hat{r}_1 \end{pmatrix}.$$

We perform the computation by constructing the matrices  $X_i$  and  $G_i$  for both species, and do the matrix multiplication using the *Python* package *numpy*[REF]. The fact that the error minimisation has an analytic solution makes it very computationally efficient, allowing us to perform many replicate calculations. However the performance of the method may be lower than other, more computationally expensive, model fitting algorithms (see discussion section 6.6). It is possible to assess the goodness of fit achieved by evaluating the error function (equation 6.32).

### 6.2.5 Examples

Here we show examples of the dynamics of both models, both with and without noise..For the holling model we also show the variability in interaction strengths during the simulations..We also present a table with the results of the GLV, comparing them to the simulation interaction strengths..

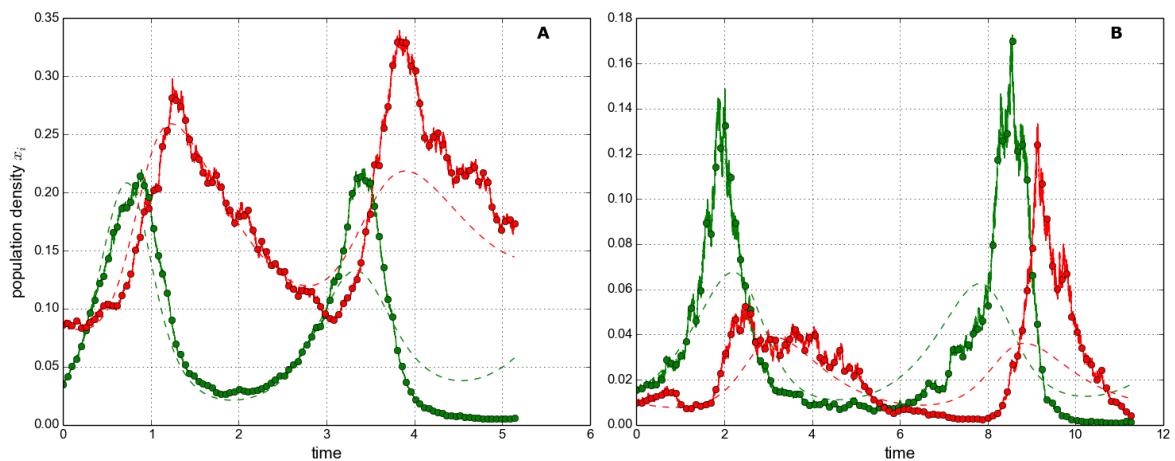


Figure 6.3: Example linear dynamics. 100 sampels. Two different parameter sets. A: noise=20. B:noise=50.

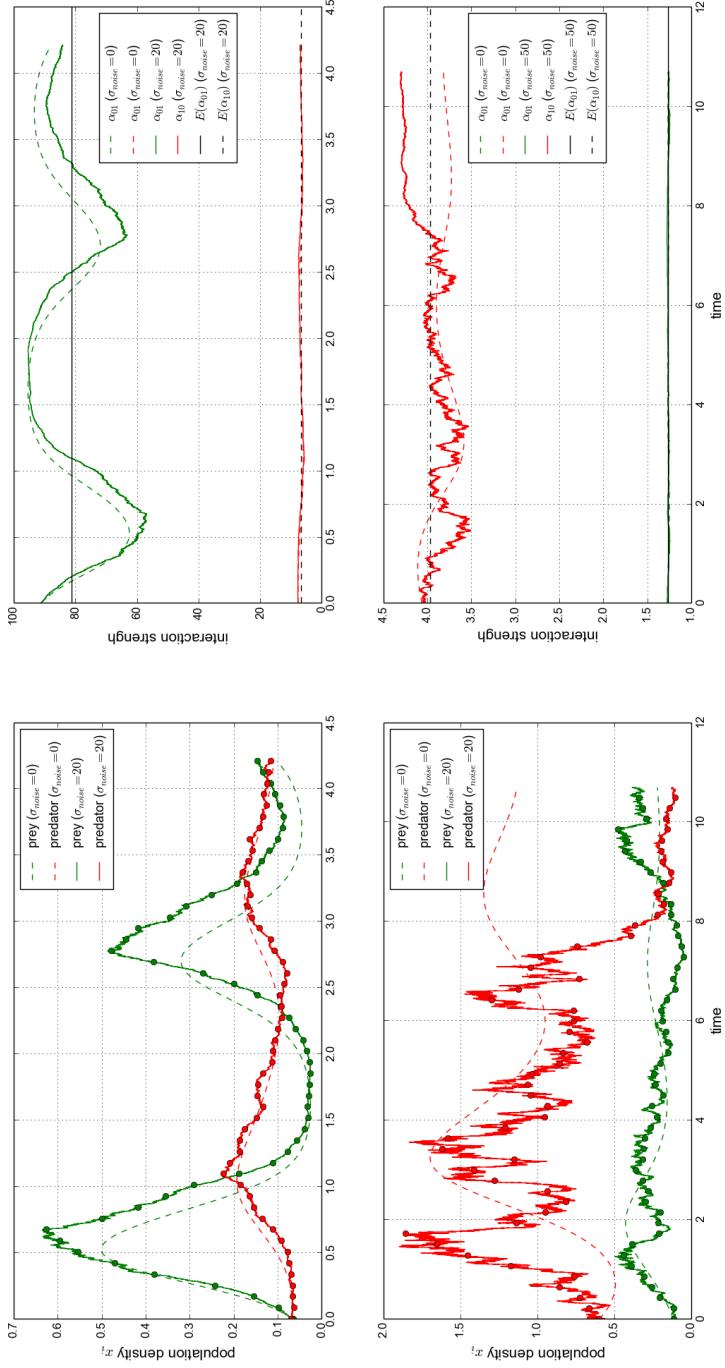


Figure 6.4: Example Holling II dynamics.

## 6.3 Results

In this sections we characterise the numerical performance of the method, described in section 6.2.4, for estimating the strength of interactions between species. The method is tested on the dynamics of two different ODE systems: a Lotka-Volterra (LV) and a Holling type II (HII) system. In the first case it is simply a test of a model fitting procedure. This is because the method works by fitting a generalised Lotka-Volterra (GLV) model to the dynamics, and he LV systems can be expressed as a GLV systems. Therefore we are simply simulating using one model, and then testing a method of estimating the model parameters form the simulated dynamics. We test the effects of noise and sampling frequency. In the second case, the HII system cannot be expressed as a GLV system. Therefore the GLV model that we fit can only approximate the dynamics and we cannot make a direct comparison between the parameters of the simulation model and the GLV model used for estimation. In this case we comapre to the mean interaction strengths (see section 6.3.2).

### 6.3.1 Linear model

Initially we run repeated simulations of the LV model using a single parameter set. We investigate how the numerical estimates of the model parameters respond to two variables: the level of noise in the simulations; and the number of samples used for estimation. Other variables are held constant using the simulation procedure described in section 6.2.3. We then generalise these results by looking at the relative error in the estimates, for repeated simulations using an ensemble of 100 selected parameter sets (as described in section 6.2.3).

**Single parameter set.** Here we can make direct comparison between model parameters. The GLV model for two species has six parameters:  $r_0, r_1, J_{00}, J_{01}, J_{10}, J_{11}$ . These correspond respectively to the following constant values of the LV system used for simulations:  $A, -1, -A, -B, C, 0$  (see equation ??). In general we find that the numerical estimates perform well at low noise intensities and poorly at high noise intensities. This is illustrated in figures 6.5 and 6.6. We also find that the estimates improve with the number of samples used, up to a point. Beyond this point the use of more samples does little to improve to estimates, and in some cases makes them worse. This behaviour is illustrated in figures 6.7 and 6.8. These patterns were found to hold across all parameter sets investiagted, but are only shown using a single parameter set here for clarity.

In panel **A** of figures 6.5 and 6.6 we see that the mean value of the estimates approaches the true value for low noise and, in panel **B** that the variance in the estimates approaches zero. This tells us that the method consistently gives a good fit of the GLV to the dynamics of the LV system, even when only 100 sample points are used (figure 6.5). As the noise intesitiy is increased the mean values of the estimates deviate from the true values, and the standard deviation in the estimates increases. Comparing the two figures we see that the response to noise is very similar

whether 100 or 10,000 samples are used. A notable exception to this is a spike in the variance in panel **B** of figure. However this appears to be a single statistically anomalous result and not part of the trend. Panel **C** of both figures shows that the error function, which is minimised by the estimation method, increases with noise for both species. This cannot be directly compared between the two plots because of the different number of samples used. However it indicates that in both cases (100 and 10,000 samples) the quality of the fit is high in the deterministic case, and decreases with noise.

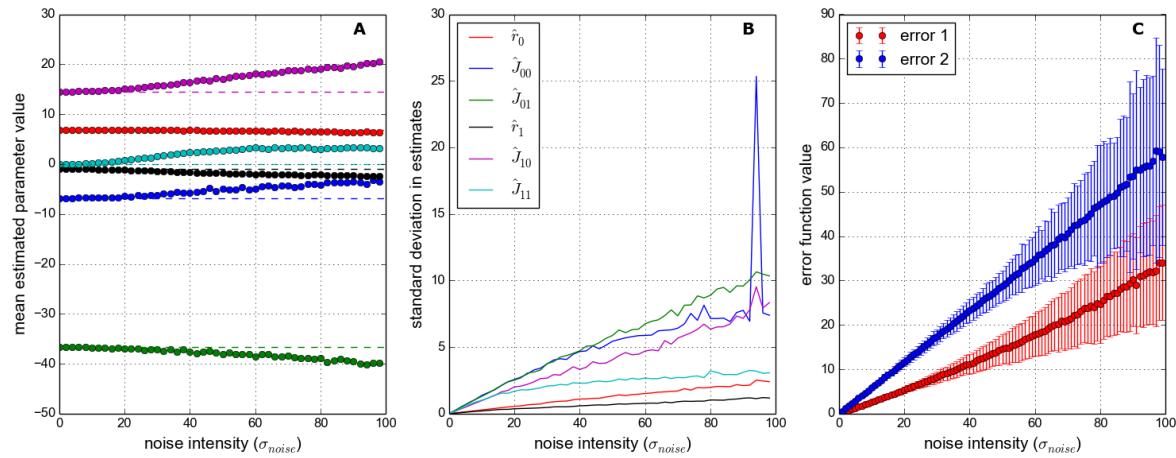


Figure 6.5: Effect of noise on numerical estimates. Here the method uses 100 samples from simulated dynamics. All simulations run using the LV model with a single parameter set. The noise intensity varies between 0 (deterministic) and 100. See section 6.2.5 for an intuition of how noisy this is. 1000 repeat simulations run at each noise level. **Panel A:** Mean estimated parameter values (each dot representing mean over 1000 repeats). The ‘true’ parameter values of the simulation model are shown by dashed lines. **Panel B:** Standard deviation in estimates. **Panel C:** Value of the error functions used in the estimation method, one for each species. The dots show the mean error, and the bars show  $\pm$  one standard deviation.

We now look at how the estimates respond to the number of samples used, in the cases of low and high noise intensity. Figure 6.7 shows the low noise case, with  $\sigma_{noise} = 10$ . In panel **A** we see that the mean value of the estimates quickly converges to close to the true parameter values, as the number of samples increases. Panel **B** shows that the standard deviation in the estimates quickly becomes small, but non-zero. Above about 32 samples there is no visible improvement in the estimates, as measured by the mean or standard deviation. In figure 6.8 we see the effect of a higher noise intensity. Here we have  $\sigma_{noise} = 50$ . Panel **A** shows that the estimates do not converge on the true parameter values, even for large numbers of samples. Also the standard deviation in the estimates, shown in panel **B**, is higher than in the low noise case. Again we find that there is little, if any, improvement in the estimates beyond about 32 samples.

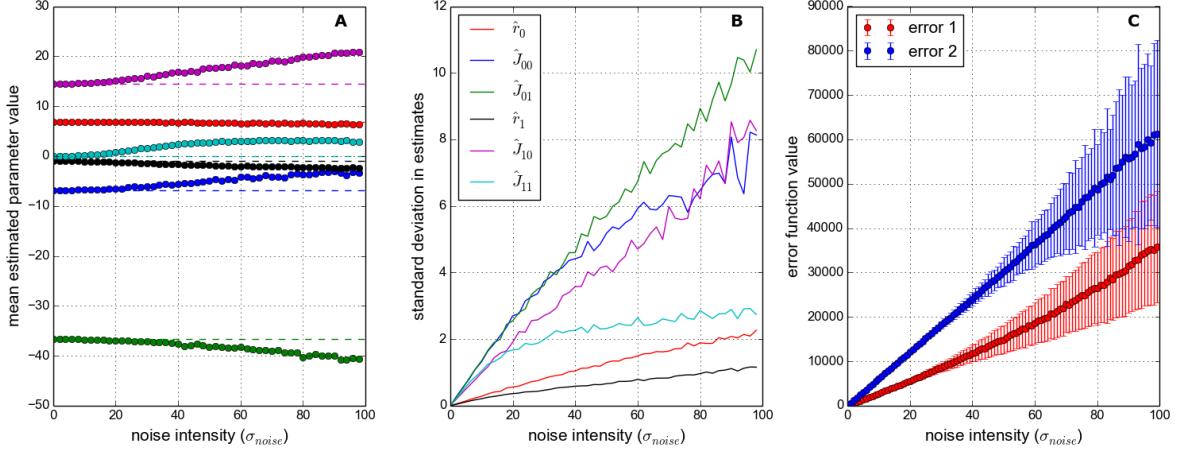


Figure 6.6: Exactly as in figure 6.5 but using 10,000 samples from the simulated dynamics.

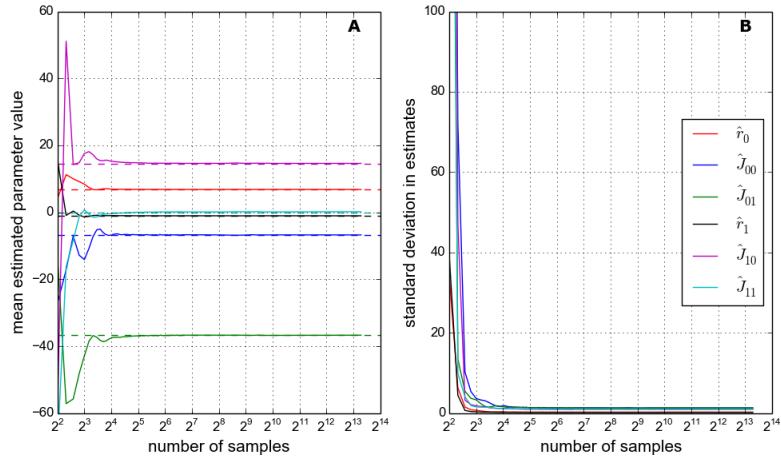


Figure 6.7: Effect of the number of samples on numerical estimates. All simulations run using the LV model with a single parameter set. The noise intensity  $\sigma_{noise} = 10$ . Number of samples ranges from 4 to 10,000. Samples drawn from simulated dynamics at equal intervals. 1000 repeat simulations for each number of samples. **Panel A:** Solid lines show mean estimated parameter values. Dashed lines show the ‘true’ parameter values of the simulation model. **Panel B:** Standard deviation in estimates.

**Ensemble of parameter sets.** Run 10 repeats for each of 100 parameter sets. In general the trends described above hold across the ensemble..

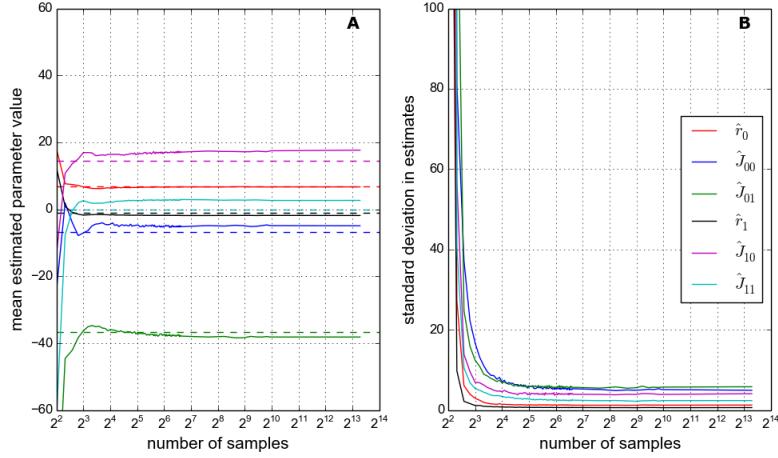


Figure 6.8: Exactly as in figure 6.7 but with noise intensity  $\sigma_{noise} = 50$ .

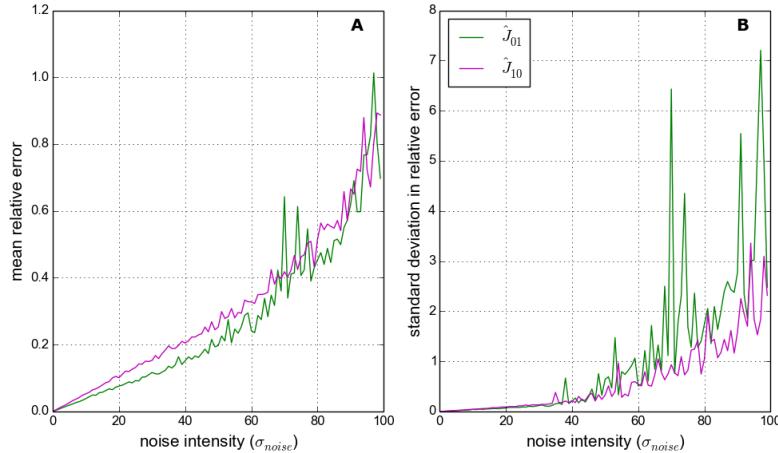


Figure 6.9: Nonsense. 1000 samples used.

### 6.3.2 Holling model

### 6.3.3 Range sampling

## 6.4 TEMP : other results

This section shows some plots which I was not planning to put into the thesis but are worth discussing..

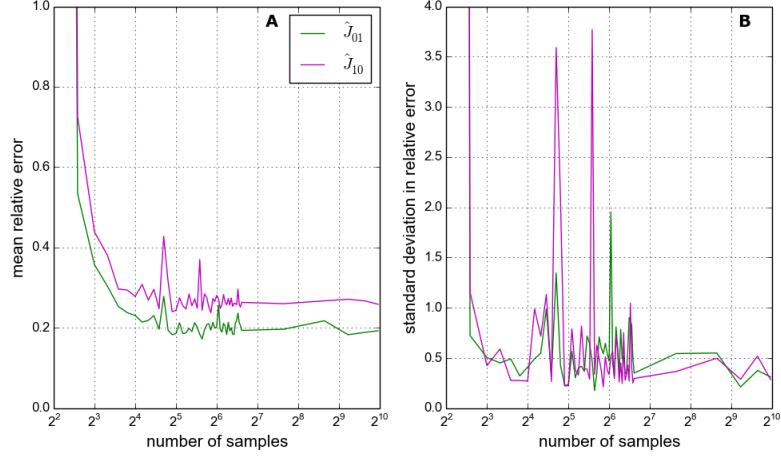


Figure 6.10: Nonsense. Noise is 50.

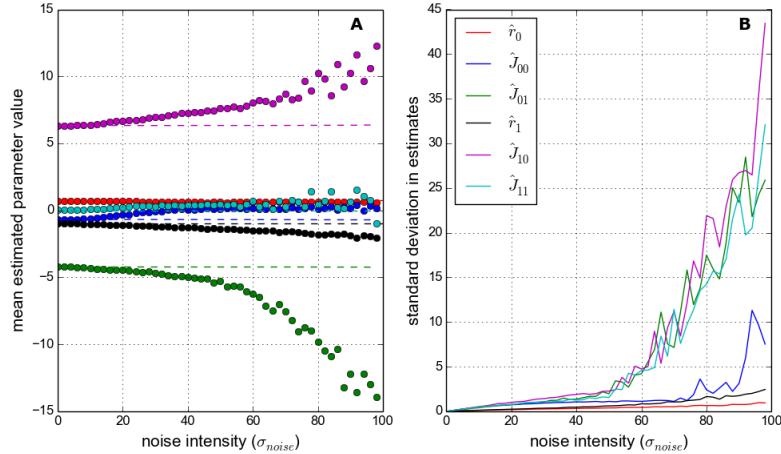


Figure 6.11: Nsamples is 10,000.

## 6.5 Application to IBM (optional)

In this section we apply the methodology for inferring species interactions to the IBM simulations. In the previous section we have seen that the method works well when fitting the GLV to two species predator-prey dynamics simulated with ODE models. In the case that the dynamics is governed by the Lotkva-Volterra equations, and in the absence of noise, fitting the GLV model produces true estimates of the underlying parameters which include the inter-specific interaction strengths. The estimates require relatively few samples in order to achieve high accuracy, and which converge on the true parameter values as the sampling intensity increases. However as noise is added to the simulations the accuracy of the estimates decreases. In particular we found that, in the presence of noise, the estimates do not converge on the true parameter values -

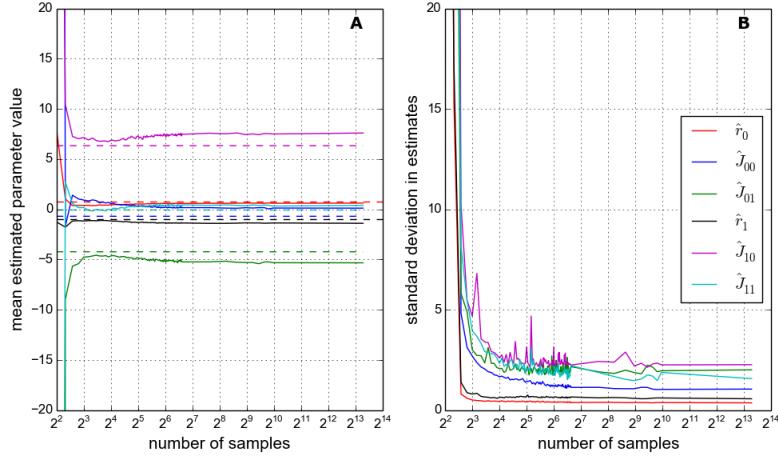


Figure 6.12: Noise is 50.

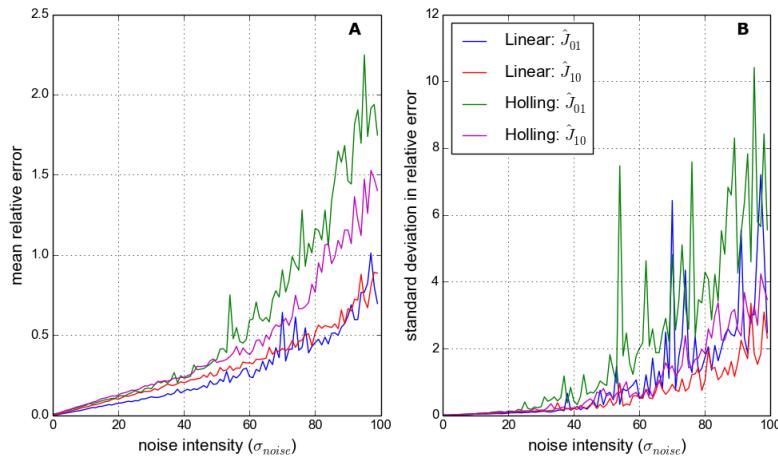


Figure 6.13: Noise is 50.

that is, noise introduces systematic error in to the estimates. In the case that the dynamics are governed by the Holling II model matters are complicated - but in general we find that the GLV can approximately capture the strength of species interactions (and also the dynamics?), provided there is not too much noise, and the FR is not too non-linear<sup>1</sup>.

Can the IBM dynamics by approximated by the GLV model? The hypothesis is that is can. Argue this..and refer to previous mention of LV type dynamics. Refer forward to testing linearity of FR. Exponential growth and decay, linear FR. However there are certain problem/factors that may hinder this approach/represent a departure from GLV dynamics...Noise, spatial effects, bioenergetic model - time delay? Immigration (one component of noise).

The issue of noise is important since, as we have shown in chapter REF, there is a strong

<sup>1</sup>This is all need to be demonstrated - WORK TO DO!

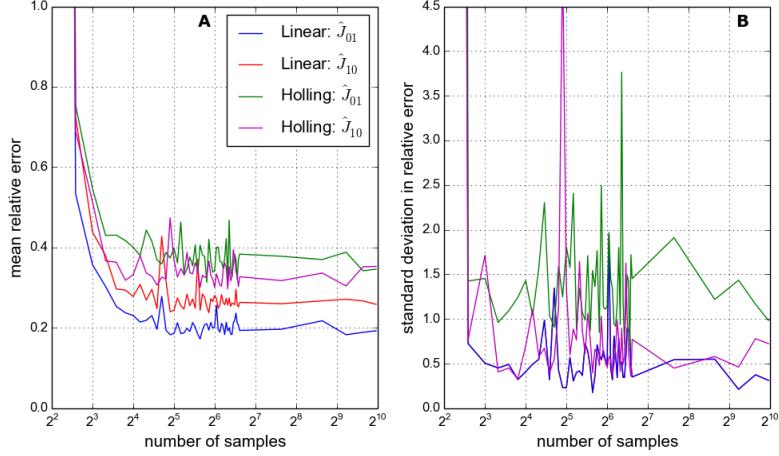


Figure 6.14: Noise is 50.

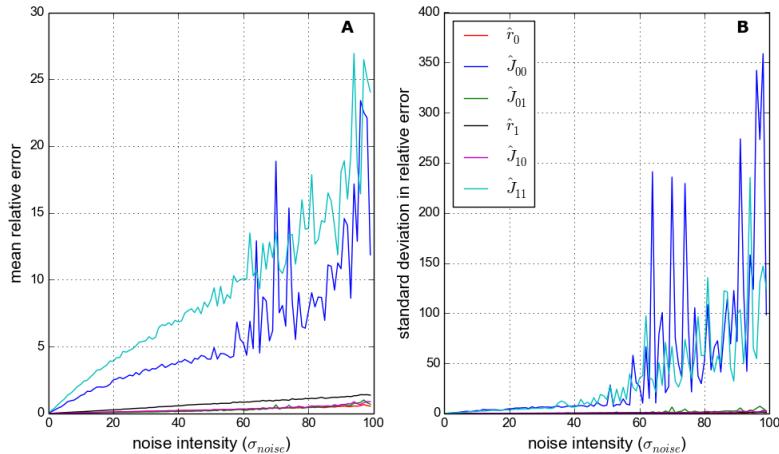


Figure 6.15: Nonsense. 1000 samples used.

stochastic component to the IBM simulations.

Modelling individuals, not biomass or energy. Is this problematic? Set herb-frac to 1. Other issues?

### 6.5.1 Testing functional response (and intrinsic growth functions)

Here we demonstrate the linearity of the predator functional response for animal predators in the second and third trophic levels (using 2sp and 3sp chain) - demonstrate that there is a slight difference. But basically linear. Other issues that may arise - high noise and low abundances create error.

We also conduct an experiment to test the intrinsic growth and death rates of plant and animal species - by setting herb-frac=0.0. Demonstrate they are well approximated by exponential.

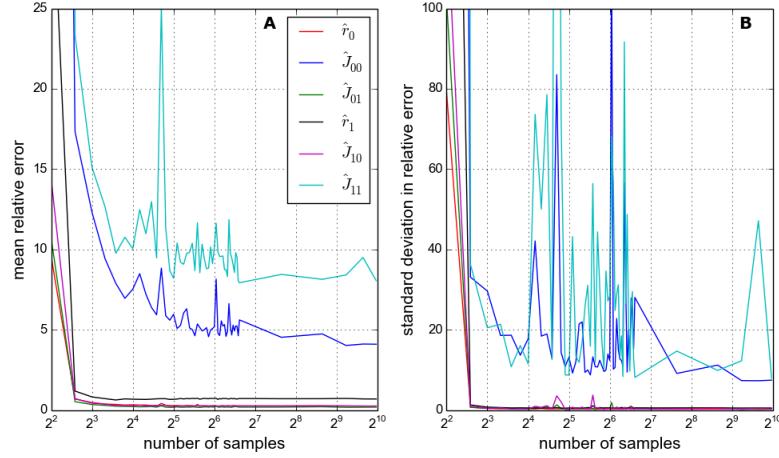


Figure 6.16: Nonsense. Noise is 50.

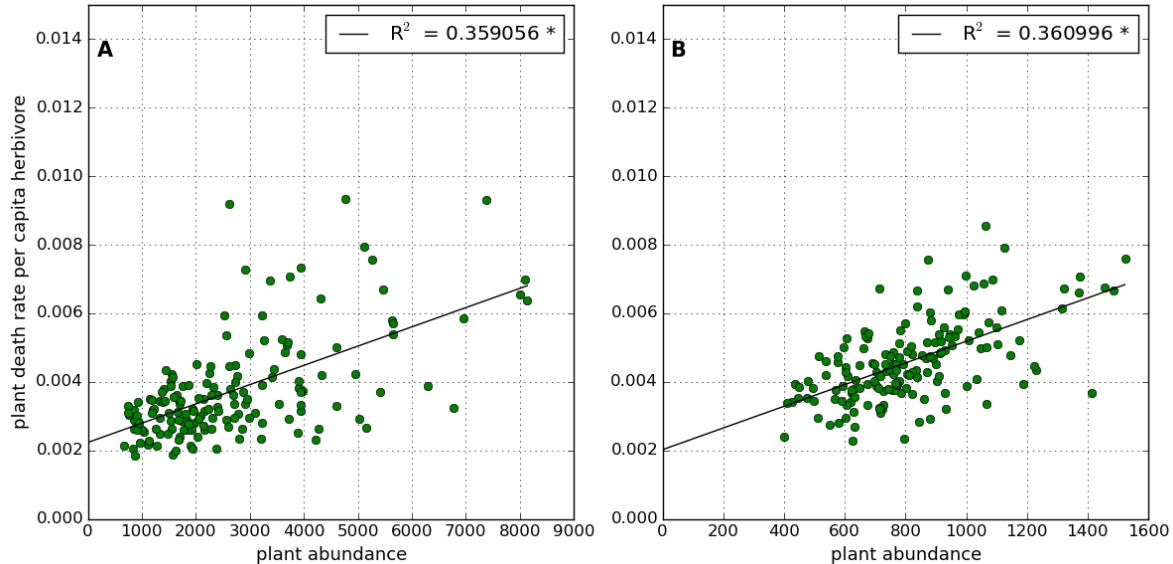


Figure 6.17: Plant functional response

In the absence of immigration.

Carrying capacity: is there evidence for density dependent birth/death? - use this fact in later analysis. Also discuss that the carrying capacity will vary with the number of species, not just a single species thing (non-pairwise interactions in competition for space - argghh!)

### 6.5.2 2 Species IBM model

IMPORTANT: carrying capacity depends on other species...introduce a new term into the model and test it?

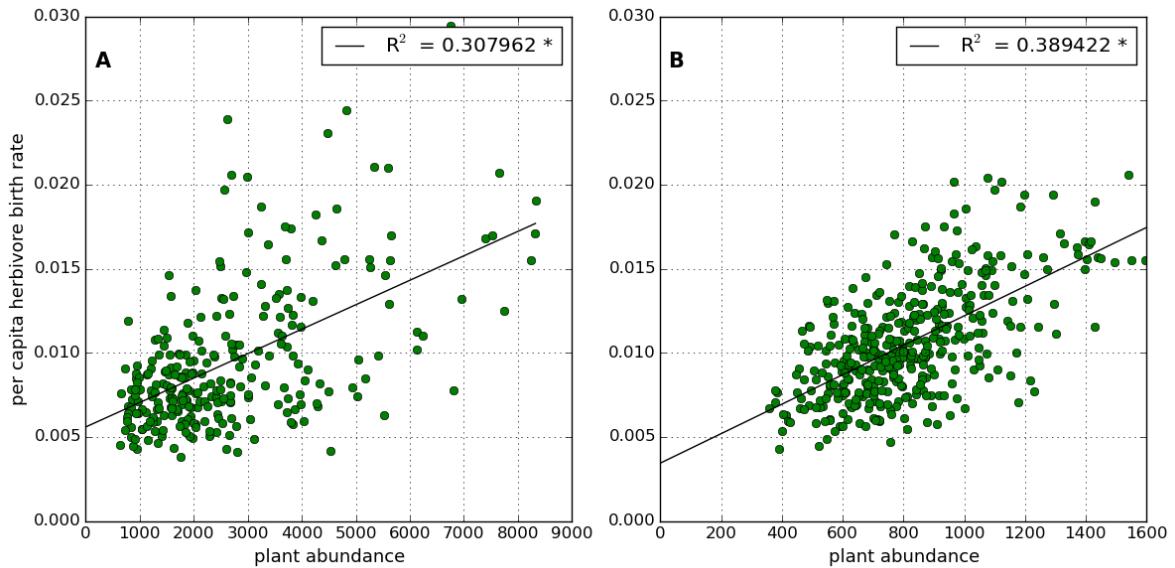


Figure 6.18: Herbivore functional response

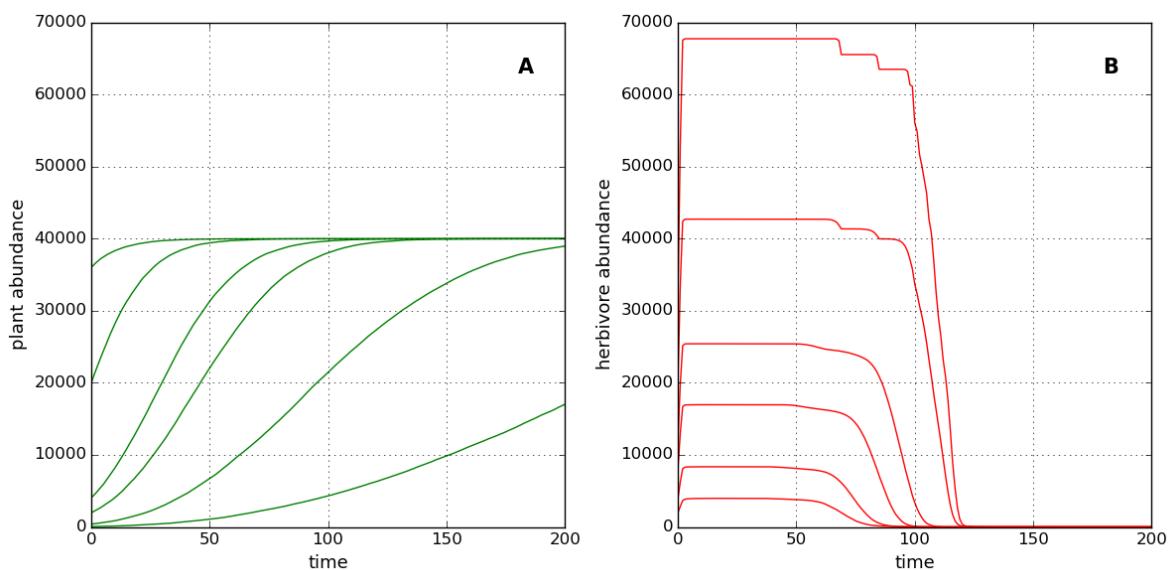


Figure 6.19: Intrinsic growth/mortality.

Define the model and what the inferred parameters represent:

- $J_{01}$ : per-capita rate consumption of the prey
- $J_{10}$ : per-capita rate reproduction of predator, due to consumption of prey. Not as well defined. But only source of predator births? Numerical response! (get REF)
- $J_{00}$ : intra-specific regulation of prey growth - see carrying capacity experiment.
- $J_{11}$ : intra-specific regulation of predator mortality? Check this. Expect zero? Or expect high number of predators means more reproduction because easier to find partner, therefore reduce mortality? Or increase birth rate. Not clear. Again SEE EXPERIMENT.
- $r_0$ : prey intrinsic growth rate. Estimate from exp?
- $r_1$ : predator intrinsic mortality rate. Estimate from exp?

So we know what values to expect, or at least the signs. We can evaluate the model fit by comparing the values of these estimates with birth/death rates from the simulations. Although not totally fair. We can also simulate GLV with the inferred parameters - does it match. Is the equilibrium the same? And check the error function of the fit. We show all this in the results section below.

**Dynamics of the model** we two species is a new thing. Here we show that with the default parameters we get relaxation-type oscillations. This is interesting, we try fitting to these in the two species case. But they become problematic for larger number of species. Argue why. Therefore we increase the reproduction rate (refer to previous chapter), which creates more WORD dynamics. We also fit to these.

PLOT: DYNAMICS UNDER IR, RR AND HL.

Also show dependence on immigration (show that it is low and what happens if it is high), concede that this is a limitation.

### 6.5.3 Extend methodology (3, 4 and 5 species)

MULTI-SPECIES DENSITY DEPENDENCE?!

This does not require much since we presented a general framework previously.

Show 3 species dynamics with the two different RR. Conclude which is better.

### 6.5.4 Results

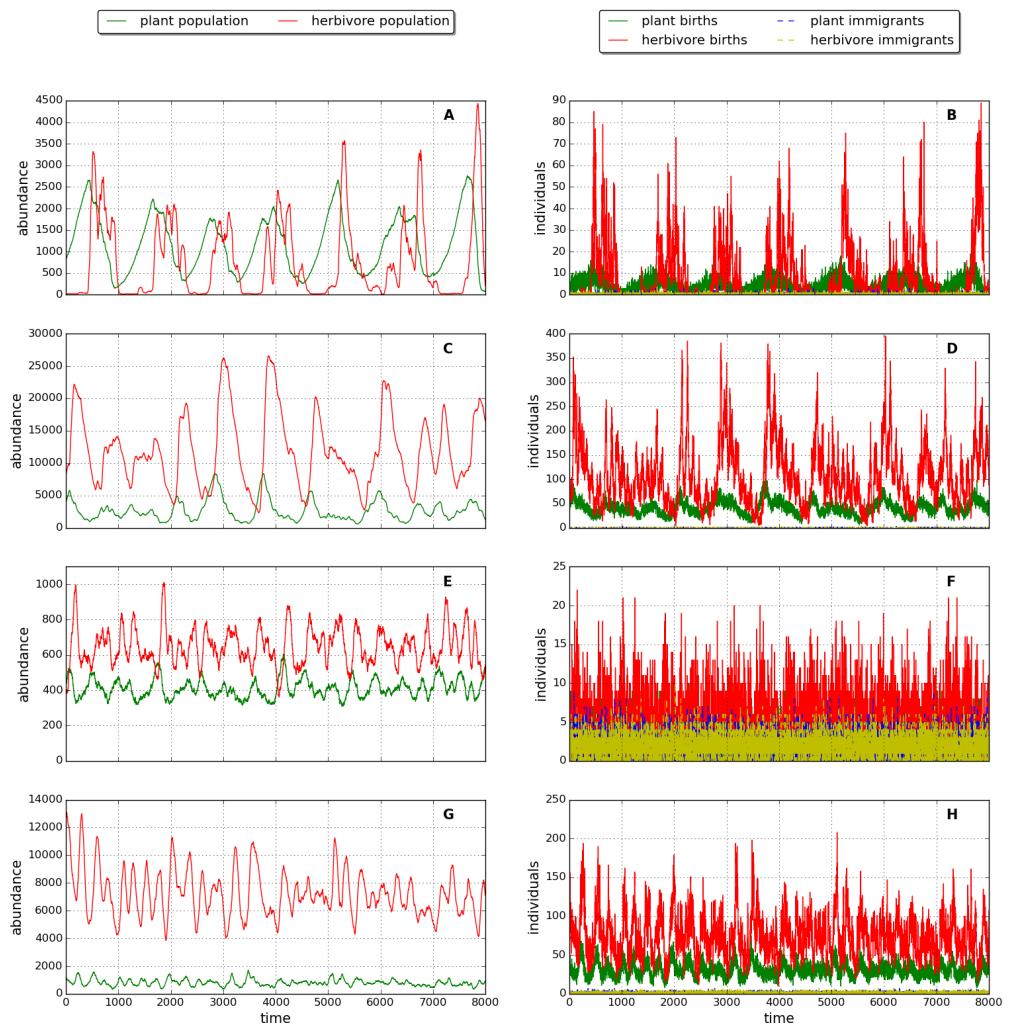


Figure 6.20: 2 species dynamics

### 6.5.5 2 species

Here we compare the results of two species.

For a single IR we look at convergence of all 6 parameters (over the ensemble) - correct signs, correct magnitudes?

Maybe repeat for other IRs and HL.

We then show rate estimates as time series and introduce quality metric for this<sup>2</sup>.

Demonstrate the quality decreases with IR and HL (box plot?). And how estimated parameters respond to the two HL scenarios (refer to previous findings). Hopefully support!

---

<sup>2</sup>Still not sure about this

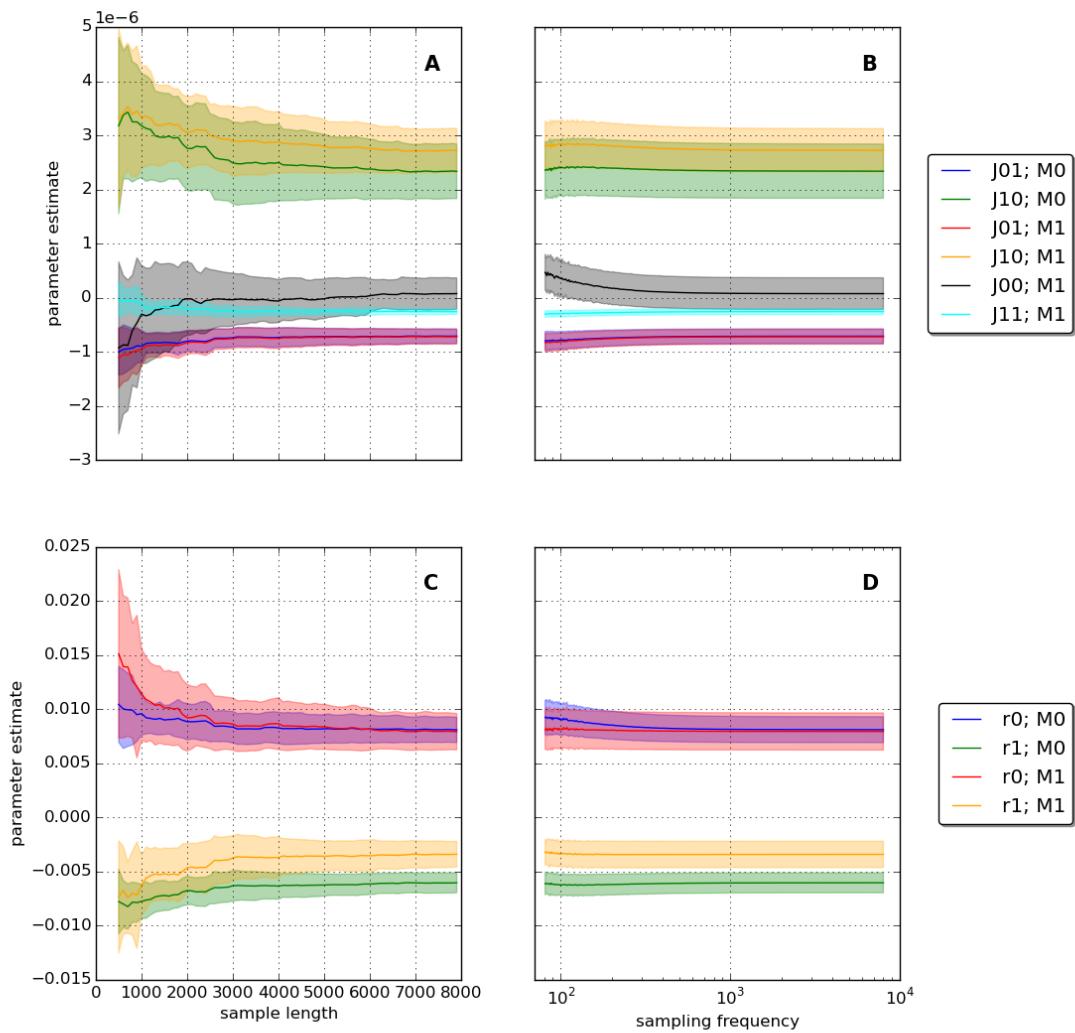


Figure 6.21: Convergence of estimates. 2 species.

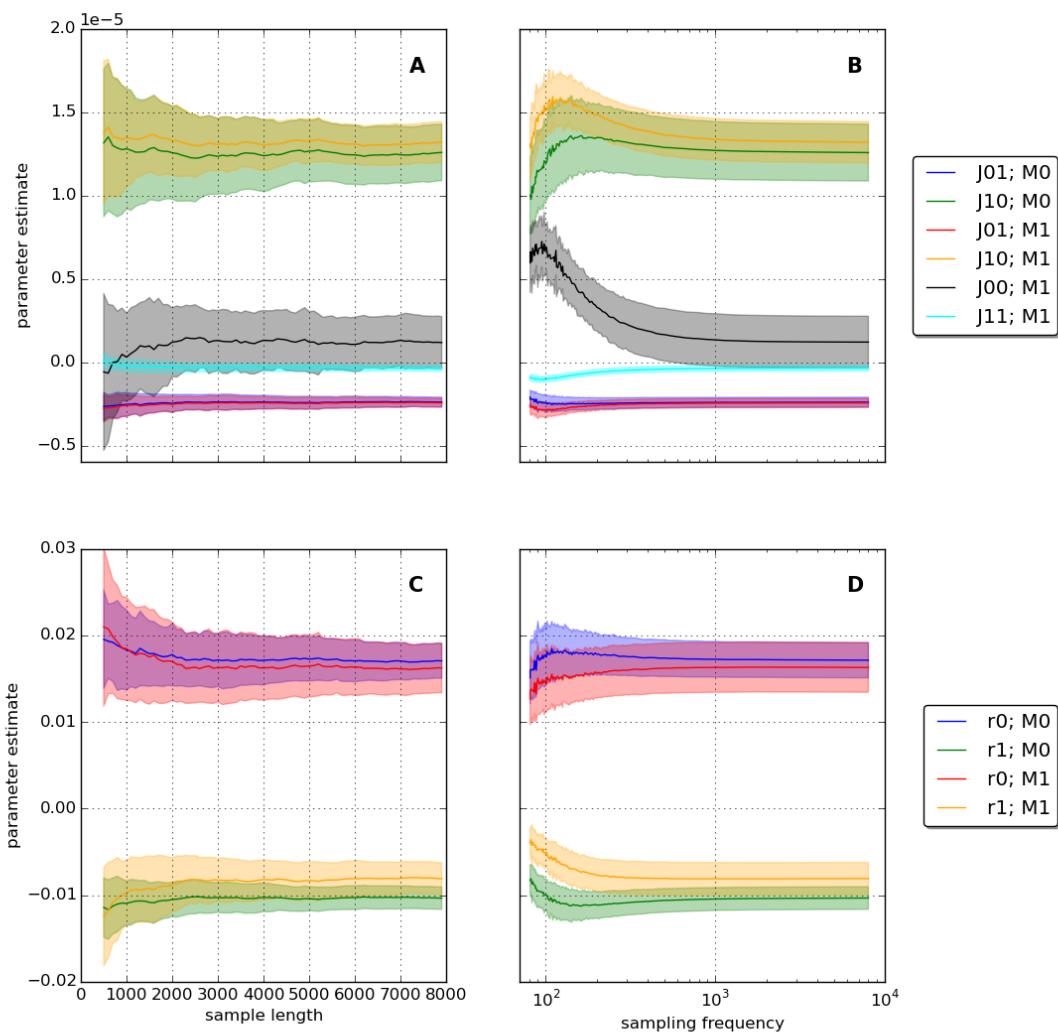


Figure 6.22: Convergence of estimates. 2 species.

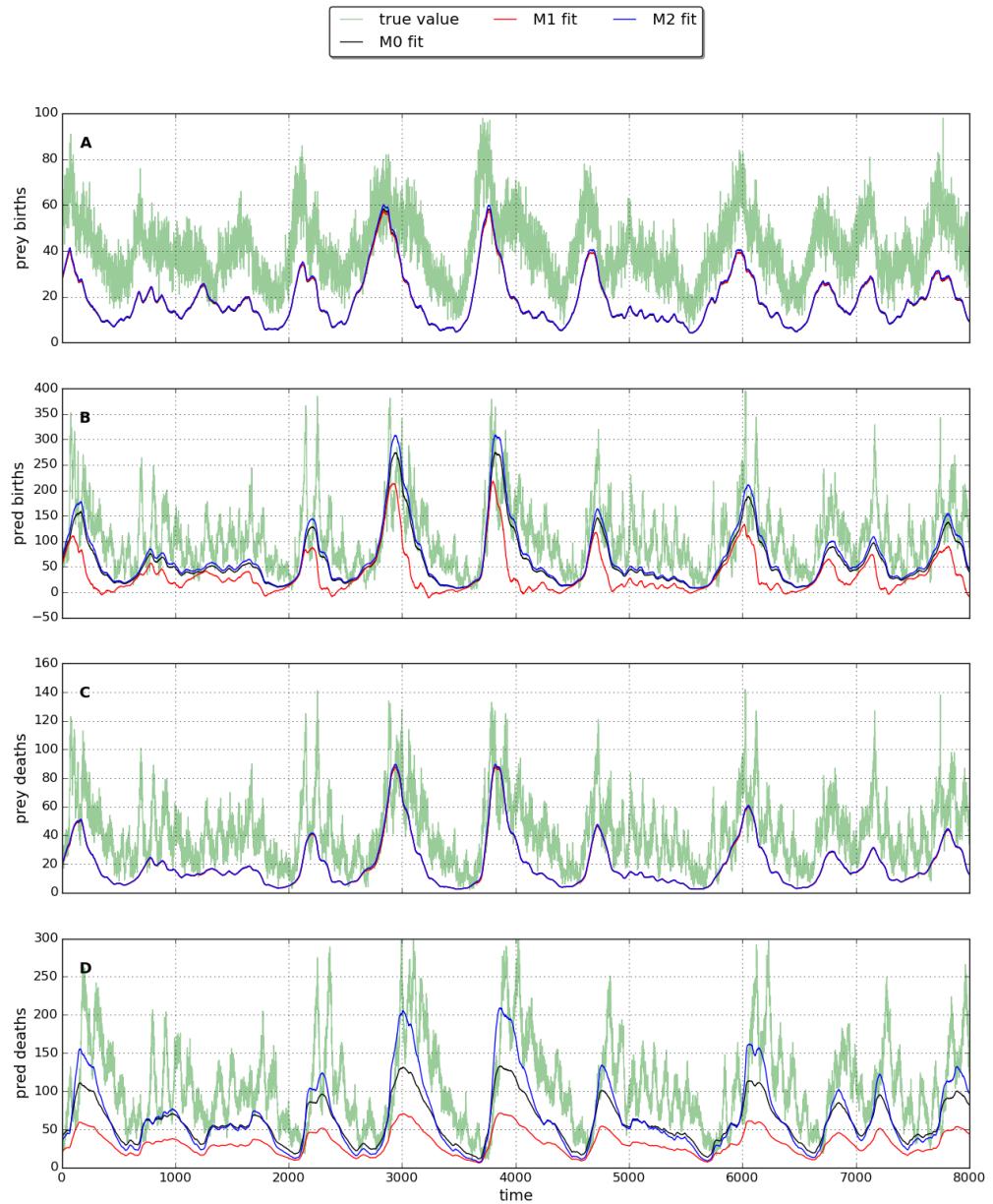


Figure 6.23: Predicted births/deaths. Low IR.

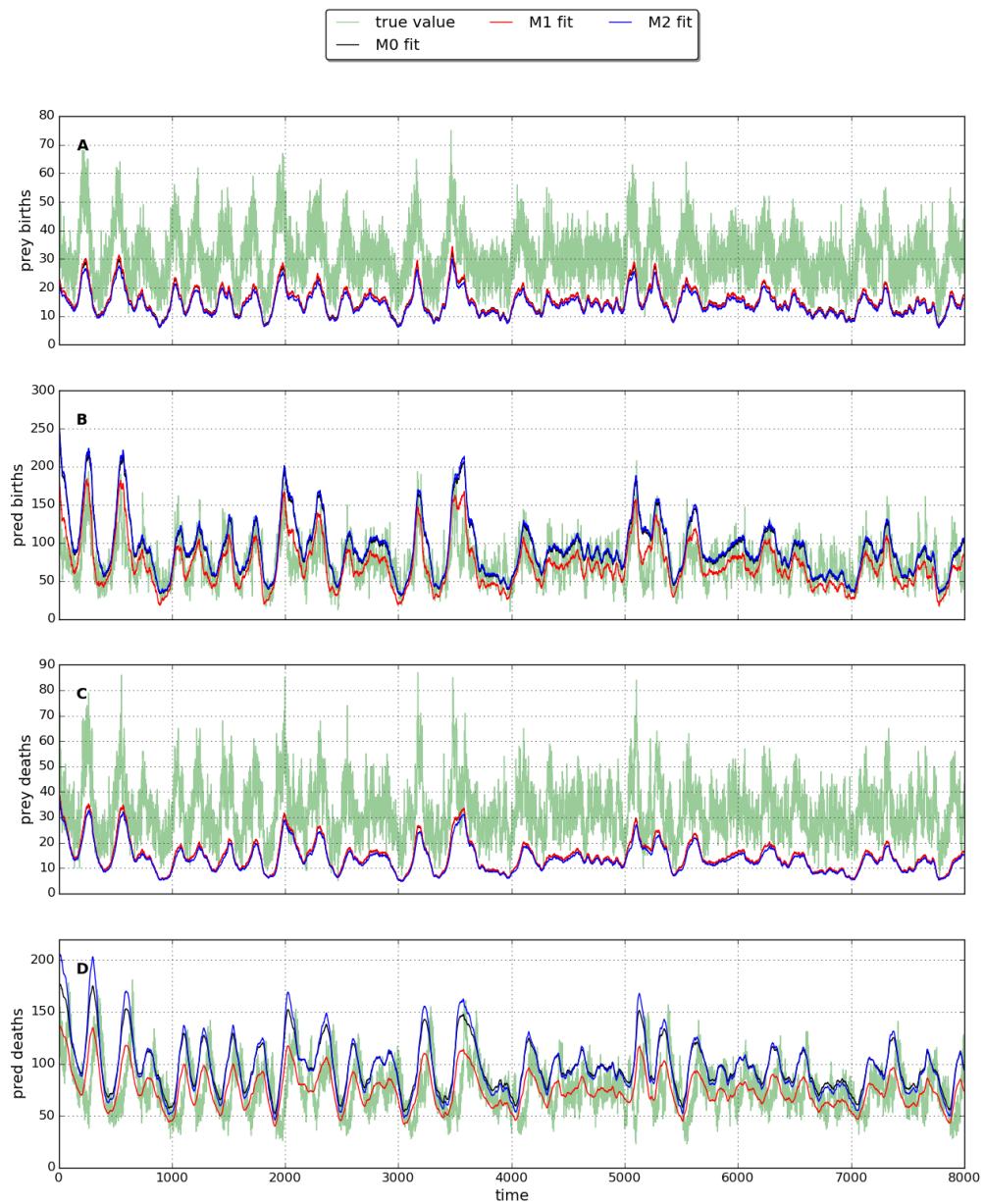


Figure 6.24: Predicted births/deaths. High IR.

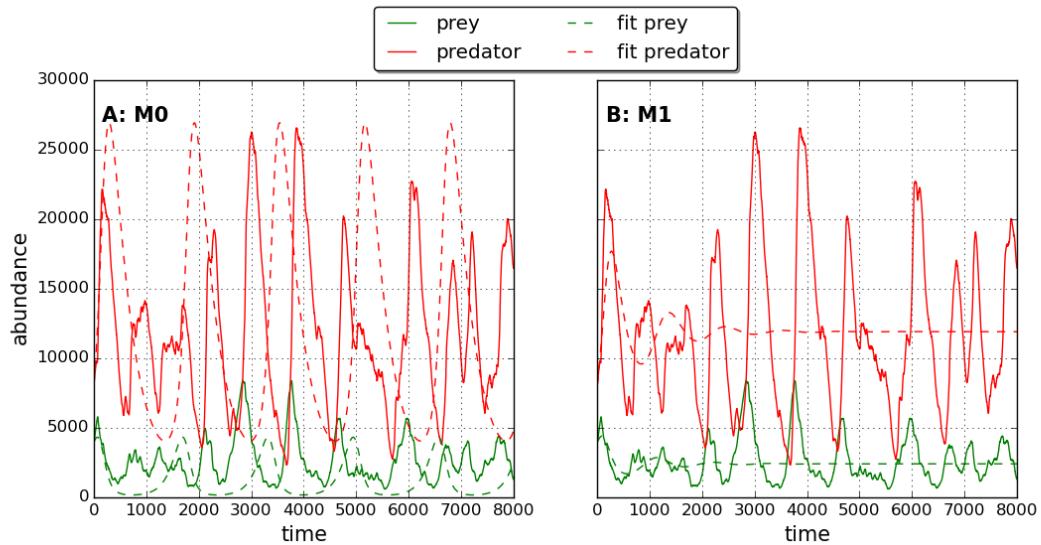


Figure 6.25: Fitted dynamics.

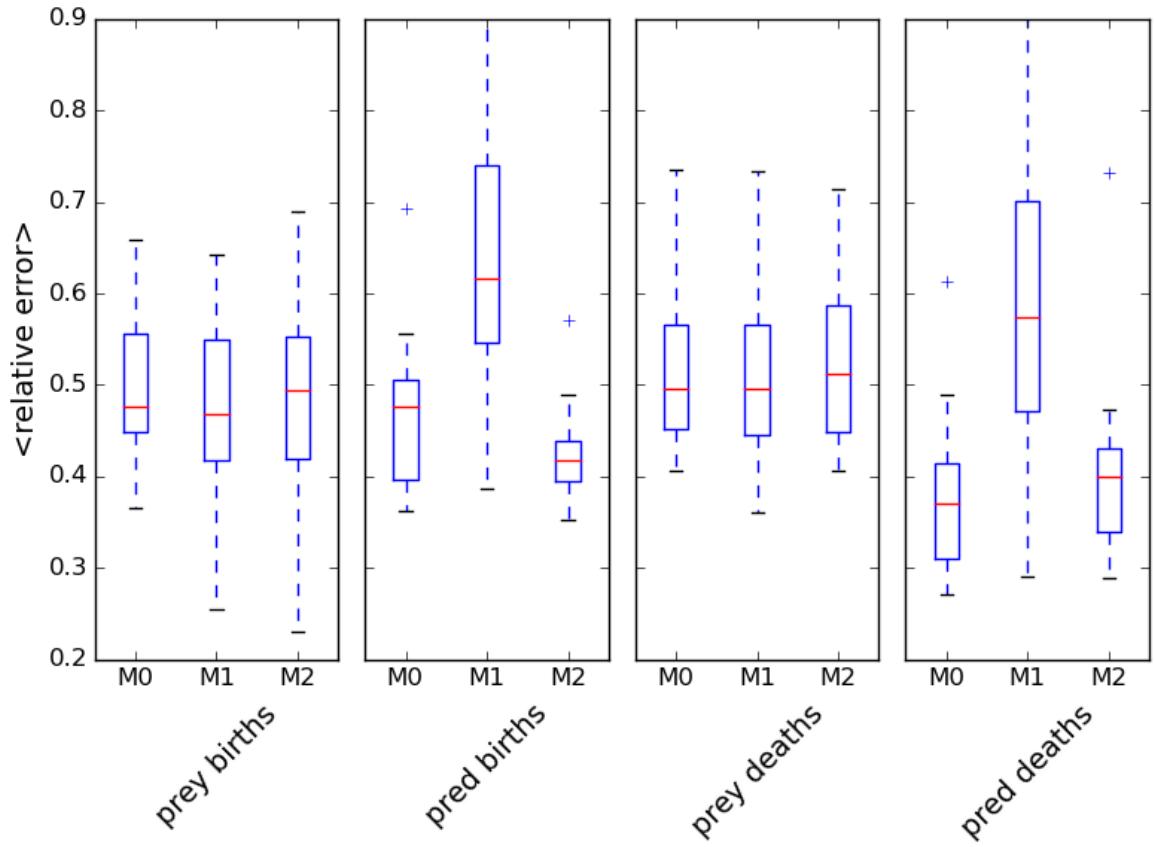


Figure 6.26: Quality. LI.

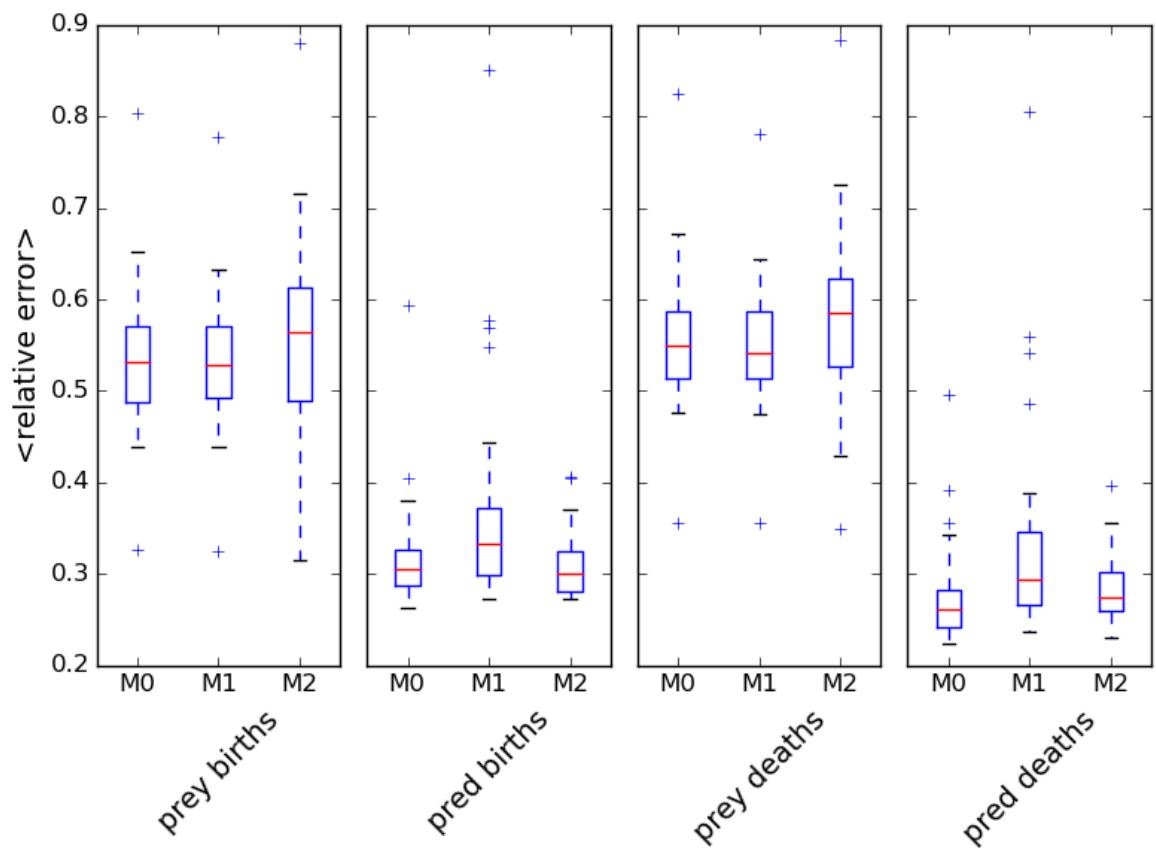


Figure 6.27: Fitted dynamics.

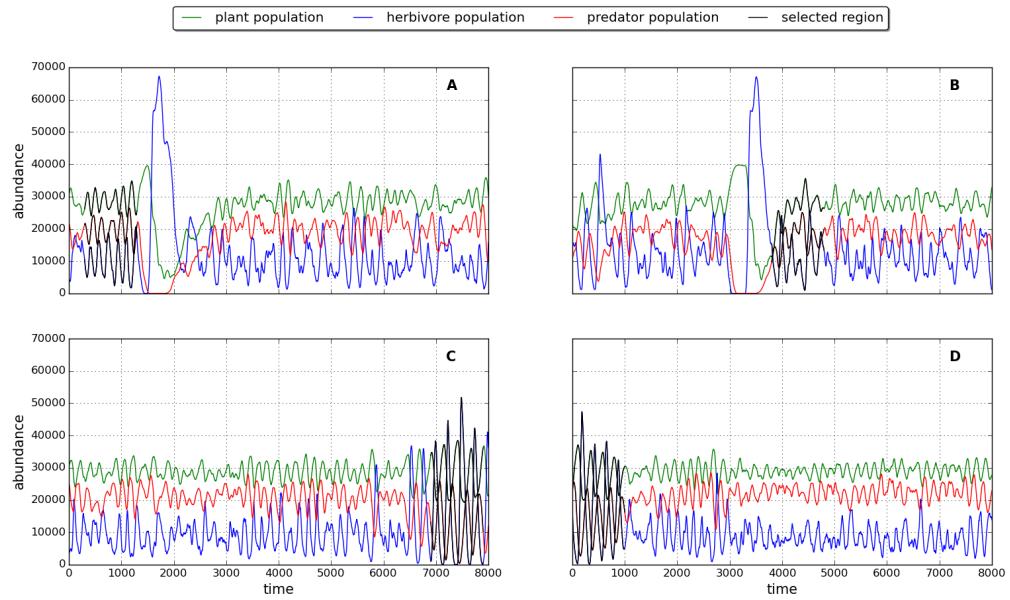


Figure 6.28: 3 species dynamics

### 6.5.6 3 species

The two species results suggest that intra-specific interactions contribute to predator deaths, whereas they contribute to prey births. This is problematic for rate estimation since we are starting from the position of not knowing which species are basal. For the purposes of calculating the rate we pretend that we know...This is in agreement with the Lotka-Volterra formulation.

The alternative convention is that intra-specific interactions are

Extra term does not improve the estimates.

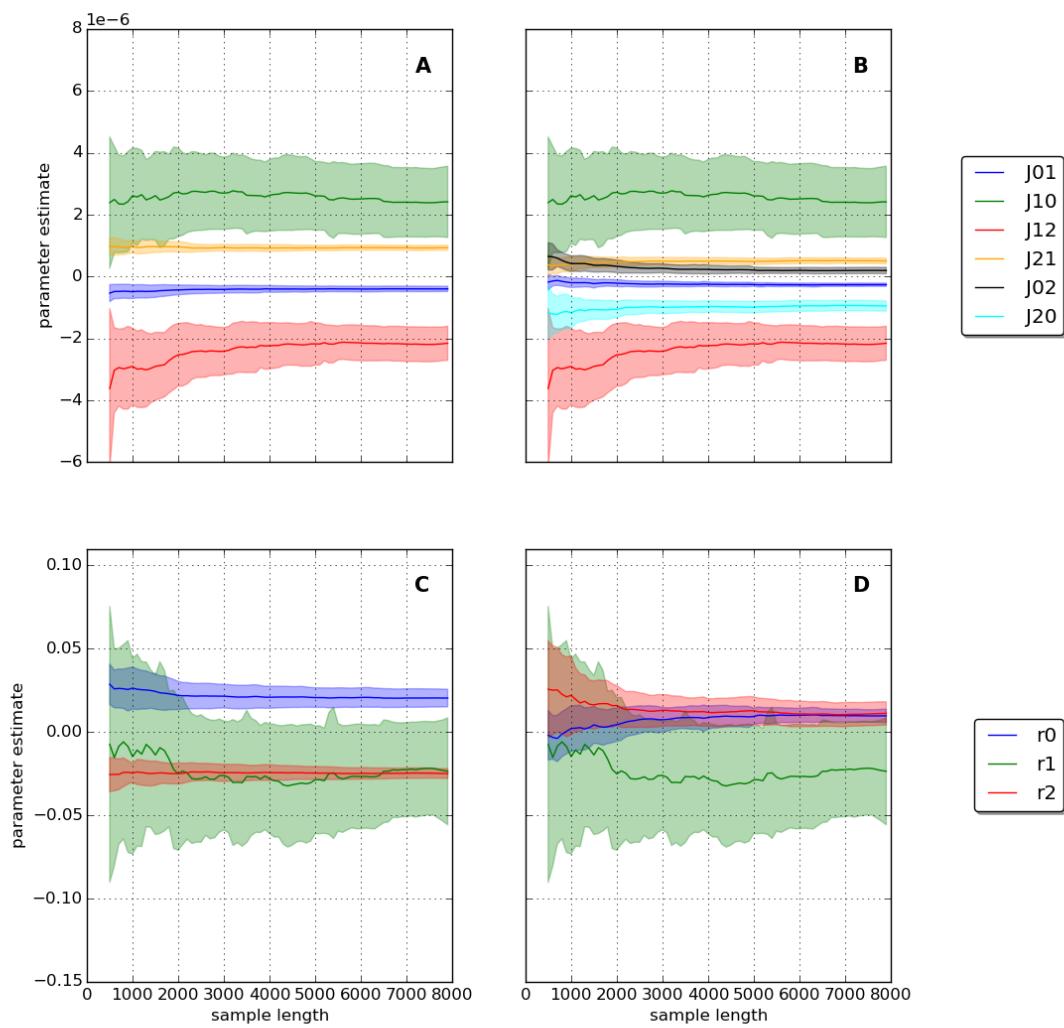


Figure 6.29: Convergence of estimates. 3 species.

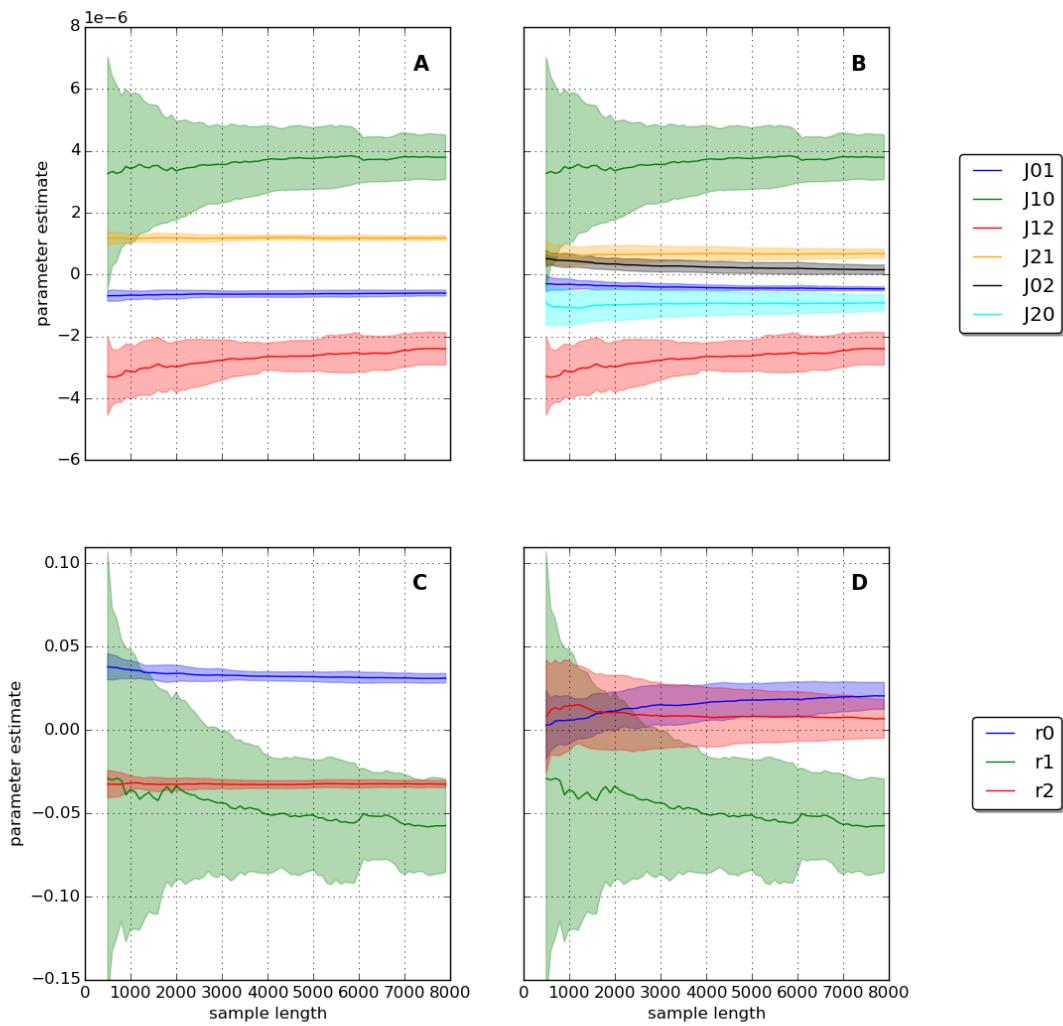


Figure 6.30: Convergence of estimates. 3 species.

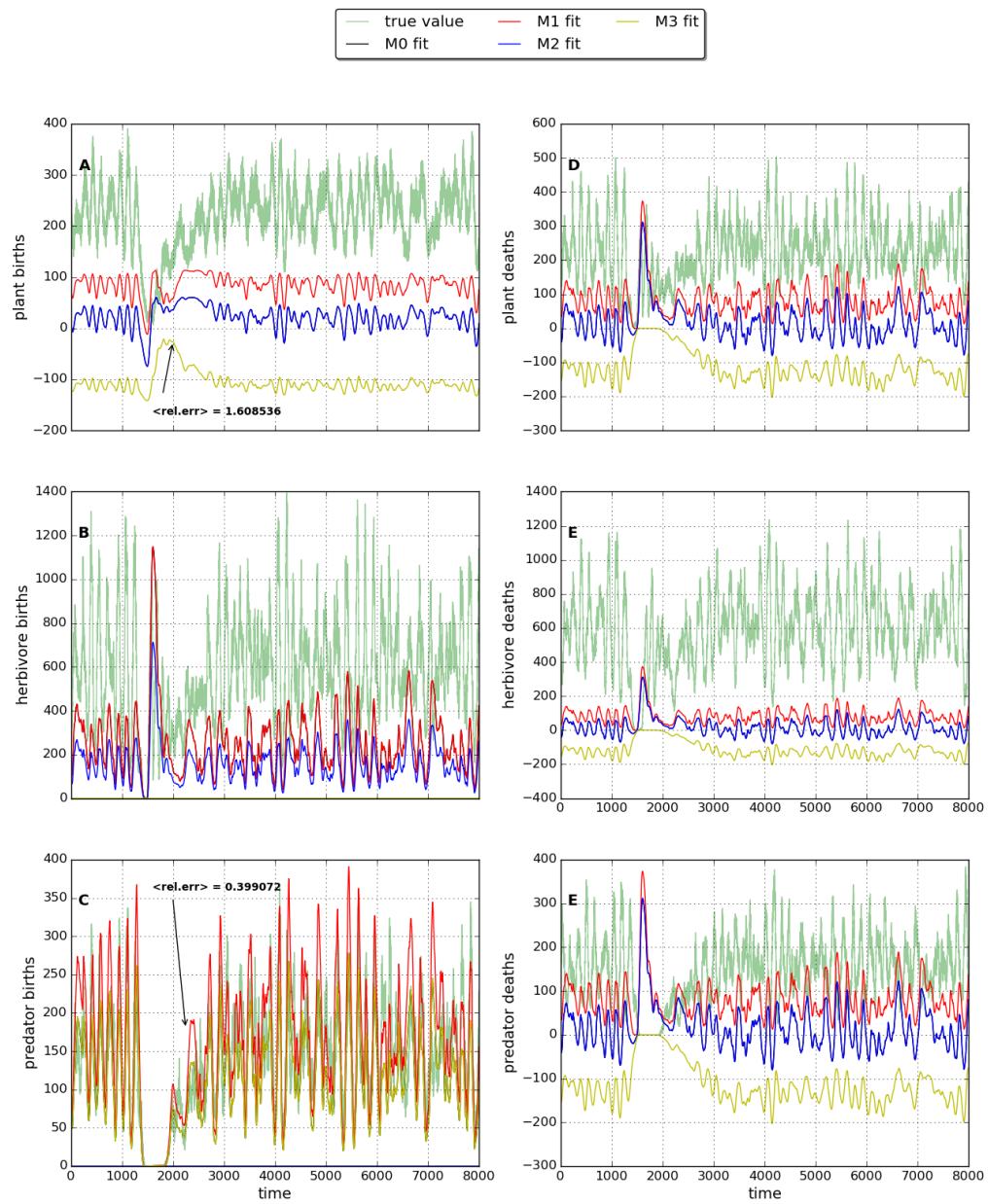


Figure 6.31: Predicted births/deaths. Low IR.

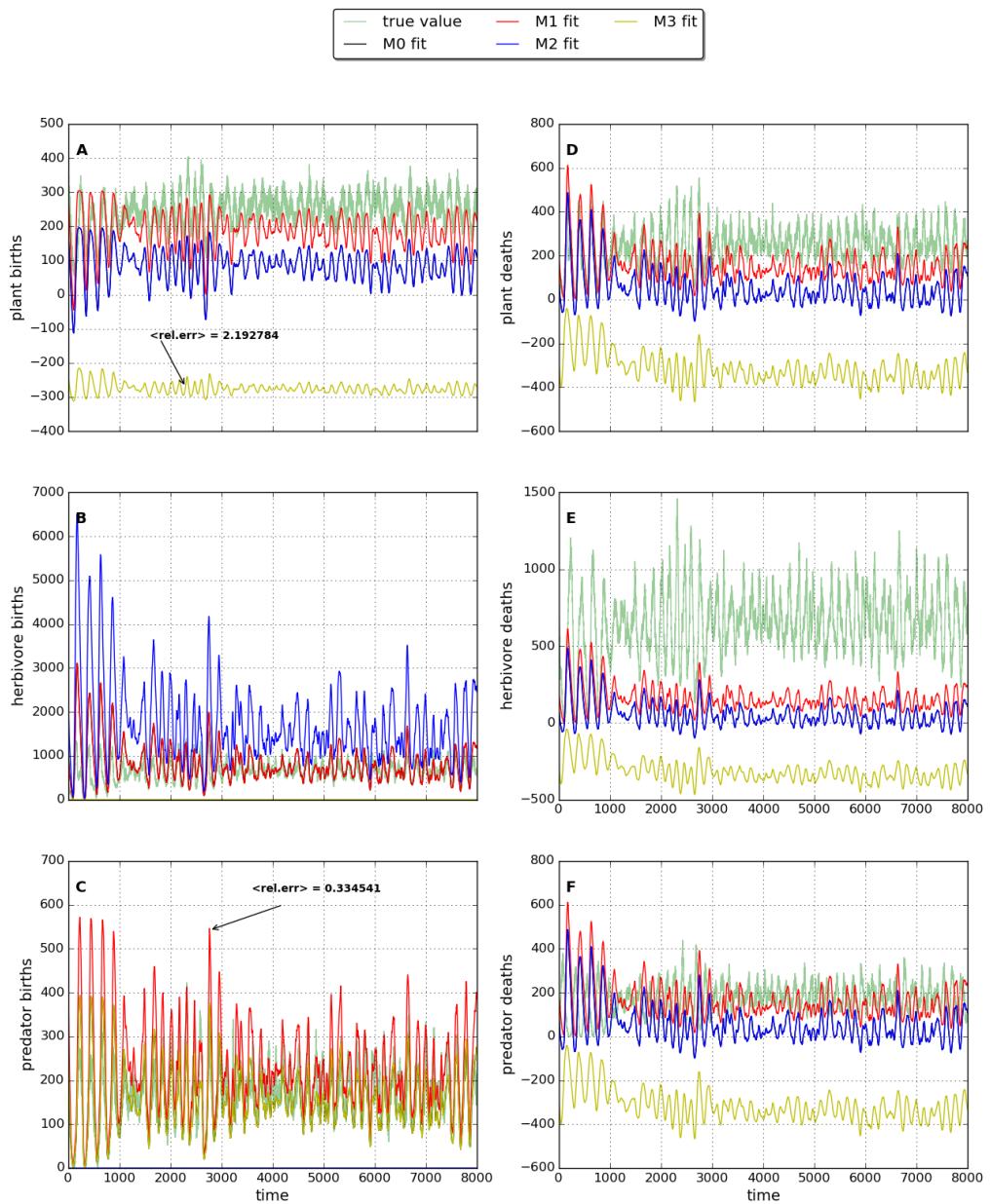


Figure 6.32: Predicted births/deaths. High IR.

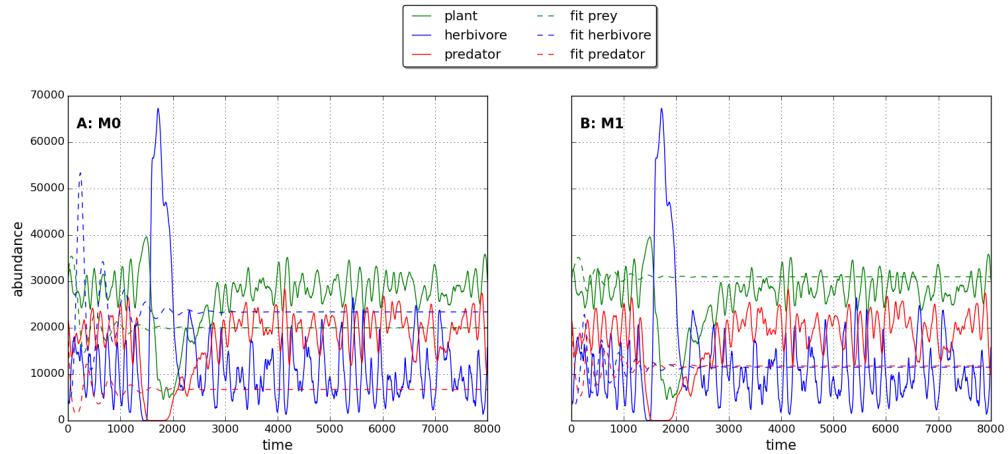


Figure 6.33: Fitted dynamics.

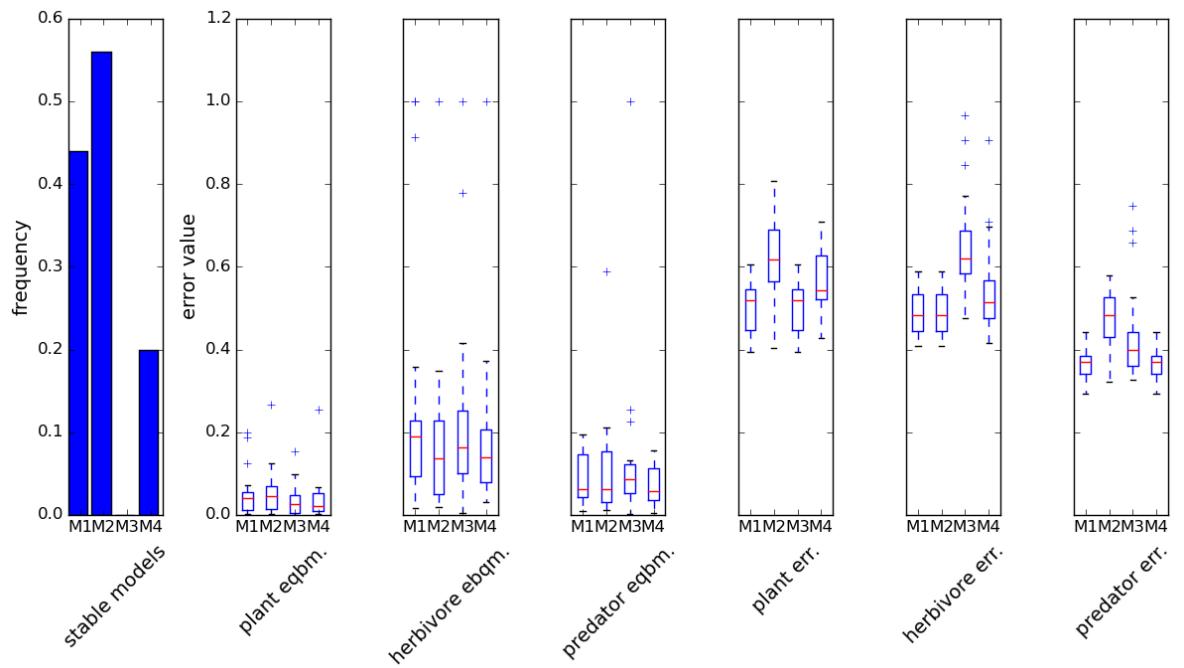


Figure 6.34: Stability and error

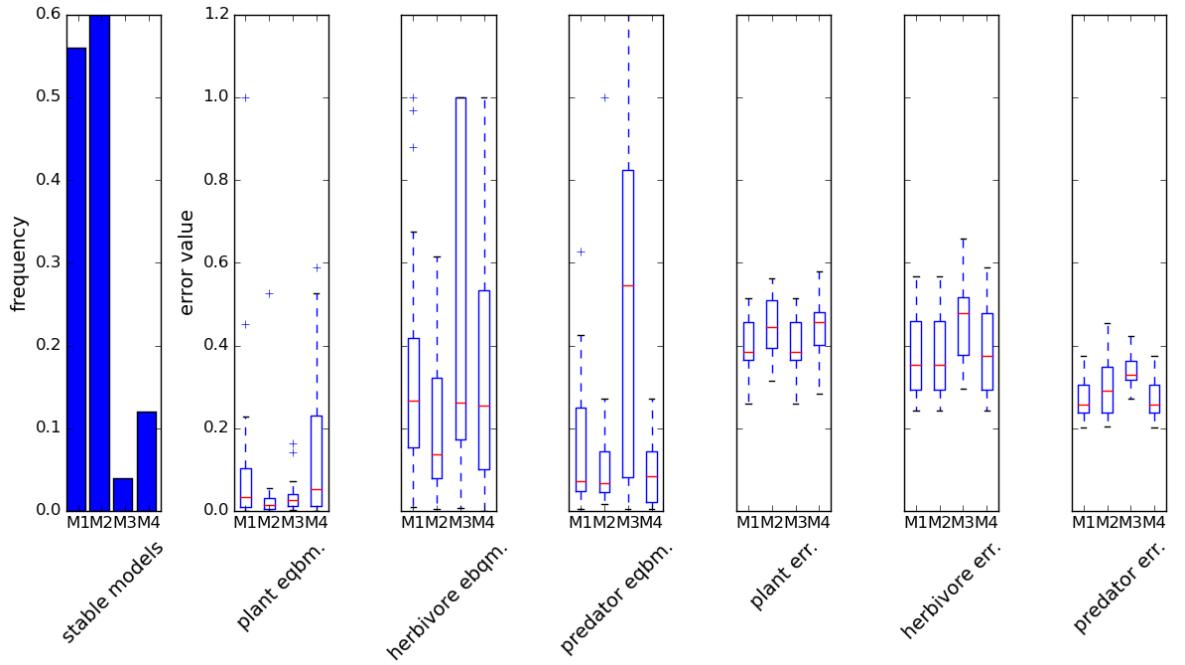


Figure 6.35: Stability and error

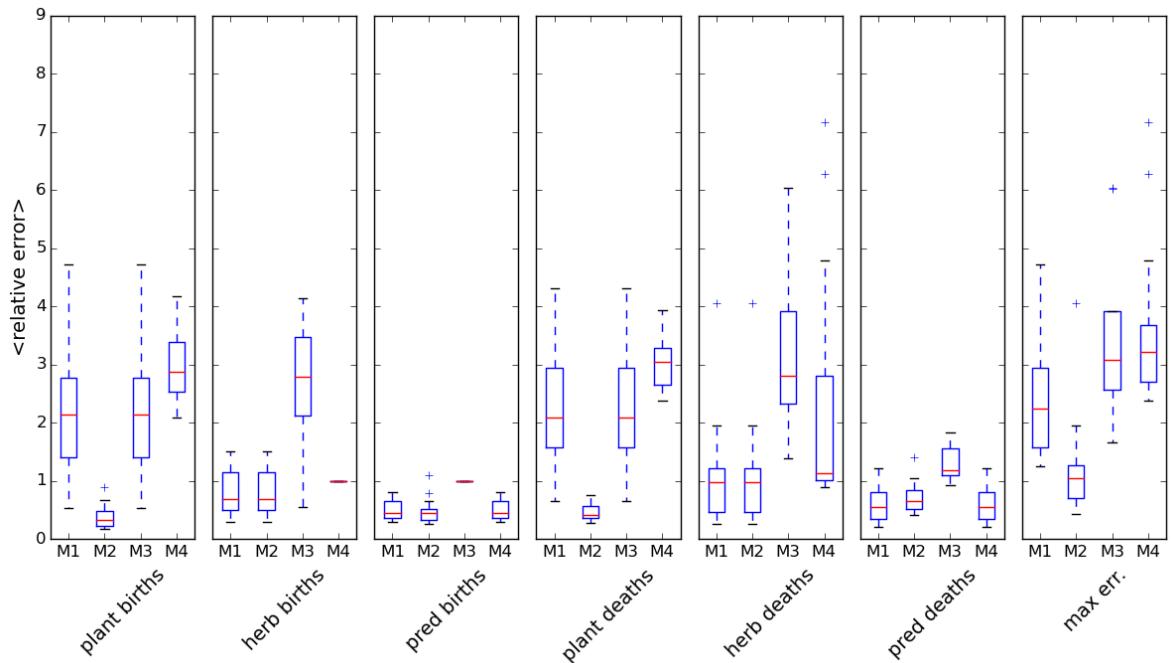


Figure 6.36: Quality 3sp LI

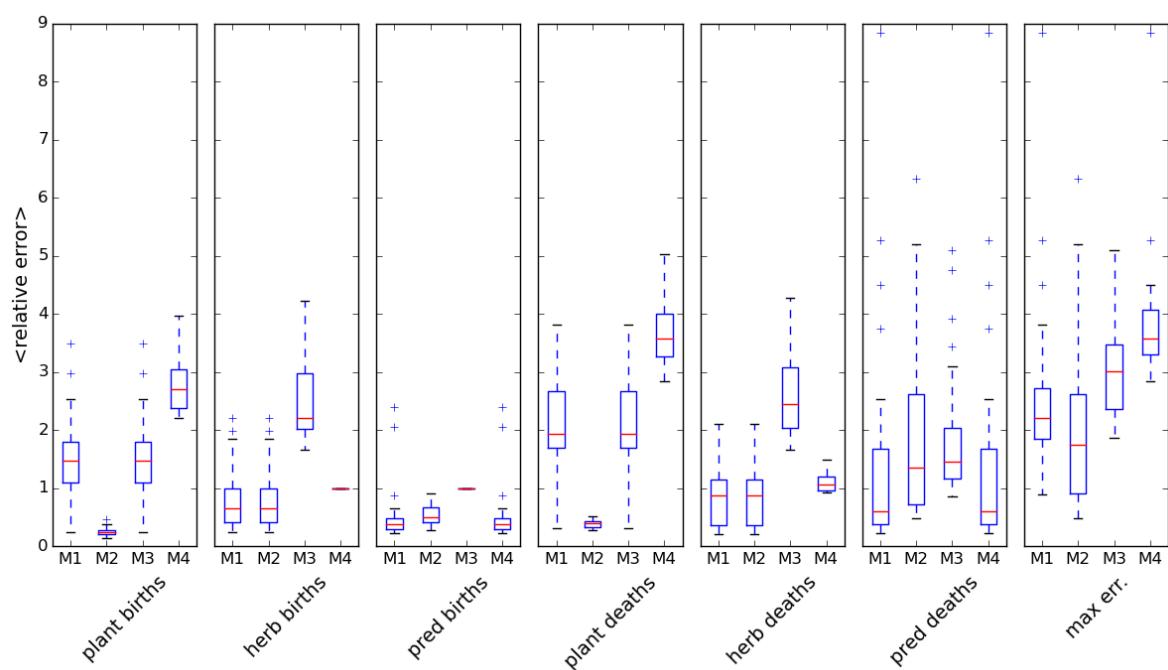


Figure 6.37: Quality 3sp HI

**6.5.7 5 species**

## 6.5. APPLICATION TO IBM (OPTIONAL)

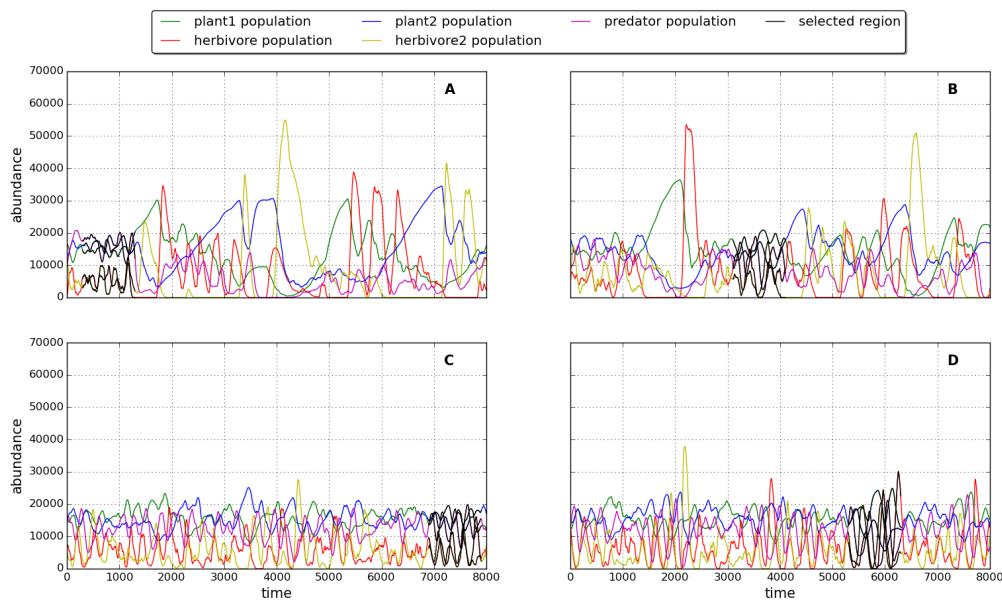


Figure 6.38: 5 species dynamics

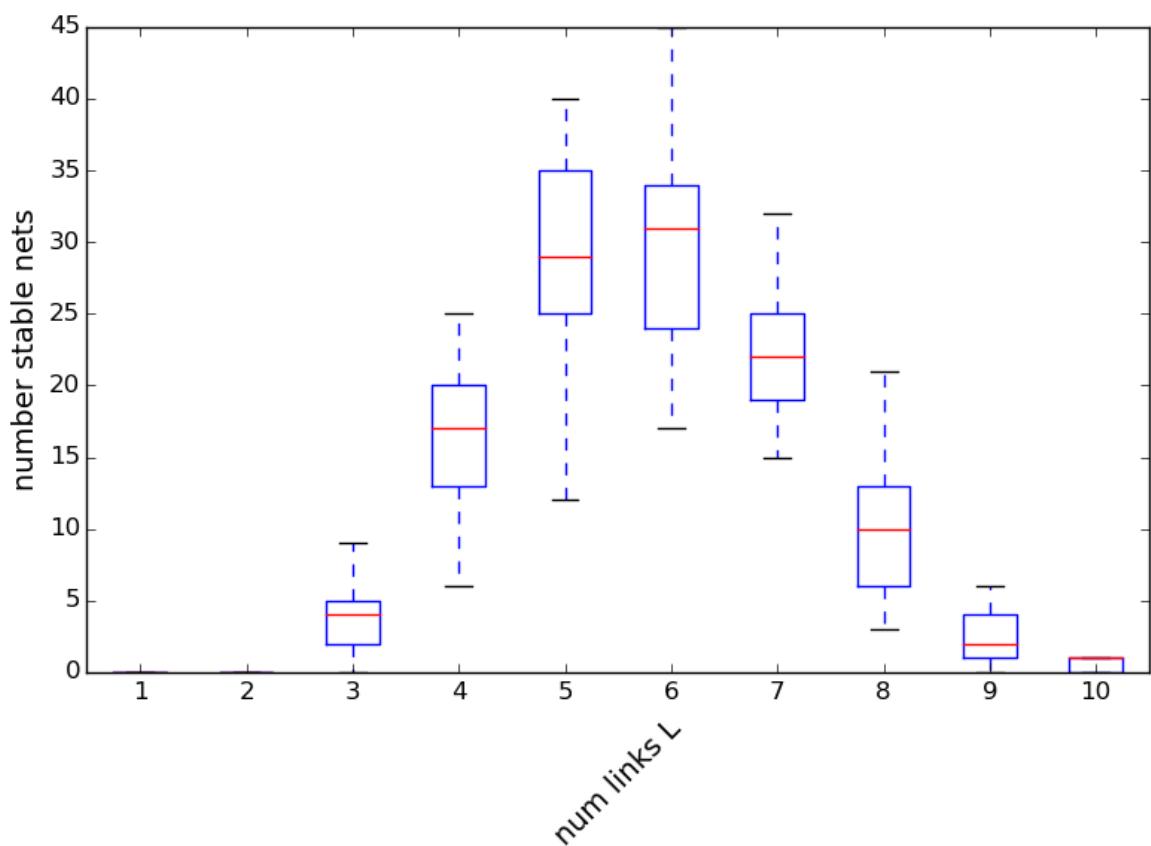


Figure 6.39: 5 species: number of stable models

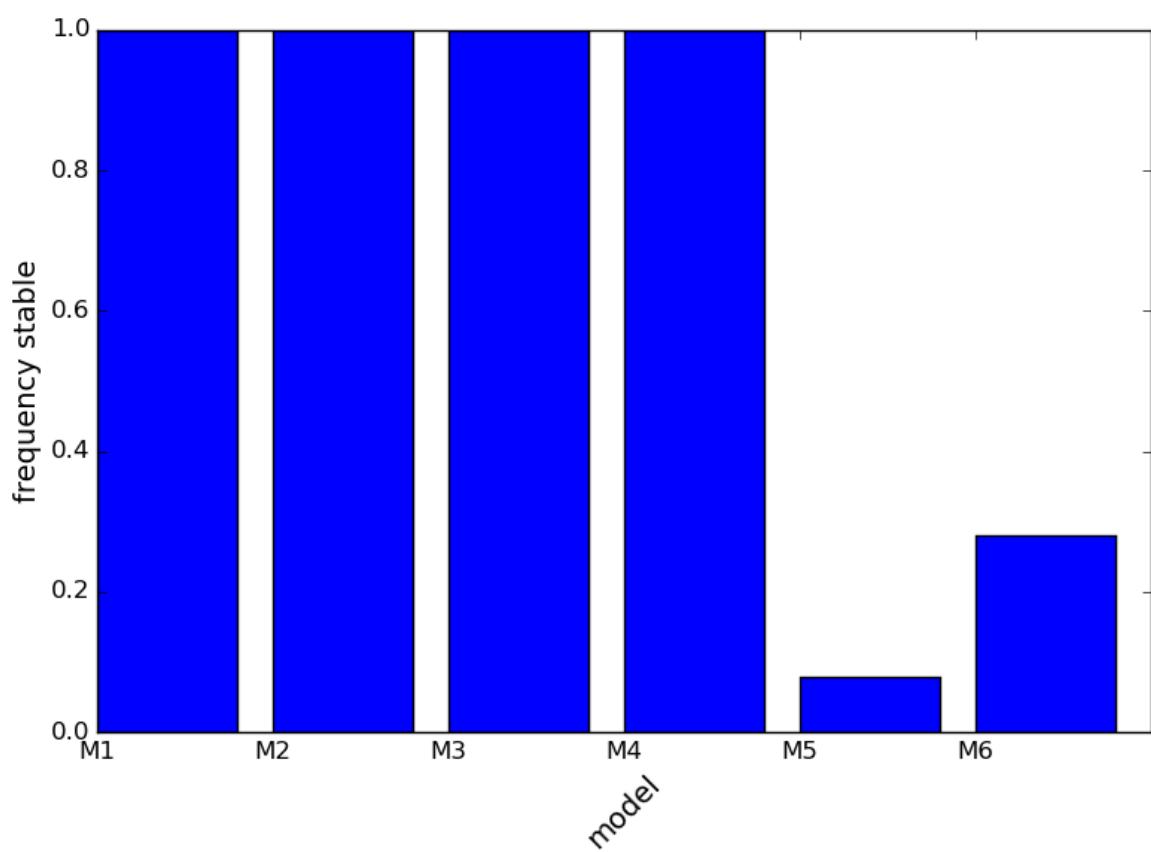


Figure 6.40: 5 species: stability of selected models

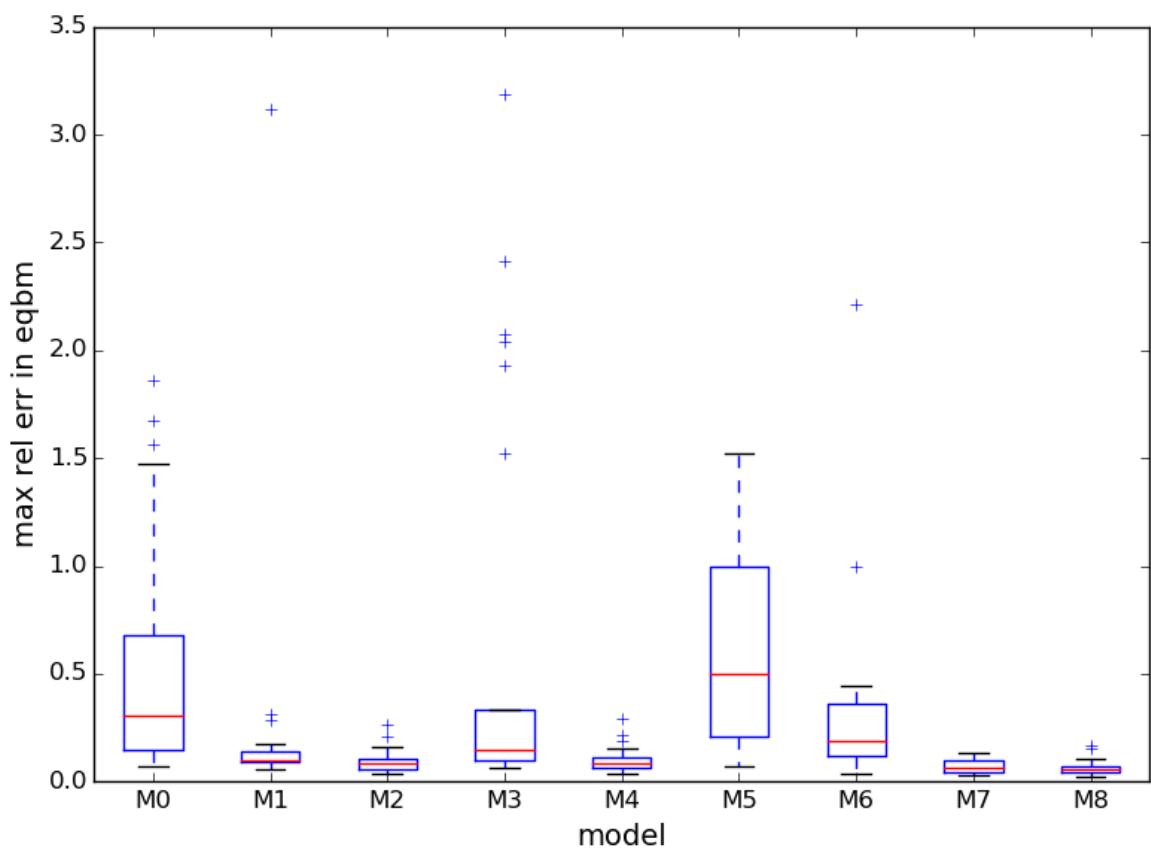


Figure 6.41: 5 species: error in equilibrium

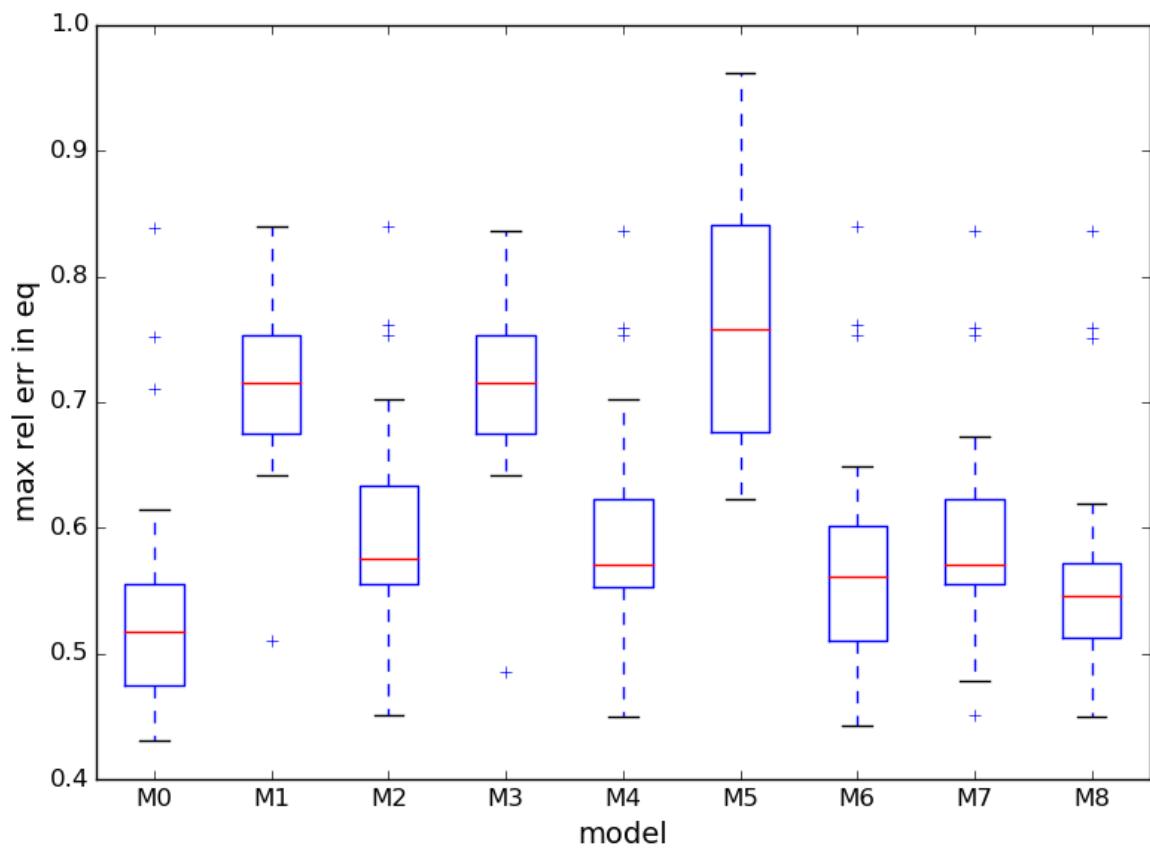


Figure 6.42: 5 species: error in gradient fit function

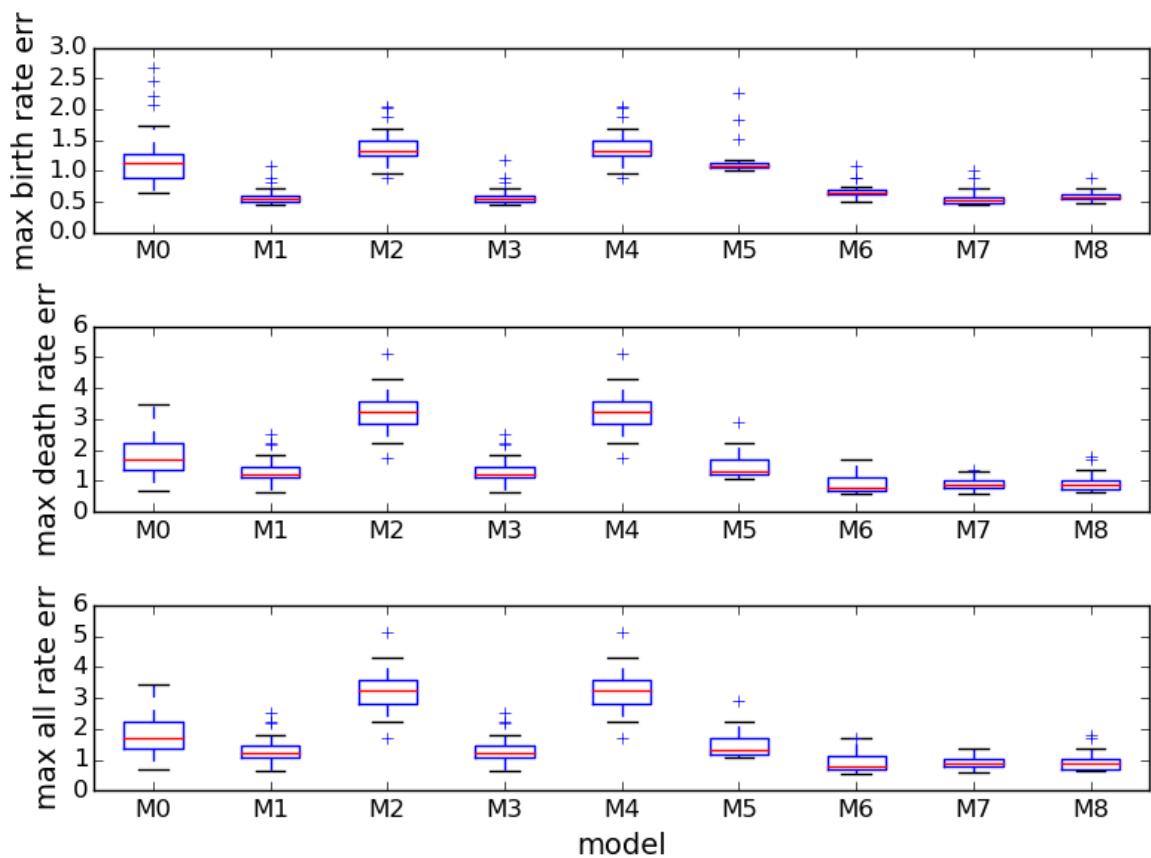


Figure 6.43: 5 species: errors in rate estimates

## 6.5. APPLICATION TO IBM (OPTIONAL)

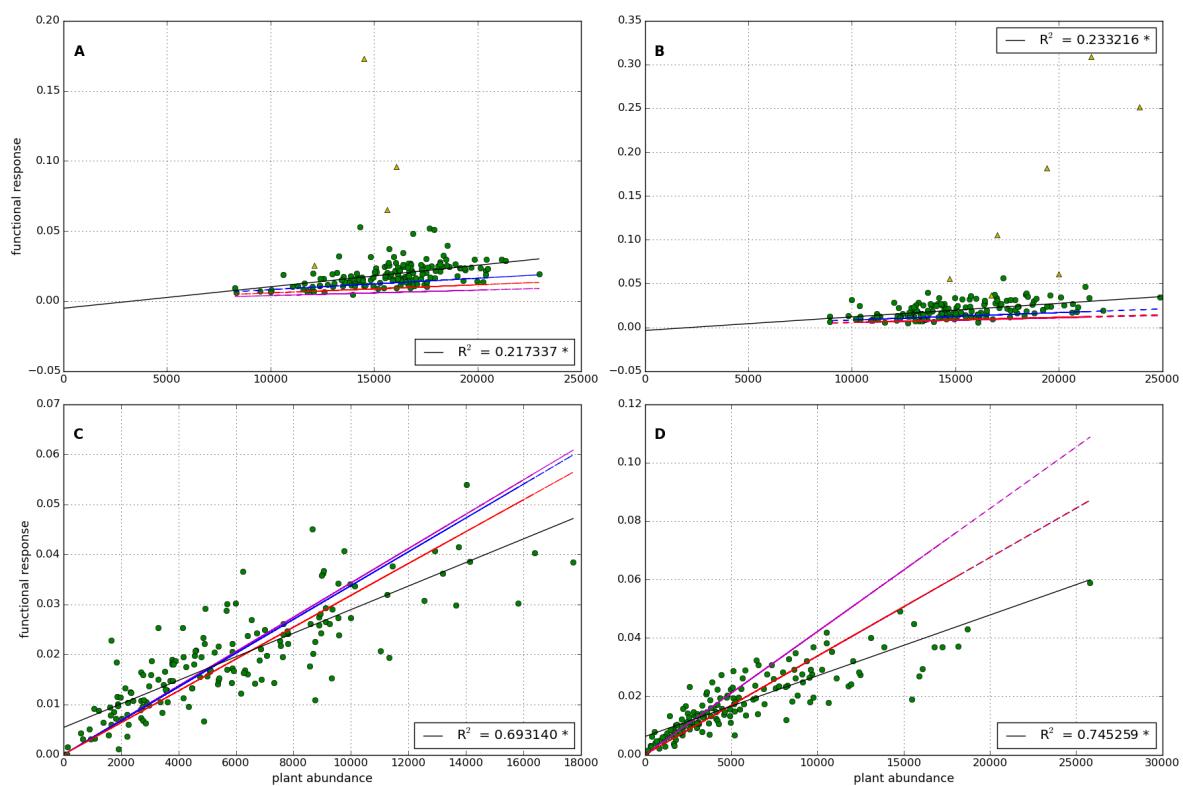


Figure 6.44: 5 species: functional response (High IR)

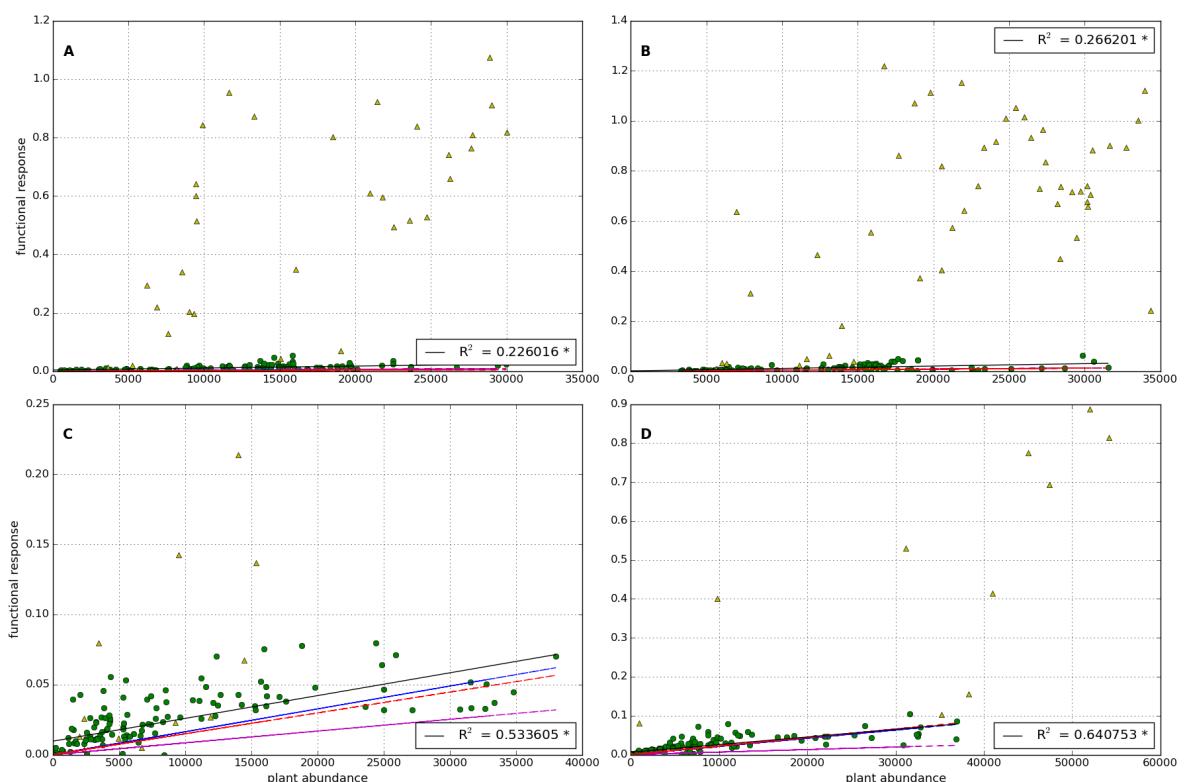


Figure 6.45: 5 species: functional response (Low IR)

## 6.6 Discussion

Points referenced in text above, make sure to discuss them!..

- Discuss how this methodology could be used on empirical data...
- Limitations of ODE models (non-spatial, response to debate on FR)
- Possibility of extending to more than two species systems (if this is actually done then change ref in text above)
- discussion of other forms of FR (not H) - or discussed already in intro?
- good GLV fit to LV even with 100 sample points - realistic?
- how good is the method of model fitting. Discuss more computationally expensive options (mentioned in section on Timme method)
- Spatial heterogeneity - we do not explore this here, but acknowledge that it represents a source of error. Gives example in extremis - 2 species only interacting on boundary. Also we know that there is some level of spatial aggregation, especially at low IR - possibly show one plot of this?
- Could introduce prey handling to IBM to create non-linear FR.
- Our method could be used to pick out coupling to other variables e.g environmental (temperature) if expressed in a certain way.
- In real application would not have luxury of selecting the section of time series with best fit! Would be lucky to have 1000 time points at all!



## BIBLIOGRAPHY

- [1] D. AI, P. DESJARDINS-PROULX, C. CHU, AND G. WANG, *Immigration, local dispersal limitation, and the repeatability of community composition under neutral and niche dynamics*, (2012).
- [2] M. ALBRECHT, P. DUELLI, B. SCHMID, AND C. B. MÜLLER, *Interaction diversity within quantified insect food webs in restored and adjacent intensively managed meadows*, Journal of Animal Ecology, 76 (2007), pp. 1015–1025.
- [3] M. ALMEIDA-NETO, P. GUIMARAES, P. R. GUIMARÃES, R. D. LOYOLA, AND W. ULRICH, *A consistent metric for nestedness analysis in ecological systems: reconciling concept and measurement*, Oikos, 117 (2008), pp. 1227–1239.
- [4] E. ANDERSON, Z. BAI, C. BISCHOF, S. BLACKFORD, J. DEMMEL, J. DONGARRA, J. DU CROZ, A. GREENBAUM, S. HAMMARLING, A. MCKENNEY, AND D. SORENSEN, *LAPACK Users' Guide*, Society for Industrial and Applied Mathematics, Philadelphia, PA, third ed., 1999.
- [5] H. P. ANDREASSEN, P. GLORVIGEN, A. RÉMY, AND R. A. IMS, *New views on how population-intrinsic and community-extrinsic processes interact during the vole population cycles*, Oikos, 122 (2013), pp. 507–515.
- [6] T. APARICIO, E. F. POZO, AND D. SAURA, *Detecting determinism using recurrence quantification analysis: Three test procedures*, Journal of Economic Behavior & Organization, 65 (2008), pp. 768–787.
- [7] J.-F. ARNOLDI, M. LOREAU, AND B. HAEGEMAN, *Resilience, reactivity and variability: A mathematical comparison of ecological stability measures*, Journal of theoretical biology, 389 (2016), pp. 47–59.
- [8] C. BANAŠEK-RICHTER, M.-F. CATTIN, AND L.-F. BERSIER, *Sampling effects and the robustness of quantitative and qualitative food-web descriptors*, Journal of Theoretical Biology, 226 (2004), pp. 23–32.
- [9] F. BARRAQUAND, *Functional responses and predator-prey models: a critique of ratio dependence*, Theoretical ecology, 7 (2014), pp. 3–20.

## BIBLIOGRAPHY

---

- [10] J. BASCOMPTE AND P. JORDANO, *Plant-animal mutualistic networks: the architecture of biodiversity*, Annual Review of Ecology, Evolution, and Systematics, (2007), pp. 567–593.
- [11] E. L. BERLOW, A.-M. NEUTEL, J. E. COHEN, P. C. DE RUITER, B. EBENMAN, M. EMMERSON, J. W. FOX, V. A. JANSEN, J. IWAN JONES, G. D. KOKKORIS, ET AL., *Interaction strengths in food webs: issues and opportunities*, Journal of animal ecology, 73 (2004), pp. 585–598.
- [12] L.-F. BERSIER, C. BANASÅEK-RICHTER, AND M.-F. CATTIN, *Quantitative descriptors of food-web matrices*, Ecology, 83 (2002), pp. 2394–2407.
- [13] N. BLÜTHGEN, J. FRÜND, D. P. VÁZQUEZ, AND F. MENZEL, *What do interaction network metrics tell us about specialization and biological traits*, Ecology, 89 (2008), pp. 3387–3399.
- [14] N. BLÜTHGEN, F. MENZEL, AND N. BLÜTHGEN, *Measuring specialization in species interaction networks*, BMC ecology, 6 (2006), p. 1.
- [15] D. BOSSIO, M. S. GIRVAN, L. VERCHOT, J. BULLIMORE, T. BORELLI, A. ALBRECHT, K. SCOW, A. S. BALL, J. PRETTY, AND A. M. OSBORN, *Soil microbial community response to land use change in an agricultural landscape of western kenya*, Microbial ecology, 49 (2005), pp. 50–62.
- [16] B. C RALL, C. GUILL, AND U. BROSE, *Food-web connectance and predator interference dampen the paradox of enrichment*, Oikos, 117 (2008), pp. 202–213.
- [17] M. W. CADOTTE, *Dispersal and species diversity: A meta-analysis*, The American Naturalist, 167 (2006), pp. 913–924.
- [18] V. CAMPBELL, G. MURPHY, AND T. N. ROMANUK, *Experimental design and the outcome and interpretation of diversity–stability relations*, Oikos, 120 (2011), pp. 399–408.
- [19] R. CONDIT, R. A. CHISHOLM, AND S. P. HUBBELL, *Thirty years of forest census at barro colorado and the importance of immigration in maintaining diversity*, (2012).
- [20] K. Z. COYTE, J. SCHLUTER, AND K. R. FOSTER, *The ecology of the microbiome: Networks, competition, and stability*, Science, 350 (2015), pp. 663–666.
- [21] J. P. CRUTCHFIELD AND K. YOUNG, *Inferring statistical complexity*, Physical Review Letters, 63 (1989), p. 105.
- [22] M. R. DALE AND M.-J. FORTIN, *Spatial analysis: a guide for ecologists*, Cambridge University Press, 2014.

- [23] C. DARWIN AND W. F. BYNUM, *The origin of species by means of natural selection: or, the preservation of favored races in the struggle for life*, AL Burt, 2009.
- [24] R. A. DESHARNAIS, *Population dynamics and laboratory ecology*, vol. 37, Elsevier, 2005.
- [25] A. DOBSON, D. LODGE, J. ALDER, G. S. CUMMING, J. KEYMER, J. MCGLADE, H. MOONEY, J. A. RUSAK, O. SALA, V. WOLTERS, ET AL., *Habitat loss, trophic collapse, and the decline of ecosystem services*, *Ecology*, 87 (2006), pp. 1915–1924.
- [26] I. DONOHUE, O. L. PETCHEY, J. M. MONTOYA, A. L. JACKSON, L. McNALLY, M. VIANA, K. HEALY, M. LURGI, N. E. O'CONNOR, AND M. C. EMMERSON, *On the dimensionality of ecological stability*, *Ecology letters*, 16 (2013), pp. 421–429.
- [27] C. F. DORMANN, B. GRUBER, AND J. FRÜND, *Introducing the bipartite package: analysing ecological networks*, *interaction*, 1 (2008), pp. 0–2413793.
- [28] K. L. DRURY, J. D. SUTER, J. B. RENDALL, A. M. KRAMER, AND J. M. DRAKE, *Immigration can destabilize tri-trophic interactions: implications for conservation of top predators*, *Theoretical Ecology*, (2015), pp. 1–12.
- [29] J. A. DUNNE, R. J. WILLIAMS, AND N. D. MARTINEZ, *Food-web structure and network theory: the role of connectance and size*, *Proceedings of the National Academy of Sciences*, 99 (2002), pp. 12917–12922.
- [30] C. DYTHAM, *The effect of habitat destruction pattern on species persistence: a cellular model*, *Oikos*, (1995), pp. 340–344.
- [31] D. M. EVANS, M. J. POCOCK, AND J. MEMMOTT, *The robustness of a network of ecological networks to habitat loss*, *Ecology letters*, 16 (2013), pp. 844–852.
- [32] E. S. EVELEIGH, K. S. MCCANN, P. C. McCARTHY, S. J. POLLOCK, C. J. LUCAROTTI, B. MORIN, G. A. McDougall, D. B. STRONGMAN, J. T. HUBER, J. UMBANHOWAR, ET AL., *Fluctuations in density of an outbreak species drive diversity cascades in food webs*, *Proceedings of the National Academy of Sciences*, 104 (2007), pp. 16976–16981.
- [33] R. M. EWERS, R. K. DIDHAM, L. FAHRIG, G. FERRAZ, A. HECTOR, R. D. HOLT, V. KPOS, G. REYNOLDS, W. SINUN, J. L. SNADDON, ET AL., *A large-scale forest fragmentation experiment: the stability of altered forest ecosystems project*, *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, 366 (2011), pp. 3292–3302.
- [34] J. A. FOLEY, R. DEFRIES, G. P. ASNER, C. BARFORD, G. BONAN, S. R. CARPENTER, F. S. CHAPIN, M. T. COE, G. C. DAILY, H. K. GIBBS, ET AL., *Global consequences of land use*, *science*, 309 (2005), pp. 570–574.

## BIBLIOGRAPHY

---

- [35] C. FONTAINE, P. R. GUIMARÃES, S. KÉFI, N. LOEUILLE, J. MEMMOTT, W. H. VAN DER PUTTEN, F. J. VAN VEEN, AND E. THÉBAULT, *The ecological and evolutionary implications of merging different types of networks*, Ecology Letters, 14 (2011), pp. 1170–1181.
- [36] M. A. FORTUNA AND J. BASCOMPTE, *Habitat loss and the structure of plant-animal mutualistic networks*, Ecology Letters, 9 (2006), pp. 281–286.
- [37] M. A. FORTUNA, A. KRISHNA, AND J. BASCOMPTE, *Habitat loss and the disassembly of mutualistic networks*, Oikos, 122 (2013), pp. 938–942.
- [38] B. L. FOSTER, T. L. DICKSON, C. A. MURPHY, I. S. KAREL, AND V. H. SMITH, *Propagule pools mediate community assembly and diversity-ecosystem regulation along a grassland productivity gradient*, Journal of Ecology, 92 (2004), pp. 435–449.
- [39] R. P. FRECKLETON, A. R. WATKINSON, AND M. REES, *Measuring the importance of competition in plant communities*, Journal of ecology, 97 (2009), pp. 379–384.
- [40] C. GARDINER, *Blue crystal*, 2016.
- [41] G. F. GAUSE, *Experimental analysis of vito volterra,Àôs mathematical theory of the struggle for existence*, Science, 79 (1934), pp. 16–17.
- [42] A. GONZALEZ, B. RAYFIELD, AND Z. LINDO, *The disentangled bank: how loss of habitat fragments and disassembles ecological networks*, American Journal of Botany, 98 (2011), pp. 503–516.
- [43] V. GRIMM AND S. F. RAILSBACK, *Individual-based modeling and ecology*, Princeton university press, 2013.
- [44] N. M. HADDAD, L. A. BRUDVIG, J. CLOBERT, K. F. DAVIES, A. GONZALEZ, R. D. HOLT, T. E. LOVEJOY, J. O. SEXTON, M. P. AUSTIN, C. D. COLLINS, ET AL., *Habitat fragmentation and its lasting impact on earth,Àôs ecosystems*, Science Advances, 1 (2015), p. e1500052.
- [45] M. HAGEN, W. D. KISSLING, C. RASMUSSEN, D. CARSTENSEN, Y. DUPONT, C. KAISER-BUNBURY, E. O'GORMAN, J. OLESEN, M. DE AGUIAR, L. BROWN, ET AL., *Biodiversity, species interactions and ecological networks in a fragmented world*, Advances in Ecological Research, 46 (2012), pp. 89–120.
- [46] I. HANSKI, G. A. ZURITA, M. I. BELLOCQ, AND J. RYBICKI, *Species-fragmented area relationship*, Proceedings of the National Academy of Sciences, 110 (2013), pp. 12715–12720.

- [47] S. HAZLEDINE, *The Analysis and modelling of chaotic calcium oscillations*, PhD thesis, Department for Computational Systems Biology, John Innes Centre, 2009.
- [48] M. HILL AND H. CASWELL, *Habitat fragmentation and extinction thresholds on fractal landscapes*, Ecology Letters, 2 (1999), pp. 121–127.
- [49] M. HOLYOAK AND S. P. LAWLER, *Persistence of an extinction-prone predator-prey interaction through metapopulation dynamics*, Ecology, (1996), pp. 1867–1879.
- [50] H. P. HSU, *Schaum's outline of theory and problems of probability, random variables, and random processes*, McGraw-Hill, 1997.
- [51] G. E. HUTCHINSON, *The paradox of the plankton*, The American Naturalist, 95 (1961), pp. 137–145.
- [52] T. C. INGS, J. M. MONTOYA, J. BASCOMPTE, N. BLÜTHGEN, L. BROWN, C. F. DORMANN, F. EDWARDS, D. FIGUEROA, U. JACOB, J. I. JONES, ET AL., *Review: Ecological networks—beyond food webs*, Journal of Animal Ecology, 78 (2009), pp. 253–269.
- [53] H. I. JAGER, E. A. CARR, AND R. A. EFROYMSON, *Simulated effects of habitat loss and fragmentation on a solitary mustelid predator*, Ecological Modelling, 191 (2006), pp. 416–430.
- [54] V. A. JANSEN AND G. D. KOKKORIS, *Complexity and stability revisited*, Ecology Letters, 6 (2003), pp. 498–502.
- [55] D. H. JANZEN, *The deflowering of central America*, Nat. Hist, 83 (1974), pp. 48–53.
- [56] O. P. JUDSON, *The rise of the individual-based model in ecology*, Trends in Ecology & Evolution, 9 (1994), pp. 9–14.
- [57] R. KAARTINEN AND T. ROSLIN, *Shrinking by numbers: landscape context affects the species composition but not the quantitative structure of local food webs*, Journal of Animal Ecology, 80 (2011), pp. 622–631.
- [58] S. KÉFI, E. L. BERLOW, E. A. WIETERS, S. A. NAVARRETE, O. L. PETCHEY, S. A. WOOD, A. BOIT, L. N. JOPPA, K. D. LAFFERTY, R. J. WILLIAMS, ET AL., *More than a meal, integrating non-feeding interactions into food webs*, Ecology letters, 15 (2012), pp. 291–300.
- [59] C. KLAUSMEIER, *Habitat destruction and extinction in competitive and mutualistic meta-communities*, Ecology Letters, 4 (2001), pp. 57–63.
- [60] T. M. KNIGHT, M. W. MCCOY, J. M. CHASE, K. A. MCCOY, AND R. D. HOLT, *Trophic cascades across ecosystems*, Nature, 437 (2005), pp. 880–883.

## BIBLIOGRAPHY

---

- [61] G. KOKKORIS, A. TROUMBIS, AND J. LAWTON, *Patterns of species interaction strength in assembled theoretical competition communities*, Ecology Letters, 2 (1999), pp. 70–74.
- [62] N. M. KORFANTA, W. D. NEWMARK, AND M. J. KAUFFMAN, *Long-term demographic consequences of habitat fragmentation to a tropical understory bird community*, Ecology, 93 (2012), pp. 2548–2559.
- [63] C. J. KREBS, *Of lemmings and snowshoe hares: the ecology of northern canada*, Proceedings of the Royal Society of London B: Biological Sciences, 278 (2011), pp. 481–489.
- [64] C. KREMEN, N. M. WILLIAMS, M. A. AIZEN, B. GEMMILL-HERREN, G. LEBUHN, R. MINCKLEY, L. PACKER, S. G. POTTS, I. STEFFAN-DEWENTER, D. P. VAZQUEZ, ET AL., *Pollination and other ecosystem services produced by mobile organisms: a conceptual framework for the effects of land-use change*, Ecology Letters, 10 (2007), pp. 299–314.
- [65] D. KWIATKOWSKI, P. C. PHILLIPS, P. SCHMIDT, AND Y. SHIN, *Testing the null hypothesis of stationarity against the alternative of a unit root: How sure are we that economic time series have a unit root?*, Journal of econometrics, 54 (1992), pp. 159–178.
- [66] J. LADYMAN, J. LAMBERT, AND K. WIESNER, *What is a complex system?*, European Journal for Philosophy of Science, 3 (2013), pp. 33–67.
- [67] M. LOREAU AND N. MOUQUET, *Immigration and the maintenance of local species diversity*, The American Naturalist, 154 (1999), pp. 427–440.
- [68] M. LURGI, D. MONTOYA, AND J. M. MONTOYA, *The effects of space and diversity of interaction types on the stability of complex ecological networks*, Theoretical Ecology, (2015), pp. 1–11.
- [69] R. MAC NALLY, *Use of the abundance spectrum and relative-abundance distributions to analyze assemblage change in massively altered landscapes*, The American Naturalist, 170 (2007), pp. 319–330.
- [70] N. MARWAN, M. C. ROMANO, M. THIEL, AND J. KURTHS, *Recurrence plots for the analysis of complex systems*, Physics reports, 438 (2007), pp. 237–329.
- [71] R. M. MAY, *Will a large complex system be stable?*, Nature, 238 (1972), pp. 413–414.
- [72] ———, *Patterns in multi-species communities*, Theoretical ecology: principles and applications, (1981), pp. 197–227.
- [73] K. MCCANN, A. HASTINGS, AND G. R. HUXEL, *Weak trophic interactions and the balance of nature*, Nature, 395 (1998), pp. 794–798.

- [74] B. J. MCGILL, R. S. ETIENNE, J. S. GRAY, D. ALONSO, M. J. ANDERSON, H. K. BENECHA, M. DORNELAS, B. J. ENQUIST, J. L. GREEN, F. HE, ET AL., *Species abundance distributions: moving beyond single prediction theories to integration within an ecological framework*, Ecology letters, 10 (2007), pp. 995–1015.
- [75] C. MCWILLIAMS, *Ibm animations*, 2016.
- [76] C. J. MELIÁN AND J. BASCOMPTE, *Food web structure and habitat loss*, Ecology Letters, 5 (2002), pp. 37–46.
- [77] J. MEMMOTT, R. GIBSON, L. CARVALHEIRO, K. HENSON, R. HELENO, M. LOPEZARAIZA, S. PEARCE, AND S. PEARCE, *The conservation of ecological interactions*, Insect Conservation Biology. The Royal Entomological Society, London, (2007), pp. 226–44.
- [78] D. MONTOYA, M. YALLOP, AND J. MEMMOTT, *Functional group diversity increases with modularity in complex food webs*, Nature communications, 6 (2015).
- [79] J. MONTOYA, J. ARNOLDI, AND M. LOREUA, *Resilience, invariability, and ecological stability across levels of organization*, WHICH JOURNAL, (in press).
- [80] J. M. MONTOYA, S. L. PIMM, AND R. V. SOLÉ, *Ecological networks and their fragility*, Nature, 442 (2006), pp. 259–264.
- [81] A. MOUGI AND M. KONDŌH, *Diversity of interaction types and ecological community stability*, Science, 337 (2012), pp. 349–351.
- [82] N. MOUQUET, P. LEADLEY, J. MÉRIGUET, AND M. LOREAU, *Immigration and local competition in herbaceous plant communities: a three-year seed-sowing experiment*, Oikos, 104 (2004), pp. 77–90.
- [83] P. J. MUMBY, I. CHOLLETT, Y.-M. BOZEC, AND N. H. WOLFF, *Ecological resilience, robustness and vulnerability: how do these concepts benefit ecosystem management?*, Current Opinion in Environmental Sustainability, 7 (2014), pp. 22–27.
- [84] A.-M. NEUTEL, J. A. HEESTERBEEK, AND P. C. DE RUITER, *Stability in real food webs: weak links in long loops*, Science, 296 (2002), pp. 1120–1123.
- [85] C. NORRIS, *Quantum theory and the flight from realism: Philosophical responses to quantum mechanics*, Routledge, 2002.
- [86] E. J. O'GORMAN AND M. C. EMMERSON, *Perturbations to trophic interactions and the stability of complex food webs*, Proceedings of the National Academy of Sciences, 106 (2009), pp. 13393–13398.

## BIBLIOGRAPHY

---

- [87] J. OKSANEN, R. KINTD, P. LEGENDRE, B. OÖHARA, M. H. H. STEVENS, M. J. OKSANEN, AND M. SUGGESTS, *The vegan package*, Community ecology package, 10 (2007).
- [88] O. OVASKAINEN, K. SATO, J. BASCOMPTE, AND I. HANSKI, *Metapopulation models for extinction threshold in spatially correlated landscapes*, Journal of Theoretical Biology, 215 (2002), pp. 95–108.
- [89] C. PARMESAN, N. RYRHOLM, C. STEFANESCU, J. K. HILL, C. D. THOMAS, H. DESCIMON, B. HUNTLEY, L. KAILA, J. KULLBERG, T. TAMMARU, ET AL., *Poleward shifts in geographical ranges of butterfly species associated with regional warming*, Nature, 399 (1999), pp. 579–583.
- [90] C. PARMESAN AND G. YOHE, *A globally coherent fingerprint of climate change impacts across natural systems*, Nature, 421 (2003), pp. 37–42.
- [91] M. J. POCOCK, D. M. EVANS, AND J. MEMMOTT, *The robustness and restoration of a network of ecological networks*, Science, 335 (2012), pp. 973–977.
- [92] M. PRIESTLEY AND T. S. RAO, *A test for non-stationarity of time-series*, Journal of the Royal Statistical Society. Series B (Methodological), (1969), pp. 140–149.
- [93] T. PUETTKER, A. A. BUENO, C. DOS SANTOS DE BARROS, S. SOMMER, AND R. PARDINI, *Immigration rates in fragmented landscapes—empirical evidence for the importance of habitat amount for species persistence*, PLoS One, 6 (2011), p. e27963.
- [94] R CORE TEAM, *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria, 2013.
- [95] D. RAFFAELLI, *How extinction patterns affect ecosystems*, Science, 306 (2004), pp. 1141–1142.
- [96] W. J. RIPPLE AND R. L. BESCHTA, *Trophic cascades in yellowstone: The first 15 years after wolf reintroduction*, Biological Conservation, 145 (2012), pp. 205–213.
- [97] S. E. SAID AND D. A. DICKEY, *Testing for unit roots in autoregressive-moving average models of unknown order*, Biometrika, 71 (1984), pp. 599–607.
- [98] A. SAUVE, C. FONTAINE, AND E. THÉBAULT, *Structure–stability relationships in networks combining mutualistic and antagonistic interactions*, Oikos, 123 (2014), pp. 378–384.
- [99] S. G. SHANDILYA AND M. TIMME, *Inferring network topology from complex dynamics*, New Journal of Physics, 13 (2011), p. 013004.

- [100] A. R. SINCLAIR, D. CHITTY, C. I. STEFAN, AND C. J. KREBS, *Mammal population cycles: evidence for intrinsic differences during snowshoe hare cycles*, Canadian Journal of Zoology, 81 (2003), pp. 216–220.
- [101] R. V. SOLE AND J. M. MONTOYA, *Ecological network meltdown from habitat loss and fragmentation*, Ecological Networks: Linking Structure to Dynamics in Food Webs, (2006), pp. 305–323.
- [102] B. J. SPIESMAN AND B. D. INOUYE, *Habitat loss alters the architecture of plant-pollinator interaction networks*, Ecology, 94 (2013), pp. 2688–2696.
- [103] P. STANICZENKO, O. T. LEWIS, N. S. JONES, AND F. REED-TSOCHAS, *Structural dynamics and robustness of food webs*, Ecology Letters, 13 (2010), pp. 891–899.
- [104] N. C. STENSETH, W. FALCK, O. N. BJØRNSTAD, AND C. J. KREBS, *Population regulation in snowshoe hare and canadian lynx: asymmetric food web configurations between hare and lynx*, Proceedings of the National Academy of Sciences, 94 (1997), pp. 5147–5152.
- [105] D. STOUFFER, J. CAMACHO, R. GUIMERA, C. NG, AND L. NUNES AMARAL, *Quantitative patterns in the structure of model and empirical food webs*, Ecology, 86 (2005), pp. 1301–1311.
- [106] A. TEYSSÈDRE AND A. ROBERT, *Contrasting effects of habitat reduction, conversion and alteration on neutral and non neutral biological communities*, Oikos, 123 (2014), pp. 857–865.
- [107] E. THÉBAULT AND C. FONTAINE, *Stability of ecological communities and the architecture of mutualistic and trophic networks*, Science, 329 (2010), pp. 853–856.
- [108] R. M. THOMPSON, U. BROSE, J. A. DUNNE, R. O. HALL, S. HLADYZ, R. L. KITCHING, N. D. MARTINEZ, H. RANTALA, T. N. ROMANUK, D. B. STOUFFER, ET AL., *Food webs: reconciling the structure and function of biodiversity*, Trends in ecology & evolution, 27 (2012), pp. 689–697.
- [109] D. TILMAN, R. M. MAY, C. L. LEHMAN, AND M. A. NOWAK, *Habitat destruction and the extinction debt*, (1994).
- [110] J. TRAVIS, *Climate change and habitat destruction: a deadly anthropogenic cocktail*, Proceedings of the Royal Society of London B: Biological Sciences, 270 (2003), pp. 467–473.
- [111] J. M. TYLIANAKIS, T. TSCHARNTKE, AND O. T. LEWIS, *Habitat modification alters the structure of tropical host-parasitoid food webs*, Nature, 445 (2007), pp. 202–205.

## BIBLIOGRAPHY

---

- [112] A. VALIENTE-BANUET, M. A. AIZEN, J. M. ALCÁNTARA, J. ARROYO, A. COCUCCI, M. GALETTI, M. B. GARCÍA, D. GARCÍA, J. M. GÓMEZ, P. JORDANO, ET AL., *Beyond species loss: the extinction of ecological interactions in a changing world*, Functional Ecology, 29 (2015), pp. 299–307.
- [113] C. VAN ALTEA, L. HEMERIK, AND P. C. DE RUITER, *Food web stability and weighted connectance: the complexity-stability debate revisited*, Theoretical Ecology, (2016), pp. 1–10.
- [114] B. H. WARREN, D. SIMBERLOFF, R. E. RICKLEFS, R. AGUILÉE, F. L. CONDAMINE, D. GRAVEL, H. MORLON, N. MOUQUET, J. ROSINDELL, J. CASQUET, ET AL., *Islands as model systems in ecology and evolution: prospects fifty years after macarthur-wilson*, Ecology letters, 18 (2015), pp. 200–217.
- [115] G. B. WEST, J. H. BROWN, AND B. J. ENQUIST, *A general model for the origin of allometric scaling laws in biology*, Science, 276 (1997), pp. 122–126.
- [116] R. H. WHITTAKER, *Dominance and diversity in land plant communities numerical relations of species express the importance of competition in community function and evolution*, Science, 147 (1965), pp. 250–260.
- [117] R. J. WILLIAMS AND N. D. MARTINEZ, *Simple rules yield complex food webs*, Nature, 404 (2000), pp. 180–183.
- [118] ———, *Success and its limits among structural models of complex food webs*, Journal of Animal Ecology, 77 (2008), pp. 512–519.
- [119] J. B. WILSON, *Methods for fitting dominance/diversity curves*, Journal of Vegetation Science, 2 (1991), pp. 35–46.
- [120] K. A. WITH AND A. W. KING, *Extinction thresholds for species in fractal landscapes*, Conservation Biology, 13 (1999), pp. 314–326.
- [121] C. E. ZARTMAN AND H. E. NASCIMENTO, *Are habitat-tracking metacommunities dispersal limited? inferences from abundance-occupancy patterns of epiphylls in amazonian forest fragments*, Biological Conservation, 127 (2006), pp. 46–54.